Analyzing the impact of the NCAA Selection committee's new quadrant system

Bruce A. Reinig[a,*] and Ira Horowitz[b]
[a]*Management Information Systems, Fowler College of Business, San Diego
State University, San Diego, CA, USA*
[b]*Information Systems and Operations Management, Warrington College of Business,
University of Florida, Gainesville, FL USA*

Abstract. In 2018 the NCAA introduced a new metric, namely a four-Quadrant system intended to adjust for home-court advantage when assessing quality wins in the selection and seeding process for the Division I Men's Basketball Tournament. We apply a linear programming procedure for ranking potential candidates for the tournament, based upon the traditional criteria, show how the rankings change with the inclusion of the Quadrant metric, and conclude that the metric has a substantial impact on the choices. We go on to show that the new ranking and choices are much more in sync with the NCAA Selection Committee's actual ranking and choices than had been the case absent the metric, leading to the inference that the metric played an important role in the Committee's decision making.

1. Introduction

The NCAA Division I Men's Basketball Tournament is a knockout (single-loss-elimination), six-round, three-week annual event that has historically been held in March, culminating in a championship game played on the first Monday evening in April. Over the years, the 64-team format that has been in place since 1985 has undergone a series of changes of varying import. The most recent and impactful change became effective with the 2011 tournament, when 68 of the then 339 Division I teams were invited to participate. But eight of the 68 were required to play in four so-called play-in games that would reduce the final set of competitors to the traditional 64-team format.

The 64 finalists are slotted into four brackets: nominally, East, West, Midwest, and South. In each bracket the teams are seeded 1 through 16, with the "1" judged to be the strongest and the "16" the weakest. The slot-

ting is done so as to allow the four 1 seeds to play their first-round games in the neutral-court venues that are closest to their home courts. Insofar as is feasible, similar courtesies are next extended to the 2 seeds, and so on down the line.

In the first-round pairings, the seed numbers sum to 17. Thus the presumed largest mismatches pair the four top seeds and the four bottom seeds, and the presumed nail-biters pair the four 8 s and the four 9 s. The 2018 tournament marked the first time in its 35-year history under the current format that a 16 – the University of Maryland (Baltimore County) – defeated a 1 – the University of Virginia. If "chalk" holds, in the sense that the higher seeds always prevail, the sum of the seed numbers in the second-round pairings will equal 9, and so forth in subsequent rounds, with the four top seeds paired in the penultimate round and the two surviving 1's then competing for the championship. Chalk, however, is honored largely in the breach at tournament time, thanks to both the inherent uncertainty in *any* sporting event and the inherent fallibility of the ten-person Selection Committee charged with selecting and seeding the 68 invitees. Hence the colloquial reference to the tourna-

*Corresponding author: Dr. Bruce A. Reinig, Fowler College of Business, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-8234, USA. Tel.: +1 619 594 3032; E-mail: breinig@sdsu.edu.

ment as March Madness aka The Big Dance. Indeed, it is not at all uncommon for the betting market to favor a lower seed over a higher seed, as the market agents do not necessarily share the Committee's judgments.

And those agents are not alone. The list of invitees and where they are seeded, along with the list of the scorned, are rarely without controversy, as there are no *definitive* rules guiding the Committee's decisions, beyond the strict requirement that invitations to the dance must be issued to the champions of each of the 32 basketball-playing conferences. The additional 36 invitations – the at-large selections - are issued at the Committee's discretion, which prior to the 2018 tournament were guided only by some vague NCAA mandates to the effect that attention should be paid to such factors as strength of schedule (SOS), a ratings percentage index (RPI) purporting to measure a team's performance, given the performance of its opponents and that of *their* opponents, a team's win percentage, and the "quality" of its wins and losses. To help the Selection Committee differentiate between high-quality and low-quality wins and losses, the NCAA provided Team Sheets that separate team schedules into four quadrants based on the quality of their opponent. In 2017 for example, the Team Sheets separated each team schedule based on games played against teams with RPIs ranked in the Top 50, RPIs ranked 51 to 100, RPIs ranked 101 to 200, and teams with RPI ranks below 201. In addition to identifying factors that should be attended to, the NCAA specifically mandates that factors such as conference affiliation or the reputation of the coach should be excluded from consideration. The Committee members are also encouraged to watch as many games as possible, especially those of teams to which they have been assigned, and to supplement those data with their eyeball tests.

Absent definitive guidance as to *how large a role* each of the non-barred factors should play in the Committee's deliberations and collective judgments, Reinig and Horowitz (2018), henceforth RH, proffered a mathematical programming (MP) approach to assigning weights to the pertinent factors which would guide both the selection of the at-large teams and the ranking and consequent seeding of all 68 invitees. The 2011 rule that required the 65th through the 68th-ranked teams to meet in one pair of play-in games whose winners would fill two 16 seeds, and the four lowest-ranked at-large teams to meet in the second pair of play-in games, would still apply. As emphasized by RH, the resulting recommendations

were not intended to *replace* the Committee's judgments, but rather to provide them a *raison d'etre* for some of their choices and an impetus for articulating the reasons for going in another direction in others.

RH carry out their analysis and provide *ex post* recommendations for the 2012 through 2016 tournaments, which they show align quite well with the respective Selection Committees' selections and seeding. More critically, they demonstrate the feasibility of their approach by applying it *ex ante* to the formal Selection Sunday announcement of the Committee's decisions. When doing so, they identified only a single case in which their approach yielded a markedly different result from that of the Committee - that in which it ranked Illinois State as the 32nd at-large selection, which as such would have merited an invitation, as opposed to the Committee's choice of Marquette, which the MP approach had ranked 39th.

Then, in early February 2018, in time for implementation by the Selection Committee in making its selections for the 2018 tournament, the NCAA announced a major alteration to the system, one intended to adjust for home-court advantage: namely, the heretofore informally and selectively applied principle that all else being equal, a road win is superior to a win at a neutral site, which is in turn superior to a home win. That is, some wins are better than others and some losses are worse than others, and a systematic recognition of such, which historically entered into the Committee's deliberations in an unsystematic and *ad hoc* basis, was now formalized through the metric of "Quadrant" wins and losses wherein four Quadrants are defined in the Team Sheets that summarize the data provided to the Committee members. There, for example, Quadrant 1 (Q1) win-loss results are those compiled at home against teams ranked 1 to 30 in accordance with RPI, plus those compiled on neutral courts against teams ranked 1 to 50 by RPI, plus those compiled on the road against teams ranked 1 to 75 by RPI. At the other extreme, Q4 win-loss results are those compiled at home against teams ranked 161 or higher by RPI, plus those compiled on neutral courts against teams ranked 201 or higher, plus those complied on the opponent's court against teams ranked 241 or higher by RPI. Individual game outcomes are listed within the opponent's Quadrant by RPI, and color-coordinated to identify non-conference games and game site, and to highlight losses. Thus, for example, the 2018 Team Sheet for Virginia shows its lowest RPI-ranked opponent in Quadrant 1 to have been 61st-ranked Virginia Tech, a road opponent that it defeated in Blacksburg,

Virginia on January 3, 2018. Virginia's February 10, 2018 home victory over that same 61st-ranked Virginia Tech in Charlottesville, Virginia, however, is listed in Quadrant 2. Thus, greater merit is given to beating Virginia Tech on the road than at home.

The addition of the new Quadrant metric in the evaluation process drew mixed reviews in the media, with some pundits viewing it as long overdue and others criticizing it because it continues to place great reliance on RPI, seen by some as a flawed measure. Paul and Wilson (2015), in particular, provide a critical analysis of its use, and identify its principal flaw as its focus on wins and losses and, unlike the Sagarin ratings, its failure to consider margin of victory. But did this new metric actually change anything? That is, were the Committee's 2018 selections any different from those that would have obtained in its absence? Unfortunately, there is no way of *knowing* that, short of interviewing the individual members and soliciting their impressions. But an alternative approach is to proceed in the by now standard vein pioneered by Coleman and Lynch (2001; 2009), and followed up in Shapiro et al. (2009), Coleman et al. (2010; 2016), Leman et al. (2014), and Paul and Wilson (2015), which uses logit/probit regression modelling in the attempt to identify the factors that had a statistically-significant impact on the likelihood that a school did or did not receive an at-large tournament bid. For reasons that are discussed at the end of Section 4 below, that approach was unrewarding in the present application.

What we *can* establish, however, is how the introduction of this new metric would have altered the selections and ranking and seeds obtained through the application of the MP approach, and concurrently attempt to *infer* its impact on the Committee's decisions. Doing so is the purpose of this paper.

Specifically, we use the RH approach to analyze the 2018 tournament both with and without using the new Quadrant system to determine what, if any, impact the Quadrant system has had on NCAA selections and bracket rankings. It will be shown that while absent the Quadrant data the recommendations of the MP approach still align quite well with the Quadrant-based decisions, incorporating those data into the MP analysis do indeed alter the MP-based recommendations and bring them even more in sync with the Committee's choices. The latter suggests, in turn, that the Committee *also* incorporated these new data into its decision processes. What our analysis further suggests, however, is that the Committee members' subjective impressions of the teams and what would

seem to be their non-systematic interpretations of the data resulted in numerous inconsistencies in the final team rankings, considered in a systematic sequence of paired comparisons.

2. A brief overview of the mathematical programming approach

RH (2018, p. 188) focus on a small but comprehensive set of seven quantifiable factors that summarize expert opinions of reporters and coaches, along with specific team-performance measures, in the team-ranking process. Those factors are: Associated Press (AP) Poll Points; USA Today Coaches Poll Points; Win Percentage; RPI; SOS; Wins against teams in the Basketball Power Index (BPI) Top 50, or Quality Wins; and Losses to Teams Other than Those in the BPI Top 50, or Non-Quality Losses.

The procedure assigns a non-negative weight to each of these factors. To assure the non-negativity of the weight assigned to non-quality losses, these are multiplied by negative one. To eliminate the influence of unit-of-measurement in the factor-weighting process, the factors are normalized prior to processing the data. Then, with the normalized value of factor X_i ($i = 1, \ldots, I$) for team k ($k = 1, \ldots, K$) denoted x_{ki}, the implied linear evaluation of team k is $V_k = \sum w_i x_{ki}$, where $\sum w_i = 1$ and $w_i \geq 0$.

Team k is said to dominate team j and hence should be ranked at least as high as team j when $x_{ki} \geq x_{ji}$ for all values of i, with the strict inequality holding for at least one value of i (e.g., Loomba and Turban, 1974, p. 232). Thus, for any pair of relationships in which team k dominates team j, it is necessarily true that $V_k - V_j = \sum (w_i(x_{ki} - x_{ji}) = S_{kj} \geq 0$. To solve the problem, RH select as their objective function the minimization of the maximum $Z = S_M \geq S_{kj}$ ($k = 1, \ldots, K$), which converts this into a readily-solved linear programming (LP) problem. From the transitivity property, when team j in turn dominates team h, it is necessarily true that team k does as well. Thus it is not necessary to specify that $S_{kh} \geq 0$. The LP process will see to it that it does.

3. The approach applied to the 2018 tournament

We applied the RH approach to the 2018 tournament data in three stages. In the first stage, which was implemented for data through February 26, we

selected 80 teams from the full set of now 351 Division I teams, less three teams ruled ineligible for having failed the NCAA's APL (academic) requirements. The set initially included the teams that rated highest in BPI to which we subsequently added what were at the time the best-performing teams from conferences that are generally regarded as being (informally) limited to one bid, and that failed to land a team in the former set. The resulting sample comprised 80 teams. The solution to the problem assured that we would be able to pluck 36 at-large teams from the set, once each of the 32 automatic bids had been determined.

The procedure was repeated the following week using the data through March 5. By this time, however, five conference champions had already been determined: Michigan (Big Ten), Loyola-Chicago (Missouri Valley), Murray State (Ohio Valley), Lipscomb (Atlantic Sun), and Radford (Big South). The former three had qualified for inclusion in the set by virtue of their BPI standing. As was the case in the second go-round, we juggled some teams from the earlier set to take advantage of the previous week's results, which included the conclusion of the regular season in several conferences as well as the early rounds of conference tournaments in several others. The results of the new trial, in particular, suggested that we might have overlooked a team such as Penn State or Utah that was peaking at the right time. We also took advantage of the insights of such so-called and well-known "bracketologists" as Joe Lunardi (ESPN) and Jerry Palm (CBS), to name two, to help assure that we were not overlooking reasonable candidates for an invitation to the dance. Although the procedure itself does not constrain the number of teams that can be considered, from a practical computational standpoint it makes sense to eliminate at the outset those teams that would seem to have eliminated themselves by virtue of their failure to win enough games. Thus, our final file and application on Selection Sunday, which used the data through March 10 and the win-loss results of March 11, included 84 teams. Since tournament play in five conferences was still taking place on that day, the set included at least two teams from each of the latter conferences, which is why both Harvard and Pennsylvania (Ivy League), and both Georgia State and UT-Arlington (Sun Belt) were included.

We then re-ran our analysis using the Team Sheet ranked data used by the Committee. Thus, for example, in place of W%, we use $Y_5 =$ W% *rank*. And since a higher win percentage implies a lower-numbered rank, switching to ranked data means multiplying the

data by negative one, so as to assure the assignment of non-negative weights to all the factor ranks. We also drop the poll data in favor of the POM (Ken Pomeroy) $= Y_1$ and SAG (Jeff Sagarin) $= Y_2$ ratings, both of which are widely used and cited. We chose *not to* incorporate either KPI (Key Performance Indicator) or SOR (Strength of Record), since neither of these metrics enjoys either status. In place of RPI we use the Average RPI of the teams Beaten = Average RPI Wins $= Y_3$ and Average RPI of the teams Lost To = Average RPI Loss $= Y_4$. SOS is introduced in the ranked format as Y_6. We do not separately consider in-conference data and performance. Finally, we normalize all these ranked data, with the normalized rank of Y_k ($k = 1, \ldots, 6$) denoted y_k. The normalization to eliminate the influence of unit-of-measurement is applied even with the ranked data since, for example, the means range from 59.36 (y_1) to 163.62 (y_3), and the standard deviations range from 45.64 (y_5) to 90.65 (y_6). The latter do not reflect multiplying by negative one prior to processing. In the next section we extend our analysis to include the Quadrant data. In the interest of parsimony, the results both with and without the Quadrant data are presented in five basic tables, labeled accordingly Table 1 through Table 5. The immediate focus and discussion, however, is confined to the results for the six-factor Non-Quadrant analysis, as so labeled in the tables.

From the first row in Table 1 it is seen that with six salient factors the 84-team final sample of schools results in 1,225 dominant rules - that is, conditions imposed on the ranking of the teams - of which 434 are non-redundant. The first row of Table 2 provides the minimax weights attached to each of the six factors (variables), $Y_{1, \ldots,} Y_6$. We remark *en passant*, that consistent with the historical results and those for the 2017 tournament (RH, 2018, pp. 184-185) RPI Wins $= Y_3$ is assigned a zero weight, but RPI Losses $= Y_4$, which reflects quality losses, is assigned a large weight of $w_4 = 0.4774$. The latter is marginally higher than the $w_5 = 0.4516$ weight that is assigned to win percentage. SOS also enters into the process, but with a relatively minor role: $w_6 = 0.0710$. As was generally the case with the two polls, neither the Pomeroy nor the Sagarin ratings came into play. Thus a coach that wants to impress the LP scheme would be well advised to play a tough schedule, win a lot of those games, and suffer his losses to high-performing teams.

As seen from the first and third columns of Table 3, there are seven principal differences between the 36 "implied" and actual at-large invitations, where the

Table 1
Rules and objective function for the two variable sets

Team Set	Teams Analyzed	Dominant Rules	Non-Redundant Rules	Minimax Objective function
Non-Quadrant (six variables)	84	1,225	434	2.47
Quadrant Expansion (10 variables)	84	1,137	477	2.15

Table 2
Minimax solution weights for the two variable sets

Team Set	POM	SAG	RPIW	RPIL	WP	SOS	Q1W	Q1&2W	Q3L	Q3&4L
Non-Quadrant (six variables)	0.0000	0.0000	0.0000	0.4774	0.4516	0.0710	—	—	—	—
Quadrant Expansion (10 variables)	0.0000	0.0000	0.0000	0.2598	0.2896	0.0000	0.1105	0.3400	0.0000	0.0000

former comprise the 36 at-large teams that are ranked highest by the LP method, applied to the 84 sample teams, using the first row of weights in Table 2. Seven discrepancies is quite large by the standards of RH (2018). Specifically, the LP approach would have recommended non-invitees Middle Tennessee, St. Mary's, Nebraska, Louisville, Mississippi State, Southern California, and Utah, while bypassing the Committee's selections of Missouri, Alabama, Providence, Texas, Oklahoma, Arizona State, and Syracuse.

Table 4 lists the 68 invitees, the source of the invitation, and the NCAA rank. The fourth (numerical) column gives each team's rank as determined through the minimax weights applied to all 84 teams. Finally, the third row in each of the two blocs of Table 5 summarizes these results in two non-parametric correlations: the generally more conservative Kendall's tau = 0.731 and Spearman's rho = 0.901, both of which are statistically significant at $p < 0.0001$. In light of the strength of the correspondence, and in the attempt to avoid further cluttering up of the table, we have not included either the actual or the implied seeds, the derivation of which is a straight-forward task that we leave to the particularly involved reader.

4. The RH approach applied anew with the quadrant data included

The Quadrant data are introduced through four additional factors: the first two of which are intended to reflect Quality Wins:

Y_7 = Quadrant 1 Wins;
Y_8 = Quadrant 1 and Quadrant 2 Wins;
The second two reflect non-quality losses:
Y_9 = Quadrant 3 Losses;

Y_{10} = Quadrant 3 and Quadrant 4 Losses.

The inclusion of the Quadrant data has three principal effects. First, as a technical matter, and as seen in the second row of Table 1, it results in 88 fewer dominant rules (because domination is more difficult to achieve with additional factors to consider), but 43 additional rules are non-redundant (for precisely the same reason). Second, and more pertinently, these data essentially cut in half the weights assigned to RPI Losses and W% and eliminate SOS in its entirety. In place of the latter two factors, the new results give the highest weight, $w_8 = 0.3400$, to Quadrant 1 plus Quadrant 2 wins – that is quality wins - and a lesser weight of $w_7 = 0.1105$ to Quadrant 1 wins alone. It is thus immediately apparent that the Quadrant Wins data, but not the Quadrant Loss data, have an important impact on the LP results, and the factors that the results suggest will be important in the LP recommendations. These results, however, continue to bear the message that at-large bids will hinge upon coaches playing a tough schedule and winning most of those games. Third, as is apparent from the second column of Table 3, including the Quadrant Wins data brings the LP recommendations much more in line with the Committee's choices. Specifically, to all intents and purposes the two lists overlap and only differ in that the LP approach would still recommend Southern California and Middle Tennessee for at-large bids instead of the Committee's choices of Arizona State and Syracuse. That is, we've gone from seven different selections down to two different selections, for at-large bids. This is also seen in the second column of rankings in Table 4, which show that all but two of the Quadrant-based LP-ranked Top 68 teams correspond to those so–ranked by the Selection Committee, which necessarily corresponds to the two-team difference in the at-large selections.

Table 3
At-Large Selection Results for Quadrant Expansion and Non-Quadrant
Expansion compared to NCAA 2018 Selection Committee

Team	NCAA At-Large Only Rank	LP Quadrant Expansion Rank	LP Non-Quadrant Expansion Rank
Xavier	1	2	2
North Carolina	2	1	11
Duke	3	5	5
Purdue	4	4	3
Michigan St.	5	9	1
Tennessee	6	7	4
Texas Tech	7	8	15
Auburn	8	10	9
Wichita St.	9	6	6
West Virginia	10	3	12
Clemson	11	11	8
Ohio St.	12	13	10
Florida	13	12	35
Miami (FL)	14	18	13
Houston	15	14	23
TCU	16	23	18
Texas A&M	17	15	29
Arkansas	18	17	17
Nevada	19	16	14
Rhode Island	20	21	7
Seton Hall	21	26	27
Creighton	22	31	21
Virginia Tech	23	22	22
Missouri	24	28	——
Florida St.	25	29	33
Kansas St.	26	19	16
NC State	27	25	30
Alabama	28	33	——
Providence	29	27	——
Butler	30	32	36
Texas	31	36	——
Oklahoma	32	35	——
UCLA	33	30	31
St. Bonaventure	34	20	20
Arizona St.	35	——	——
Syracuse	36	——	——
Southern California	——	24	32
Middle Tenn.	——	34	19
Saint Mary's (CA)	——	——	24
Nebraska	——	——	25
Louisville	——	——	26
Mississippi St.	——	——	28
Utah	——	——	34

And in a closely-related vein, but from the somewhat different perspective of the non-parametric correlations of Table 5, it is seen that the Kendall coefficient increases to $\tau = 0.892$, and the Spearman coefficient increases to $\rho = 0.981$, when the Quadrant data determine both the at-large selections and the overall ranking of all 68 invitees.

We also applied the logit-regression approach, in two stages, to the 52-team sample of at-large candidates (i.e., teams receiving automatic bids were removed from the set) that survived our initial cut

for the RH analysis. In the first stage, as independent variables we included the six basic factors running from $Y_1 = \text{POM}$ to $Y_6 = \text{SOS}$. In the second stage we added the four Quadrant metrics, $Y_{7, \ldots}, Y_{10}$. In first-stage solution, only $Y_5 = \text{W\%}$ rank has a statistically-significant ($p = 0.029$) parameter estimate ($w_5 = 0.054$). Five of the estimates have the "correct" positive sign, but $w_1 = -0.085$ ($p = 0.316$), if taken seriously would imply that being well-regarded by Pomeroy is a detrimental to an aspirant's hopes. In the second-stage solution, none of the parameters

Table 4

NCAA Selection Committee Tournament Rankings compared to LP Solutions for Quadrant and Non-Quadrants Variable Sets

Team	NCAA Bid	NCAA Rank	LP Quad. Exp. Rank	LP Non-Quad Rank	Team	Bid	NCAA Rank	LP Quad. Exp. Rank	LP Non-Quad Rank
Virginia	Auto	1	1	4	NC State	At-Large	35	33	42
Villanova	Auto	2	3	7	Alabama	At-Large	36	41	——
Kansas	Auto	3	2	18	Providence	At-Large	37	35	——
Xavier	At-Large	4	5	3	Butler	At-Large	38	40	48
North Carolina	At-Large	5	4	17	Texas	At-Large	39	44	——
Duke	At-Large	6	9	9	Oklahoma	At-Large	40	43	——
Purdue	At-Large	7	8	5	UCLA	At-Large	41	38	43
Cincinnati	Auto	8	6	2	St. Bonaventure	At-Large	42	28	30
Michigan St.	At-Large	9	16	1	Arizona St.	At-Large	43	——	——
Tennessee	At-Large	10	13	8	Syracuse	At-Large	44	——	——
Michigan	Auto	11	14	11	San Diego St.	Auto	45	51	53
Texas Tech	At-Large	12	15	23	Loyola Chicago	Auto	46	50	50
Auburn	At-Large	13	17	14	New Mexico St.	Auto	47	45	33
Wichita St.	At-Large	14	11	10	Davidson	Auto	48	48	52
Gonzaga	Auto	15	19	6	South Dakota St.	Auto	49	46	24
Arizona	Auto	16	10	19	Murray St.	Auto	50	47	27
Kentucky	Auto	17	12	16	Buffalo	Auto	51	49	37
West Virginia	At-Large	18	7	20	UNCG	Auto	52	55	60
Clemson	At-Large	19	18	13	Col. of Charleston	Auto	53	53	56
Ohio St.	At-Large	20	21	15	Marshall	Auto	54	52	54
Florida	At-Large	21	20	47	Bucknell	Auto	55	56	51
Miami (FL)	At-Large	22	26	21	Montana	Auto	56	54	49
Houston	At-Large	23	22	34	Wright St.	Auto	57	59	61
TCU	At-Large	24	31	28	SFA	Auto	58	57	57
Texas A&M	At-Large	25	23	41	Lipscomb	Auto	59	61	58
Arkansas	At-Large	26	25	26	Georgia St.	Auto	60	63	64
Nevada	At-Large	27	24	22	Cal St. Fullerton	Auto	61	62	62
Rhode Island	At-Large	28	29	12	Iona	Auto	62	64	63
Seton Hall	At-Large	29	34	39	UMBC	Auto	63	60	59
Creighton	At-Large	30	39	31	Penn	Auto	64	58	55
Virginia Tech	At-Large	31	30	32	Radford	Auto	65	65	65
Missouri	At-Large	32	36	——	LIU Brooklyn	Auto	66	67	67
Florida St.	At-Large	33	37	45	NC Central	Auto	67	68	68
Kansas St.	At-Large	34	27	25	Texas Southern	Auto	68	66	66

Notes: The LP Quad. Expansion solution includes (ranks) Southern California (#32) and Middle Tenn. (#42). The LP Non-Quad solution includes (ranks) Middle Tenn. (#29), Saint Mary's (CA) (#35), Nebraska (#36), Louisville (#38), Mississippi St. (#40), Southern California (#44), Utah (#46).

Table 5

Non-parametric correlations between the NCAA Selection Committee's 68-team ranking and the implied ranking using the LP Quadrant Expansion and Non-quadrants weights

Correlation Measure (N = 68 teams)		NCAA Rank	Quadrant Expansion	Non-Quadrant
Kendall's tau	NCAA Rank	1.000		
	Quadrant Expansion	0.892	1.000	
	Non-Quadrant	0.731	0.752	1.000
Spearman's rho	NCAA Rank	1.000		
	Quadrant Expansion	0.981	1.000	
	Non-Quadrant	0.901	0.911	1.000

are statistically significant ($p \leq 0.978$), and four of the estimates have the "wrong" negative sign.

The logit approach fails, at least in this particular application and we suspect falls short in some of its earlier applications as well, because of a serious collinearity problem that manifests itself in two ways:

notably, the coefficient estimates are highly unstable and shift in both sign and order of magnitude depending upon which factors are included in the equation. Thus, each of the ten binary-logit regressions run with just one factor included as an independent variable results in a "correctly-signed" parameter estimate,

eight of which are highly significant ($p \leq 0.007$), with two only somewhat less-so (two-tail p-values: $0.050 < p < 0.093$); one-tail p-value : $p < 0.05$). Further, none of the estimates for the basic six factors exceed 0.104, while the estimates for the four Quadrant metrics range from 0.597 to 0.934. By contrast, with all ten factors in the equation, the parameter estimates vary in absolute magnitude from a low of 0.261 to a high of 114.727 and as mentioned above, four have the wrong sign. Assuredly, none of this makes it a hopeless problem, so it becomes incumbent upon the analyst – perhaps abetted by a stepwise procedure - to decide when to stop either eliminating or adding factors and to live with the implied collinearity. The problem is scarcely unique to the present application, although it would seem to be especially vexing in this application given the presence of many performance measures all trending in the same direction.

Rather than pursue variable reduction, or propose a model with classic symptoms of multicollinearity, we take a linear programming approach that allows us to maintain all available variables and assign interpretable weights which are necessarily in the "right" direction. Doing so comes at the cost of forgoing tests of statistical significance for each coefficient, however, we analyze the resulting solution by the degree to which it accurately mirrors that of the committee and use that similarity as basis for testing for statistical significance.

5. Consistency in the selection committee's ranking

The genesis for the MP approach was a paper by Becker et al. (1997), which used LP to determine a set of weights in a weighted-average evaluation scheme that would produce aggregate ratings for, in the present context, a set of 348 eligible NCAA Tournament candidates such that one team would rank first and another 348th. Zappe et al (1993) extended this approach by applying it to Bill James' ranking of the 10 greatest Major League Baseball players at each of the eight regular positions (James, 1988). Their intention was to see whether a set of weights exists that is consistent with James' ranking, which in fact was the case, implying that James was consistent in his ranking process. It is now our intention to see whether there exists a set of weights for the ten invitation-determining factors that produces a ranking consistent with the Selection Committee's

ranking of its 68 invitees. Specifically, we attempted to determine a set of weights that rates Virginia first, Villanova second, Kansas third, and so forth, as seen in Table 4, based on the set of 67 constraints of the form $V_k - V_j = \sum(w_i (x_{ki} - x_{ji}) = S_{kj} \geq 0$, where $i = k + 1$. Thus, the first constraint states that the overall evaluation of Virginia is at least as great as the evaluation of Villanova, and the second states that the evaluation of Villanova is at least great as the evaluation of Kansas, and so forth. There is no feasible solution to this problem!

Indeed, we invoked several different formulations of the objective function and were repeatedly rebuffed by what we suspect to be some combination of the members failing to employ a systematic overall evaluation scheme, their considering factors other than those for which the data are provided on the Team Sheets, and the members liberal use of the eye-ball test and their in-season observations from watching specific teams to which they'd been assigned.

For further insight, we eliminated the non-negativity part of the constraint and re-ran the program using a maximin objective function, since with negative S_{kj}'s we want to minimize the extent to which the constraints are violated. The solution that we obtained produced zero weights for Y_1 and Y_3, and only three weights in excess of 0.12 – those for Y_5, Y_6, and Y_7. Thirty-seven of the "slack" variables are positive, which implies consistency with the Committee's overall ranking of the teams in those particular paired comparisons, and 30 are negative, which belies the relative ranking of that particular pair. While acknowledging consistency to be the hobgoblin of little minds, we nonetheless confess to finding the lack thereof to be troubling where tournament ranks and the general public's confidence in the ranking process, are concerned.

6. Conclusions

Our primary goal in undertaking this project was to determine the impact of the introduction of the NCAA's new Quadrant metric on the selections and ranking of the teams invited to participate in the 2018 Men's Basketball Tournament - when the decision-making process is guided by the mathematical programming approached proffered by Reinig and Horowitz (2018). An important secondary goal was to infer its impact on the Selection Committee's decisions. We believe we have accomplished both goals.

In the former regard, we have demonstrated that the model's rankings were dramatically altered when the metric was incorporated into the analysis. In the latter regard we demonstrated that those rankings were much more in sync with those of the committee than they had been in its absence, which suggested it did indeed impact the Committee's decisions.

In a related vein, we showed that there is no set of (linear) factor weights that can produce the Committee's rankings, because of a combination of human frailty and the fact that some basketball teams are better than others in some aspects of the game, or at least in some quantified performance aspects of the game, and worse in others. This demonstration reinforces our confidence in the mathematical approach to the ranking process, which is based on the establishment of the dominant relationships that obtain among the teams, such that one team is said to be definitively superior to another, only when it performs at least as well as the other in all aspects of the game that enter into the overall-performance-evaluation process. As an unintended bonus, we hope to have convinced others as well.

References

Becker, R.A., Derby, L, McGill, R. & Wilks, A.R., 1987, 'Analysis of data from the places rated almanac', *American Statistician*, *41*(3), 169-186.

Coleman, B.J. & Lynch, A.K., 2001, Identifying the NCAA tournament dance card, *INTERFACES*, *31*(3), 76-86.

Coleman, B.J., Dumond, J.M. & Lynch, A.K., 2016, An easily implemented and accurate model for predicting NCAA tournament at-large bids, *Journal of Sports Analytics*, *2*(2), 121-132.

Coleman, B.J., Dumond, J.M., & Lynch, A.K., 2010, Evidence of bias in NCAA tournament selection and seeding, *Managerial and Decision Economics*, *31*(7), 431-452.

Coleman, B.J. & Lynch, A.K., 2009, NCAA tournament games: The real nitty-gritty, *Journal of Quantitative Analysis in Sports*, *5*(3), Article 8: doi: 10.2202/1559-0410.1165

James, B., 1988, The Bill James Historical Baseball Abstract. (Villard Books, NY).

Leman, S.C., House, L., Szarka, J. & Nelson, H., 2014, Life on the bubble: Who's in and who's out of March madness? *Journal of Quantitative Analysis in Sports*, *10*(3), 315-328.

Loomba, N.P. & Turbin, E., 1974, Applied Programming for Management. (Holt, Rinehart, & Winston, NY).

Paul, R.J. & Wilson, M., 2015, Political correctness, selection bias, and the NCAA basketball tournament, *Journal of Sports Economics*, *16*(2), 201-213.

Reinig, B.A. & Horowitz, I., 2018, Using mathematical programming to select and seed teams for the NCAA tournament, *INTERFACES*, *48*(3), 181-188.

Shapiro, S.L., Drayer, J., Dwyer, B. & Morse, A.L., 2009, Punching a ticket to the big dance: A critical analysis of at-large selection into the NCAA Division I men's basketball tournament, *Journal of Issues in Intercollegiate Athletics*, 2(March), 46-63.

Zappe. C., Webster, W., & Horowitz, I., 1993, Using linear programming to determine post-facto consistency in performance evaluations of Major League Baseball players, *INTERFACES*, *23*(6), 107-113.