E C O N O M I C S   B U L L E T I N

A Cheap Ticket to the Dance: Systematic Bias in College Basketball's Ratings Percentage Index

Shane Sanders
Assistant Professor, Auburn University Montgomery

Abstract

A contest model is constructed to examine the existence of conference bias in college basketball's Ratings Percentage Index (RPI). Though a general RPI bias has been identified in previous literature, this is the first study to address whether the bias is random or systematic in nature. Within the theoretical model, the RPI is shown to be systematically biased against teams in high ability conferences, even when all teams play to expectation and can be transitively compared. Further, the bias can prevent the RPI from producing an ordinal mapping from revealed team ability level to the real number line. Given the longevity of the controversial RPI as the NCAA's primary measure of team ability, these results may indicate that the NCAA is serving a demand for team heterogeneity in selecting for the NCAA Men's Basketball Tournament.

I would like to thank Yang-Ming Chang, Dong Li, Bhavneet Walia, and Dennis L. Weisman for helpful comments. Also, thank you to Tomas Sjostrom and an anonymous referee for invaluable suggestions. The usual caveat applies.
Citation: Sanders, Shane, (2007) "A Cheap Ticket to the Dance: Systematic Bias in College Basketball's Ratings Percentage Index." *Economics Bulletin*, Vol. 4, No. 34 pp. 1-7
Submitted: August 26, 2007. **Accepted:** September 12, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume4/EB-07D80023A.pdf

1. Introduction

The Ratings Percentage Index (RPI) is the prominent measure of team ability level in NCAA Division I Men's Basketball. The RPI essentially uses information from a team's revealed performance to numerically describe that team's ability. Since 1981, the RPI has aided the NCAA Division I Men's Basketball Committee in selecting and seeding NCAA tournament teams. Thus, the index is generally valued for its ability to comprehensively and unambiguously rank college basketball teams. Given the enormous monetary and emotional value of a marginal NCAA Tournament berth, it is important to understand the methodology of the RPI and determine whether it is systematically biased.

Important methodological distinctions exist between the RPI and team ranking systems in other sports. Therefore, we cannot develop a satisfactory characterization of the RPI based on past studies of other ranking systems. The RPI as a subject of study has heretofore been the domain of college basketball analysts and a few empirical researchers. With varying degrees of underlying thought, college basketball analysts have devoted many words toward characterizing the RPI. We will invoke some of their observations to motivate the model. In an empirical study, Harville (2003) finds that the RPI is a biased predictor of team performance in the post-season. Further, he shows that a least-squares approach performs better than the RPI in predicting post-season tournament outcomes. West (2007) outlines previously established methods for rating NCAA basketball teams. He shows ratings based on logistic regression to have a closer fit with observed tournament data than simulations based on a Bradley-Terry approach. Despite these significant findings, previous literature falls short of identifying the specific nature of the RPI's bias. For instance, is the bias random or systematic in nature? This is an important question, as a long-standing, systematic bias might suggest that the NCAA uses the measure to promote a particular post-season agenda. In this paper, we identify a potential source of systematic RPI bias. Further, we explore which type of team is hurt and which is helped by such a bias.

The RPI value for a given team is a weighted average of that team's winning proportion (.25), the average winning proportion of the team's opponents (.5), and the average winning proportion of the team's opponents' opponents (.25).[1] By measuring the latter two winning proportions, the RPI is commonly believed to control for a team's strength of schedule. Noted college basketball analyst Jerry Palm states, "(The RPI) is a measure of strength of schedule and how a team does against that schedule" (Keri 2007). However, the measure is problematic in that a large proportion of Division I college basketball games take place between conference opponents. As conferences are essentially hierarchical groups within college basketball, we expect a strong positive correlation between a team's ability level, the ability level of a team's opponents, and the ability level of a team's opponent's opponents. Thus, it likely requires more ability for a team in a major conference to achieve a particular season winning proportion than a team in a mid-major conference.[2] Similarly, we expect that it requires more ability, on average,

[1] The NCAA Men's Basketball Committee has recently adjusted the RPI such that away wins are more valuable than home wins (NCAA).

[2] There are two conference tiers in college basketball that are relevant to the NCAA tournament discussion- major (i.e., first tier) and mid-major (i.e., second tier). Though the distinction is unofficial, it is commonly accepted that six of the 32 Division I conferences are "majors," while the remainder are "mid-majors" or "low-majors." Despite their relative scarcity, majors have earned 97 of 108 Tournament Final Four spots during the RPI era.

for a major conference team's opponents to achieve a particular season winning proportion than a mid-major conference team's opponents. Lastly, given the effect of the conference season, we expect that it requires more ability for a major conference team's opponents' opponents to achieve a particular season winning proportion than a mid-major conference team's opponents' opponents. Analyst Jon Scott (2007) addresses this point:

> "The major reason the RPI is a poor model for determining team strength is because it is too simplistic to reliably differentiate teams and relies completely on the assumption that winning percentage is a valid indicator of how strong a team is. Comparing only the won-loss percentage of the last place team of a power conference with the won-loss percentage of a low-level conference champion, with no regard for the schedule each school actually played, one would be completely misled as to which team was the stronger."

If the RPI is conference biased, we might expect a mid-major conference team that is strong relative to its conference opponents to sometimes achieve a higher RPI than a major conference team that is mediocre or weak relative to its conference opponents, *even when the major conference team has more ability and has exhibited this in games against strong mid-major conference teams*. The possibility of a positive "mid-major bias" in the RPI has implications upon the merit of NCAA tournament selection. Though mid-major teams are slightly outnumbered in the NCAA Tournament, the NCAA Division I Men's Basketball Committee may be stocking the Tournament with more mid-major conference teams than are merited.[3]

In this paper, we construct a contest model of a college basketball season. We assume there are four team-types within college basketball, where a team-type is defined as a group of teams sharing the same ability level. A team's expected likelihood of victory in a game is determined by its ability level relative to the opponent's ability level. That is, if a team of type i were to play a team of type j, the probabilities of victory would be given respectively by

$$p_i = \frac{t_i}{t_i + t_j} \quad \text{and} \quad p_j = \frac{t_j}{t_i + t_j}, \text{ where } t_{i(j)} = \text{ability level for team-type } i(j).^4 \qquad (1)$$

We focus on the ranking of two college basketball teams, each of a distinct type. Specifically, we wish to ascertain whether a more talented team might earn a *lower* RPI due to the nature of its conference schedule. If every team type plays to its expected level and all team types can be transitively compared during the course of a season, an unbiased measure of team ability would always come up with a correct ordinal ranking of team-types. However, we find that, even under such conditions, the RPI does not necessarily represent an ordinal mapping from revealed team ability level to the real number line. The RPI is flawed even in the face of perfect information in that it systematically undervalues the ability level of some team types and overvalues the ability level of other team types. In expectation, the measure rewards top teams

[3] The Men's Basketball Committee selects 34 of the 65 tournament teams. The remaining 31 teams enter through automatic berths. The Committee also has the responsibility of seeding all tournament teams.

[4] As in Amegashie and Kutsoati (2005), the equations in (1) represent additive form contest success functions. For alternative forms, see, e.g., Tullock (1980) or Hirshleifer (1989).

in low-ability conferences at the expense of mediocre-to-bottom teams in high-ability conferences.

2. The Model

In this simple model of a college basketball season, we assume that there are four types of teams, where a team-type is a set of teams sharing the same ability level. We consider the ranking of two teams, each of a distinct type. Specifically, Team $i(\in(2,3))$ is of type *i*. Each team inhabits a distinct conference. Team 2 inhabits a conference in which all teams are of type 1 or 2. Team 3 resides in a conference in which all teams are of type 3 or 4. Given a cursory examination as to the relative magnitude of conference schedules, we assume that half of a team's games are against conference opponents, and remaining games are inter-conference in nature. Thus, a team from one type might play an opponent of the same type or an opponent of a different type. For simplicity, we assume that each team within a type plays the same types of opponents with the same relative frequency. Let the ability level differ across types such that

$$t_1 > t_2 > t_3 > t_4 . \tag{2}$$

Further, assume that teams of type 1 play two-thirds of games against other type 1 teams and remaining games against type 2 teams. Also, type 2 teams play type 1, 2, and 3 opponents with equal frequency such that their conference schedule is identical to that of a type 1 team. Similarly, type 3 teams play type 2, 3, and 4 opponents with equal frequency. Lastly, type 4 teams play two-thirds of games against other type 4 teams and remaining games against type 3 teams such that their conference schedule is identical to that of a type 3 team.

Thus, our model allows for the existence of heterogeneous schedules across teams, conference play, and inter-conference play. The implication of these distributional assumptions is that conference play causes a team to remain in the neighborhood of its own type. This is not the only scheduling distribution one could consider but is useful and sufficiently plausible for a general evaluation of the RPI. Given these assumptions, we can tractably examine the RPI's ability to interpret the heterogeneous experience of each team type in a setting where teams perform to expectation and can be transitively compared. Given these scheduling assumptions, the expected winning proportion for each team-type is as follows:

$$E(w_1) = \frac{\left[2\left(\frac{t_1}{t_1+t_1}\right) + \frac{t_1}{t_1+t_2}\right]}{3} = \frac{\left[1 + \frac{t_1}{t_1+t_2}\right]}{3}; \quad E(w_2) = \frac{\left[\frac{1}{2} + \frac{t_2}{t_1+t_2} + \frac{t_2}{t_2+t_3}\right]}{3} \tag{3a}$$

$$E(w_3) = \frac{\left[\frac{1}{2} + \frac{t_3}{t_2+t_3} + \frac{t_3}{t_3+t_4}\right]}{3}; \qquad E(w_4) = \frac{\left[1 + \frac{t_4}{t_3+t_4}\right]}{3} \tag{3b}$$

From the expected winning proportions in (3), we can calculate the expected RPI for Teams 2 and 3, respectively, as follows

$$E(RPI_2) = 0.25(E(w_2)) + 0.5\left(\frac{E(w_1) + E(w_2) + E(w_3)}{3}\right) + 0.25\left(\frac{3E(w_1) + 3E(w_2) + 2E(w_3) + E(w_4)}{9}\right)$$

$$E(RPI_2) = \frac{E(w_1)}{4} + \frac{E(w_2)}{2} + \frac{2E(w_3)}{9} + \frac{E(w_4)}{36}. \quad \text{[5]} \tag{4a}$$

Similarly,

$$E(RPI_3) = \frac{E(w_1)}{36} + \frac{2E(w_2)}{9} + \frac{E(w_3)}{2} + \frac{E(w_4)}{4}. \tag{4b}$$

Subtracting (4b) from (4a), we can determine whether Team 2's expected RPI is unambiguously greater.

$$[E(RPI_2) - E(RPI_3)] = \frac{\left[\frac{8t_1}{t_1 + t_2} + \frac{10t_2}{t_1 + t_2} + \frac{10t_2}{t_2 + t_3} - \frac{10t_3}{t_2 + t_3} - \frac{10t_3}{t_3 + t_4} - \frac{8t_4}{t_3 + t_4}\right]}{108} \tag{5}$$

By virtue of each team's defined talent level, we know the following:

$$\frac{8t_1}{t_1 + t_2} > \frac{8t_4}{t_3 + t_4}; \quad \frac{10t_2}{t_2 + t_3} > \frac{10t_3}{t_2 + t_3}; \text{ but } \frac{10t_2}{t_1 + t_2} < \frac{10t_3}{t_3 + t_4}.$$

For $E(RPI_2) > E(RPI_3)$, the first two inequalities must dominate the third. However, we cannot be certain that this is the case. This brings us to the following proposition:

Proposition 1: *It is indeterminate, a priori, whether $E(RPI_2) > E(RPI_3)$ (i.e., team 2 outranks team 3 in RPI), despite the fact that (i) $t_2 > t_3$, (ii) teams play to expectation, and (iii) information exists concerning relative performance in the intersecting portion of these two teams' schedules.*

[5] Please see Appendix A for a more detailed calculation.

More generally, it is clear that

$$\frac{\left[E\left(RPI_2\right)-E\left(RPI_3\right)\right]}{E\left(RPI_3\right)}<\left(\frac{t_2-t_3}{t_3}\right) \quad . \tag{6}$$

Inequality (6), which is derived in Appendix B, leads us to the following proposition:

Proposition 2: *As a measure of team ability, the RPI value always underestimates team 2's ability in relation to team 3.*

From these two propositions, we conclude that the RPI pulls team-types 2 and 3 relatively closer to one another than would a meritorious measure. In some cases, this bias is sufficient to cause team-type 3 to achieve a higher RPI ranking than team-type 2.

3. Simulation

This section presents a counter-example that reveals, under reasonable conditions, the indeterminacy found in Proposition 1. The example is based on the same scheduling assumptions as those made in the previous section.

Let $t_1 = 10$, $t_2 = 6$, $t_3 = 5.5$, $t_4 = 3$. Given our additive form contest success function, the implications of these attributed talent levels are as follows:

$$\frac{t_1}{t_1+t_2}=\frac{10}{16}=.625 \qquad \frac{t_1}{t_1+t_3}=\frac{10}{15.5}\approx .645 \qquad \frac{t_1}{t_1+t_4}=\frac{10}{13}\approx .769$$

$$\frac{t_2}{t_2+t_3}=\frac{6}{11.5}\approx .522 \qquad \frac{t_2}{t_2+t_4}=\frac{6}{9}\approx .667 \qquad \frac{t_3}{t_3+t_4}=\frac{5.5}{8.5}\approx .647$$

In other words, these values appear collectively plausible within our simple model of a college basketball season.

Then, $E\left(RPI_2\right)\approx .5011$ and $E\left(RPI_3\right)\approx .5021$. In this case, $E\left(RPI_2\right)< E\left(RPI_3\right)$. In Appendix C, the reader will find a contingency table that examines relative RPI values for team-types two and three given different values for t_3 (i.e., the output of this simulation compared with that of three others). Given the scheduling assumptions of the model, we find from Appendix C that $E\left(RPI_2\right)< E\left(RPI_3\right)$ as long as t_3 is sufficiently close to t_2. As expected, Appendix C also shows that the bias found in Proposition 2 is present no matter the simulated value of t_3.

4. Stand By Your Measure: The NCAA's Continued Use of the RPI

Given the longstanding controversy surrounding the RPI's validity, it is fair to question the NCAA's motive in creating and standing behind this measure since 1981. It is not a stretch to posit that the NCAA does not hold a national tournament to determine the best team in the country. The deep field and single elimination structure of the NCAA Tournament makes this point clear. If the Tournament were meant to be a sort of hypothesis test on each of the 65 teams in the field, with the null hypothesis being that a particular team is not the nation's best, the likelihood of Type II error for most teams is almost negligible. That is to say, if the NCAA's purpose in holding a tournament were to determine the nation's most talented team, it would be better served to create a multiple-elimination or series structure featuring a small number of contending teams (i.e., evaluate candidate teams relative to one another based on a sample size larger than one).

What, then, is the purpose of the NCAA Men's Basketball Tournament from the perspective of the NCAA? Television revenues from the Tournament go to NCAA member schools. Further, the NCAA is comprised primarily of representatives from member schools. Hence, we can assume that the NCAA functions as a profit-maximizing firm (on behalf of member schools in this case) in making Tournament decisions. As such, the NCAA will choose the Tournament structure that maximizes fan interest and excitement, *ceteris paribus*. It is likely that the current Tournament structure achieves this goal because it gives the progression of games an exciting, sudden death feeling, and it also allows more teams to enter the field. The latter fact encourages the Cinderella story (or the David and Goliath story) to develop itself and lend a touch of drama to Tournament contests. How does the RPI fit into all of this? The RPI may also perpetuate the David and Goliath story within the NCAA tournament by allowing more relatively unknown mid-major teams to enter the field. The drama generated from this team heterogeneity within the tournament field may be sufficiently valuable, in terms of fan interest, to cause the Selection Committee to discount team ability in certain marginal cases.

We can consider the NCAA's Tournament selection decision when fans care only about talent as compared to when fans care about both talent and team heterogeneity. Let $t = \left(h - h^{\alpha}\right)$, where t is the talent level of the Tournament field , h represents the degree to which non-major conference teams are represented in the Tournament, and $\alpha(>1)$ is a parameter that helps determine the tradeoff between talent and heterogeneity when selecting a Tournament field. If the Selection Committee chooses exclusively major conference teams, it will not optimize talent level for the Tournament. After all, several mid-major conference teams *are* deserving of Tournament status. Further, if the Selection Committee chooses "too many" mid-major teams, it does not maximize the talent level of the Tournament because it is crowding qualified major conference teams out of the field. According to the above equation, talent level for the Tournament is maximized at a point where there is some mid-major team representation.
If the NCAA wished simply to maximize talent level when selecting the Tournament field (i.e., fans care only about talent level), their problem would look as follows:

$$\underset{h}{Max} \quad t = \left(h - h^{\alpha}\right) \text{ where } t \geq 0,\ h \geq 0,\ \alpha > 1.$$

This objective function leads us to the following selection allocations:

$$h^* = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} \qquad t^* = \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha-1}} - \left(\frac{1}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} \tag{7}$$

Given that $\alpha > 1$ both allocations in (7) must be greater than zero.

If fan interest in the Tournament is determined both by talent level and heterogeneity, then the Committee's problem might appear as follows:

$$\underset{h}{Max}\ \ th = \left(h - h^{\alpha}\right)h\text{, where } t \geq 0,\ h \geq 0.$$

This functional form implies that the Selection Committee chooses the field that maximizes the product of talent and heterogeneity. Thus, fans value a Tournament with more talent or one with more heterogeneity, *ceteris paribus*. However, fans have no interest in the Tournament if teams are completely untalented or completely homogeneous with respect to conference origin. Fans might view a completely homogeneous tournament as an iteration upon the regular season, for instance. Note that the NCAA Tournament Committee is selecting talent from a willing pool rather than hiring it. The Committee need not pay a wage in selecting talent. The NCAA does pay an implicit cost in selecting talent however due to the tradeoff between talent and heterogeneity. This objective function leads us to the following selection allocations:

$$h^{**} = \left(\frac{2}{\alpha+1}\right)^{\frac{1}{\alpha-1}} \qquad t^{**} = \left(\frac{2}{\alpha+1}\right)^{\frac{1}{\alpha-1}} - \left(\frac{2}{\alpha+1}\right)^{\frac{\alpha}{\alpha-1}} \tag{8}$$

Again, both allocations must be greater than zero. Comparing (8) to (7), we find that $h^{**} > h^*$ and $t^{**} < t^*$. As long as fans have a taste for heterogeneity, the NCAA will choose a level of Tournament heterogeneity larger than that which maximizes talent level. Thus, the RPI, with its systematic bias, may well be a subtle tool calibrated to produce a "fan's tournament" while allowing the NCAA to maintain credibility. As has been shown in the case of the NBA's white player premium (see Burdekin et al. 2005), talent is not the only taste entertained by sports fans, and sporting outcomes are partly contingent upon these peripheral tastes.

5. Conclusion

Within the model, we have shown that the RPI is not only biased, but systematically so. The measure punishes teams according to the ability level of their conference such that type 2 teams and type 3 teams are pulled relatively closer to one another. This effect may be sufficient to cause team-type 3 to achieve a higher RPI ranking than team-type 2, even when team-type two is more talented, teams play to expectation, and each team can be transitively compared across schedules. The presence of a long-standing, systematic bias, even in the face of controversy, may suggest an underlying NCAA agenda. Indeed, the paper shows that, when fans value heterogeneity, the NCAA must choose less talent and more team heterogeneity than a meritorious selection process would feature in order to maximize fan interest. As the RPI is only one in a set of Tournament selection tools, a future study might ascertain to what degree selected Tournament fields reinforce systematic biases within the RPI.

6. Appendix Section

6.1 Appendix A

A detailed calculation of $E(RPI_2)$:

$$E(RPI_2) = 0.25(E(w_2)) + 0.5\left(\frac{E(w_1) + E(w_2) + E(w_3)}{3}\right) + 0.25\left(\frac{3E(w_1) + 3E(w_2) + 2E(w_3) + E(w_4)}{9}\right)$$

In other words, the RPI value for team-type two is a weighted average of its average winning proportion (0.25), the average winning proportion of its opponents (0.5), and the average winning proportion of its opponents' opponents (0.25). This calculation leads us to the following equation

$$E(RPI_2) = \frac{E(w_1)}{4} + \frac{E(w_2)}{2} + \frac{2E(w_3)}{9} + \frac{E(w_4)}{36}.$$

6.2 Appendix B

Show that $\frac{E(RPI_2) - E(RPI_3)}{E(RPI_3)} < \frac{t_2 - t_3}{t_3}$.

$$\frac{E(RPI_2) - E(RPI_3)}{E(RPI_3)} = \frac{E(RPI_2)}{E(RPI_3)} - 1 = \left(\frac{\left(\frac{10t_2}{(t_2 + t_3)} + 8 + \frac{2t_2}{t_1 + t_2}\right)}{\left(\frac{10t_3}{(t_2 + t_3)} + 8 + \frac{2t_3}{t_3 + t_4}\right)} - 1\right)$$

and

$$\frac{t_2}{t_2 - t_3} = \left(\frac{t_2}{t_3} - 1\right) = \left(\frac{\left(\frac{10t_2}{(t_2 + t_3)}\right)}{\left(\frac{10t_3}{(t_2 + t_3)}\right)} - 1\right)$$

$$\left(\frac{\left(\frac{10t_2}{(t_2 + t_3)} + 8 + \frac{2t_2}{t_1 + t_2}\right)}{\left(\frac{10t_3}{(t_2 + t_3)} + 8 + \frac{2t_3}{t_3 + t_4}\right)} - 1\right) < \left(\frac{\left(\frac{10t_2}{(t_2 + t_3)}\right)}{\left(\frac{10t_3}{(t_2 + t_3)}\right)} - 1\right) \quad \text{by } t_2 > t_3 \text{ and } \frac{t_2}{t_1 + t_2} < \frac{t_3}{t_3 + t_4}.$$

$$\therefore \quad \frac{E(RPI_2) - E(RPI_3)}{E(RPI_3)} < \frac{t_2 - t_3}{t_3}.$$

6.3 Appendix C

Four simulations under the schedule assumed in the model:

(t_1, t_2, t_3, t_4)	$E(RPI_2)$	$E(RPI_3)$	$\frac{[E(RPI_2) - E(RPI_3)]}{E(RPI_3)} < \left(\frac{t_2 - t_3}{t_3}\right)$
(10, 6, 5.75, 3)	.5007	.5040	Yes
(10, 6, 5.50, 3)	.5011	.5021	Yes
(10, 6, 5.25, 3)	.5015	.5001	Yes
(10, 6, 5.00, 3)	.5019	.4981	Yes

7. References

Amegashie, J. and E. Kutsoati (2005) “Rematches in Boxing and Other Sporting Events” *Journal of Sports Economics* **6**, 401-411.

Burdekin, R., Hossfeld, R., and J. Smith (2005) “Are Fans Becoming Indifferent to Race? Evidence From the 1990s” *Journal of Sports Economics* **6**, 144-159.

Harville, D. (2003) “The Selection or Seeding of College Basketball or Football Teams for Postseason Competition” *Journal of the American Statistical Association* **98**, 17-27.

Hirshleifer, J. (1989) “Conflict and rent-Seeking success functions: ratio vs. difference models of relative success” *Public Choice* **63**, 101-112.

Keri, J. (2007) “Pac-10 emerging as toughest conference” *The New York Sun*. Retrieved April 6, 2007, from the World Wide Web: http://www.nysun.com/article/46197

National Collegiate Athletic Association (2006, October 5) “Tournament History” Retrieved April 12, 2007, from the World Wide Web: http://www.ncaasports.com/basketball/mens/story/9033549

Scott, J. (2007) “The ratings percentage index myth.” Retrieved May 22, 2007 from the World Wide Web: http://www.bigbluehistory.net/bb/rpi.html

Tullock, G. (1980) “Efficient rent seeking” in *Toward A Theory of the Rent-Seeking Society* by J. M. Buchanan, R.D. Tollison and G. Tullock (Eds.) College Station: Texas A&M University Press, 97-112.

West, B. (2006) “A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament” *Journal of Quantitative Analysis in Sports* **2**, http://www.bepress.com/jqas/vol2/iss3/3