Insights from the LRMC Method for NCAA Tournament Prediction

Mark Brown
City College, City University of New York,
New York, NY, USA 10031
Email: cybergarf@aol.com

Paul Kvam, George Nemhauser, Joel Sokol[1]
Georgia Institute of Technology,
Atlanta, GA, USA 30332-0205
Email: {pkvam,gnemhaus,jsokol}@isye.gatech.edu

Abstract

We compare the ability of over 100 different procedures for ranking NCAA basketball teams to correctly predict outcomes of NCAA Tournament games. Over the nine-year period for which we have data, the Bayesian LRMC method appears to be the best in terms of number of games correctly predicted, and the difference is statistically significant in almost all cases. We also use LRMC analysis to develop insights into NCAA basketball. We demonstrate that contrary to popular opinion, home court advantage is statistically almost the same across all teams except perhaps Denver. We use home-and-home data to show that teams that win close games are not much better than teams that lose close games (i.e., good teams don't win more than their share of close games). Finally, we use the Bayesian LRMC model to resolve the apparent paradox that home-court advantage is only about 4 points, yet 20-point home winners are more than 50% likely to lose to the same opponent on the road, and even 50-point home winners often lose to the same opponent on the road.

1 Introduction

The National Collegiate Athletic Association Division I men's basketball tournament (the "NCAA Tournament"), a 68-team single-elimination tournament played each year to determine the national champion at the end of the season, is the most wagered-upon sporting event in the United States. In [1], we introduced the LRMC method for predicting winners of NCAA Tournament games. LRMC is a two-stage process. First, for every game played between two Division I teams that season prior to the NCAA Tournament, a logistic regression model is used to estimate the likelihood that the winner is a better team than the loser. Those likelihoods are then used to develop transition probabilities in a Markov chain whose steady-state probabilities imply an overall ranking of college basketball teams. Since NCAA Tournament games are played on neutral courts, LRMC assumes that the higher-ranked team will win each game. [1] demonstrates that LRMC outperforms the most well-known rankings at predicting the outcomes of individual NCAA Tournament games (as well as in bracket-style scoring). In [2], we demonstrated that LRMC can be improved by using an empirical Bayes model in place of the logistic regression.

Both [1] and [2] concentrated on the mathematical details of the models; [1] compared LRMC's performance with just a few of the other most well-known rankings, and [2] primarily compared the improved version of LRMC with the original. In Section 2 of this paper, we compare over 100 different rankings in terms of their ability to predict outcomes of NCAA Tournament games. In the course of our work in [1] and [2], our models have also provided insight into commonly-discussed questions: which team has the largest home court advantage, whether good teams know how to win close games, and why home-and-home results in conference are often so surprising. In Section 3, we will discuss these insights in the context of our LRMC models.

2 Prediction quality

In this section, we make empirical comparisons between LRMC and over 100 other rankings on NCAA Tournament game data from 2003-2011 (a total of 579 games). Since 2003, [3] has recorded dozens of pre-NCAA Tournament rankings of Division I basketball teams, including polls, algorithmic rankings, expert rankings, etc. Using [3] and other data sources, we made retrospective comparisons of prediction quality between the LRMC methods and over 100 others for which data was available. The rankings and data sources are listed in the Appendix.

Our comparisons were done game-by-game: for each NCAA Tournament game played from 2003-2011, a ranking earned one point if it ranked the game's winner higher than the game's loser, and half a point if its rankings of the game's winner and loser were equal. Teams that were not ranked (e.g., teams that received no votes in a poll) were considered equal, but below all ranked teams.

[1] Corresponding author

However, just looking at each ranking's raw fraction of points scored would be misleading, for two reasons. First, some rankings were not tracked in all years. The number of upsets varies significantly from year to year, so that the average number of correct predictions ranged from 50.3 out of 64 (79%) in 2007 to 42.8 out of 67 (64%) in 2011. So, a ranking untracked in a low-upset year would look worse than it should (having missed out on a year when it could've scored high) and a ranking untracked in a high-year would look better than it should.

Moreover, while most rankings are complete (i.e., they cover all 300+ teams), some are partial rankings with as few as just the top 25, guaranteeing that they will give no prediction for a number of NCAA Tournament games. As we show below, the winner of this type of game is harder than average to predict, so looking at the fraction of games predicted correctly will overstate the quality of partial rankings.

In order to deal with these comparison difficulties, we look at ranking quality in four different ways, each of which is designed to account for some of the difficulties. The first comparison is the simplest. For each year, we order the rankings by points scored: the ranking with the most points is in 1st place, etc. Where there are ties, the average is used. (For example, if 9 rankings tied for 9th-17th place, each would earn (9+17)/2=13th place.) However, when counting the number of times a ranking was 1st, 2nd, etc., all rankings get the benefit of the doubt (so all rankings tied for 9th-17th place would be counted as top-10). Table 1 shows the results of this comparison. The full table is shown in the Appendix; below, we list only the rankings that appear in the top-10 in at least two seasons. The Bayesian LRMC ranking looks best in this comparison, with 4 top-1, 5 top-2, and 6 top-5 and top-10 scores in 9 seasons, and an average score of 11.6. LRMC Classic is second-best, followed by Cheong, Pomeroy, and TSR.

Table 1. Rankings with at least two top-10 appearances, with total years tracked ("n") and top 1, 2, 5, and 10 finishes.

RANK-ING	**n**	**Top 1-2-5-10**	**Avg**	**RANK-ING**	**n**	**Top 1-2-5-10**	**Avg**	**RANK-ING**	**N**	**Top 1-2-5-10**	**Avg**	**RANK-ING**	**n**	**Top 1-2-5-10**	**Avg**
BLRMC	9	4-5-6-6	11.6	DWH	5	0-0-1-2	20.9	BOB	9	0-0-2-2	26.3	SEED	9	0-1-2-2	33.3
LRMCC	9	1-2-4-5	14.3	WIL	6	0-1-2-2	21.5	GRN	8	1-1-2-2	24.4	DC	7	0-0-1-2	26.3
CNG	8	0-0-2-4	15.8	VEG	9	0-0-3-3	21.7	ECK	5	0-0-1-2	24.7	KLK	5	0-0-1-2	27.0
POM	9	1-1-4-4	17.4	WLK	9	0-0-2-3	21.8	PRED	9	0-0-2-4	24.9	RPI	9	0-0-2-3	28.8
TSR	5	0-0-2-2	17.8	DES	6	0-0-0-3	22.5	LRMC0	9	2-3-3-4	25.1	USA	9	0-1-1-2	31.3
MOR	9	1-2-3-5	19.2	ELO	9	1-1-2-2	22.9	JEN	4	0-0-1-2	25.6	LYN	7	0-1-1-2	33.3
AVG	9	1-1-3-4	20.2	WOL	9	0-0-2-3	23.6	BIH	8	0-0-1-2	26.0				

For the next two comparisons, we scaled each ranking's point total in two ways. (1) For each year that a ranking was not tracked, its estimated performance is the highest point total that year minus the ranking's average shortfall from the highest total (over all years it was tracked). (2) For each year that a ranking was not tracked, its estimated performance is the average point total that year plus the ranking's average difference from the average point total (over all years it was tracked). These two methods have complementary strengths and weaknesses. In the first method, a ranking's score could be hurt if a single other ranking was high, but it would not be affected by the presence of many good or bad rankings in the data set. In the second method, the presence of many good or bad rankings could affect the score of a ranking, but it would not be affected much by a single other ranking's success.

For both of these methods, we report only results for rankings that were tracked in at least two of the nine years. When we included rankings tracked in just one of the nine years, we observed many paradoxes[2]. For example, AUS, PEQ, PTS, TRP, and TMR (each tracked only in 2011) would all have higher scaled point totals than POM (tracked all nine seasons), despite POM having predicted 2 or 3 more games correctly in the only year they were all tracked.

Table 2. Estimated performance (out of 579 games) scaled to each year's highest point total, 2003-11.

RANK-ING	**n**	**Pre**	**RANK-ING**	**n**	**Pre**	**RANK-ING**	**n**	**Pre**	**RANK-ING**	**n**	**Pre**	**RANK-ING**	**N**	**Pre**	**RANK-ING**	**n**	**Pre**
BLRMC	9	437	PKL	2	419	OMY	2	415	RTH	9	411	RPI	9	407	DUN	6	401
CNG	8	426	ROH	3	419	WLK	9	414	SEED	9	411	APE	9	407	STH	4	401
LRMCC	9	425	DES	6	418	SEL	7	414	CPA	3	410	HKB	2	406	MKV	3	398
KMV	2	424	AVG	9	417	DWH	5	414	BIH	8	410	REW	4	406	SAP	3	398
LYD	2	424	ISR	4	417	SAG	9	413	WOB	8	410	USA	9	405	IMS	2	397

[2] The famous impossibility theorem of [4] guarantees the possibility of paradoxes. Our comparisons fit into [4]'s voting theory framework by considering ranking systems as candidates and years as voters who cast partial preference ballots.

ACU	3	424	SPW	4	417	WOL	9	413	KPK	3	410	RTB	3	404	GRS	2	397
POM	9	421	DC2	3	416	BOB	9	413	MAS	5	408	BPI	3	404	STR	3	395
WIL	6	421	DCI	3	416	NOL	3	413	COL	9	408	LYN	7	404	ARG	3	395
MOR	9	420	VEG	9	416	DC	7	413	DOL	9	408	SCR	4	403	SIM	6	391
LRMC0	9	419	TSR	5	415	PRED	9	413	PGH	4	408	ECK	5	403	TRX	3	382
DOK	6	419	PIG	5	415	KRA	4	412	AP	9	408	RTR	5	403	CPR	3	380
JEN	4	419	KLK	5	415	MB	6	412	ENT	3	407	SAU	3	401	TW	2	379
BD	2	419	ELO	9	415	GRN	8	411	USAE	9	407	HER	3	401	PCT	9	367
REN	2	419	RT	2	415												

Table 2 shows the results of scaling to the top-scoring ranking of each year. Bayesian LRMC again scores at the top, with 437 correct predictions out of 579 (75.5%), followed by Cheong (73.5%) and LRMC Classic (73.4%). In contrast, [2] uses data on Las Vegas favorites to give a rough estimate of 76-77% as a reasonable upper bound on the predictive ability of any pre-tournament ranking system.

Table 3. Estimated (scaled) number of games above or below average (out of 579 games), 2003-11.

RANK-ING	**n**	**+/-**	**RANK-ING**	**n**	**+/-**	**RANK-ING**	**n**	**+/-**	**RANK-ING**	**n**	**+/-**	**RANK-ING**	**n**	**+/-**	**RANK-ING**	**n**	**+/-**
BLRMC	9	27	JEN	4	4	SAU	3	2	IMS	2	1	ECK	5	-1	APE	9	-3
LRMCC	9	15	SAG	9	3	CPA	3	1	SEED	9	1	PGH	4	-1	DUN	6	-4
CNG	8	11	WOL	9	3	BD	2	1	PKL	2	1	OMY	2	-2	STH	4	-4
POM	9	11	BOB	9	3	RTH	9	1	MKV	3	1	COL	9	-2	SCR	4	-5
MOR	9	10	ISR	4	3	ENT	3	1	ROH	3	0	DOL	9	-2	USA	9	-5
LRMC0	9	9	GRN	8	3	BIH	8	1	KRA	4	0	STR	3	-2	RTR	5	-6
AVG	9	7	KMV	2	3	WOB	8	1	GRS	2	0	AP	9	-2	LYN	7	-6
VEG	9	6	PRED	9	3	SEL	7	1	NOL	3	0	REW	4	-2	SAP	3	-7
ELO	9	5	DWH	5	3	DC2	3	1	HKB	2	0	USAE	9	-3	TW	2	-7
TSR	5	4	SPW	4	3	DCI	3	1	KPK	3	-1	RPI	9	-3	CPR	3	-9
WIL	6	4	LYD	2	2	REN	2	1	MB	6	-1	ARG	3	-3	TRX	3	-9
WLK	9	4	PIG	5	2	MAS	5	1	HER	3	-1	RTB	3	-3	SIM	6	-16
DOK	6	4	ACU	3	2	RT	2	1	DC	7	-1	BPI	3	-3	PCT	9	-43
DES	6	4	KLK	5	2												

Table 3 shows the estimated number of correct predictions above or below the average from 2003-11. Bayesian LRMC again has the top results, 27 games above average, followed by LRMC Classic (15 games), Cheong (11 games), Pomeroy (11 games), and Moore (10 games). In other words, someone using Bayesian LRMC to predict game winners would've made 27 more correct predictions than the average. On the other hand, using team winning percentage to predict game winners was 43 games below the average. This is an effect of having such disparate conference strengths (and more generally, schedule strengths) from team to team; at the request of a television network that broadcasts NBA playoff games we did the same analysis on the NBA (where every team plays each other, and schedule strength is much more balanced) and winning percentage was almost as good a predictor as LRMC.

The first three methods of comparison all have Bayesian LRMC, LRMC Classic, Cheong, and Pomeroy near the top. These methods are likely to benefit rankings that rank all 300+ teams each year, since games between two teams of equal rank gain the ranking system one-half point, equal to what one would expect by just flipping a coin for those games. (As we show below, that's approximately what we observe the success rate to be for the AP and USA Today polls, but it might underestimate the success rate of other methods if they were to extend to more teams.)

In order to compare only games where a prediction could be explicitly made from each ranking, we look at *disagreement sets*. For each pair of rankings, a game between team *i* and team *j* is in the pair of rankings' disagreement set if one of the two rankings has *i* higher than *j* and the other has *j* higher than *i*. We can then use a one-tailed McNemar's test to measure the significant level of one method's superiority over another. This is the same approach used in [1] and [2]. Since Bayesian LRMC appears to have the best predictive performance in the comparisons above, in this extended abstract we report only the McNemar's test results for Bayesian LRMC against all other methods.

Before dealing with partial rankings, we first present McNemar's test results for complete rankings. Table 4 shows the significance results for Bayesian LRMC over each complete ranking whose disagreement set with Bayesian LRMC contains at least 20 games. The table shows the number of seasons each ranking was tracked, the "win-loss" record of

Bayesian LRMC in the disagreement set, the fraction of games in the disagreement set, and the McNemar p-value. For example, the well-knows Sagarin Predictor rankings (PRED) were tracked in all nine years, and disagreed with Bayesian LRMC in 9% of all games (54 disagreements). In those 54 disagreements, the team Bayesian LRMC ranked higher won 39 times and the team PRED ranked higher won 15 times. The 39-15 "win-loss" record yields a McNemar p-value of 0.001, indicating that it is very likely Bayesian LRMC is a better predictor than PRED.

Table 4. Significance levels for Bayesian LRMC over each complete ranking[3] with at least 20 disagreements, 2003-11.

RANKING	n	W-L	Pct	pval	RANKING	n	W-L	Pct	pval	RANKING	n	W-L	Pct	pval	RANKING	n	W-L	Pct	pval
LYN	7	43-17	13%	.001	WLK	9	49-26	13%	.005	KLK	5	27-13	13%	.019	SAU	3	17-8	13%	.054
SIM	6	53-24	20%	.001	DC	7	53-29	18%	.005	PGH	4	31-16	18%	.020	SEL	7	40-26	15%	.054
PRED	9	39-15	9%	.001	BOB	9	55-31	15%	.006	CPA	3	15-5	10%	.021	CNG	8	26-15	8%	.059
DUN	6	51-24	19%	.001	GRN	8	47-25	14%	.006	MKV	3	15-5	10%	.021	WIL	6	42-28	18%	.060
DOL	9	59-30	15%	.001	RTR	5	41-21	19%	.008	PIG	5	36-20	17%	.022	VEG	9	38-25	11%	.065
CPR	3	32-12	23%	.002	KPK	3	23-9	16%	.010	POM	9	36-20	10%	.022	SPW	4	28-17	17%	.068
RPI	9	66-36	18%	.002	WOL	9	63-39	18%	.011	KRA	4	32-17	19%	.022	ROH	3	19-10	15%	.068
BIH	8	55-28	16%	.002	MAS	5	30-14	14%	.011	LRMCC	9	23-11	6%	.029	REN	2	16-8	19%	.076
SCR	4	33-13	18%	.002	SAP	3	30-14	23%	.011	JEN	4	17-7	9%	.032	DWH	5	32-21	17%	.084
COL	9	64-35	17%	.002	MB	6	47-27	19%	.013	STR	3	24-12	19%	.033	NOL	3	19-11	15%	.100
SAG	9	46-22	12%	.002	AVG	9	50-30	14%	.016	MOR	9	48-31	14%	.036	ISR	4	25-16	16%	.106
STH	4	28-10	15%	.003	ARG	3	17-6	12%	.017	RTB	3	22-11	17%	.040	TSR	5	33-23	18%	.114
WOB	8	60-33	18%	.003	DOK	6	30-15	12%	.018	BPI	3	23-12	18%	.045	KMV	2	15-9	18%	.154
RTH	9	57-31	15%	.004	REW	4	34-18	20%	.018	HER	3	16-7	12%	.047	DCI	3	21-14	18%	.155
ECK	5	36-16	16%	.004	ELO	9	62-40	18%	.019	LRMC0	9	61-43	18%	.048	PKL	2	15-10	20%	.212
TW	2	25-9	26%	.005															

As the results in Table 4 show, Bayesian LRMC is significantly better at predicting NCAA Tournament games than almost all of the other complete rankings with sufficient data. Among well-known ranking methods, Bayesian LRMC appears to be significantly better that Sagarin's predictor (PRED, p=0.001) and standard ranking (SAG, p=.002), Pomeroy's rankings (POM, p=0.022), and the NCAA's own RPI (p=0.002), as well as the Las Vegas Favorite (VEG, p=0.065). It also outperforms LRMC Classic (LRMCC, p=0.029) and LRMC0 (p=0.048). Given the recent interest in consensus methods and the "wisdom of crowds", it is also notable that Bayesian LRMC outperformed the consensus ranking (AVG, p=0.016).

In the Appendix, we show the significance results for Bayesian LRMC against complete rankings with fewer than 20 disagreements. Obviously, it is hard to draw conclusions from most of this data since, for example, even a win-loss record of 12-6 (against HKB) does not have enough observations to be statistically significant even at the 0.1 level.

We now move on to comparisons between Bayesian LRMC and partial rankings. There are nine partial rankings in our data set. The AP and USA Today polls and DeSimone rank only the top 25 teams each year. Trexler ranks the top 40 teams each year, and extending the AP and USA Today rankings to all teams that get votes puts each at an average of slightly more than 40 teams each year. AccuRating ranked about 120 teams in each of the three years it was tracked. And although winning percentage and the NCAA Tournament seeds in principle predict a winner for each game, there were enough ties for each one to qualify as a partial ranking.[4]

Table 5 shows the McNemar's test results for Bayesian LRMC against each of the partial rankings. Bayesian LRMC appears to be significantly better than all but two. Of special interest is the DeSimone ranking (DES), which over six years disagreed with Bayesian LRMC 37 times and was correct 18 of those times.

Table 5. Significance levels for Bayesian LRMC over partial rankings, 2003-11.

[3] In rare cases, there were ties even for rankings that covered every team. For example, there were three "pick-em" games in which the Las Vegas point spread was zero. We consider a ranking to be "complete" for the purpose of including it in this table if it gives predictions for at least 98% of NCAA Tournament games.

[4] Teams with the same seed can play each other in the initial round (one or four games) and in the final two rounds when same-seeded teams from different regions play each other (e.g., the #2 seed from two different regions). The NCAA explicitly ranks all four #1 seeds, so we use that ranking to infer a predicted winner between two #1 seeds.

RANKING	n	W-L	Pct	pval	RANKING	N	W-L	Pct	pval	RANKING	n	W-L	Pct	pval	RANKING	n	W-L	Pct	pval
PCT	9	118-48	29%	.000	TRX	3	28-11	21%	.005	USA	9	39-24	13%	.038	ACU	3	18-13	17%	.237
APE	9	55-28	15%	.002	SEED	9	54-31	15%	.008	AP	9	37-23	12%	.046	DES	6	19-18	12%	.500
USAE	9	54-28	15%	.003															

However, just as treating games between unranked teams as ties might tend to make complete rankings look better, ignoring the results of games between unranked teams might tend to make partial rankings look better. To test this hypothesis, we looked at the two major polls, the Associated Press poll (AP) and the USA Today poll (USA). Table 6 summarizes the results. In both cases, the polls' predictive performance was much worse in games where the higher-ranked team had at least one poll vote but was not in the top-25 (in fact, the polls' performance was not statistically different from a 50/50 coin flip), while Bayesian LRMC's performance was only slightly worse in those games than in games involving a team that ranked in the polls' top-25. This suggests that ranking teams beyond the top 25 might be more difficult, and therefore the value of a good complete ranking is more than that of a partial ranking whose performance in top-25 games is similar.

Table 6. Degradation of polls' performance when extended beyond top-25, 2003-11.

RANKING	Poll, top-25 gms		Poll, other gms		BLRMC, top-25		BLRMC, other		BLRMC vs poll, other games	pval
	W-L	Pct	W-L	Pct	W-L	Pct	W-L	Pct		
AP	364-128	74.0%	32-34	48.5%	378-114	76.8%	45-21	68.2%	18-5	.005
USA	360-130	73.5%	34-29	54.0%	375-115	76.5%	45-18	71.4%	15-4	.010

Finally, since the Bayesian LRMC results are retroactive from 2003-9, we looked separately at the years 2010-11. In those two years, Bayesian LRMC still had the top predictive ability, with 96 correct predictions. LRMC Classic was 2nd with 94 correct predictions, followed by PRED (92), USAE (91.5), and DES (91.5). Bayesian LRMC had the top score in 2010, and was one correct prediction behind co-leaders LRMC Classic and POM in 2011.

3 Additional insights from LRMC

3.1 Which home court gives the biggest advantage?

In popular opinion, the biggest home court advantages belong to schools that win a lot of games not just at home, but on the road as well. Duke, Kansas, Syracuse, Kentucky, North Carolina, etc. show up over and over in lists of the top home court advantages (see, for example, [5,6,7,8,9]). However, a least-squares fit gives different results. From 2003-2011 the teams with the highest average home-court advantage were Denver, Texas Tech, and Ohio. Duke ranks 43rd, just ahead of Wright State. Kansas is 61st, North Carolina 82nd, Syracuse 203rd, and Kentucky 232nd. Two-tailed t-tests show that with the exception of Denver, no team's home-court advantage is significantly ($p \leq 0.05$) higher than even half of the other teams' advantage, nor is any team's advantage significantly less than half of the others'. We also observed that the correlation between vectors of home-court advantages from one season to the next is just 0.08. So, a team's home-court advantage in one season means very little when predicting its home-court advantage the next season. Overall, with perhaps the exception of Denver, home court advantage seems to be fairly equal from team to team.

So, why do teams like Duke, Kansas, etc. have a reputation for having such a big home-court advantage? The simple explanation is that winning on those home courts is particularly difficult, but not because of the court itself; in most years beating those teams anywhere is a difficult task. It might also be that the psychology of home-court advantage is relative, and tied to the visiting team (or outside observers interested in the possibility of an upset). The relative probability of beating a team on the road compared to beating them at home increases as the quality of the opposing team increases; the Bayesian model estimates that it if team i is 3 points better than team j, then j's chance of beating i are 2 times less on i's home court than on j's. If i is 11 points better, then j's chances are 3 times less on i's home court, and if i is 20 points better, then j's chances are 20 times less on i's home court.

3.2 Which teams "know how to win the close games"?

The idea that better teams have an ability to win close games is one of the standard clichés of sports. Googling phrases like "know how to win close games", "win the close games", etc. generates over a million results [10], from media stories about teams that have the ability to win close games to dueling analyses about whether such an ability exists. They span a wide range of sports; the first page of "win close games" alone includes Premier League soccer, match-play table tennis, high school and college basketball, and NFL football.

To test the truth of the cliché in college basketball, we analyzed the home-and-home conference data that we used in [1] to calibrate the logistic regression piece of LRMC. In most conferences, many or all pairs of teams play each other twice in a season, once on each team's home court. [1] compiled data on all such pairs over four years (1999-2000 through 2002-2003, nearly 9000 games). We compared the results of teams that won at home by three or fewer points (i.e., one shot) with the results of teams that lost at home by three or fewer points. If better teams have the ability to win close games, then the teams that won the close home games should have a significantly higher road winning percentage against the same opponents than would the teams that lost the close games. However, the data show that this is not the case. Close home winners won about 35% of their road rematches, and close home losers won about 33%. Both our logistic regression model and our Bayesian model predict a difference of 36%-33%. Thus, the cliché appears to be false in NCAA basketball.

3.3 What does a 20-point win show?

The home-and-home game data in [1] seemed to highlight a paradox. The magnitude of home-court advantage is generally agreed to be approximately 4 points [2]. So, one might expect that when a team wins on its home court by 8 points, it would be approximately 50% likely to beat the same opponent on the road. More formally, if h is the home-court advantage and m_{ij} is the "true" difference (measured in points) between teams i and j, then team i should have an advantage of $m_{ij}+h$ at home and $m_{ij}-h$ on the road. Since h=4, observing the home-court result $m_{ij}+h=8$ would imply that $m_{ij}=4$ and $m_{ij}-h=0$, so the "expected" road result would be a toss-up. However, the empirical data show only a 40% chance that an 8-point home winner would beat the same team on the road; the break-even point of 50% is not reached until team i wins at home by 22 points (implying a home-court advantage of h=11 points). Equally surprising is that of the 10 teams in our data set that won a home game by 50 or more points, only 6 of them won their road rematch.

Our Bayesian analysis was able to both resolve this apparent paradox and explain why 50-point road winners were so much less successful than one would expect.

In [2], we model team strengths as normally distributed with mean zero and variance τ^2, so the prior distribution of the difference m_{ij} between two teams i and j is normally distributed with mean zero and variance $2\tau^2$. However, the actual outcome of each game also includes some random variation (since when two teams play each other more than once, even on the same court, they certainly do not finish with the same margin of victory every time). [11] found that the difference between Las Vegas betting lines and actual margins of victory is approximately normally distributed with mean zero and variance 11, and we [1] found the same amount variability when comparing fits of LRMC steady-state probability differences to margins of victory. So, our Bayesian model estimates that the margin of victory of team i over team j on i's home court to be $X_{ij} \sim N(m_{ij}+h,\sigma^2)$, where σ=11 points.

After fitting our model to the data, we deduced that τ=4.26 [1]. Notice that the variance in the random effects in each game is σ^2=121, which is much higher than the variance in the prior distribution of team strengths, $2\tau^2$=36. So the paradox is explained by the fact that in a 20-point home win, it is likely that most of the margin of victory is due to randomness, and much less than half is due to the difference in team strength.

As we show in [2], the model gives the following estimate for the probability that team i will beat team j on the road given an x-point win at home:

$$\Pr(i \text{ beats } j \text{ on road} \mid i \text{ beat } j \text{ by } x \text{ points at home}) = \Phi\left(\frac{x}{\sigma}\frac{2\tau^2}{\sqrt{(\sigma^2+2\tau^2)(\sigma^2+4\tau^2)}} - \frac{h}{\sigma}\sqrt{\frac{\sigma^2+4\tau^2}{\sigma^2+2\tau^2}}\right). \qquad (1)$$

So, if team i beat team j by x=20 points at home, (1) implies that its probability of beating team j on the road is 49.0%, which agrees almost exactly with the value of 49.1% observed in smoothed home-and-home data [1]. The value of 49.0% is also the same probability that would be expected if all we knew was that team i was m_{ij}=3.7 points better than team j. In other words, team i's 20-point victory tells us only that we should update our prior estimate from i and j having equal team strengths to i being 3.7 points better than j. (Of course, for a team that wins game after game by 20 points, a model jointly-conditioned on all of those results will deduce that the outcomes are probably due to differences in team strength – but such an instance is unlikely, because the variability data suggest that if a team is 20 points better than all of its opponents, it would win about 1/6 of its games by 31 or more points and about 1/6 by 9 or fewer points.)

In addition to explaining the home court paradox, our model also predicts the break-even point for road games at about 22 points, and predicts that a 50-point home winner is only about 70% likely to beat the same team on the road, so the empirical observation that 6 out of 10 such teams won their road games is not unlikely after all.

4 Acknowledgments

Over the years, a number of Georgia Tech undergraduates have contributed to our LRMC project, including Kristine Johnson, Harold Rivner, Blaine Deluca, Pete Kriengsiri, Andy Jones, Tucker Moore, Trironk Kiatkungwanglai, Dara Thach, Holly Matera, Jared Norton, Katie Whitehead, and Blake Pierce.

5 References

[1] P. H. Kvam and J. Sokol, "A Logistic Regression/Markov Chain Model for College Basketball Rankings," Naval Research Logistics, vol. 53, no. 8, pp. 788-803, December 2006.
[2] M. Brown and J. Sokol, "An Improved LRMC Method for NCAA Basketball Prediction," Journal of Quantitative Analysis in Sports, vol. 6, no. 3, article 4, May 2010.
[3] K. Massey College Basketball Ranking Comparison, http://masseyratings.com/cb/compare.htm last retrieved April 1, 2011.
[4] K.J. Arrow, "A Difficulty in the Concept of Social Welfare", J. Political Econom., vol. 58, no. 4, pp. 328-346, 1950.
[5] http://www.sbrforum.com/ncaa-basketball/news/duke-s-cameron-indoor-tops-toughest-courts-ncaa-basketball-a-20428/ retrieved on January 5, 2012.
[6] http://bleacherreport.com/articles/530463-dukes-cameron-indoor-and-the-top-20-home-court-advantages-in-college-hoops retrieved on January 5, 2012.
[7] http://www.oddsshark.com/betting-101/ncaab/handicapping-college-basketball-rivalries retrieved on January 5, 2012.
[8] http://collegebasketball.rivals.com/content.asp?CID=1044138 retrieved on January 5, 2012.
[9] http://www.pastapadre.com/2009/04/03/unc-named-toughest-home-court retrieved on January 5, 2012.
[10] Google search engine. www.google.com, results retrieved on January 5, 2012.
[11] D. Breiter and B. Carlin, "How to Play Office Pools If You Must," Chance, vol. 10, no. 1, pp. 5-11, 1997.
[12] J. Sagarin. Jeff Sagarin NCAA Basketball Rankings. http://www.usatoday.com/sports/sagarin.htm 2003-2011.
[13] ESPN college basketball rankings. http://espn.go.com/mens-college-basketball/rankings last retrieved April 1, 2011.
[14] Wikipedia. http://en.wikipedia.org/wiki/2003_NCAA_Men's_Division_I_Basketball_Tournament through http://en.wikipedia.org/wiki/2011_NCAA_Men's_Division_I_Basketball_Tournament, last retrieved January 5, 2012.
[15] Statfox college basketball betting lines. http://www.statfox.com/cbb/logs/default.htm last retrieved October 30, 2012.
[16] VegasInsider college basketball betting lines. http://www.vegasinsider.com/college-basketball/ last retrieved October 30, 2012.
[17] LRMC rankings. http://www.lrmc.gatech.edu last retrieved January 5, 2012.

6 Appendix

6.1 Data sources

Table 7 lists all of the rankings for which [3] had comparative ranking data, along with the years each ranking's data was available.

Table 7. List of rankings available from [3], with years each ranking was available.

ACU	AccuRatings 2005-7	ISR	ISR 2007-10	RIS	Rishi 2005
ARG	ARGH 2003-5	JCI	JCI 2006	ROG	Rogers 2011
AP	Associated Press 2003-11	JEN	Jens 2008-11	ROH	Rohde 2005-7
AUS	Austin Sports 2011	JON	Jones 2004	RTH	Rothman 2003-11
BCM	BC Moore 2005	KEL	Kellner 2010	RT	RoundTable 2010-11
BIH	Bihl 2003-8,10-11	KMV	Kellner MOV 2010-11	RTB	Roundtable BCS 2009-11
BOB	Bobcat 2003-11	KPK	Kirkpatrick 2006,10-11	RPI	RPI 2003-11
BD	Boyer-DeSimone 2004-5	KLK	Kislanko ISOV 2005,7-10	RTP	RT Power 2011
BPI	BPI Sports 2009-11	KOS	Kosmo 2011	SAG	Sagarin 2003-11
BKM	Bracket Matrix 2008	KRA	Krach 2007-8,10-11	SAP	SAP 2005-7
BRZ	Brzezinski 2003	LYD	Lloyd 2007-8	SAU	Sauceda 2003-4,10

CNG	Cheong 2004-11	LYN	Lynch 2004-10	SCR	Score Card 2005-8
CMV	Coleman's MinV 2005	MKV	Markov 2003-4,10	SEL	Self 2003-7,9,11
COL	Colley 2003-11	MB	Massey BCS 2006-11	SIM	Simpson 2004-9
CPA	CPA 2008-10	MAS	Massey MOV 2003-5,10-11	SPW	Snapper's World 2008-11
CPR	CPA Retro 2008-10	MGY	Montgomery 2004	STH	Sport Theory 2008-11
DC	Dance Card 2005-11	MOR	Moore 2003-11	STR	Strength 2003,8-9
DC2	Daniel Curry 2 2009-11	NOL	Nolan 2009-11	TMR	Team Rankings 2011
DCI	Daniel Curry Index 2009-11	NOR	Nordness 2004	TSR	The Sports Report 2003-5,7-8
DES	DeSimone 2004-5,8-11	OMY	O'Malley 2009,11	TRX	Trexler 2006,9-10
DOK	Dokter Entropy 2006-11	PH	Pick Hoops 2006	TW	Tulsa World 2010-11
DOL	Dolphin 2003-11	PKL	Pickle 2006-7	UCS	UCS 2007
DUN	Dunkel 2003,6,8-11	PIG	Pigskin 2006,8-11	USA	USA Today Coaches 2003-11
DWH	DWHoops 2003-7	PTS	Pointshare 2011	WTE	White 2003
ECK	Eck 2003-6,8	PEQ	Pollard EQ 2011	WLK	Whitlock 2003-11
ERD	Elrod 2003	POM	Pomeroy 2003-11	WIL	Wilson 2005,7-11
ENT	Entropy 2003-5	PGH	Pugh 2008-11	WOB	Wobus 2003-8,10-11
GRS	Gindin 2003-4	RM	Random Monkeys 2004	WOL	Wolfe 2003-11
GRN	Greenfield 2003-10	RTR	Real Time RPI 2007-11	AVG[5]	Average of all 2003-11
HER	Hermrats 2003-5	RSE	Reese 2008		
HOL	Holland 2003	REI	Reid 2004		
HKB	Huckleberry 2009-10	REN	Rennie 2005-6		
IMS	Imes 2003,8	REW	Rewards 2008-11		

In addition to the rankings reported by [3], we tested eight additional rankings and two other predictors, all shown in Table 8. Three were LRMC rankings: Bayesian LRMC, LRMC Classic, and LRMC0 (a version that does not take margin of victory into account; although it is less accurate, we make this available to the NCAA since they as a matter of principle do not use rankings that rely on margin of victory). Four others were related to rankings available on [3]: Sagarin's Predictor and Elo-Chess rankings were found in [12], and the extended AP and USA Today polls (including all teams receiving votes, not just the top 25) were taken from [13]. The final ranking we tested was a naïve ranking based on teams' winning percentage. The two predictors we included in our tests were the NCAA Tournament seed of each team, and the Las Vegas favorite. (Because all but two of the 105 prediction methods we tested are rankings, we use the word "rankings" generically throughout this paper.) Data for all ten additions were available for all nine years from 2003-11.

Table 8. Additional rankings tested, along with source. (Abbreviations are our own.)

BLRMC	Bayesian LRMC [2,17]	PRED	Sagarin Predictor [12]	SEED	NCAA Tournament Seeds [14]
LRMCC	LRMC Classic [1,17]	APE	Associated Press Extended [13]	VEG	Las Vegas Favorite [15] (2003-8) and [16] (2009-11)
LRMC0	LRMC0 [1,17]	USAE	USA Today Coaches Extended [13]		
ELO	Sagarin Elo-Chess [12]	PCT	Winning Percentage [13]		

6.2 Full ranking placement data

For completeness, Table 9 is a version of Table 1 that includes all 105 rankings.

Table 9. Rankings sorted by average score, with total years tracked ("n") and top 1, 2, 5, and 10 finishes.

RANK-ING	n	Top 1-2-5-10	Avg	RANK-ING	n	Top 1-2-5-10	Avg	RANK-ING	N	Top 1-2-5-10	Avg	RANK-ING	n	Top 1-2-5-10	Avg
BLRMC	9	4-5-6-6	11.6	ELO	9	1-1-2-2	22.9	DC	7	0-0-1-2	26.3	LYN	7	0-1-1-2	33.3
AUS	1	0-0-0-1	13.0	JCI	1	0-0-0-0	23.0	DCI	3	0-0-1-1	26.5	SCR	4	0-0-0-0	33.5
PTS	1	0-0-0-1	13.0	PKL	2	0-0-0-1	23.3	ROH	3	0-0-1-1	26.5	REW	4	0-0-0-1	34.8
PEQ	1	0-0-0-1	13.0	SAG	9	0-0-1-1	23.4	KLK	5	0-0-1-2	27.0	SIM	6	0-0-0-0	35.7
LRMCC	9	1-2-4-5	14.3	GRS	2	0-0-0-0	23.5	COL	9	0-0-1-1	27.8	STH	4	0-0-0-0	35.8
PH	1	0-0-0-0	14.5	WOL	9	0-0-2-3	23.5	HER	3	0-0-1-1	27.8	UCS	1	0-0-0-0	37.0
CNG	8	0-0-2-4	15.8	ISR	4	0-0-0-1	23.6	BKM	1	0-0-0-0	28.0	RTR	5	0-0-0-0	37.7

[5] This abbreviation is our own; [3] does not provide one. All other abbreviations in this table are from [3].

LYD	2	0-0-0-0	16.5	MAS	5	0-0-0-1	23.8	RSE	1	0-0-0-0	28.0	BRZ	1	0-0-0-0	38.0
POM	9	1-1-4-4	17.4	PIG	5	0-0-0-1	24.0	CPA	3	0-0-0-1	28.2	ERD	1	0-0-0-0	38.0
KMV	2	0-0-0-1	17.5	SPW	4	0-0-1-1	24.1	DOL	9	0-0-1-1	28.6	WTE	1	0-0-0-0	38.0
TSR	5	0-0-2-2	17.8	BOB	9	0-0-2-2	24.3	RPI	9	0-0-2-3	28.8	CPR	3	1-1-1-1	38.5
KEL	1	0-0-0-0	18.0	MKV	3	0-0-0-0	24.3	STR	3	0-0-1-1	29.8	BCM	1	0-0-0-0	39.0
REN	2	0-0-0-1	18.5	GRN	8	1-1-2-2	24.4	USAE	9	0-1-1-1	29.9	OMY	2	0-0-0-0	39.8
ACU	3	0-0-1-1	18.7	ECK	5	0-0-1-2	24.7	AP	9	0-0-1-1	30.0	BPI	3	0-0-0-0	40.2
ENT	3	0-0-0-1	19.2	SEL	7	0-0-0-0	24.7	APE	9	0-0-0-0	30.3	HOL	1	0-0-0-0	41.5
MOR	9	1-2-3-5	19.2	RTH	9	0-0-1-1	24.8	DUN	6	0-0-0-0	30.4	SAP	3	0-0-0-0	41.8
AVG	9	1-1-3-4	20.2	PRED	9	0-0-2-4	24.9	KRA	4	0-0-1-1	30.5	REI	1	0-0-0-0	42.0
SAU	3	0-0-0-0	20.5	RTP	1	0-0-0-0	25.0	KPK	3	0-0-0-0	31.2	RTB	3	0-0-0-0	42.8
DWH	5	0-0-1-2	20.9	TMR	1	0-0-0-0	25.0	RT	2	0-0-0-1	31.3	TRX	3	0-0-0-0	43.5
IMS	2	0-0-0-0	21.0	LRMC0	9	2-3-3-4	25.1	USA	9	0-1-1-2	31.3	NOR	1	0-0-0-0	46.0
WIL	6	0-1-2-2	21.5	MB	6	0-0-1-1	25.2	HKB	2	0-0-0-0	31.8	RIS	1	0-0-0-0	46.0
DOK	6	0-0-1-1	21.6	WOB	8	0-0-1-1	25.6	ARG	3	0-0-0-0	31.8	PCT	9	0-0-0-0	47.1
VEG	9	0-0-3-3	21.7	JEN	4	0-0-1-2	25.6	NOL	3	0-0-0-1	32.8	TW	2	0-0-0-0	49.8
WLK	9	0-0-2-3	21.8	BIH	8	0-0-1-2	26.0	JON	1	0-0-0-0	33.0	CMV	1	0-0-0-0	50.0
RM	1	0-0-0-0	22.0	SEED	9	0-1-2-2	26.0	MGY	1	0-0-0-0	33.0	KOS	1	0-0-0-0	52.5
DES	6	0-0-0-3	22.5	DC2	3	0-0-0-1	26.2	PGH	4	0-0-0-0	33.3	ROG	1	0-0-0-0	52.5
BD	2	0-0-1-1	22.8												

6.3 Significance levels for rankings with little data

For completeness, Table 10 shows significance test results against Bayesian LRMC even for rankings with little comparative data.

Table 10. Significance levels for Bayesian LRMC over each complete ranking with fewer than 20 disagreements, 2003-11.

RANK-ING	**n**	**W-L**	**Pct**	**pval**	**RANK-ING**	**n**	**W-L**	**Pct**	**pval**	**RANK-ING**	**n**	**W-L**	**Pct**	**pval**	**RANK-ING**	**n**	**W-L**	**Pct**	**pval**
CMV	1	13-2	23%	.004	RSE	1	8-2	16%	.055	LYD	2	8-4	9%	.194	BD	2	10-8	14%	.407
WTE	1	7-0	11%	.008	RIS	1	8-2	16%	.055	ROG	1	8-4	18%	.194	AUS	1	4-3	10%	.500
IMS	2	13-3	13%	.011	ENT	3	12-5	9%	.072	KOS	1	9-5	21%	.212	PTS	1	4-3	10%	.500
BRZ	1	8-1	14%	.020	DC2	3	13-6	10%	.084	GRS	2	10-6	13%	.227	REI	1	6-5	17%	.500
HOL	1	11-3	22%	.029	NOR	1	7-2	14%	.090	OMY	2	8-5	10%	.291	PEQ	1	3-2	7%	.500
RT	2	11-3	11%	.029	HKB	2	12-6	14%	.119	UCS	1	5-3	13%	.363	JON	1	5-5	16%	.623
BCM	1	5-0	8%	.031	JCI	1	9-4	20%	.133	TMR	1	7-5	18%	.387	MGY	1	5-5	16%	.623
ERD	1	9-2	17%	.033	KEL	1	9-4	20%	.133	RTP	1	8-6	21%	.395	RM	1	4-5	14%	.746
BKM	1	7-1	13%	.035	PH	1	7-3	16%	.172										