Supplementary Materials

1 Simulation study S1

We conduct two experiments as described in Kolde et al. (Robust rank aggregation for gene list integration and meta-analysis, Bioinformatics, 2012). All ranking lists consisted of 1000 items out of which 5% were preferentially ranked at the top. We regard the top 50 items as the informative items and the remaining as the noised items. The means of the informative items are sampled from the exponential distribution with scale 1/2, and means of all the noised items are 0. We randomly draw the value for each item from normal distribution with mean as specified and unit variance, and then rank the items according to the sample values. We compared five rank aggregation methods: Mean, Median, RRA, BIRRA, MM and EMM, in the following two experiments. Each experiment is replicated for 100 times.

In the first experiment, we generated 10 such top-50 ranking lists. As seen from the receiver operating characteristic (ROC) curves in Figure 1 (a), all five methods performed similarly well. Their Area Under Curve (AUC) scores were 0.903 ± 0.029 (Mean), 0.768 ± 0.036 (Median), 0.681 ± 0.045 (RRA), 0.871 ± 0.035 (BIRRA), 0.904 ± 0.029 (MM) and 0.903 ± 0.029 (EMM).

In the second experiment, we studied the robustness of the algorithms against noise. We generated 10 top-50 rankings lists as described before and 30 randomly ordered top-50 ranking lists. In this case, their performance varied, as shown in Figure 1 (b). Their AUC scores were 0.817 ± 0.031 (Mean), 0.501 ± 0.043 (Median), 0.627 ± 0.037 (RRA), 0.802 ± 0.037 (BIRRA),

0.819 ± 0.031 (MM) and 0.849 ± 0.032 (EMM).

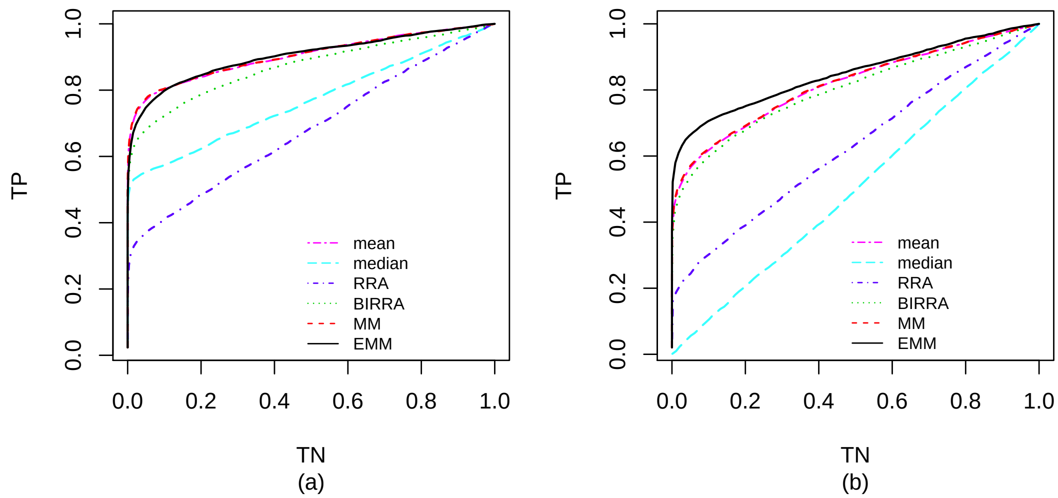

Figure 1: ROC curves from Simulation study S1. The vertical axis shows the true positive rate. The horizontal axis shows the false positive rate.

From the above two experiments, we can see that compared with other approaches, the average rank, BIRRA, MM and EMM had more robust performance in those scenarios. In the second experiment, when there are multiple noisy ranking lists, EMM performed better than other methods.

2 Simulation study S2

In this simulation study, we consider the scenarios that ϕ_i does not follow $\phi(1-\alpha^i)$ as assumed in the main text, but it is either constant, linear or bell-shaped, as shown in Figure 2. When $\phi_i \sim \text{dbeta}^*(a, b)$, it means that $\phi_i \propto f(\frac{i}{30})$, where $f(\cdot)$ is the probability density function of the beta(a, b) distribution. In all three cases, the mean of ϕ_i is 0.5.

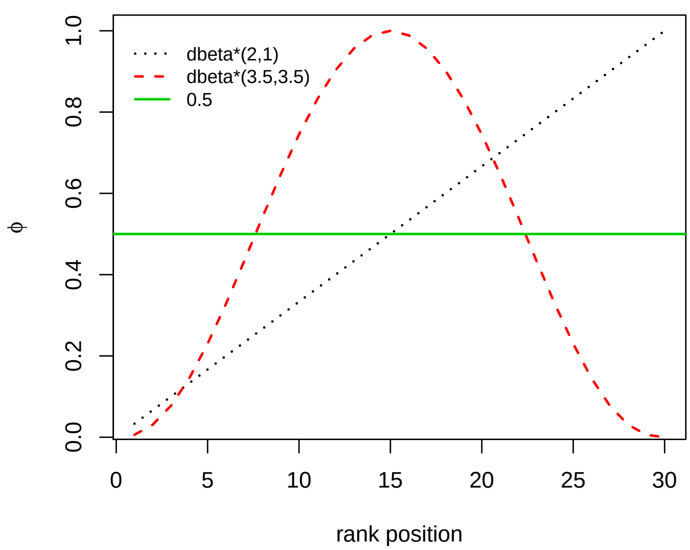

Figure 2: Different function forms of ϕ_i in Simulation study S2.

Let $U = \{1, 2, ..., 100\}$ and $\boldsymbol{\pi}_0 = (1, 2, ..., 100)$. We independently generated m top-30 ranking lists from the following model (Equation 1). We assume the quality of ranking lists vary and sample ω_l from $Unif(0.6, 1)$ with probability 0.5; otherwise, we sample it from $Unif(0, 0.6)$. We generated a couple of top-30 datasets by setting $m = 4,\ 10,\ 20$. Under each simulation setting, we generated 100 synthetic data sets and reported the mean of the Kendall tau distance at different rank positions. The results are summarized in Figure 3. We can see that EMM outperformed the alternatives with a smaller Kendall tau distance at different positions in those cases, showing that EMM is quite robust for aggregating ranking lists of varying quality, though ϕ_i is not the same as assumed in the main text. When the number of ranking lists became larger, the performance of all of the methods improved.

$$\begin{aligned} p(\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_m|\boldsymbol{\pi}_0, \phi, \alpha, \omega) &= \prod_{l=1}^{m}\prod_{i=1}^{30} p\left(V_i(\boldsymbol{\pi}_l, \boldsymbol{\pi}_0) = v_i|\phi, \alpha, \omega\right) \\ &= \prod_{l=1}^{m}\prod_{i=1}^{30}\left\{\omega_l \frac{\phi_i^{v_{li}}}{Z(\phi_i, i)} + (1-\omega_l)\frac{1}{n-i+1}\right\}, \\ &\quad 0 < \omega_l < 1,\ l = 1, 2, ..., m. \end{aligned} \tag{1}$$

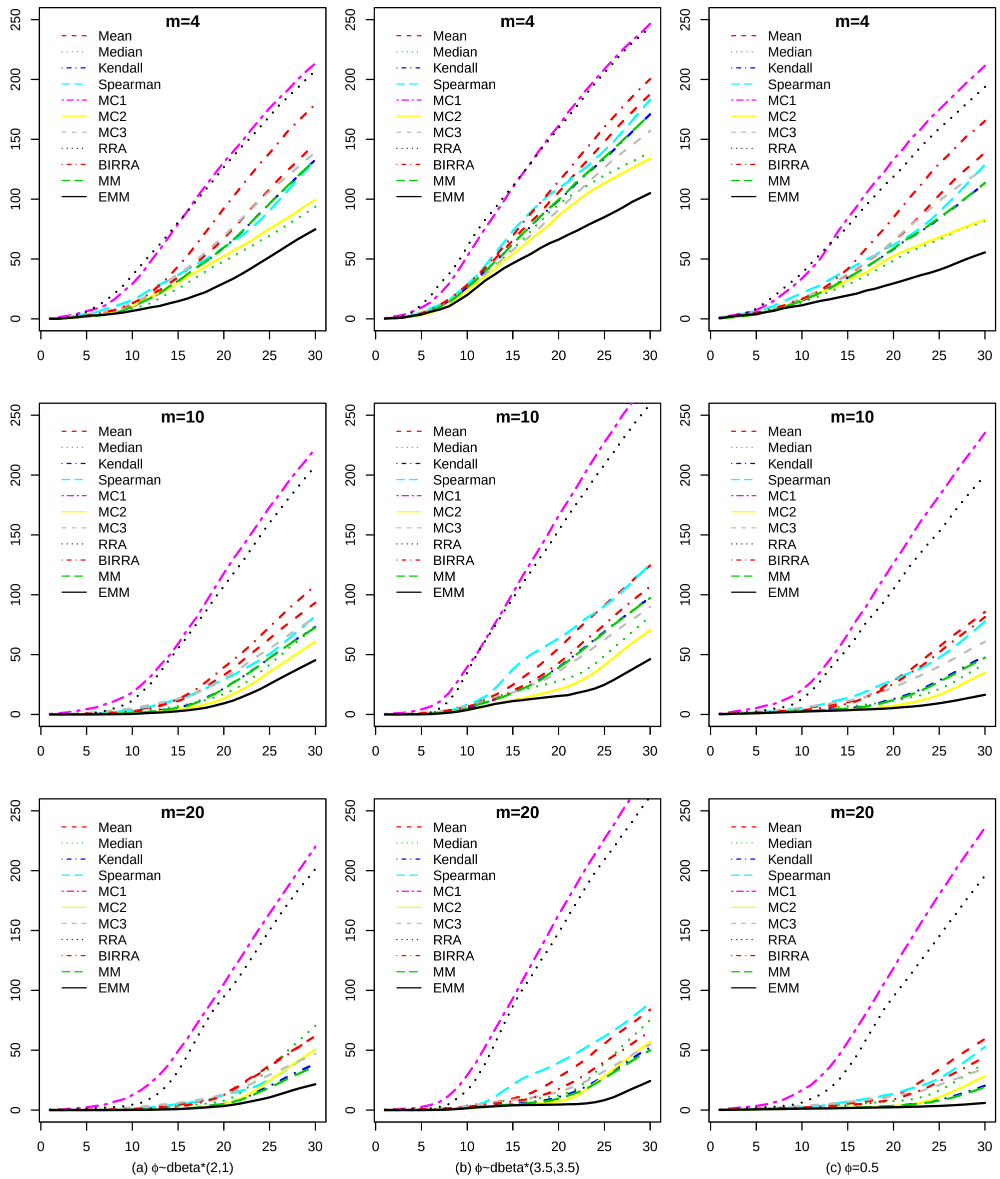

Figure 3: Stagewise Kendall tau distance of consensus rankings in Simulation study S2.

3 An extra real application: aggregating NBA team rankings

This example is about aggregating the multiple rankings of NBA teams and was studied by Deng et al. (Bayesian aggregation of order-based rank data, Journal of the American Statistical Association, 2014). They collected 34 rankings, including 6 professional rankings and 28 amateur rankings, for the 30 NBA teams in the 2011-2012 season. The professional rankings were obtained from six professional news agencies, such as NBA.com and ESPN.com. The amateur rankings were given by the Harvard graduate students, who were asked to select the best eight NBA teams of this season. The students classified themselves into one of the four groups in the survey: (1) avid fans, (2) fans, (3) infrequent watchers, and (4) non-interested individuals. In Table 1, the first 16 teams entered the 2011-2012 playoffs. Assuming we do not know any prior information about the quality of the rankings, we aim to predict the teams that entered the final and evaluate the quality and stability of the individual rankings.

We first evaluated the correlation amongst the 34 rankings, and their rank coefficients are shown in Figure 4. The rank coefficients of the paired rankings are calculated up to the eighth ranking stage. We can see that the professional rankings were detected as highly correlated. We applied both EMM and HEMM to fit the data, and the aggregated ranking lists produced by different methods are shown in Table 2. We can see that both EMM and HEMM exhibited a quite satisfactory performance in predicting the top 16 teams that would make the playoffs, while making just one mistake: including the Trail Blazers instead of the Jazz. Of the 12 competing methods, the median rank, the Markov-chain-based methods, and RRA performed worse than the others by making several mistakes. The results of the other methods were similar to those of Deng et al. (2014), as they made the same predictions for the top 15 teams.

Next, we examined the qualities of the rankings given by different groups. From Figure 5, it is observed that the MLE of the ω of the professional and avid fan groups (except S_5) was 1, much larger than the values for other groups, indicating that their rankings indeed had better quality. S_5 was only accurate in predicting the ranks of a few top teams; thus, its estimated ω was not as large as that of the other avid fans. The means of ω in different groups appeared to be consistent with what their group label implies. In terms of ranking stability, the MLE of α of the professional and avid fan groups were smaller compared with the other three groups, suggesting that they were more stable.

For the specific ranks of teams, there were a few disagreements between the consensus ranking of HEMM and that of the other methods. For instance, most of the aggregation methods ranked the "Lakers" as the second best team, whereas HEMM ranked the "Thunder" as the second best team. This was in accordance with the professional and avid fan groups, who tended to rank the "Thunder" above the "Lakers."

Table 1: Rankings of the 30 NBA teams collected in the study by Deng et al.(2014)

		Professionals						Avaid Fans						Fans						Infrequent watchers							Non-interested individuals								
N.o.	Team	P_1	P_2	P_3	P_4	P_5	P_6	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}
1	Heat	1	2	1	1	1	1	1	1	2	3	1	3	-	1	2	1	3	1	3	1	4	1	4	2	1	2	-	4	-	-	1	1	1	2
2	Thunder	3	3	2	3	2	3	2	2	3	2	-	2	-	-	4	2	7	4	-	-	-	2	-	-	2	-	-	-	-	2	-	-	-	-
3	Spurs	7	10	11	5	8	7	6	5	5	5	-	6	-	6	5	4	-	-	5	-	8	6	3	6	-	-	-	-	-	-	-	-	-	-
4	Celtics	5	11	10	9	9	5	4	8	1	4	2	5	2	3	1	3	4	3	-	4	2	3	2	-	4	4	-	2	-	-	-	2	-	-
5	Clippers	8	5	6	10	5	6	-	6	-	8	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	4	-
6	Lakers	6	7	7	6	6	8	3	7	6	1	3	1	1	2	7	7	1	2	1	2	1	4	5	1	3	1	5	1	-	-	8	4	2	-
7	Pacers	14	13	14	14	13	12	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	5	8	-	-
8	76ers	15	16	13	15	15	15	-	-	-	-	-	-	-	4	-	-	8	6	-	-	-	-	-	5	6	-	7	6	6	4	3	3	-	-
9	Mavericks	2	1	3	2	4	4	-	-	7	7	-	8	-	-	8	6	6	5	-	8	3	-	1	-	-	-	3	-	-	-	4	-	5	-
10	Bulls	4	4	4	4	3	2	5	4	8	6	4	4	3	-	3	-	-	8	4	-	-	-	-	-	-	3	-	3	3	5	-	5	-	1
11	Knicks	9	6	9	8	7	13	-	3	4	-	5	-	-	-	-	-	2	-	-	-	-	-	-	4	7	8	-	-	2	8	-	7	7	6
12	Grizzlies	10	8	8	7	11	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	-	-	-	-
13	Nuggets	19	9	5	13	10	9	-	-	-	-	8	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	8	-
14	Magic	11	12	17	11	14	11	-	-	-	-	-	-	6	-	-	-	-	-	-	5	-	-	-	-	5	-	4	-	-	-	6	6	-	-
15	Hawks	12	18	12	18	12	18	7	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	8	7	-	-	5	-	-	-	-	8
16	Jazz	18	23	26	27	28	19	-	-	-	-	-	-	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
17	TrailBlazers	13	14	15	12	16	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	Rockets	21	15	16	16	17	17	-	-	-	-	-	-	5	7	6	-	-	-	2	6	-	-	-	3	-	5	-	5	-	3	-	-	-	7
19	Bucks	16	17	20	17	20	16	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	6	-	-	-	-
20	Suns	20	22	19	21	19	21	-	-	-	-	-	-	-	8	-	-	-	-	7	7	6	8	8	-	-	-	-	-	-	-	7	-	-	-
21	Nets	17	19	24	20	24	23	-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	3	5
22	Warriors	22	21	23	19	22	20	-	-	-	-	6	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
23	Timberwolves	23	20	22	22	23	24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	7	-	-	-	-	-	-	4	7	-	-	-	4
24	Hornets	27	28	18	23	18	25	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	Pistons	25	25	25	24	25	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8	-	-	-	8	-	-	2	-	-	-
26	Kings	29	24	21	26	21	26	-	-	-	-	-	-	4	5	-	-	-	-	8	-	5	5	-	-	-	-	-	-	1	-	-	-	-	-
27	Wizards	28	27	28	25	27	27	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	7	-	-	-	-	6	-
28	Raptors	24	26	29	28	30	28	-	-	-	-	-	-	-	-	-	8	-	-	-	-	-	-	-	-	-	-	8	-	-	1	-	-	-	-
29	Cavaliers	26	29	27	29	26	29	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-
30	Bobcats	30	30	30	30	29	30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	-	-	-

Table 2: Results of different rank aggregation methods in the NBA teams study.

N.o.	Team	Mean	Median	Kendall	Spearman	MC1	MC2	MC3	RRA	BIRRA	MM	EMM	HEMM
1	Heat	1	1	1	1	1	1	1	2	1	1	1	1
2	Thunder	4	5	3	3	7	4	5	6	3	3	4	2
3	Spurs	7	7	8	7	5	7	7	4	7	8	7	6
4	Celtics	3	2	4	4	3	3	3	7	4	4	3	4
5	Clippers	9	12	9	9	11	10	11	10	9	9	8	9
6	Lakers	2	3	2	2	2	2	2	5	2	2	2	3
7	Pacers	15	**22**	16	13	**19**	**18**	15	**21**	14	16	14	13
8	76ers	10	14	10	15	**18**	9	9	9	13	10	16	15
9	Mavericks	6	6	6	6	6	6	6	3	6	6	5	7
10	Bulls	5	4	5	5	4	5	4	1	5	5	6	5
11	Knicks	8	**24**	7	8	10	8	8	8	8	7	9	8
12	Grizzlies	11	10	11	10	**21**	12	**17**	**20**	10	15	11	10
13	Nuggets	14	16	14	11	**25**	16	14	**17**	12	13	10	11
14	Magic	12	**27**	12	12	9	11	12	11	11	12	12	12
15	Hawks	16	**18**	15	16	**20**	**20**	13	15	16	14	13	16
16	Jazz	**24**	**17**	**25**	**23**	**22**	**30**	**27**	**29**	**24**	**25**	**26**	**24**
17	TrailBlazers	17	8	18	14	30	19	23	30	15	18	15	14
18	Rockets	13	23	13	17	8	15	10	12	17	11	17	17
19	Bucks	18	28	19	18	17	27	21	25	18	19	18	18
20	Suns	19	26	17	19	24	26	16	13	19	17	19	19
21	Nets	20	30	23	20	16	17	20	19	20	23	21	21
22	Warriors	21	29	20	21	15	23	22	22	21	22	20	20
23	Timberwolves	22	11	22	22	14	14	19	16	22	21	22	22
24	Hornets	26	19	26	26	28	25	28	28	25	26	24	23
25	Pistons	25	25	24	25	23	21	24	26	26	24	23	25
26	Kings	23	20	21	24	12	13	18	14	23	20	25	26
27	Wizards	27	21	27	27	26	28	25	18	29	27	27	27
28	Raptors	28	13	28	28	13	22	26	27	27	28	28	28
29	Cavaliers	29	15	29	29	27	29	29	23	28	29	29	29
30	Bobcats	30	9	30	30	29	24	30	24	30	30	30	30

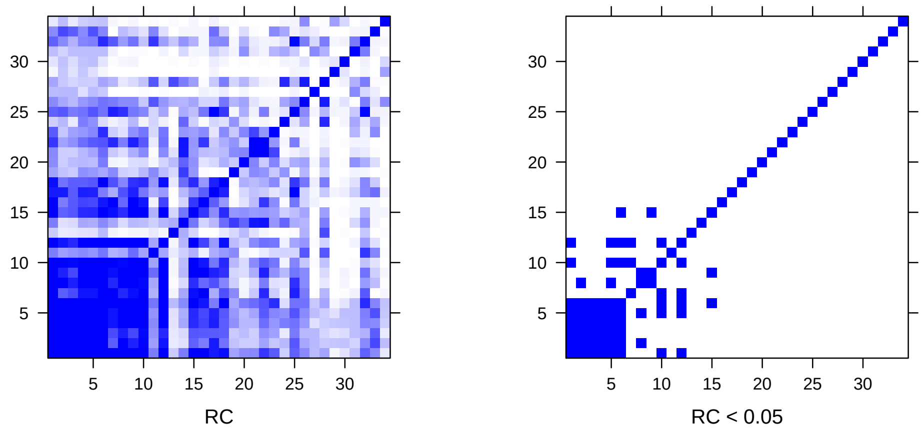

Figure 4: Correlation matrix of the NBA team rankings using the rank coefficient statistic.

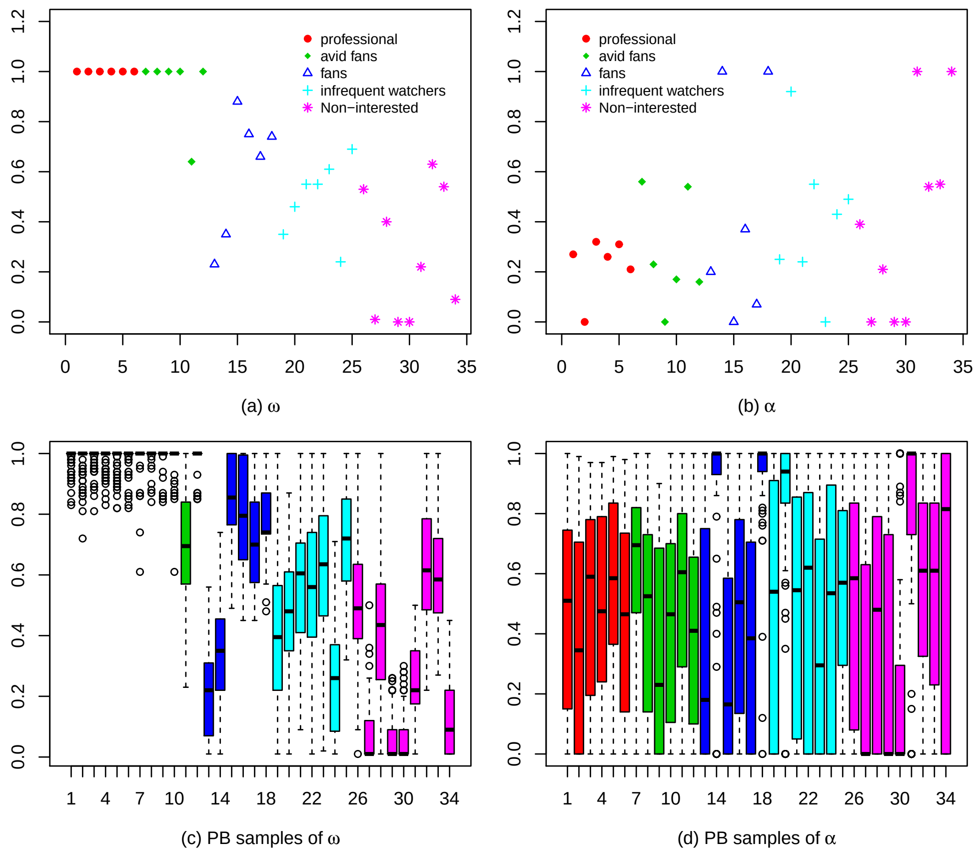

Figure 5: The MLE of the parameters from the real data (top) and parametric bootstrap samples (bottom) in the NBA team study.