ISSN 1750-9823 (print)
International Journal of Sports Science and Engineering
Vol. 07 (2013) No. 01, pp. 036-050

Investigation of Round Robin Sports Tournaments by the Analytic Hierarchy Process

Dr. M.A.A. Cox [+]

School of Psychology, Ridley Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, England.

(*Received June 7, 2012, accepted September 9, 2012*)

Abstract. It has been suggested that the analytic hierarchy process (AHP) can be employed to interpret the results of sports tournaments [1] with particular reference to evidence of cheating. This manuscript presents a first attempt to use this approach on round robin tournaments and evaluate its effectiveness. It is applied to two real world examples in one of which some irregularities are known to have occurred. The ideal method for combining the results for each pair of teams (home/away) is investigated. The aim is to assess if the AHP is an effective approach to this problem and to use the results to investigate any irregularities in a teams' performance. The analysis failed to detect any evidence of non-sporting behaviour. A number of consistency criteria are described and exhaustively compared for three attributes (3×3 matrices).

Keywords: Analytic Hierarchy Process (AHP), Binary AHP, Sport.

1. Introduction

A simple model is explored to investigate inconsistency in sports results. This is particularly relevant in view of recent cheating allegations in the Guinness Premiership, popularly referred to as "bloodgate" implicating team 4 below. Any search on the web will reveal the numerous ramifications of this case; see for example The Telegraph [2]. The approach is applied to all regular season matches in the Guinness Premiership (rugby football union) 2007/8 [3] and the Barclays Premier League (association football), 2007/8 [4]. Both of these are round robin (or play all) tournaments, each pair of teams playing twice (home and away). If the method is effective then the team implicated in "bloodgate" should be indicated in the Guinness Premiership analysis.

The Analytic Hierarchy Process (AHP) has been widely employed as an aid to decision making both by practitioners and academics. The AHP was introduced by Saaty [5] and more fully described in Saaty [6]. Its foundation was as an eigenvector method, analysing a reciprocal symmetric matrix. The entries of the matrix are the decision makers' preferences of n alternatives with respect to some criterion, when pair wise comparisons are performed. This procedure is now briefly reviewed.

2. A Brief Introduction of the AHP

Table 1. AHP Verbal Scale

Numerical scale	Verbal Scale of Saaty
1	Equal importance of both elements.
3	Moderate importance of one element over another.
5	Strong importance of one element over another.
7	Very strong importance of one element over another.
9	Extreme strong importance of one element over another.

In the AHP procedure the decision maker is required to make pair wise comparisons between n alternatives based on a ratio scale. The choices are made from the integers between 1 and 9 and their reciprocals $\left(I=\left\{\frac{1}{9},\frac{1}{8},\frac{1}{7},\frac{1}{6},\frac{1}{5},\frac{1}{4},\frac{1}{3},\frac{1}{2},1,2,3,4,5,6,7,8,9\right\}\right)$ see Saaty [6]. Here 1 equates to equal relative importance while 9 equates to extreme relative importance. The verbal equivalents are displayed in Table 1.

[+] *E-mail address*: mike.cox@newcastle.ac.uk

An alternative would be to employ an exponential scale [7]. The resulting decisions are summarised in a square matrix (A: $a_{ij} \in I$ $i=1,\ldots,n$ $j=1,\ldots,n$) of dimension n. It is reciprocal symmetric that is $a_{ij} = \frac{1}{a_{ji}}$. It is assumed that $A_{ij} = \frac{v_i}{v_j}$ where the weights represent the probability of adopting choice i where $\sum_{i=1}^{n} v_i = 1$. We then have the matrix equation $A\underline{v}=n\underline{v}$ which equates to an eigen problem. In addition, A is consistent because the following condition is satisfied: $a_{ik} = a_{ij}\, a_{jk}$, $i, j, k = 1,\ldots,n$.

In a general decision-making environment, we cannot give the precise values of the v_i/v_j but only estimates of them. In which case, traditionally, the weights are obtained by solving the eigen problem, $A\underline{v}=\lambda_{max}\underline{v}$, where λ_{max} is the maximum eigenvalue. The priority vector of the alternatives is the right eigenvector associated with the dominant eigenvalue. An alternate approach is to adopt the row geometric mean $\left(v_i = \frac{1}{N}\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n} \right)$ with an appropriate normalisation term $\left(N : \sum_{i=1}^{n} v_i = 1 \right)$ [8]. For a detailed review of various alternate approaches to this numerical problem see Triantaphyllou et al. [9] who employ a least squares approach and Choo and Wedley [10] who consider 18 methods for estimating the preference vector.

3. Applications to Sports Problems

The AHP is the sole approach adopted here. Alternative methods may be employed, for example, Barnett and Hilditch [11] examined data on the use of artificial pitch surfaces in the English football league. They assessed if there was a possible home team advantage gained on such pitches. A statistical analysis of the end-of-season results for the four divisions over the period 1981-1999, suggested that there is indeed such an advantage and that it is of a sufficient scale to be a cause for concern. The authors concluded, that a fuller study should be undertaken when the results of more matches becomes available. A parametric model was adopted by Dixon and Coles [12], which was fitted to English league and cup football data from 1992 to 1995. The model was used to exploit potential inefficiencies in the association football betting market, and used bookmakers' odds from 1995 to 1996. The technique was based on a Poisson regression model. Maximum likelihood estimates were obtained and shown to give a positive return when used as the basis of a betting strategy. Both of these papers have led to their own research threads.

Following Nishizawa [1] the AHP is employed to investigate sports results. In this case an element a_{ij} ($i \neq j$) of the comparison matrix A takes one of only three values [13], either ϑ (win), 1 (draw) or $1/\vartheta$ (loss) and is called the binary AHP. Nishizawa [1] adopts $\vartheta = 2$, Boginski et al. [14] choose $\vartheta = 1.2$, these alternatives will be investigated below. The latter authors also adopted the strategy of coding conflicting results (a win and a loss against the same opponents) as a draw.

In the United Kingdom this approach is particularly valuable since team comparisons are far from straight forward. For instance we have the vagaries of weather, since these are outdoor winter sports. The occurrence of injuries within a fixed squad of players may lead to straightened circumstances. Also in rugby union there is the regular absence of key players when required by the national team. There is the annual series of autumn and winter internationals, further exacerbated by the world cup every four years. The final league position may not be a fair reflection of a team’s calibre (potential).

In the following sections the basic problem is introduced with a number of associated criteria of consistency. Having explored how the round robin property may be incorporated, the choice of the parameter in the binary AHP is explored. The problem is then generalised to employ both a continuous and a discrete scale for the AHP. Two examples are examined in depth. In addition the criteria of consistency are exhaustively examined for all 3×3 binary AHP matrices.

4. Criteria of Consistency

The traditional measure of consistency is the consistency index (CI). If the maximum eigenvalue of an AHP matrix A of order n is λ_{max} then $CI = \frac{\lambda_{\max} - n}{n-1}$ and is compared to the random consistency index to give the consistency ratio (*CR*). A small ratio is desirable. However Alonso and Lamata [15] have some reservations about this approach, deeming it inflexible and restrictive as the size of the matrix increases. The values of the random index have been tabulated for $n = 3,\ldots,15$ ([16] see also an extensive review [17]) while for larger values of n, $CI = -0.799 + 0.596\,n - 0.0518\,n^2 + 0.00152\,n^3$, ($R^2 = 99.4\%$, using regression on the conventional estimates) can be used as an approximation.

Nishizawa [1] adopted a criterion for consistency based on the number of cycles in a graph associated with A. A vertex matrix (V) is constructed $v_{ij} = \begin{cases} 1 & \text{if } a_{ij} > 1 \\ 0 & \text{otherwise} \end{cases} : i = 1,\ldots,n \quad j = 1,\ldots,n$ where this represents a directed graph. Inconsistency is represented by a cycle in this graph (a defeats b, b defeats c and c defeats a). To measure inconsistency, cycles of length three are identified. To aid in this an S matrix is constructed where $s_{ik} = \sum_{j=1}^{n} v_{ij} v_{jk} v_{ki} : i = 1,\ldots,n \quad k = 1,\ldots,n$. To examine cycles of length three associated with teams i, j and k consider $s_{ijk} = \begin{pmatrix} s_{ii} & s_{ij} & s_{ik} \\ s_{ji} & s_{jj} & s_{jk} \\ s_{ki} & s_{kj} & s_{kk} \end{pmatrix}$ and $i > j > k$, a cycle and hence an inconsistency is indicated if any of the six row/column sums do not vanish. Nishizawa [1] suggests a consistency improving method by attempting to correct these inconsistent cycles. It may be wise to check the raw data for these matches; however no ad-hoc correction should take place since they truly record the outcome of the match. Here the number of distinct triad cycles are simply noted (C_N) with a maximum value $n(n-1)(n-2)/6$.

As an alternative the Kwiesielewicz and van Uden [18] criterion for inconsistency is based on a lack of transitivity in the judgements. Where transitivity implies that a relation holds between three elements, such that if it holds between the first and second and it also holds between the second and third it must necessarily hold between the first and third. In other words a violation of $a_{ij} \times a_{jk} = a_{ik}$. They identified six cases, which are summarised in Table 2, illustrated with the special case of a binary matrix.

Table 2. A Catalogue Of Inconsistencies With Examples For A Binary Matrix

Case	General Case			Binary Matrix		
	a_{ij}	a_{ik}	a_{jk}	a_{ij}	a_{ik}	a_{jk}
1	> 1	< 1	> 1	ϑ	$1/\vartheta$	ϑ
2	< 1	> 1	< 1	$1/\vartheta$	ϑ	$1/\vartheta$
3	= 1	> 1	< 1	1	ϑ	$1/\vartheta$
4	= 1	< 1	> 1	1	$1/\vartheta$	ϑ
5	= 1	= 1	< 1	1	1	$1/\vartheta$
6	= 1	= 1	> 1	1	1	ϑ

Their algorithm is summarised as

Either $\ln(a_{ij})\ln(a_{ik}) \le 0$ and $\ln(a_{ik})\ln(a_{jk}) < 0$

Or $\ln(a_{ij}) = 0$ and $\ln(a_{ik}) = 0$ and $\ln(a_{jk}) \ne 0$

$i=1,\ldots,n \; j=1,\ldots,n \; k=1,\ldots,n \; i \ne j \ne k$

where both imply an inconsistency and span the six cases which could arise. Here the number of inconsistencies is simply counted (C_{KU}) with a maximum value of $n(n-1)(n-2)$. Note that only the first two cases in Table 2 exhibit a cycle [1].

In the case of a binary matrix each cell may take one of three values $(\vartheta, 1, 1/\vartheta)$. For a three by three matrix the three free cells may be chosen in $3^3 = 27$ ways, the diagonal values being one and the remainder found by reciprocal symmetry. Of these possibilities, 7 produce a perfect fit ($\lambda_{max} = n = 3$ in this case), 6 are presented in Table 2 and for the remainder (14) no conclusion can be reached. For instance if $a_{ij} < 1$ and $a_{ik} < 1$ no firm conclusion can be reached about a_{jk}. In the terms used here, team j defeats team i as does team k, however this implies nothing about the likely outcome of the match between j and k.

To provide a concrete example the results from the Guinness Premiership for 2007/8, a season during which some misdemeanours were known to occur is considered. This detailed, example is employed to address some key questions. The number of inconsistencies is considered. A decision must be made concerning which matrix to analyse. Given all teams play twice there are duplicate matches and, maybe, contradictory results. Does the choice of the parameter (θ) in the binary AHP matter? The AHP usually employs a discrete scale (1,2,…,9), is this sensible, or should the full (continuous) interval be employed? These questions are now addressed.

5. The First Example

There is some conflict between the two approaches employed for identifying inconsistencies. Essentially they are measuring different "inconsistencies" within a single framework Consider the matrix in Table 3. These are the results of the Guinness Premiership with the method of Boginski et al. [14] employed for conflicting results, that is win/loss matches are merged. More formally if $a_{ij} \neq 1/a_{ji}$ then $a_{ij} = a_{ji} = 1$ is adopted. The team identification and final rankings are summarised in Appendix 1.

Table 3. Typical Analytic Hierarchy Matrix

Team	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	½	½	1	1	2	2	1	1	1	2
2	½	1	½	½	½	½	½	½	1	½	½	1
3	2	2	1	1	½	1	1	1	1	1	2	½
4	2	2	1	1	1	1	1	1	1	1	2	2
5	1	2	2	1	1	2	½	1	1	1	2	2
6	1	2	1	1	½	1	1	2	1	1	1	2
7	½	2	1	1	2	1	1	2	1	½	1	½
8	½	2	1	1	1	½	½	1	2	1	1	1
9	1	1	1	1	1	1	1	½	1	1	1	2
10	1	2	1	1	1	1	2	1	1	1	1	1
11	1	2	½	½	½	1	1	1	1	1	1	1
12	½	1	2	½	½	½	2	1	½	1	1	1

In this case the Nishizawa [1] approach identifies 2 inconsistent cycles ((1,3,12) and (5,7,12)) from a possible total of 220. The procedure of Kwiesielewicz and van Uden [18] identifies 222 inconsistencies from a possible 1,320. For example in addition to those seen above the triple (1,5,3) is also observed.

On examining the teams that gave rise to inconsistencies little can be concluded from C_N since there are only two cycles. For C_{KU} teams 7 and 12 featured in the most inconsistencies, the ranking sequence being (7,12,8,9,3,5,1,11,10,4,6,2). The only interesting entry was for team 2 which only featured 28 times. This team would appear to be the most consistent. Unfortunately the target team (4) lies towards the consistent end of the scale.

Analysing the AHP problem for this example, the dominant eigenvalue is 12.77, with normalised vector (0.09,0.05,0.09,0.10,0.10,0.09,0.09,0.08,0.08,0.09,0.07,0.07), which orders the teams (2,11,12,8,9,7,3,10,6,1,4,5) with consistency index 0.07 and consistency ratio 0.05 which is deemed acceptable. Note that following end-of-season playoffs the ordering was (5,6,4,1,10,3,7,9,11,8,12,2) with the third and fourth teams being tied. For consistency with the AHP calculation this list should be reversed giving (2,12,8,11,9,7,3,10,1,4,6,5). On comparing the two orderings, this is quite a satisfactory outcome. It should be bourn in mind that the points awarded by the league also include additional information, such as the number of tries scored and points difference.

The two approaches for measuring inconsistency are inherently different. In fact the final four cases in Table 2 fail to produce any cycles. To see this consider the third case, $a_{ij} = 1$, $a_{ik} > 1$ *and* $a_{jk} < 1$ which gives $A = \begin{pmatrix} 1 & 1 & \vartheta \\ 1 & 1 & 1/\vartheta \\ 1/\vartheta & \vartheta & 1 \end{pmatrix}$, hence $V = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and finally $S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ which clearly identifies no cycles. An exhaustive comparison of C_N and C_{KU} for all 3×3 binary matrices is presented in Appendix 2.

With repeated fixtures, the first question to address is, which matrix to analyse? How should the information be represented? Is a single matrix sufficient or should a pair of matrices be utilised?

6. Which Matrix to Analyse?

In order to employ the AHP, the matrix must be reciprocal symmetric $\left(a_{ij} = \frac{1}{a_{ji}}\right)$. Within the tournaments explored here a_{ij} would be the result with team i at "home" while a_{ji} is for the "away" leg, versus j. These may well violate the reciprocity condition. There appear to be three options to address this problem. To analyse the first half of the season, then forming the matching symmetric result in the summary matrix. Then a similar strategy is adopted for the second half of the season. The approach of Boginski et al. [14] (merging win/loss matches and if $a_{ij} \neq 1/a_{ji}$ then adopting $a_{ij} = a_{ji} = 1$) did not appear to work well (Table 3 and related calculations). Thus the chronological split, which necessarily preserves more information, is pursued. Saaty [19] suggests an approach in which the geometric mean of the two candidate vectors (for the first and second half of the season) is taken when combining results from a series of matrices.

The season is structured so that if in the first half of the season team *a* plays team *b* (*a* versus *b*), then the reverse fixture (*b* versus *a*) takes place in the second half of the season. This is the most natural division to preserve all the data. It is reasonable to assume that by adopting a shorter (half season) time scale, the results will also be more consistent.

On splitting the results chronologically the first half of the season resulted in $C_N = 40$ a ratio of 0.18 and $C_{KU} = 260$ a ratio of 0.20. The dominant eigenvalue was 13.99 with normalised vector (0.10,0.05,0.10,0.09,0.11,0.11,0.07,0.06,0.08,0.09,0.08,0.07) with associated orders (2,8,7,12,9,11,10,4,3,1,5,6). The consistency index was 0.18 with a consistency ratio of 0.12, which is deemed unacceptable by comparison to the expected consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

The second half of the season gave $C_N = 40$ a ratio of 0.18 and $C_{KU} = 262$ a ratio of 0.20. The dominant eigenvalue was 13.97 with normalised vector (0.09,0.04,0.08,0.10,0.10,0.06,0.10,0.09,0.08,0.10,0.07,0.06) with associated orders (2,6,12,11,3,9,1,8,5,10,7,4). The consistency index was 0.18 with a consistency ratio of 0.12, which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

Neither set of results appears ideal. On merging the vectors by taking the geometric mean of the two candidate vectors [19] the result is (0.10,0.05,0.09,0.10,0.10,0.08,0.08,0.08,0.08,0.09,0.07,0.07) with associated orders (2,12,11,8,9,7,6,3,10,1,4,5).

Since the solutions appear problematic, producing different sequences, the choices made will be investigated. There does appear to be some conflict in the literature about the choice of θ. This is the key parameter employed in the binary AHP; recall that previous authors have chosen *1.2* and *2* without any discussion of their relative merits.

7. Does θ matter?

The classic AHP scale includes the values [1,9] plus their reciprocals as the available scale. In view of this values of θ for the binary AHP in this range are examined, extending the simple choice of *1.2* or *2*. Since the value of 1 corresponds to a draw, the lowest value investigated was 1.2. The method of Boginski et al. [14] is employed for conflicting results (if $a_{ij} \neq 1/a_{ji}$ then $a_{ij} = a_{ji} = 1$) in the Guinness Premiership data. The key statistics are displayed in Table 4 (for $\theta = 2$ these are the results reported for Table 3) with team ranking and the final column (True) being the actual result in Table 5.

The correlation (in Table 5) is employed as a measure of the difference between the team ranking, for a given θ, and the result actually observed. For the example considered here $\theta = 1.2$ gives the best match, which deteriorates as θ increases. A small value of θ seems preferable. However an alternative to the binary AHP is to employ a mapping to a continuous scale.

Table 4. Key Statistics As θ Varies

θ	C_N	Ratio	C_{KU}	Ratio	λ_{max}	CI	CR	Conclusion
1.2	2	0.0091	222	0.1682	12.05	0.0046	0.0032	acceptable
2	2	0.0091	222	0.1682	12.77	0.0703	0.0483	acceptable
3	2	0.0091	222	0.1682	14.10	0.1904	0.1308	unacceptable
4	2	0.0091	222	0.1682	15.58	0.3258	0.2238	unacceptable
5	2	0.0091	222	0.1682	17.16	0.4687	0.3219	unacceptable
6	2	0.0091	222	0.1682	18.78	0.6160	0.4231	unacceptable
7	2	0.0091	222	0.1682	20.43	0.7661	0.5262	unacceptable
8	2	0.0091	222	0.1682	22.10	0.9182	0.6306	unacceptable
9	2	0.0091	222	0.1682	23.79	1.0717	0.7361	unacceptable

Table 5. Ranking Of Teams As θ Varies

	Θ									
Team	**1.2**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**True**
1	3	3	3	3	3	3	3	3	3	3.5
2	12	12	12	12	12	12	12	12	12	12
3	6	6	6	6	6	6	6	6	6	6
4	2	2	2	2	2	2	2	2	2	3.5
5	1	1	1	1	1	1	1	1	1	1
6	4	4	4	5	5	5	5	5	5	2
7	7	7	5	4	4	4	4	4	4	7
8	9	9	9	10	10	10	10	10	10	10
9	8	8	8	8	9	9	9	9	9	8
10	5	5	7	7	7	7	7	8	8	5
11	10	11	11	11	11	11	11	11	11	9
12	11	10	10	9	8	8	8	7	7	11
Correlation	0.97	0.96	0.93	0.89	0.87	0.87	0.87	0.82	0.82	

8. Continuous Scale

To replace θ in the binary AHP, for a specific match in the Guinness Premiership, the absolute value of the score difference is mapped onto the conventional AHP scale, in this case [0,70]→[1,9.5] (the maximum score difference being, Leicester 73 Bristol 3 on 25 April 2009). This is done employing a simple linear equation which maps an absolute score difference (0→1 and 70→9.5), that is *AHP = 1.00 + 0.121 Abs. Score diff.* (where $0.121 \approx \frac{9.5-1}{70}$).

On splitting the results chronologically the first half of the season resulted in C_N = 28 a ratio of 0.13 and C_{KU} = 270 a ratio of 0.20. The dominant eigenvalue was 15.52 with normalised vector (0.09,0.04,0.09,0.09,0.12,0.14,0.07,0.05,0.09,0.08,0.08,0.06) with associated orders (2,8,12,7,10,11,1,4,3,9,5,6). The consistency index was 0.32 with a consistency ratio of 0.22, which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17]. To produce the rank comparable to the league position in Table 6 the procedure is to recognise that team 1 lies at position 7, team 2 at 1 and so on, producing a list (7,1,9,8,11,12,4,2,10,5,6,3) then the sequence is reversed by effectively subtracting each entry from 13 to give (6,12,4,5,2,1,9,11,3,8,7,10). These are the values presented in Table 6.

The second half of the season gave C_N = *26* a ratio of 0.12 and C_{KU} = 276 a ratio of 0.21. The dominant eigenvalue was 15.93 with normalised vector (0.09,0.03,0.07,0.12,0.13,0.08,0.10,0.08,0.08,0.12,0.06,0.06) with associated orders (2,12,11,3,8,9,6,1,7,4,10,5). The consistency index was 0.18 with a consistency ratio of 0.25, which is deemed unacceptable by comparison to the consistency index generated by Tummala and

Wan [16] and Tummala and Ling [17].

Neither set of results appears ideal. On merging the vectors by taking the geometric mean of the two candidate vectors [19] the result is (0.09,0.03,0.08,0.10,0.12,0.10,0.08,0.06,0.08,0.10,0.07,0.06) with associated orders (2,12,8,11,3,7,9,1,10,4,6,5). Which does bear some similarity to the true positions. See Table 6 for a comparison of the results.

Table 6. Team Positions For The Continuous And Discrete Models

Team	Continuous			Discrete			True
	First Half	Second Half	Merged	First Half	Second Half	Merged	
1	6	5	5	7	5	4	3.5
2	12	12	12	12	12	12	12
3	4	9	8	4	9	7	6
4	5	3	3	5	3	3	3.5
5	2	1	1	2	1	1	1
6	1	6	2	1	4	2	2
7	9	4	7	8	6	9	7
8	11	8	10	11	10	10	10
9	3	7	6	3	7	6	8
10	8	2	4	9	2	5	5
11	7	10	9	6	8	8	9
12	10	11	11	10	11	11	11
Correlation	0.795	0.820	0.960	0.743	0.904	0.963	

Clearly the merged vector in both cases provides the best solution, being closest the true values.

Is a continuous scale, considered above, consistent with the AHP philosophy? Would restricting choices to (1,2,…,9) and their reciprocals be more effective? To see if any improvement can be achieved the previous continuous mapping is adjusted to the more conventional discrete scale.

9. Discrete Scale

Adopting a discrete scale and splitting the results for the Guinness Premiership chronologically, the first half of the season resulted in $C_N = 6$ a ratio of 0.03 and $C_{KU} = 280$ a ratio of 0.21. The dominant eigenvalue was 14.29 with normalised vector (0.08,0.05,0.09,0.09,0.11,0.13,0.08,0.06,0.09,0.07,0.09,0.07) with associated orders (2,8,12,10,7,1,11,4,3,9,5,6). The consistency index was 0.20 with a consistency ratio of 0.14, which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

The second half of the season gave $C_N = 4$ a ratio of 0.02 and $C_{KU} = 212$ a ratio of 0.16. The dominant eigenvalue was 14.34 with normalised vector (0.10,0.04,0.07,0.11,0.12,0.10,0.08,0.07,0.08,0.11,0.07,0.05) with associated orders (2,12,8,3,11,9,7,1,6,4,10,5). The consistency index was 0.21 with a consistency ratio of 0.15, which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

Neither set of results appears ideal on comparison to the true positions. On merging the vectors by taking the geometric mean of the two candidate vectors [19] the result is

(0.09,0.04,0.08,0.10,0.12,0.11,0.08,0.06,0.09,0.09,0.08,0.06)

with associated orders (2,12,8,7,11,3,9,10,1,4,6,5). This bears some similarity to the true positions.

Table 6 contains a comparison of the results. The discrete approach appears marginally superior, and probably the most natural to adopt. The frequency with which the teams appear in C_{KU} , a measure of their inconsistency is summarised in Table 7.

It would appear that overall teams 2 and 6 are the most consistent, while the remainder are very similar in their performance. It might be argued (with the benefit of hindsight, "bloodgate"), that team 4 was more inconsistent in the second half of the season. However team 3 shows the greatest change closely followed by teams 5, 9 and 10.

Table 7. Inconsistencies Found By C_{KU}

First Half			Second Half		
Team	Count	Percent	Team	Count	Percent
6	54	6.43	2	30	4.72
8	60	7.14	5	38	5.97
3	60	7.14	10	42	6.60
11	64	7.62	6	46	7.23
12	64	7.62	9	46	7.23
2	70	8.33	11	52	8.18
1	70	8.33	8	60	9.43
4	72	8.57	12	60	9.43
10	78	9.29	7	60	9.43
5	82	9.76	1	62	9.75
7	82	9.76	4	66	10.38
9	84	10.00	3	74	11.64
Total	840			636	

10.Conclusion Of The First Example

The correlations in Table 6 would suggest that the ideal approach is to use a discrete scale [1,9]. On examining the contradictions [18], which are summarised in Table 7, the most frequently occurring team between the two halves of the season is not consistent. There appears little evidence for any misdemeanours, however some were known to occur. The teams exhibiting the greatest change between the two halves of the season were not that which was known to be cheating! The identified offences occurred during the latter part of the season, when the pressure to win games became greatest.

A second example, in which no irregularity was reported, is now briefly considered. The results are taken from the English Football Association. This is also a round robin tournament. However, when compared to the Rugby Union, scores are generally much lower numerically and ties are extremely common.

11.A Second Example

A similar approach is now applied to the results for the Barclays Premier League season, 2008/9. The team details and final rankings are summarised in Appendix 1.

To replace θ for a specific match in the Barclays Premiership, the absolute value of the score difference is mapped onto the conventional AHP scale, in this case $[0,6]\rightarrow[1,9.5]$. The mapping was adjusted to the more conventional discrete scale (1,2,…,8,9).

On splitting the results chronologically the first half of the season resulted in $C_N = 261$ a ratio of 0.23 and $C_{KU} = 1{,}566$ a ratio of 0.23. The dominant eigenvalue was 41.52 with normalised vector (0.06,0.07,0.03,0.04,0.08,0.05,0.06,0.04,0.06,0.06,0.08,0.03,0.05,0.04,0.04,0.04,0.06,0.03,0.03,0.04) with associated orders (18,3,19,12,16,20,15,8,14,4,6,13,10,17,7,1,9,2,5,11). The consistency index was 1.13 with a consistency ratio of 0.48 (obtained using the cubic approximation), which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

The second half of the season gave $C_N = 216$ a ratio of 0.19 and $C_{KU} = 1{,}296$ a ratio of 0.19. The dominant eigenvalue was 39.23 with normalised vector (0.08,0.04,0.04,0.03,0.06,0.09,0.05,0.03,0.08,0.05,0.09,0.04,0.04,0.04,0.05,0.03,0.05,0.03,0.05,0.03) with associated orders (18,16,8,4,20,13,3,2,14,12,15,19,10,17,7,5,1,9,11,6). The consistency index was 1.01 with a consistency ratio of 0.43 (obtained using the cubic approximation), which is deemed unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

Neither set of results appears ideal. On merging the vectors by taking the geometric mean of the two candidate vectors [19] the result is

(0.07,0.05,0.03,0.04,0.07,0.07,0.06,0.03,0.07,0.05,0.08,0.04,0.04,0.04,0.04,0.03,0.05,0.03,0.04,0.03)

with associated orders (18,16,3,20,8,4,12,19,14,15,13,2,10,17,7,6,1,5,9,11). Not that close to the true values. Table 8 contains a comparison of the results.

Table 8. Comparison Of The Team Ranking For The Barclays Premier League

Team	First Half	Second Half	Merged	True
1	5	4	4	4
2	3	13	9	6
3	19	14	18	15
4	11	17	15	13
5	2	5	3	3
6	10	1	5	5
7	6	6	6	7
8	13	18	16	17
9	4	3	2	2
10	8	8	8	10
11	1	2	1	1
12	17	11	14	19
13	9	15	10	18
14	12	12	12	14
15	14	10	11	12
16	16	19	19	16
17	7	7	7	8
18	20	20	20	20
19	18	9	13	9
20	15	16	17	11
Correlation	0.80	0.84	0.86	

Again the merged vector provides the best match to the true result.

On examining the contradictions that are summarised in Tables 9 and 10 the most frequently occurring team between the two halves of the season is not consistent. However teams 11 and 17 hold an identical position, while team 1 exhibits by far the greatest change.

Table 9. C_N For The Barclays Premier League

First Half			Second Half		
Team	Count	Percent	Team	Count	Percent
11	21	2.68	11	14	2.16
5	28	3.58	1	25	3.86
4	31	3.96	18	25	3.86
2	32	4.09	3	26	4.01
18	33	4.21	5	26	4.01
14	37	4.73	8	31	4.78
20	37	4.73	10	31	4.78
16	38	4.85	4	32	4.94
6	39	4.98	15	32	4.94
8	39	4.98	20	32	4.94
3	40	5.11	2	33	5.09
9	40	5.11	6	34	5.25
10	40	5.11	13	34	5.25
19	41	5.24	14	35	5.40
12	42	5.36	9	36	5.56
13	46	5.87	19	36	5.56
1	49	6.26	12	37	5.71
7	49	6.26	16	39	6.02
15	50	6.39	7	44	6.79
17	51	6.51	17	46	7.10
Total	783			648	

Table 10. C_{KU} For The Barclays Premier League

First Half			Second Half		
Team	Count	Percent	Team	Count	Percent
11	126	2.68	11	84	2.16
5	168	3.58	1	150	3.86
4	186	3.96	18	150	3.86
2	192	4.09	3	156	4.01
18	198	4.21	5	156	4.01
14	222	4.73	8	186	4.78
20	222	4.73	10	186	4.78
16	228	4.85	4	192	4.94
6	234	4.98	15	192	4.94
8	234	4.98	20	192	4.94
3	240	5.11	2	198	5.09
9	240	5.11	6	204	5.25
10	240	5.11	13	204	5.25
19	246	5.24	14	210	5.40
12	252	5.36	9	216	5.56
13	276	5.87	19	216	5.56
1	294	6.26	12	222	5.71
7	294	6.26	16	234	6.02
15	300	6.39	7	264	6.79
17	306	6.51	17	276	7.10
Total	4698			3888	

Clearly the percentages reported in Tables 9 and 10 are identical; the consistency checks are apparently in harmony. This finding is probably fortuitous given the results in Table 4 and the related discussion. There seems to be no clear reason why C_N should equal C_{KU} in this case.

It would appear that team 17 offers the most violations, being the most inconsistent. Ranking the remaining teams gives the sequence (17,7,15,13,12,16,19,9,6,14,1,10,8,20,2,3,4,18,5,11).

It may be argued that since the maximum score difference was only 6 the full range of the AHP is not utilised. Since low scoring games dominate and draws are not unusual this data set will not be subjected to further consideration. Interest now returns to the Guinness Premiership.

12. Possible Model Extensions

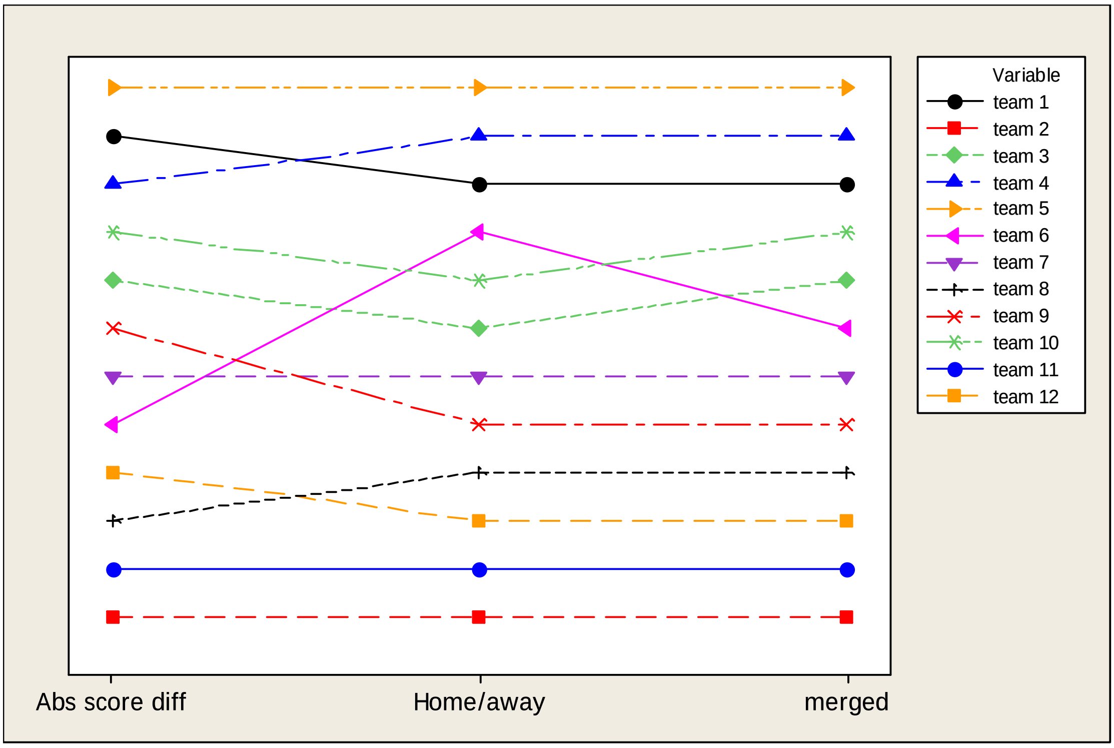

Figure 1. Parallel Coordinates Plot Comparing The Team Positions

In considering the absolute score difference, an important factor that has not been considered is the, so-called, home team advantage. This is now explored. The same chronological split is employed and a home win assigned a value of $\vartheta = 2$. The analysis parallels that for score difference. The previous problems recur; with the consistency criterion suggesting that the results are unacceptable by comparison to the consistency index generated by Tummala and Wan [16] and Tummala and Ling [17].

The original solution for the absolute score difference can be compared to that obtained on simply modelling home/away wins. These two solutions are then merged using the geometric mean of the two candidate vectors [19]. A useful presentation (Figure 1) is the parallel coordinates plot (see [20] and more recent papers by the same author). The most successful team (5) is shown at the top of the figure and is denoted by a right pointing triangle, while the least successful team (2) forms the base of the figure and is denoted by a square. The remaining teams are shown in rank order.

The team positions are remarkably consistent with only team 6 switching drastically. However this effect is diminished on plotting the relevant vectors (Figure 2) as opposed to the ranks.

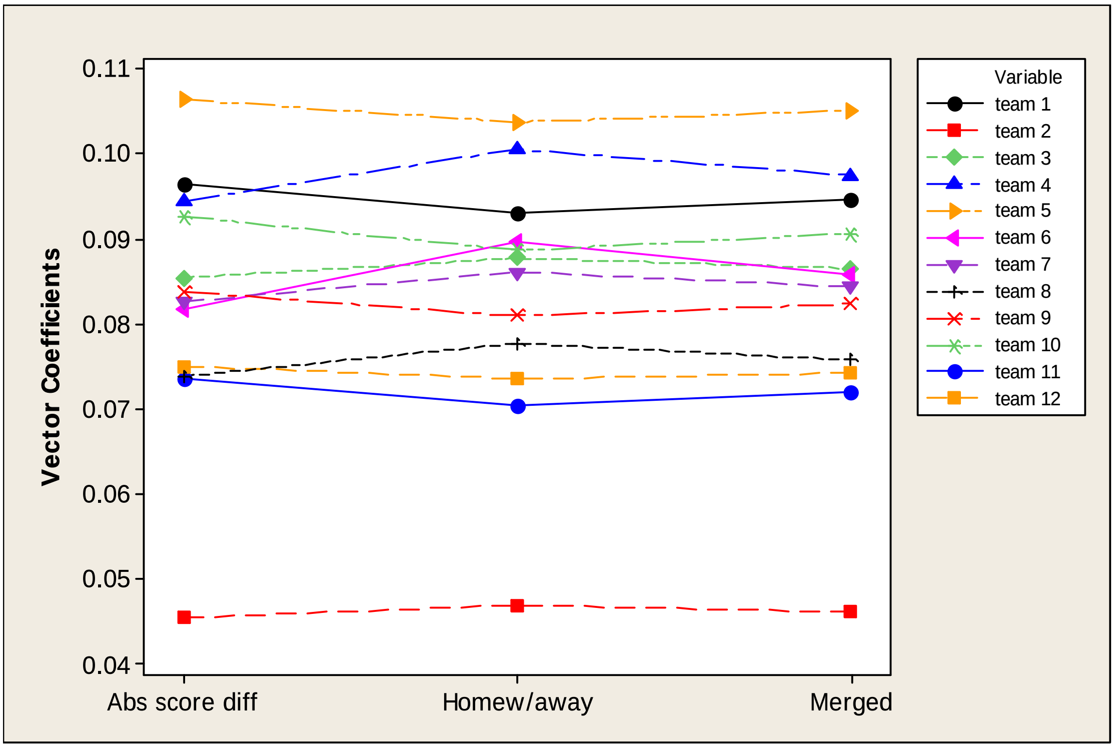

Figure 2. Parallel Coordinates Plot Comparing The Vectors

Building a more complex model has not realised radically different results. Various other extensions to the model might be applied. However unless these are motivated by the underlying problem it might suggest desperation in seeking any model that might work.

13.Conclusion

The aim of this manuscript was to assess the effectiveness of the AHP when employed to model the performance of teams in round robin sports tournaments. In particular will the AHP assist in investigating any evidence of inconsistencies, which might be indicative of a teams cheating. The "jury is still out"; there is no evidence of non-sporting behaviour in the examples considered.

There is evidence of serious events [2] that affected the Guinness Premiership results. These are not indicated by the results presented here. Maybe no major cheating took place. It could be that the Guinness Premiership results were not materially affected by the illegal strategy one team adopted. It might be suggested that results from the Barclays Premier League lacks discrimination, most matches exhibit very low scores with a high proportion of draws,

Alternately, the methods reviewed and utilised here do not work! This should not be taken as a critique of the AHP, simply recognition that alternate approaches ([11] or [14]) would probably be more appropriate. It should be recalled that the absolute score difference has been modelled as well as consideration of home team advantage. This may not a sensible application of the AHP; hopefully this investigation will encourage further work, in particular the adoption of alternate models. The numerous applications of the AHP techniques as indicated by the vast array of published papers stress its applicability. In addition the thorough investigation of inconsistency techniques should aid in their wider utilisation.

Of particular mention of an application of the AHP is its use for ranking sports teams [21] who initially applied the method to soccer. This was a more traditional (appropriate) use of the methodology, it utilised six criteria

i. the facility

ii. the coach

iii. the players

iv. the fans

v. the previous seasons performance

vi. the current performance

which were considered by an expert to construct pair wise comparisons and hence to compare the teams. It is also necessary when applying the procedure to provide a weighting between these individual scales. Unfortunately this extensive information is not available in our case. At even the most basic level, the approach may not be employed for the data considered here. For instance the same teams do not arise for criteria (v) and (vi). The approach was also applied to basketball [22], in this case 4 experts utilised 6 criteria. Interestingly the problem criteria (v and vi) were replaced by an assessment of the "team's tradition".

Many models may be employed to examine this form of data; typically these are special cases of the generalised linear model. One particular example is the Bradley-Terry model, on which there is an extensive literature. A bibliography on the method of paired comparisons [23] lists several hundred entries. The book by David [24] does provide a widespread examination of paired-comparison models, although it does not focus exclusively on the Bradley-Terry model. A more recent review of the model is given by Simons and Yao [25]. Care should be taken when adopting this approach since not all software allows for ties [26].

Clearly there are numerous successful applications of numerical techniques to sports data, although the AHP may not be an ideal approach to round robin tournaments.

14. References

[1] Nishizawa K. A consistency improving method in binary AHP. *Journal Of The Operations Research Society Of Japan*. 1995, **38**(1): 21-33.

[2] Kelso, P. 2009, The Telegraph, 26 August 2009, http://www.telegraph.co.uk/sport/rugbyunion/club/6089556/Bloodgate-European-Rugby-Cups-damning-judgment.html accessed May 2012.

[3] ESPN scrum, 2009, http://www.scrum.com/premiership/rugby/series/26028.html?template=results accessed May 2012.

[4] Wiki, 2009, http://en.wikipedia.org/wiki/2008%E2%80%9309_Premier_League accessed May 2012 see also http://soccernet.espn.go.com/news/story?id=546813andcc=5739 accessed May 2012.

[5] Saaty, T.L. Scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*. 1977, **15**(3): 234-281.

[6] Saaty, T.L. *The analytic hierarchy process*. New York: McGraw-Hill, 1980.

[7] Triantaphyllou E., Lootsma F.A., Pardalos P.M. and Mann S.H. On the evaluation and application of different scales for quantifying pairwise comparisons in fuzzy sets. *Journal of Multi-Criteria Decision Analysis*. 1994, **3**(3): 133-155.

[8] Barzilai, J. Deriving weights from pairwise comparison matrices. *Journal of the Operational Research Society*. 1997, **48**: 1226-1232.

[9] Triantaphyllou E., Pardalos P.M. and Mann, S.H. A minimization approach to membership evaluation in fuzzy-sets and error analysis. *Journal Of Optimization Theory And Applications*. 1990, **66**(2): 275-287.

[10] Choo, E.U. and Wedley, W.C. A common framework for deriving preference values from pairwise comparison matrices. *Computers and Operations Research*. 2004, **31**: 893-908.

[11] Barnett V. and Hilditch S. The effect of an artificial pitch surface on home team performance in football (soccer). *Journal Of The Royal Statistical Society Series A-Statistics In Society*. 1993, **156**: 39-50.

[12] Dixon M.J. and Coles S.G. Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics-Journal Of The Royal Statistical Society Series C.* 1997, **46**(2): 265-280.

[13] Takahashi I. AHP applied to binary and ternary comparisons. *Journal Of The Operations Research Society Of Japan*. 1990, **33**(3): 199-206.

[14] Boginski V., Butenko S. and Pardalos P.M. Matrix-based methods for college football orderings, In: *Economics, Management and Optimization in Sports*. Butenko S., Gil-Lafuente J. and Pardalos P. (eds.). Berlin: Springer-Verlag. 2004.

[15] Alonso J.A. and Lamata M.T. Consistency in the analytic hierarchy process: A new approach. *International Journal Of Uncertainty Fuzziness And Knowledge-Based Systems*. 2006, **14**(4): 445-459.

[16] Tummala V.M.R. and Wan Y. On the mean random inconsistency index of analytic hierarchy process (AHP). *Computers and Industrial Engineering*. 1994, **27**(1-4): 401-404.

[17] Tummala V.M.R. and Ling H. A note on the computation of the mean random consistency index of the analytic hierarchy process (AHP). *Theory And Decision***.** 1998, **44**(3): 221-230.

[18] Kwiesielewicz M. and van Uden E. Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Computers and Operations Research*. 2004, **31**: 713-719.

[19] Saaty, T.L. The seven pillars of the analytic hierarchy process. *Lecture Notes In Economics And Mathematical Systems*. 2001, **507**: 15-37.

[20] Inselberg. A. The plane with parallel coordinates. *The Visual Computer.* 1985, **1**(4): 69–91.

[21] Sinuany-Stern, Z. Ranking of sports teams via the AHP. *Journal of the Operational Research Society.* 1988, **39**(7): 661-667.

[22] Sinuany-Stern Z., Israeli Y. and Bar-Eli M. Application of the analytic hierarchy process for the evaluation of basketball teams. *Int. J. Sport Management and Marketing*. 2006, **1**(3): 193-207.

[23] Davidson R.R. and Farquhar P.H. A bibliography on the method of paired comparisons. *Biometrics.* 1976, **32**: 241-252.

[24] David H.A. *The method of paired comparisons*(2nd ed). Oxford Univ. Press, 1988.

[25] [Simons G. and Yao Y.-C. Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.* 1999, **27**: 1041-1060.

[26] Firth D. Bradley-Terry models in R. *Journal of Statistical Software.* 2005, **12**(1): 1-12.

Appendix 1 – Team Identification

The teams participating in the Guinness Premiership in the, 2008/9 season are shown in Table 11.

Table 11. Teams Participating In The Guinness Premiership

Team	Identifier	Position	Final Position
Bath	1	4	3.5
Bristol	2	12	12
Gloucester	3	6	6
Harlequins	4	3	3.5
Leicester	5	1	1
London Irish	6	2	2
London Wasps	7	7	7
Newcastle	8	10	10
Northampton	9	8	8
Sale	10	5	5
Saracens	11	9	9
Worcester	12	11	11

The position shown is the teams ranking at the end of the full set of fixtures, while the final position shows the result after the top four teams participated in an end of season play off.

The teams participating in the Barclays Premier League, 2008/9 season are shown in Table 12, plus their ranking at the end of the full set of fixtures.

Table 12.Teams Participating In The Barclays Premier League

Team	Identifier	Position
Arsenal	1	4
Aston Villa	2	6
Blackburn Rovers	3	15
Bolton Wanderers	4	13
Chelsea	5	3
Everton	6	5
Fulham	7	7
Hull City	8	17
Liverpool	9	2
Manchester City	10	10
Manchester United	11	1
Middlesbrough	12	19
Newcastle United	13	18
Portsmouth	14	14
Stoke City	15	12
Sunderland	16	16
Tottenham Hotspur	17	8
West Bromwich Albion	18	20
West Ham United	19	9
Wigan Athletic	20	11

Appendix 2 – A Comparison Of C_N And C_{KU} For All 3×3 Binary Matrices

A simple trial is conducted on all binary 3×3 matrices to compare C_N and C_{KU} for $\vartheta = 2$ some 27 (3^3) options. The results are summarised in Table 13, where the matrix entries selected (a_{12}, a_{23}, a_{31}) are those used by Tummala and Ling [17] and Tummala and Wan [16] in their presentations.

Table 13. A Comparison Of C_N And C_{KU} For All 3×3 Binary Matrices

Identifier	a_{12}	a_{23}	a_{31}	C_N	C_{KU}	λ_{max}	CI	CR
1	0.5	0.5	0.5	1.00	1.00	3.50	0.25	0.50
2	0.5	0.5	1	0.00	0.33	3.22	0.11	0.22
3	0.5	0.5	2	0.00	0.00	3.05	0.03	0.05
4	0.5	1	0.5	0.00	0.33	3.22	0.11	0.22
5	0.5	1	1	0.00	0.33	3.05	0.03	0.05
6	0.5	1	2	0.00	0.00	3.00	0.00	0.00
7	0.5	2	0.5	0.00	0.00	3.05	0.03	0.05
8	0.5	2	1	0.00	0.00	3.00	0.00	0.00
9	0.5	2	2	0.00	0.00	3.05	0.03	0.05
10	1	0.5	0.5	0.00	0.33	3.22	0.11	0.22
11	1	0.5	1	0.00	0.33	3.05	0.03	0.05
12	1	0.5	2	0.00	0.00	3.00	0.00	0.00
13	1	1	0.5	0.00	0.33	3.05	0.03	0.05
14	1	1	1	0.00	0.00	3.00	0.00	0.00
15	1	1	2	0.00	0.33	3.05	0.03	0.05
16	1	2	0.5	0.00	0.00	3.00	0.00	0.00
17	1	2	1	0.00	0.33	3.05	0.03	0.05
18	1	2	2	0.00	0.33	3.22	0.11	0.22
19	2	0.5	0.5	0.00	0.00	3.05	0.03	0.05
20	2	0.5	1	0.00	0.00	3.00	0.00	0.00
21	2	0.5	2	0.00	0.00	3.05	0.03	0.05
22	2	1	0.5	0.00	0.00	3.00	0.00	0.00
23	2	1	1	0.00	0.33	3.05	0.03	0.05
24	2	1	2	0.00	0.33	3.22	0.11	0.22
25	2	2	0.5	0.00	0.00	3.05	0.03	0.05
26	2	2	1	0.00	0.33	3.22	0.11	0.22
27	2	2	2	1.00	1.00	3.50	0.25	0.50

Where identifiers 27, 1, 11, 18, 20, 17 correspond to cases 1,…,6 in Table 2 and identifiers 6, 8, 12, 14, 16, 20, 22 provide an exact fit ($\lambda_{max} = n = 3$).

Clearly, even in these simple cases C_N is not consistent with C_{KU}. Obviously C_N and C_{KU} are independent of ϑ, this is not the case for the eigenvalue and related indices, for which exact results are available. Using the notation of Tummala and Ling [17] and Tummala and Wan [17], consider the matrix $A = \begin{pmatrix} 1 & x & 1/z \\ 1/x & 1 & y \\ z & 1/y & 1 \end{pmatrix}$ adopting the row geometric mean to estimate the priority vector [8] with an appropriate normalisation term gives $v = \frac{1}{N}\begin{pmatrix} \left(\frac{x}{z}\right)^{1/3} \\ \left(\frac{y}{x}\right)^{1/3} \\ \left(\frac{z}{y}\right)^{1/3} \end{pmatrix}$ where $N = \left(\frac{x}{z}\right)^{1/3} + \left(\frac{y}{x}\right)^{1/3} + \left(\frac{z}{y}\right)^{1/3}$. Solving $A\underline{v} = \lambda \underline{v}$ gives $\lambda_{\max} = 1 + (xyz)^{1/3} + (xyz)^{-1/3}$ as previously reported and $CI = \frac{1}{2}(xyz)^{1/3} + \frac{1}{2}(xyz)^{-1/3} - 1$.