COMPARISON ACCURACY – IMPLICATIONS FOR DERIVING PRIORITIES AND CONSISTENCY

William C. Wedley
Beedie School of Business
Simon Fraser University
Burnaby, B. C. CANADA
E-mail: wedley@sfu.ca

ABSTRACT

Some paired comparisons are more accurate than others. Which ones? If we knew, then we could use the more accurate comparisons to derive the priority vector and maybe discard the less accurate ones. Using experiments where the true comparison values are known, this study investigates which comparisons are more accurate and whether a priority vector should be based upon partial information. In general, comparisons to the least dominant alternative are found to be more accurate. It is recommended that the first $n-1$ comparisons be made in reference to the least dominant alternative. Thereafter, changes in priorities and degree of predicted consistency can be used to determine whether the comparison process can be stopped. An ancillary finding is that consistency alone is not a reliable indicator of accuracy. In order to achieve accuracy, the DM must be both knowledgeable about the task being evaluated and conscientious in making evaluations.

Keywords: Comparison accuracy, incomplete comparisons, priorities, consistency.

1. Introduction

Each comparison matrix requires $n(n-1)/2$ comparisons. A typical problem involves many matrices. A common complaint about AHP/ANP is that too many comparisons are required. A procedure that reduces the number of required comparisons would make AHP/ANP more efficient and satisfying.

2. Literature Review

Concern for the large number of comparisons goes back to the early days of AHP. Harker (1987a, b, c) was instrumental in providing procedures for reduced comparisons. He developed (1) two methods for determining missing values in partially completed matrices, (2) a gradient method for choosing the next comparison that would provide the greatest information and (3) stopping rules for terminating evaluations before all comparisons are done. Wedley (1993) utilized Harker's ideas to develop regression equations that predict the final consistency ratio if the DM proceeded to complete all comparisons. This information is useful as an adjunct for Harker's stopping rules.

3. Hypotheses/Objectives

In spite of the roadmap provided by Harker, very few software procedures have been designed to reduce the number of comparisons. It is the hypothesis of this paper that some comparisons are more accurate than others. With knowledge about accuracy, the objective is to make recommendations that improve the comparison process.

4. Research Design/Methodology/Data

In order to determine comparison accuracy, we must work with examples that have known true answers. Let $V=(v_1,v_2 \ldots v_n)^T$ be a vector of ratio values for some property of n objects. $M=[v_i/v_j]$ will be the perfectly accurate comparison matrix. Let $T=[t_{ij}]$ be the DM's matrix of comparison estimates such that $t_{ij} \approx v_i/v_j$, $T \approx M$. From T, an AHP priority vector is derived, $P=(p_1,p_2 \ldots p_n)^T$, with $N=[p_i/p_j]$. T-M is then used to measure comparison accuracy while N-M is used to measure priority accuracy.

Empirical data for this study was provided by Bernasconi et al (2010). They conducted three estimation experiments (distances, probability, and rainfall) where the true answer was known to the experimenters.

5. Data Analysis

Each empirical experiment had 5 alternatives that were evaluated by 69 subjects. Since each T requires 10 direct comparisons the total number of comparison values for each experiment is 1380 (690 direct and 690 reciprocal). Analysis of average absolute and proportional errors revealed the following important findings:

1. The distance experiment has the best accuracy. Closely behind is the probability experiment. In comparison, the rainfall experiment has very poor accuracy.
2. Reciprocal comparisons have lower error values because the more dominant item is the unit for comparison. Conversely, direct comparisons have larger error values because they have the less dominant item as the unit.
3. In matrix columns with large magnitude units, absolute errors are small. Those columns have more reciprocal comparisons.
4. Proportional errors adjust for differences in magnitude of the column units. Proportional results indicate that accuracy is better in columns with lower magnitude units.

To reveal accuracy relationships, multiple regressions were conducted with the following variables: (1) true magnitude v_i/v_j (2) experiment type (Distance=0), (3) column magnitude effect (smallest unit column, $P_1=0$) and (4) order of comparison entry. Those variables were used in three regressions: (1) all comparisons, (2) direct comparisons and (3) direct comparisons exclusive of rainfall data (Table 1). Important findings are:

1. There is positive correlation between absolute errors and the true magnitude of an object. This effect disappears with proportional calculations. Particularly for direct comparisons, larger magnitude items exhibit lower proportional error.
2. Positive coefficients for the probability and rainfall experiments indicate that they have larger errors than the distance experiment.
3. Compared to the smallest unit column ($P_1=0$), other columns with larger units of measure tend to have larger errors (i.e. coefficients are positive).

4. The positive coefficients for comparison order signifies that proportional errors are larger for later comparisons, although the effect is not significant when rainfall data are excluded.

Table 1 Multiple Regression Results from Three Perspectives

	ABSOLUTE ERRORS			PROPORTIONAL ERRORS		
	All Comparisons	**Direct Comparisons**	**Direct without Rainfall**	**All Comparisons**	**Direct Comparisons**	**Direct without Rainfall**
# Observations	4140	2070	1380	4140	2070	1380
Multiple R	0.470	0.477	0.213	0.480	0.746	0.450
R-Square	0.221	0.228	0.045	0.230	0.557	0.202
Standard Error	1.569	1.792	1.651	3.152	3.222	2.337
Variable	*Coefficients*			*Coefficients*		
Constant	-0.362	1.085	1.024	-1.568	*0.344*	0.997
True Comparison Magnitude	0.310	0.053	0.063	*0.015*	-0.212	-0.213
Probability = 1 (Distance=0)	*0.087*	0.115	*0.146*	0.284	0.428	0.550
Rainfall =1 (Distance=0)	1.504	1.074	n.a.	2.744	2.052	n.a.
P_2=1 (P_1=0)	0.417	0.531	0.671	0.346	*0.171*	*0.057*
P_3 = 1 (P_1=0)	1.289	1.272	0.825	1.997	1.661	*0.322*
P_4 = 1 (P_1=0)	*0.139*	0.576	0.625	0.552	1.586	1.267
P_5 = 1 (P_1=0)	0.716	2.299	1.725	3.121	9.683	5.910
Comparison Order	0.029	-0.042	*-0.030*	0.160	0.066	*0.012*

p<0.05 for all coefficients except *underlined and italicized* which are not significant (p>0.05)

Besides comparison accuracy, priority vector accuracy can be analyzed from M-N. The priority vectors considered included the eigenvector method based upon all comparisons and 7 other methods based upon just 4 spanning tree comparisons (see Table 2 for a listing). For the distance experiment, n-1 comparisons to the lowest magnitude unit (P_1) had statistically lower error, even better than the eigenvector method. For probability data, the eigenvector solution was the best, but not statistically better than the best n-1 method ($P_{diagonal}$).

Table 2 Frequency of Vectors being Most Accurate or Tied for Most Accurate

Vector Source	Mean Absolute Errors			Total #, %	Mean Absolute Proportion Errors			Total #, %
	Distance	Probability	Rainfall		Distance	Probability	Rainfall	
P_1 Lowest Unit Column	38	36	7	81 (38.8%)	36	33	7	76 (36.7%)
P_2 2nd Lowest Unit Column	9	5	5	19 (9.1%)	12	3	5	20 (9.7%)
P_3 3rd Lowest Unit Column	5	2	24	31 (14.8%)	6	3	23	32 (15.5%)
P_4 4th Lowest Unit Column	3	2	20	25 (12.0%)	3	2	27	32 (15.5%)
P_5 Largest Unit Column	5	3	8	16 (7.7%)	2	3	3	8 (3.9%)
$P_{diagonal}$	2	12	2	16 (7.7%)	2	9	3	14 (6.8%)
P_{random}	5	4	2	11 (5.3%)	6	4	4	14 (6.8%)
$P_{eigenvector}$ (all comparisons)	7	7	4	18 (8.6%)	7	13	3	23 (11.0%)

Table 2 summarizes how many times a method is most accurate or tied as most accurate. Notice that P_1 from the lowest unit column was best most often. Of 207 comparison matrices, P_1 was best 81 times for MAE (39%) and 76 times for MAPE (37%). Only 7 of those successes came from the less reliable rainfall data. Successes for P_3 and P_4 came almost exclusively from the rainfall data set. Interestingly, the eigenvector routine was only supreme about 9-11% of the time.

6. Discussion and Limitations

Unlike most studies where V is unknown, this study has the luxury of being able to measure the true error of comparisons and priority vectors. Proportional results that adjust for comparison magnitude indicate that comparisons to the lowest magnitude object are the most accurate. The reason for this behavioral tendency is not clear. Perhaps before doing comparisons, people subconsciously rank the objects in their mind or think of the extreme points of the range.

In three other studies, (Wedley et al, 1993; Carmone et al 1997; Wedley, 2009) support was found for using the lowest ranked object as the unit for the first n-1 comparisons. However, it should be noted that the Carmone et al (1997) data were simulated. If the relationship to the least dominant object holds for simulated data, then it raises a question as to whether or not the cause is behavioral.

A useful check upon whether the empirical results are behavioral would be to simulate the same proportional error for each matrix of this study. Perturbing M in this manner would reflect random inconsistency by the DM. These random values, with no behavioral effect, would then act as a check upon the empirical data.

Knowing V and true errors also provides the opportunity to analyze the consistency ratio and its ability to detect the inaccurate and unreliable data. Table 3 summarizes the results.

Table 3 Regression Statistics for the Eigenvector solutions and Consistency Ratios

Type of	MAE			MAPE		
Statistic	Distance	Probability	Rainfall	Distance	Probability	Rainfall
Average CR	0.067	0.148	0.080	0.067	0.148	0.080
Standard Deviation of CR	0.086	0.228	0.107	0.086	0.228	0.107
Average Error of Eigenvector	0.862	0.943	2.204	0.432	0.560	3.151
Regression CR=f(Eigenvector Errors)						
r	0.69	0.45	0.03	0.81	0.43	0.03
R-Square	0.47	0.21	0.00	0.66	0.18	0.00
F-Ratio	59.47	17.34	0.08	130.20	14.93	0.07
p-Value	< 0.0001	< 0.0001	0.7794	< 0.0001	0.0003	0.7903

The distance and rainfall experiments have good consistency while the probability experiment has modestly inflated inconsistency. From the distance and probability data, we can see that reasonably good accuracy occurs when the consistency ratio is less than 0.20. But the most important finding is from the rainfall experiment. There, participants

were extremely inaccurate but quite consistent. This illustrates that the consistency ratio is not a reliable indicator of accuracy. Participants in the rainfall experiment produced matrices of acceptable CR, yet they were unaware that their priorities are quite inaccurate. This reinforces a point often overlooked in AHP studies – the person making the comparisons must be both knowledgeable about the property being considered and diligent in making good estimates.

7. Conclusions

Since comparisons to the least dominant object tend to be more accurate, we should identify and give preference to those judgments. Two ways are suggested to identify the least dominant object. One method is to roughly rank the objects before comparisons are undertaken. The other method is to calculate initial priorities after n-1 comparisons.

Somewhere between n-1 and complete n(n-1)2 comparisons is a good place to stop. Exactly where is uncertain, but that decision could be left to the DM. After n-1 comparisons, priorities and step statistics can be calculated. If step statistics appear stable, they could be displayed to the DM along with the priorities at that stage. If reasonable, the DM could accept the results and terminate comparisons. Ultimately, it is the DM who decides whether his/her priorities are acceptable. This discretion avoids the inefficiency and cost of making unnecessary comparisons.

8. Key References

Bernasconi M., Choirat, C. and Seri, R. (2010). The Analytic Hierarchy Process and Theory of Measurement, *Management Science*, 56 (4), 699-711.

Carmone, F.J., Kara, A. and Zanakis, S H. (1997). A Monte Carlo investigation of incomplete pairwise comparisons matrices in AHP, *European Journal of Operational Research*, 102, 538-553.

Harker, P. T. (1987a). Shortening the comparison process in the AHP, *Mathematical Modelling,* 8, 139-141

Harker, P. T. (1987b). Alternative modes of questioning in the analytic hierarchy process, *Mathematical Modelling*, 9, 3-5, 353-360.

Harker, P. T. (1987c). Incomplete pairwise comparison in the analytic hierarchy process, *Mathematical Modelling* 9 (11), 837-848.

Wedley, W. C. (1993). Consistency prediction for incomplete AHP matrices *Mathematical and Computer Modelling*, 17 (4/5), 151-161.

Wedley, W. C. (2009). Fewer comparisons – Efficiency via sufficient redundancy" Proceedings of the Tenth International Symposium on the Analytic Hierarchy Process, Pittsburgh, PA, July 30 - August 2, http://www.isahp.org/2009Proceedings/index.htm, Accessed April 2014.