Effectiveness Analysis of Deriving Priority Vectors from Reciprocal Pairwise Comparison Matrices*

Bajwa, G., Choo**, E.U. and Wedley, W.C.

1. Introduction

In evaluating n competing alternatives $A_1,...,A_n$ under a given criterion, it is natural to use the framework of pairwise comparisons represented by a nxn square matrix from which a set of priority values for the alternatives is derived. The basic assumptions are that there exist priority values $v_1,...,v_n$ such that v_i represents the preference intensity of A_i (i=1,...,n) and the decision maker is able to provide t_{ij} (>0) as the answer to a simple question on pairwise comparison of v_i with v_j for all i,j=1,...,n. The priority values $v_1,...,v_n$ are assumed to be measured in some ratio scale. Then t_{ij} is an approximation of v_i/v_j provided by the decision maker, with some errors, for all i,j=1,...,n. The nxn matrix $T=[t_{ij}]$ is the reciprocal pairwise comparison judgment matrix with unity along the diagonal. By making the matrix reciprocal (that is, $t_{ij}=1/ t_{ji}$ for all i,j =1,...,n), the number of paired comparisons is n(n-1)/2.

The problem is to determine $w_1,...,w_n$ which estimate the priority values $v_1,...,v_n$ respectively from T. The matrix T is said to be consistent if $t_{ij}t_{jk}=t_{ik}$ for all i,j,k=1,2…n. If T is consistent, then the rank of T is one and the priority values, unique up to a positive multiple, are readily given by any column of T. But in general, T is inconsistent as $v_1,...,v_n$ are not explicitly known to the decision maker and thus there is a need for a method to estimate the priority values. The main challenge is how to reconcile the inevitable inconsistency of the pairwise comparison matrix elicited from the decision makers in real-world practical applications.

Many methods for estimating the priority values from the pairwise comparison judgment matrix have been proposed and their effectiveness comparatively evaluated (Basat 1991, Blankmeyer 1987, Bryson 1995, Budescu, Zwick & Rapoport 1986, Buede 1996, Carmone, Kara & Zanakis 1997, Carriere & Finster 1992, Chandran, Golden and Wasil 2005, Choo and Wedley 2004, Cook & Kress 1988, Fichtner 1986, Genest & Rivest 1994, Golany & Kress 1993, Herman & Koczkodaj 1996, Jensen 1984, Krovak 1987, Lin 2006, Lin 2007, Lootsma 1996, Mikhailov 2000, Mikhailov 2004, Ra 1999, Saaty & Vargas 1984, Williams & Crawford 1985, Zahedi 1986, Zanakis & Solomon 1998). It appears that none of the estimating methods for deriving priority vectors is universally superior over the others in all respects. Usually each method performs best in its own underlying criterion of effectiveness.

In this paper, we compare seven direct methods for estimating ratio scaled priority values from reciprocal pairwise comparison judgment matrices. The right eigenvector method (Saaty 1980), geometric mean method (Crawford & Williams 1985), normalized column mean method (Zahedi 1986), logarithmic least absolute error method (Cook & Kress 1988) and weighted least square method (Chu, Kalaba & Spingarn 1979) are selected because they can be solved easily and were shown to have good effectiveness and desirable analytical properties (Barzilai 1996, Budescu, Zwick & Rapoport 1986, Fichtner 1986, Golany & Kress 1993, Krovak 1987, Lin 2007, Saaty & Vargas 1984, Zahedi 1986). The simple column mean method and the chainwise geometric mean method (Ra 1999) are selected for its utter simplicity.

A simulated experiment is conducted to compare these methods under four measures of effectiveness: mean square error (MSE), mean absolute deviation (MAD), mean central conformity (MCC) and mean rank violation (MRV). These four effectiveness measures were also used in previous studies (Choo and Wedley 2004, Lin 2007, Zahedi 1986, Golany and Kress 1993). The problem sizes used in our simulation study are n=4,5,7. It is assumed that $t_{ij}=(v_i/v_j)+e_{ij}$ and the error terms e_{ij} (i,j=1,...,n & i>j) have independent normal distributions with zero means, where $T=[t_{ij}]$ is the nxn reciprocal pairwise comparison judgment matrix elicited from the decision maker and $v_1,...,v_n$ are the true priority values. Unlike previous studies in which $v_1,...,v_n$ were randomly generated, we adopted a novel design with specific types of priority vectors $v=[v_1,...,v_n]$ selected to represent the harder cases of "no obvious best alternative" and the easier cases of "two equal best alternatives" in $A_1,...,A_n$.

The seven estimating methods to be evaluated are described in some details in the next section. The four effectiveness measures used to evaluate the methods are given in section 3. In section 4, we describe the design of the simulation experiment. The simulation results are presented in section 5. Finally we give some concluding remarks.

2. Review of Seven Estimating Methods

If the pairwise comparison judgment matrix $T=[t_{ij}]$ is error free, then the true priority vector $v=[v_1,...,v_n]$ is readily given by any column of T. However, practical or real life judgment matrices contain inconsistencies and thus the estimated $w=[w_1,...,w_n]$ is only "close" to $v=[v_1,...,v_n]$. Most estimating methods of priority values use column or row information to minimize some form of distance between $[t_{ij}]$ and $[w_i/w_j]$ (Choo and Wedley 2004). A notable exception is the right eigenvector method that uses stabilized calculations from the interaction of row and column information.

Most methods have the following desirable properties: (a) preserve ranks strongly: if $t_{ik} \geq t_{jk}$ (k=1,...,n), then $w_i \geq w_j$; (b) correctness in the error free case: if T is error free, then $[w_1,...,w_n] = [v_1,...,v_n]$; (c) smoothness of result: small changes in T do not result in huge changes in $[w_1,...,w_n]$; (d) comparison order invariance: $[w_1,...,w_n]$ is independent of the order of the alternatives compared in T.

There are many direct methods for deriving ratio scaled priority vectors from reciprocal pairwise comparison matrices. Some of these direct methods are extended by adding more desirable constraints or by combining them into hybrid models (Chandran, Golden and Wasil 2005, Jones and Mardle 2004, Lin 2006, Mikhailov 2004). We now describe briefly the seven methods to be evaluated.

The Right Eigenvector Method (REV)

This method was recommended by Saaty (1980) using $Tw=\lambda_{max}w$, where λ_{max} is the principal right eigenvalue of T, to model the mathematical fact that Tv=nv is true when T is consistent. The REV method satisfies desirable properties (a), (b), (c) and (d). It always yields a unique priority vector. The normalized REV represents the relative dominance of the alternatives and is estimated by

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{k=m} \frac{T^k e}{e^T T^k e}$$, where $e^T = [1,...,1]$ is the unit n-vector.

The Geometric Mean Method (GM)

The original idea of this method was fully adapted and developed by Crawford and Williams (1985). The fundamental principle of this method is to minimize the sum of square distances of the $[\log(t_{ij})]$ from $[\log(w_i/w_j)]$

since $t_{ij} \approx w_i/w_j$. We then have $\log(t_{ij}) \approx \log(w_i)-\log(w_j) \Rightarrow [\log(w_i)-\log(w_j)-\log(t_{ij})]^2 \approx 0$. The model is to determine $[w_1,...,w_n]$ which minimizes $\square_i\square_j[\log(w_i)-\log(w_j)-\log(t_{ij})]^2$ subject to $\prod_i w_i=1$. The optimal solution, unique up to a positive multiple, is given by the geometric mean of the row elements of the matrix T: $w_i = (\square_j t_{ij})^{1/n}$, i=1,...,n. Crawford and Williams (1985) proved the validity of this method and showed the uniqueness of the optimal solution. The GM method satisfies desirable properties (a), (b), (c) and (d).

The Chainwise Geometric Mean Method (CGM)

To reduce the number of pairwise comparisons, Ra (1999) developed CGM method in which only n pairs $R_i=w_i/w_{i+1}$ (i=1,...,n-1) and $R_n=w_n/w_1$ are compared. The direct estimate d_i and indirect estimate I_i of R_i are given by: $d_i=t_{i,i+1}$ (i=1,...,n-1) and $d_n=t_{n1}$ are elicited directly from the decision-maker, while I_i is indirectly estimated as $I_i=d_i/\prod_j d_j$ for i=1 to n. Then R_i is estimated by the weighted geometric mean $G_i = d_i^{(n-1)/n} I_i^{1/n} = d_i/(\prod_j d_j)^{1/n}$ of d_i and I_i. Finally $[v_1,...,v_n]$ is estimated by: $w_i = G_i G_{i+1}...G_{n-1}$ (i=1,...,n-1) and $w_n=1$ which is then normalized to sum to unity. The CGM method satisfies desirable properties (b) and (c).

The Normalized Column Mean Method (NCM)

This estimating method is called the mean transformation method by Zahedi (1986). It is based on the fact that the priority vector is estimated by j-th column of T which can be normalized (sum to unity) to $[t_{1j}/(\sum_k t_{kj}),..., t_{nj}/(\sum_k t_{kj})]$, for j=1,...,n. By taking the average of these normalized columns as priority vector, $[w_1,...,w_n]$ is given by $w_i = (1/n)\sum_j(t_{ij}/(\sum_k t_{kj}))$, i=1,...,n. It is true that $[w_1,...,w_n]$ minimizes the distance $\sum_i\sum_j(t_{ij}/(\sum_k t_{kj})-w_i)^2$. The NCM method satisfies desirable properties (a), (b), (c) and (d).

The Simple Column Mean Method (SCM)

SCM is based on the fact that the priority vector is estimated by j-th column of T, j=1,...,n. By taking the average of these columns as priority vector, $[w_1,...,w_n]$ is given by $w_i = (1/n)\sum_j t_{ij}$, i=1,...,n. Note that the only difference between the NCM method and SCM method is that the columns of T are normalized to sum to unity in NCM before averaging. The SCM method satisfies desirable properties (a), (b), (c) and (d).

The Weighted Least Square Method (WLS)

Chu, Kalaba & Spingar (1979) introduced the WLS method to derive the priority vector. It is based on sound algebraic equations, which are conceptually easy to understand. It follows from $t_{ij} \approx w_i/w_j \Rightarrow t_{ij}w_j \approx w_i \Rightarrow (t_{ij}w_j-w_i)^2 \approx 0$ that $[w_1,...,w_n]$ should minimize $\sum_i\sum_j(t_{ij}w_j-w_i)^2$ with $\sum_i w_i=1$. The WLS method satisfies desirable properties (a), (b), (c) and (d). The final solution can be expressed in a matrix form (Chu et al 1979).

The Logarithmic Least Absolute Error Method (LLAE)

The underlying idea of the LLAE method (Cook & Kress 1988) is $t_{ij} \approx w_i/w_j \Rightarrow \log(t_{ij}) \approx \log(w_i)-\log(w_j) \Rightarrow |\log(w_i)-\log(w_j)-\log(t_{ij})| \approx 0$. It follows that $[w_1,...,w_n]$ should minimize $\sum_i\sum_j|(\log(w_i)-\log(w_j)-\log(t_{ij})|$. We note that $[w_1,...,w_n]$ can be multiplied by any positive number without changing the total deviation. Thus we may assume that $w_i>1$,

$i=1,2,...,n$. This minimization problem can be rewritten as a linear goal programming problem. The LLAE method satisfies desirable properties (a), (b) and (d).

3. Four Effectiveness Measures for Evaluation

In this paper, we will evaluate the seven selected estimating methods based on the following four effectiveness measures: (1) mean square error (MSE), (2) mean absolute deviation (MAD), (3) mean central conformity (MCC), and (4) mean ranking violation (MRV). These four effectiveness measures were also used in previous studies (Choo and Wedley 2004, Lin 2006, Lin 2007, Zahedi 1986, Golany and Kress 1993). For any true priority vector $[v_1,...,v_n]$ and $t_{ij}\approx v_i/v_j$, $(i,j=1,...,n)$, different estimating methods would give different $[w_1,...,w_n]$ as approximation of the priority vector. An estimating method is deemed to be better if its solution vector $[w_1,...,w_n]$ is "closer" to $[v_1,...,v_n]$. We now describe the four effectiveness measures. Note that we have scaled the measures by 100 to avoid dealing with small decimal numbers.

Mean Square Error (MSE): $MSE = 100\sum_i(w_i-v_i)^2/n$

Mean Absolute Distance (MAD): $MAD = 100\sum_i|w_i-v_i|/n$

Mean Central Conformity (MCC): Let $[m_1,...,m_n]$ be the average of the solution vectors from the seven estimating methods applied on $T=[t_{ij}]$. An estimating method does not conform with the other six estimating methods if the mean absolute deviation $100\sum|w_i-m_i|/n$ between its solution vector $[w_1,...,w_n]$ and the average solution vector $[m_1,...,m_n]$ is large.

Mean Ranking Violation (MRV): A ranking violation occurs in $[w_1,...,w_n]$ when $v_i>v_j$ and $w_i<w_j$. MRV = (number of ranking violations)/n

4. Simulation Design

A simulation experiment is conducted to evaluate the seven selected estimating methods under the four selected effectiveness measures. The problem sizes used in the simulation study are $n=4,5,7$. These are deemed to be sufficient from a practical point of view. It is assumed that $t_{ij}=(v_i/v_j)+e_{ij}$ and the error terms e_{ij} $(i,j=1,...,n \ \& \ i>j)$ have independent identical normal distributions with zero means and standard deviation σ, where $T=[t_{ij}]$ represents the nxn reciprocal pairwise comparison judgment matrix elicited from the decision maker and $v_1,...,v_n$ are the true priority values. The normal distribution for e_{ij} is truncated at $-v_i/v_j$ and v_i/v_j to avoid negative t_{ij} values. Unlike previous studies in which $v_1,...,v_n$ were randomly generated, here specific types of priority vectors $[v_1,...,v_n]$ are selected to represent the harder cases of no obvious best alternative and the easier cases of two equal better alternatives in $A_1,...,A_n$. In particular, we use $[1/n,...,1/n]$ to represent the harder case of "no obvious best alternative" and $[2/(n+2),2/(n+2),1/(n+2),...,1/(n+2)]$ to represent the easier case of "two equal best alternatives". A wide array of real life or practical problems fall into these two cases which are in clear contrast to cases with a single dominating alternative. Two σ values (σ=0.1, 0.3) are used to control the error terms e_{ij}. For each scenario ("no obvious best alternative" or "two equal best alternatives"; σ=0.1 or 0.3): 200 simulated matrices for $T=[t_{ij}]$ are generated and the solution vectors from T by the seven estimating methods are derived for $n=4,5,7$. The corresponding effectiveness measures of the methods are computed and the average effectiveness measure of the 200 replications is then calculated for each method. The total number of replications is 2,400.

In our preliminary runs, we explored using "m obvious best alternatives" for true priority vectors **v**=[v_1,v_2,...,v_n] with different m. We also considered using the true priority vector [k/(n+2k-2),k/(n+2k-2),1/(n+2k-2),...,1/(n+2k-2)] for "two equal best alternatives" with k=2,3,...,9. To control the scope of the simulation, we decided not to use m and k as control parameters. The cases with m=0, m=2 and k=2 are deemed to be more interesting and included in the simulation because they have less obvious solutions. In particular, we did not include the easiest case of "one obvious best alternative".

All the seven methods are easy to solve and solutions can be computed by using Microsoft EXCEL. The Solver add-in was used to compute solutions for the LLAE method which is a linear goal programming model. The *Data Table* in EXCEL was be used to execute 200 simulations for each scenario except LLAE for which *Solver Table* (http://www.indiana.edu/~mgtsci/SolverTable.html) was used to get all the solutions systematically.

6. Simulation Results

The four main factors of the simulation experiment, listed in the order of importance, are: (1) seven different methods, (2) two choices of true priority vectors, (3) two choices of standard deviations for the normally distributed random errors, and (4) three choices of matrix sizes. The simulated results are presented in carefully selected two factor tables. The results for n=4, n=5 and n=7 are combined in Table 1 and Table 2. For the case of "no obvious best alternative", Scenarios #1 and #2 have σ=0.1 and σ=0.3 respectively. For the case of "two equal best alternatives", Scenarios #3 and #4 have σ=0.1 and σ=0.3 respectively. The results for the four scenarios are combined in Table 3. The results for all scenarios and all matrix sizes are combined in Table 4.

The entries in Table 1 and Table 2 are the mean effectiveness measures of 600 replications with equal sample size of 200 each for n=4,5,7. The entries in Table 3 are the mean effectiveness measures of 800 replications with equal sample size of 200 for each of the four scenarios. The entries in Table 4 are the mean effectiveness measures and the t-test values of all the 2,400 replications. In all tables, the “B” superscript in a row signifies the method with the best performance in the effectiveness measure for that row while the “W” superscript signifies the worst. A one-sided t-test at significance level of 90 percent (α=0.10) is used to compare the performance of each method with the best performing method. A method is significantly worse than the best method under each effectiveness measure unless it has an “n” superscript which means that it is not significantly different. We highlight the noticeable results for each scenario below.

"No obvious best alternative" Scenarios

From Table 1, we can see that the GM and NCM are the best methods under both the σ=0.1 and σ=0.3 scenarios for random error. NCM is either the best or tied for best method in 6 of 8 tests. GM is either the best, tied for best, or insignificantly different from the best method in 7 of the 8 tests. Here, we are using the symbol “n” in the tables, to indicate that we cannot statistically differentiate the results of a method from the best method of the row. For example, the MRV results of the GM, SCM and WLS methods under the σ=0.1 scenario are insignificantly different from the best method (NCM).

In all other cases, the methods are significantly inferior (□=0.1) to either NCM or GM, the best methods. Although REV, SCM and WLS come close to matching the best method, they are significantly inferior. CGM and to a lesser extent LLAE are markedly inferior when compared to the effectiveness measures of other methods.

"Two equal best alternatives" Scenarios

The pattern for “two equal best alternatives” is similar to the results for “no obvious best alternative”. From Table 2, we can see that GM is either the best method or insignificantly different from the best method in 7 of 8 tests. NCM is second best overall, being best on two tests and insignificantly different from the best on one test. For the minor perturbation scenario (σ=0.1), SCM is the best method on one measure and insignificantly different from the best on two measures. For the other scenario with bigger perturbation (σ=0.3), WLS and LLAE are each the best method for different measures. Again, the CGM method is markedly inferior as compared to the other methods.

We note that the mean effectiveness measures for REV are better than other methods that are insignificantly different from the best method (NCM) in the case for MRV in scenarios 1 and 3, and MAD in scenario 3. These anomalies can be explained in part by the fact that REV has smaller variance. The t-test statistics of REV are 1.73, 1.50 and 1.34 (not shown in the tables) for MRV in scenarios 1 and 3 and MAD in scenario 3, respectively. These t-test statistics are larger than the critical value of 1.28 and thus REV is significantly inferior to the best method (NCM). The t-test statistic of WLS for the MRV measure in scenario 1 is 0.894 and the t-test statistics of SCM for MAD and MRV measures in scenario 3 are 0.51 and 0.69 respectively. Thus, they are insignificantly different from the best method (NCM). If REV had larger variances, then it too might have been insignificantly different from the best method (NCM). The simulated results show that REV had a smaller variance and was still less accurate. This means it was consistently less accurate for those particular cases.

Table 1. Simulated Results for “no obvious best alternative”

	Measure	REV	GM	CGM	NCM	SCM	WLS	LLAE
Scenario #1 (σ=0.1)	**MSE**	0.00632	0.00631^{n}	0.01442^{W}	0.00630^{B}	0.00633	0.00640	0.01000
	MAD	0.60739	0.60680^{n}	0.94487^{W}	0.60659^{B}	0.60808	0.61149	0.77329
	MCC	0.11920	0.11837^{B}	0.62916^{W}	0.11881	0.12131	0.12415	0.37787
	MRV	0.97667	0.97500^{n}	3.87333^{W}	0.97167^{B}	0.97500^{n}	0.98500^{n}	2.12667
Scenario #2 (σ=0.3)	**MSE**	0.10395	0.08570	0.18629^{W}	0.08456^{B}	0.13119	0.11273	0.10757
	MAD	2.25643	2.14757^{n}	3.27624^{W}	2.14675^{B}	2.37912	2.41744	2.49306
	MCC	0.47678	0.40030^{B}	2.12436^{W}	0.43598	0.65100	0.76766	1.36673
	MRV	7.49500	7.35500^{B}	9.25167^{W}	7.35500^{B}	7.42500	7.82833	7.99167

^{B} Best one in row; ^{W} Worst one in row; ^{n} Those not being rejected by 90% Significance Test

Table 2. Simulated Results for "two equal best alternatives"

	Measure	REV	GM	CGM	NCM	SCM	WLS	LLAE
Scenario #3 (σ=0.1)	**MSE**	0.00356	0.00356	0.01090^W	0.00356	0.00344^B	0.00400	0.00423
	MAD	0.45366	0.45357^n	0.80858^W	0.45347^B	0.45555^n	0.45855	0.49570
	MCC	0.08086	0.08023^B	0.50460^W	0.08114	0.16708	0.14979	0.28899
	MRV	0.16500	0.16000^B	1.30167^W	0.16000^B	0.17000^n	0.18000	0.26500
Scenario #4 (σ=0.3)	**MSE**	0.05353	0.04509^n	0.14159^W	0.04530^n	0.06441	0.06347	0.04463^B
	MAD	1.57810	1.51327^B	2.73743^W	1.52489	1.63058	1.63659	1.56747
	MCC	0.30895	0.29005^B	1.63274^W	0.30589	0.63220	0.60882	0.98297
	MRV	2.57333	2.51667^n	4.10667^W	2.54000	2.81333	2.46667^B	2.61500

^{B} Best one in row; ^{W} Worst one in row; ^{n} Those not being rejected by 90% Significance Test

Results by Matrix Size

In Table 3, simulated results for different matrix sizes are reported. We can see that GM tends to be best for all sizes while CGM is worst for all sizes. NCM is fairly good for n=4 and n=5. We note that as matrix size increases, MSE and MAD tend to decrease while MRV tends to increase. MSE and MAD are inversely related to matrix size because smaller priorities are produced with larger sized matrices. And with smaller priorities, MRV will be lager because there will be greater opportunity for ranking violations.

Table 3. Simulated Results for Different Sizes

	Measure	REV	GM	CGM	NCM	SCM	WLS	LLAE
n=4	**MSE**	0.05741	0.05539	0.10202^W	0.05428^B	0.06023	0.06895	0.06263
	MAD	1.56440	1.54803	2.11633^W	1.54016^B	1.57662	1.65738	1.66533
	MCC	0.20616	0.18917^B	1.09004^W	0.19903	0.39437	0.40410	0.81638
	MRV	1.50625	1.48125^B	1.89000^W	1.48625^n	1.48875^n	1.52625	1.60875
n=5	**MSE**	0.04055	0.03474^n	0.08942^W	0.03432^B	0.04918	0.04751	0.04622
	MAD	1.24946	1.20757^B	1.95763^W	1.20852^n	1.30157	1.30201	1.47052
	MCC	0.25689	0.23633^B	1.26285^W	0.25138	0.40445	0.42691	0.85994
	MRV	2.33500	2.32125^n	3.46250^W	2.31375^B	2.38125	2.39375	2.96875
n=7	**MSE**	0.02757	0.01536^B	0.07346^W	0.01619	0.04462	0.02350	0.01597^n
	MAD	0.85782	0.78531^B	1.75139^W	0.80010	0.92682	0.88365	0.86129
	MCC	0.27630	0.24122^B	1.31526^W	0.25595	0.37988	0.40681	0.58611
	MRV	4.56625	4.45250^B	8.54750^W	4.47000^n	4.66750	4.67500	5.17125

^{B} Best one in row; ^{W} Worst one in row; ^{n} Those not being rejected by 90% Significance Test

Aggregate Results

Aggregated results are given in Table 4. Overall, it is obvious that the GM is the best method and the NCM method is second best. The t-test values of REV indicated that it was never superior or insignificantly different from the best on any of the tests. Quite clearly, CGM is the worst method.

7. Conclusion

Of the various tests, we place more importance on the results with σ=0.1. It stands to reason that the results are more reliable when σ is smaller. As well, we suggest more credence be placed upon MSE and MAD as measures of effectiveness. Although MCC has been used in the literature and included here, we think that minimizing error from the true priority vector is more important than conforming to the average of some less than perfect methods. Similarly, MRV uses a lower order of measurement that loses some sight of accuracy.

Table 4. Aggregated (Averages) Simulated Results

	Measure	REV	GM	CGM	NCM	SCM	WLS	LLAE
Average Over All Replicas	**MSE**	0.04184	0.03516^{n}	0.08830^{W}	0.03493^{B}	0.05134	0.04665	0.04161
	MAD	1.22389	1.18030^{B}	1.94178^{W}	1.18292	1.26834	1.28102	1.33238
	MCC	0.24645	0.22224^{B}	1.22272^{W}	0.23545	0.39290	0.41261	0.75414
	MRV	2.80250	2.75167^{B}	4.63333^{W}	2.75667^{n}	2.84583	2.86500	3.24958
t-test Value Over All Replicas	**MSE**	3.25631	0.99695	19.94396	Best	3.31159	10.49257	5.73071
	MAD	7.17861	Best	35.23288	1.50795	7.04747	15.72715	11.38196
	MCC	6.39608	Best	57.84685	8.23490	16.65985	22.26176	43.92424
	MRV	4.14057	Best	31.99002	0.42633	5.54460	6.44717	13.67957

^{B} Best one in row; ^{W} Worst one in row; ^{n} Those not being rejected by 90% Significance Test

We have evaluated seven selected methods for estimating priority values under four well known effectiveness measures by a simulated experiment. **The simulation results suggest that the GM is the best method and NCM is the second best**. CGM is the worst method. Although other methods gave close results, they were significantly inferior.

One major advantage of the best methods (GM and NCM) is that they have simple formulas for computing the solution vectors. Because of its many desirable properties (Fichtner, 1986), the GM method is quite well known. However, the NCM method has not been well recognized. It performed well in this study and has been found to be robust (Zahedi, 1986). Thus, the results provide some support for Zahedi's (1986) recommendation to use it. However, GM has performed much better in our simulation and it is also the best method recommended in Lin (2007).

It is rather surprising that although the REV method gave close results and ranked a distant third overall, it was significantly inferior to the best method in all tests. Since REV is more mathematically complicated and since it did not prove to be superior, the results do not provide support for the recommendation by Saaty & Vargas (1984) that REV be the only method to use when data are not entirely consistent. Also, the REV method is not in the four best methods recommended in the study by Lin (2007). Furthermore, Blanquero, Carrizosa and Conde (2006) showed that the REV method is not efficient in a sense of vector maximization. Practitioners should be cautioned when using the REV method popularly used in many commercial AHP software.

So which method should be used? Golany and Kress (1993) who did not find any one method to be superior than all others suggested that the choice of method should be dictated by the desired measure of effectiveness. Mathematically, different error measures support different methods. Hence, the defining question is not which method is better, but what application or criteria are more valued. We hope that the GM method, the top choices here, will receive much more attention in future research and comparative studies.

References

Barzilai, J., 1997. Deriving Weights from Pairwise Comparison Matrices. *Journal of Operational Research Society*, Vol. 48, 1226-1232.

Blanquero, R., Carrizosa, E., Conde, E., 2006. Inferring Efficient Weights from Pairwise Comparison Matrices. *Mathematical Methods of Operations Research*, Vol. 64, 271-284.

Basat, I., 1991. Inference in Pairwise Comparison Experiments Based on Ratio Scales. *Journal of Mathematical Psychology*, Vol. 35, 80-91.

Blankmeyer, E., 1987. Approaches to Consistency Adjustment. *Journal of Optimization Theory and Applications*, Vol. 54, 479-488.

Bryson, N., 1995. A Goal Programming Method for Generating Priority Vector. *Journal of Operational Research Society*, Vol. 46, 641-648.

Budescu, D.V., Zwick, R., Rapoport, A., 1986. A Comparison of the Eigenvalue Method and the Geometric Mean Procedure for Ratio Scaling. *Applied Psychological Measurement*, Vol. 10, 69-78.

Buede, D., 1996. Second Overview of the MCDM Software Market. *Journal of Multi-Criteria Decision Analysis*, Vol. 5, 312-316.

Carmone, F.J., Kara, A., Zanakis, S.H., 1997. A Monte Carlo Investigation of Incomplete Pairwise Comparisons Matrices in AHP. *European Journal of Operational Research*, Vol. 102, 538-553.

Carriere, J.F., Finster, M., 1992. Statistical Theory for the Ratio Model of Pairwise Comparisons. *Journal of Mathematical Psychology*, Vol. 36, 450-460.

Chandran, B., Golden, B., Wasil, E., 2006. Linear Programming Models for Estimating Weights in the Analytic Hierarchy Process, *Computers & Operations Research*, Vol. 32, 2235-2254.

Choo, E.U., Wedley, W.C., 2004. A Common Framework for Deriving Preference Values from Pairwise Comparison Matrices. *Computers & Operations Research*, Vol. 31, 893-908.

Chu, A.T.W., Kalaba, R.E., Spingarn, K., 1979. A Comparison of Two Methods for Determining the Weights of Belonging to Fuzzy Sets. *Journal of Optimization Theory and Applications*, Vol. 27, 531-538.

Cook, W.D., Kress, M., 1988. Deriving Weights from Pairwise Comparison Ratio matrices: An Axiomatic Approach. *European Journal of Operational Research*, Vol. 37, 355-362.

Fichtner, J., 1986. On Deriving Priority Vectors from Matrices of Pairwise Comparisons. *Socio-Economic Planning Sciences*, Vol. 20, 341-345.

Genest, C., Rivest, L.P., 1994. A Statistical Look at Saaty's Method of Estimating Pairwise Preferences Expressed on a Ratio Scale. *Journal of Mathematical Psychology*, Vol. 38, 477-496.

Golany, B., Kress, M., 1993. Multicriteria Evaluation of Methods for Obtaining Weights from Ratio-Scale Matrices. *European Journal of Operational Research*, Vol. 69, 210-220.

Harker, P.T., 1987. The Incomplete Pairwise Comparisons in the Analytic Hierarchy Process. *Mathematical Modeling*, 9, 837-848.

Herman, H.W., Koczkodaj, W.W., 1996. Monte Carlo Study of Pairwise Comparison. *Information Processing Letters*, Vol. 57, 25-29.

Jensen, R.E., 1984. An Alternative Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology*, Vol. 28, 317-332.

Jones, D.F., Mardle S.J., 2004. A Distance-Metric Methodology for the Derivation of Weights from a Pairwise Comparison Matrix. *Journal of the Operational Research Society*, Vol. 55, 869-875.

Krovak, J., 1987. Ranking Alternatives - Comparison of Different Methods Based on Binary Comparison Matrices. *European Journal of Operational Research*, Vol. 32, 86-95.

Lin, C.C., 2006. An Enhanced Goal Programming Method for Generating Priority Vectors, *Journal of the Operational Research Society*, Vol. 57, 1491-1496.

Lin, C.C., 2007. A Revised Framework for Deriving Preference Values from Pairwise Comparison Matrices. *European Journal of Operational Research*, Vol. 176, 1145-1150.

Lootsma, F.A., 1996. Model for the Relative Importance of the Criteria in the Multiplicative AHP and SMART. *European Journal of Operational Research*, Vol. 94, 467-476.

Mikhailov, L., 2000. Fuzzy Programming Method for Deriving Priorities in the Analytic Hierarchy Process. *Journal of the Operational Research Society*, Vol. 51, 341-349.

Mikhailov, L., 2004. Deriving Priorities from Ratio-Scale Comparison Judgments: A Multiple Objective Approach, *International Conference of Multiple Criteria Decision Making*, Whistler, B.C., Canada, August 6-11, 2004.

Ra, J.W., 1999. Chainwise Paired Comparisons. *Decision Sciences*, Vol. 30, 581-599.

Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw-Hill, New York.

Saaty, T.L., Vargas, L.G., 1984. Comparison of Eigenvalue, Logarithmic Least Squares and Least Squares Methods in Estimating Ratios. *Mathematical Modeling*, Vol. 5, 309-324.

Williams, C., Crawford, G., 1985. A Note on the Analysis of Subjective Judgment Matrices. *Journal of Mathematical Psychology*, Vol. 29, 387-405.

Zahedi, F., 1986. A Simulation Study of Estimation Methods in the Analytic Hierarchy Process. *Socio-Economic Planning Sciences*, Vol. 20, 347-354.

Zanakis S.H., Solomon A., Wishart N., Dublish S., 1998. Multi-attribute decision making: a simulation comparison of select methods. *European Journal of Operational Research*, Vol. 107, 507-529.