HANDLING MISSING SURVEY DATA IN AHP

Gang Kou, Yi Peng*
School of Management and Economics
University of Electronic Science and Technology of China
Chengdu, China, 610054
E-mail: pengyicd@gmail.com

Daji Ergu
School of Management and Economics
University of Electronic Science and Technology of China
Chengdu, China, 610054
and
Southwest University for Nationalities
Chengdu, China, 610200

Yong Shi
College of Information Science & Technology
University of Nebraska at Omaha
Omaha, NE 68182, USA
and
Research Center on Fictitious Economy and Data Sciences
Chinese Academy of Sciences
Beijing 100190, China

ABSTRACT

The questionnaire is a widely used to collect opinions and views of decision makers in AHP. The item scores corresponding to missing or invalid survey data have to be estimated to keep the consistency for the comparison matrix. In this paper, a scale format is used to design the surveyed score items for the comparison matrix. An induced bias matrix model (IBMM) is proposed to estimate the missing item score. The survey questionnaire can be optimized according to the importance of the surveyed questions. An example is introduced to illustrate the proposed estimation and optimization model.

Keywords: Questionnaire, AHP, Missing survey data

1. Introduction

The reciprocal pairwise comparison matrix (RPCM hereinafter) is a widely used technique in the MCDM (Pelaez and Lamata 2003)., especially in the analytic hierarchy approach (AHP) proposed by Saaty (1980, 2001, 2008). Questionnaire survey conducted by email, telephone, personal interview, or on-line, is used to collect opinions and views of decision makers in AHP. The values of comparisons in single RPCM are gathered from various questionnaire surveys. Invalid or missing item scores of a questionnaire could lead to inconsistent comparison matrix (Fedrizzi and Giove 2007; Chiclana et al. 2009; Peng et al. 2011). In

*Corresponding author

Acknowledgments This research has been supported by grants from the National Natural Science Foundation of China (#70901011, #70901015, and # 70921061)

the AHP, the score items designed for a RPCM in a questionnaire increase when the number of comparisons increase, which will reduce the response rate. Therefore, it is necessary to optimize a questionnaire format for one RPCM in the AHP, and estimate the missing comparisons of an incomplete RPCM in questionnaire while the RPCM consistency can be kept. In this paper, a scale format is used to design the surveyed items for single RPCM in questionnaire survey. Besides, an induced bias matrix model (IBMM) is proposed to estimate the missing item scores of the RPCM whilst keep the global consistency. The survey questionnaire can be optimized according to the importance of the surveyed questions.

This paper is organized as follows. The next section briefly addresses the methodologies including the description of the IBMM method and the optimization principals and design format for questionnaire design. In section 3, an example is introduced to illustrate the proposed estimation and optimization model. Section 4 concludes the paper.

2. Methodology

2.1 Estimating the missing item scores using IBMM method

In our earlier paper (Ergu et al. 2011), an induced bias matrix (IBM) is proposed to estimate the missing values in an incomplete RPCM while keep the global consistency. The structure of the IBMM is shown in Figure 1.

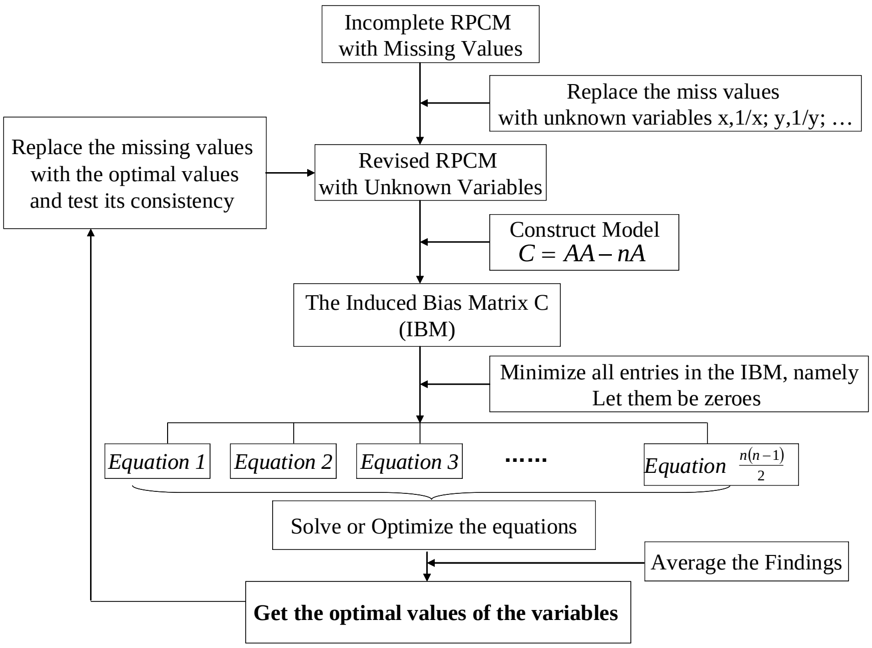


Figure 1. The structure of the IBMM for estimating the missing entries in IRPCM

The steps of the BIMM for estimating the missing values include:

Step 1: Replace the missing values with unknown variables x,1/x; y,1/y; z,1/z etc for the IRPCM and get the revised 'complete' reciprocal PCM A.

Step 2: Construct the proposed model $C = AA - nA$ and calculate the bias induced matrix C.

Step 3: Minimize all bias entries of the bias induced matrix C, that is, let all entries with unknown variables be (equal to) zeros, and get $n(n-1)/2$ number of equations.

Step 4: Optimize and solve these linear or nonlinear equations.

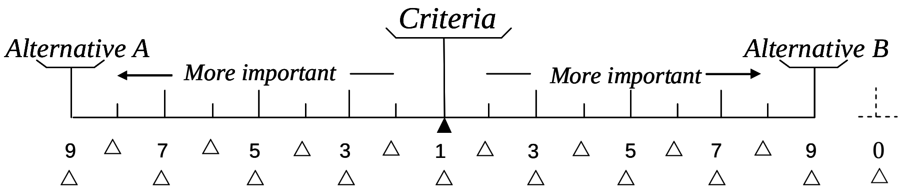
Step 5: Average all solutions in order to keep the global consistency and find the optimal values of variables.
Step 6: Replace the missing values with the optimal values and test the revised PCM's consistency in order to maintain its consistency.

2.2 The optimization principals and design format for questionnaire design

Based on the theorem of the proposed IBMM for an incomplete RPCM, some comparisons in terms of the importance can deliberately be omitted, which will reduce the number of score items in a questionnaire survey design. The missing values can be estimated using the IBMM technique once the questionnaire survey data are collected. For different needs of questionnaire design, the general optimizing principal, the format of scale and two design formats are developed.

The General Optimizing Principal: Set the score items to be omitted from 1 to $n(n-1)/4$ for each RPCM, where n is the size of the RPCM. The missing comparisons can not be located at the same row or same column if the number of missing comparisons is $n(n-1)/4$.

The Format of Scale: To express the relationship between two alternatives with respect to one criterion, the following format in 9-point scale as shown in Figure 2, is proposed to score the items for respondents. The score "0" is added to this scale denoting the uncertain item.



Note: *1 – Equal importance; 3 – Weak importance; 5 – Strong importance;*
7 – Demonstrated importance; 9 – Absolute importance;
2,4,6,7 – Intermediate values between the two adjacent judgments
0 - Uncertain
Tick " √" the corresponding score in the symbol "△"

Figure 2. The general format with uncertain designed to compare two alternatives with respect to one criteria in a questionnaire survey design

The following two types of formats are proposed to design a questionnaire survey for a RPCM in terms of the importance of score items.

Design Format 1: Use the general format with uncertain factor as shown in Figure 2 to design a questionnaire survey.
Design Format 2: Skip some score items in terms of the importance and above principals to design a questionnaire survey.

The uncertain score factor is added to the scale score item in the first design format considering the respondents' limited expertise and/or preference conflict. For the second type of design format, according to the importance of score items, we can set the missing number of comparison from 1 to $n(n-1)/4$ to reduce the number of score items.

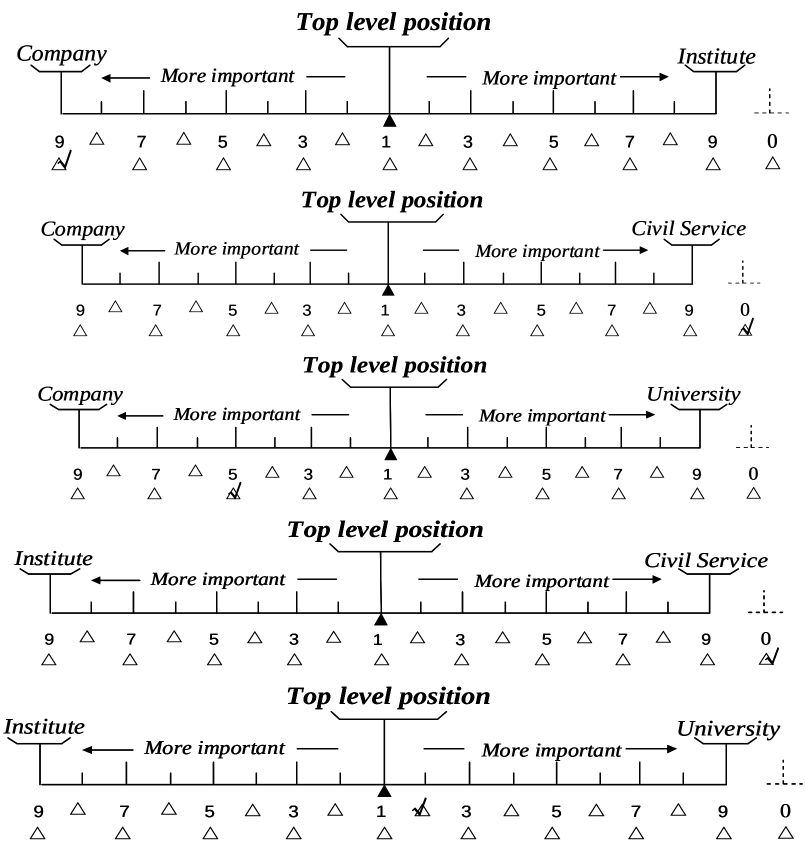
3. Illustrative example

Since the processes of designing score items and estimating the missing score item for single RPCM can be copied and applied to all others RPCM, in this section, single RPCM with four orders is used to demonstrate the proposed method. Besides, due to the principal of estimating different number of missing values is similar and repetitive, two missing comparisons are only considered in this RPCM. Assume we need to design a questionnaire to survey the doctorial candidates which kinds of job he/she will be preferred after getting the degree. Assume there are four alternatives: Industry, Research Institute, University, Government, and the corresponding RPCM matrix with $\lambda_{\max} = 4.0076$ and CR=0.0028 with respect to top level position are obtained from once questionnaire survey as shown in Table 1.

Table 1. Single RPCM with 4 orders obtained from once questionnaire survey

Top level position	Industry	Institute	Government	University
Industry	1	9	2	5
Institute	$\frac{1}{9}$	1	$\frac{1}{5}$	$\frac{1}{2}$
Government	$\frac{1}{2}$	5	1	2
University	$\frac{1}{5}$	2	$\frac{1}{2}$	1

Assume *Design format 1* is used to design the score items for this RPCM in questionnaire design, and the response from one respondent in once questionnaire is collected as shown in Figure 3.



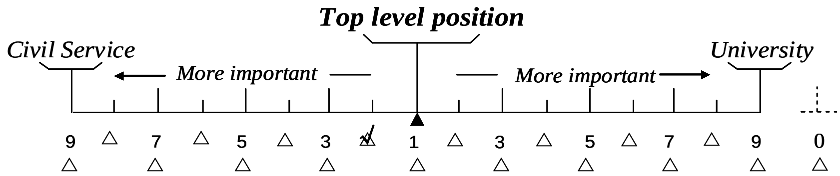


Figure 3. The response from one respondent in once questionnaire designed by *Design format 2*

The following incomplete comparison matrix as shown in Table 2 can be obtained from this response:

Table 2. The incomplete RPCM obtained from once questionnaire survey

Top level position	Industry	Institute	Government	University
Industry	1	9	×	5
Institute	$\frac{1}{9}$	1	×	$\frac{1}{2}$
Government	×	×	1	2
University	$\frac{1}{5}$	2	$\frac{1}{2}$	1

Apply the IBMM method to calculate this incomplete comparison matrix and estimate the two missing comparisons. The steps are as follows:
Step 1: Replace the two missing values with unknown variables x,1/x; y,1/y as shown below

$$\begin{pmatrix} 1 & 9 & x & 5 \\ \frac{1}{9} & 1 & y & \frac{1}{2} \\ \frac{1}{x} & \frac{1}{y} & 1 & 2 \\ \frac{1}{5} & 2 & \frac{1}{2} & 1 \end{pmatrix}$$

Step 2: Calculate the bias induced matrix C using the IBMM formula $C = AA - nA$.
Step 3: Set all entries with unknown variables be zeros in the upper triangular matrix, and get 6 number of equations.
Step 4: Solve pairwise combined systems of linear equations.
Step 5: Average all solutions and find the optimal values of variables. The results of calculation are x=2.4391 and y=0.2639. The corresponding closest scale values to the values located at 9-point scale are 2 and 1/4 respectively, which are the same or closest to the values in original RPCM, step 6 is skipped.

Likewise, we can set deliberately any two of comparisons in above RPCM to be missing, and estimate them using the IBMM. Some of the missing comparisons and the estimated values are shown in Table 3 .

Table 3. Some results of two missing comparisons and their estimated values in above RPCM

MC	EV	CV	OV	MC	EV	CV	OV
a_{12} a_{23}	9.8949 0.2561	9 1/4 or 1/5	9 1/5	a_{13} a_{14}	2.0798 4.3245	2 4 or 5	2 5

a_{14} a_{23}	4.2480 0.2366	4 or 5 1/5 or 1/4	5 1/5	a_{14} a_{24}	3.8855 0.4190	4 1/2	5 1/2
a_{23} a_{34}	0.2188 2.4391	1/5 2 or 3	1/5 2	a_{24} a_{34}	0.5458 2.6497	1/2 2 or 3	1/2 2

Note: *MC- Missing Comparisons; EV- Estimated Values;*
CV- Closest Value within 9-point Scale; OV- Original Values

4. Conclusions

In this paper, the missing item scores in comparisons for a questionnaire survey can be estimated by the proposed IBMM while the global consistency is kept. Besides, the questionnaire design can be optimized based on the theorem of the proposed IBMM. Two kinds of design format are proposed in terms of the importance of the score items, and one general optimizing principal is also proposed. In addition, a scale form with 9-point scale is proposed to design the score items in a questionnaire survey. The example of the reciprocal pairwise comparison matrix with 4 orders shows the effectiveness of the IBMM to estimate the missing comparisons and the optimization of a questionnaire design.

REFERENCES

Chiclana, F, Herrera-Viedma, E., and Alonso,S. (2009). A note on two methods for estimating missing pairwise preference values. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 39 1628-1633

Ergu,D.J., Kou,G. Peng,Y. and Shi,Y. (2011). A Simple Method to Improve the Consistency Ratio of the Pair-wise Comparison Matrix in ANP, Accepted with Minor Revision, European Journal of Operational Research

Fedrizzi, M. and Giove, S. (2007). Incomplete pairwise comparison and consistency optimization. European Journal of Operational Research, vol. 183(1), pages 303-313.

Pelaez, J.I. and Lamata, M.T. (2003). A new measure of consistency for positive reciprocal matrices. Computers & Mathematics with Applications. Vol. 46, no. 12, pp. 1839-1845.

Peng, Y., Kou, G., Wang, G., Wu, W., and Shi, Y. (2011) An AHP-Based Evaluation of Ensemble Methods for Software Defect Prediction, DOI: 10.1142/S0219622011004282, International Journal of Information Technology & Decision Making, Vol. 10, Issue: 1, 187–206.

Saaty, T.L. (1980). The Analytical Hierarchy Process. New York: McGraw-Hill.

Saaty, T.L. (2001). Deriving the AHP 1-9 Scale from First Principles, *ISAHP 2001 Proceedings*, Bern, Switzerland

Saaty, T.L. (2008). Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas. Volume 102, Number 2, 251-318, DOI: 10.1007/BF03191825