Decision Making with Uncertain Judgments: A Stochastic Formulation of the Analytic Hierarchy Process

Eugene D. Hahn
Department of Information and Decision Sciences
The Franklin P. Perdue School of Business
Salisbury University
Salisbury, MD 21801
410-548-3315 (Office)
410-546-6209 (Fax)
E-mail: edhahn@salisbury.edu

Acknowledgements: The author would like to thank the Associate Editor, two anonymous referees, and Jeff Kottemann for stimulating discussions which helped enhance the article. The research was funded in part by a research grant from the Franklin P. Perdue School of Business at Salisbury University.

Decision Making with Uncertain Judgments: A Stochastic Formulation of the Analytic Hierarchy Process

ABSTRACT

In the Analytic Hierarchy Process (AHP), priorities are derived via a deterministic method, the eigenvalue decomposition. However, judgments may be subject to error. A stochastic characterization of the pairwise comparison judgment task is provided and statistical models are introduced for deriving the underlying priorities. Specifically, a weighted hierarchical multinomial logit model is used to obtain the priorities. Inference is then conducted from the Bayesian viewpoint using Markov chain Monte Carlo methods. The stochastic methods are found to give results that are congruent with those of the eigenvector method in matrices of different sizes and different levels of inconsistency. Moreover, inferential statements can be made about the priorities when the stochastic approach is adopted, and these statements may be of considerable value to a decision maker. The methods described are fully compatible with judgments from the standard version of AHP and can be used to construct a stochastic formulation of it.

Subject Areas: Analytic Hierarchy Process, Bayesian Inference, Multi-criteria Decision Making (MCDM), Logit Modeling, Markov Chains, and Simulation.

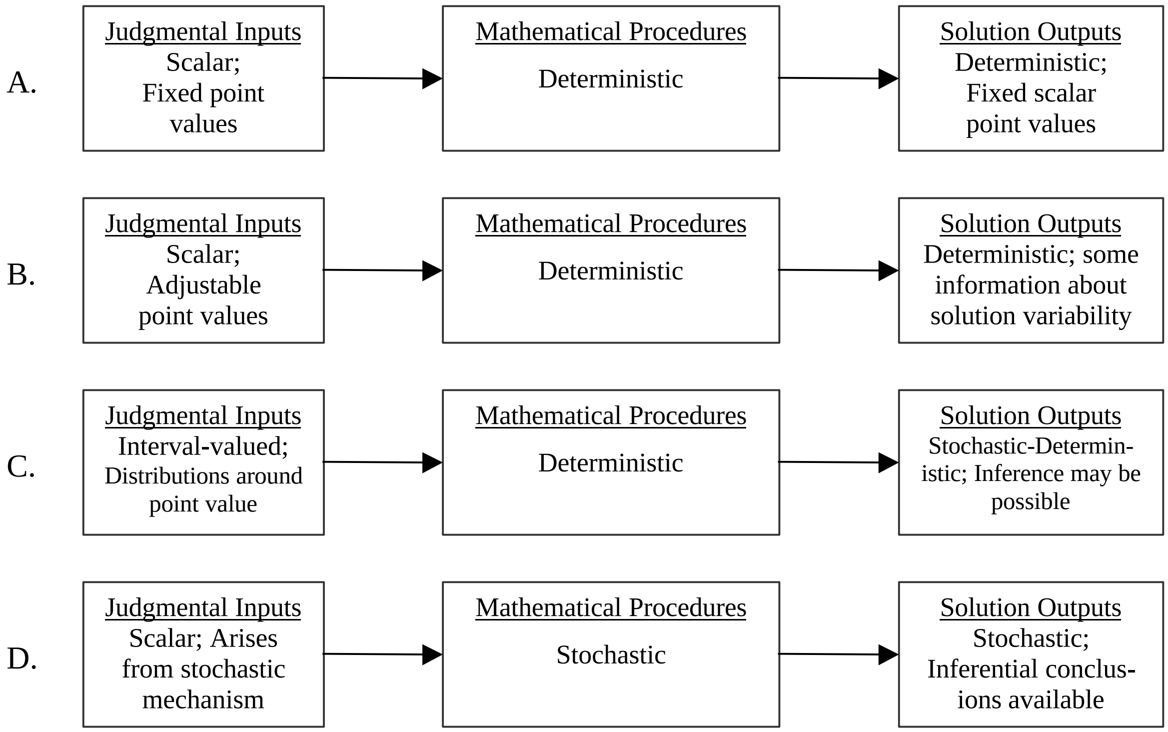
INTRODUCTION

Multi-criteria decision making methods are an important set of tools for addressing challenging business decisions as they allow the manager to better proceed in the face of uncertainty, complexity and conflicting objectives. These decision methods are typically predicated on a small set of assumptions or axioms. One particularly essential axiom is that the manager can quantify his or her preferences regarding the situation at hand (e.g., Axiom 3, Keeney, 1982; Assumptions 2a-b, Pratt, Raiffa & Schlaifer, 1964). In practice this is generally taken to mean that relative preferences can be characterized by a single scalar number or point value. For example, if the manager can indicate that one outcome is “twice” or “five times” as preferable as another, this axiom is considered to hold. When this axiom holds, we can take the manager’s numeric judgments and apply a series of mathematical operations to them to obtain a solution. As an example, consider a multi-criteria decision method such as the Analytic Hierarchy Process (AHP). In this method, the eigenvalue decomposition is applied to a matrix of numeric judgments that have been provided by the decision maker with regard to a set of alternatives. This operation yields a set of priorities which indicate the decision maker’s underlying preferences for the alternatives.

A consequence of the quantification axiom is that error in judgments is typically considered to be non-existent or negligible in multi-criteria decision making methods. Phrased differently, judgments are typically taken to be certain and thus can be represented by scalar values. Figure 1 provides a typology of multi-criteria decision making methods with regard to the quantification axiom. Figure 1 shows that Type A methods, such as AHP and multi-attribute utility theory, are ones in which point-valued judgments are processed by mathematical procedures that are predicated on judgmental certainty. A set of point values is output as the solution. Such procedures are wholly deterministic in that judgments are taken to be certain and deterministic operations are performed on the judgments. Nonetheless, the dynamic nature of business environments is an important source of what may be called “external uncertainty”. For example, a manager might be uncertain as to whether a particular competitor will enter the market, or whether a particular product will be highly desired by consumers. Here, there is some uncertainty about a possible external event or scenario. Still, one might wish to continue as before by requesting point-valued certain judgments from our decision maker and making use of a deterministic procedure. Some information can be gained by asking a set of "what if" questions and determining whether the solution changes. For example, we might wonder whether a particular course of action is still the best if a certain judgment were changed by a small amount. Techniques which incorporate this approach appear as Type B in Figure 1. The approach is typically known as *sensitivity analysis*, and is broadly applicable to many decision methodologies (e.g., Triantaphyllou & Sánchez, 1997). In using sensitivity analysis, we can determine the range of inputs for which a particular solution will hold. As such, we can get a sense of the decision's robustness.

In Type A and Type B approaches to multi-criteria decision making, error in judgments is assumed to be non-existent, or perhaps negligible. In some cases, it may well be true that error in the judgments of experts is at best negligible. However, in some cases considerable uncertainty may be associated with one or more judgments, and it is thus conceivable that these judgments may be made with some degree of error. Alternatively, slightly different attributes of mental constructs may come into focus across a set of comparisons, yielding inconsistent judgments (e.g., Genest & Rivest, 1994). Hence, there very well may be "judgment uncertainty" present. Because we may have uncertain judgments, it is of interest to examine stochastic approaches to multi-criteria decision making. One possible approach is to assess an interval-valued judgment as opposed to a scalar judgment. For example, we may elicit a probability

Figure 1 – A Typology of Multi-criteria Decision Making Methods



distribution from the decision maker instead of a single number. Then, the deterministic procedure of interest can be used. Techniques which incorporate this concept appear as Type C techniques in Figure 1. We can see that this approach is a relatively straightforward extension of the Type B approach. In particular, a set of deterministic transformations is applied to the interval-valued judgments. Using this type of approach it is possible in principle to make inferential statements regarding the alternatives. This is an important advance, as will be discussed momentarily. Finally, we may also consider techniques in which a scalar judgment is assumed to be a realization of a stochastic phenomenon. This approach is perhaps most in keeping with the notion that a certain amount of uncertainty may simply be inherent in the judgment process. More importantly, a set of powerful conclusions can be made with the use of these approaches. Here, we can make inferential statements regarding the alternatives under consideration. For example, we will be able to determine whether we can be 95% confident that two alternatives have unequal priorities. This represents a substantial advance as the decision maker can determine probabilistically what the chances are of alternative states of affairs[1]. Moreover, we can clearly see that deterministic multi-criteria decision methods are a special case of their stochastic counterparts. That is, deterministic multi-criteria methods can be obtained from stochastic methods in the limit as uncertainty in judgments tends to zero.

Below we first review a widely used method for multi-criteria decision making, the AHP. We then describe a stochastic conceptualization of the AHP (c.f. Moskowitz, Tam & Lang, 2000, for an example of a stochastic approach to multi-attribute utility theory). The method we describe is designed to facilitate decision making with uncertain judgments. We examine the performance of the method with four example matrices and then provide a more thorough examination by means of a Monte Carlo study. We then discuss techniques for interval estimation and conclude with an example in which the inference process is described in detail.

---

[1] We note that error may arise as a result of systematic distortions and biases which lead to judgments that are consistently faulty. Errors of this type may well be quite pernicious and undermine decision making substantially; however, methods for remediating these cognitive and psychological sources of error will not be discussed in this paper (see instead Arkes, 1991). In the absence of systematic error or bias, we may find that non-systematic error, random perturbations around a central "true" value, may exist. Errors of this type can be addressed through stochastic approaches, and it is this type of error that will be examined here.

ELEMENTS OF AHP

In the AHP, it is of interest to evaluate a set of alternatives with respect to various criteria or objectives. The priorities of the different alternatives are computed from a matrix of pairwise comparisons via the eigenvalue decomposition (e.g., Saaty, 1977; Saaty & Vargas, 1991). One can motivate the use of the eigenvalue decomposition as follows. Consider a set of objects, $O_1, \ldots, O_K$ that have known weights $w_1, \ldots, w_K$. We can construct a square matrix, $\boldsymbol{C}$, of pairwise comparison ratios, w_i/w_j, where all of the elements on the main diagonal, C_{ii}, are equal to one. Then, the task is to find the scalars l and the vectors $\boldsymbol{w}$ that satisfy the relation $\boldsymbol{Cw} = l\boldsymbol{w}$, or equivalently $(\boldsymbol{C} - l\boldsymbol{I})\boldsymbol{w} = \mathbf{0}$. If we take the largest value of l, then the corresponding eigenvector $\boldsymbol{w}$ will contain the original weights $(w_1, \ldots, w_K)^T$ up to a scaling constant. Often it is convenient to normalize the weights as this preserves the relative magnitudes of the weights and provides a common scale for comparative purposes. These normalized weights are called priorities. Moving beyond the realm of physical weights, if one has pairwise comparisons on a set of alternatives along a dimension such as “preference” or “importance”, the eigenvalue decomposition can again be applied. The resulting priorities indicate the relative preference or importance of the alternatives.

The eigenvalue decomposition is a deterministic mathematical approach to deriving priorities. In adopting it, error is assumed to be non-existent or negligible. However, a number of researchers have applied statistical methodologies to the AHP. In a review of the literature we discern that, despite some degree of overlap, there are at least three areas of methodological commonality. In one methodological area, authors have obtained stochastic judgments by eliciting interval judgments from decision makers as opposed to single-valued judgments (e.g., Arbel, 1989; Haines, 1998; Saaty & Vargas, 1987; see Stam & Silva, 1997, for a review). The methods associated with this methodological area typically correspond with the Type C approach described in Figure 1. Advantages of this approach include the ability to calculate the probabilities of rank reversal and in principle the ability to make inferential statements regarding the priorities. However, each judgment may require additional cognitive effort and time of the decision maker. For larger problems, the additional burden may become considerable. Another active area of inquiry is the use of least squares or maximum likelihood techniques to obtain estimates of the priority vector (e.g., de Jong, 1984; Genest & Rivest, 1994; Jensen, 1984). Frequently the focus in this area tends to be on the production of priority vector estimates which are better by some criterion than those produced by the eigenvector method. For example, if we assume a multiplicative error term is associated with pairwise comparison ratios, then a linear model formulation can be achieved by taking the natural logarithms of the parameters. This gives rise to estimates of the priority vector based on a normalized vector of geometric means. Crawford and Williams (1985) showed that such a procedure leads to estimates which produce smaller sums of squared error terms than does the eigenvector method.

In a third (and somewhat more specialized) area, stochastic judgments are obtained by considering group decision making and the concomitant variability in judgments associated with the opinions of a committee. Here, the aggregation of group judgments provides a natural linkage to the use of stochastic approaches. As such, methods for group multi-criteria decision making are somewhat more likely to have Type D characteristics. For example, Ramanathan (1997) employed the fact that normality of errors may be safely assumed when the number of judges is large (i.e., greater than 30). Basak (1991) proposed a technique in which Gamma-distributed errors were assumed and pairwise comparisons were elicited both above and below the diagonal of the judgment matrix. This technique was shown to have desirable properties when the number of judges was as small as 3. Methods for group multi-criteria decision making in which judgments are not aggregated but are kept separate comprise an interesting special case of the field, but because they are less broadly applicable they will not be discussed at length here.

The Type D approach in the current paper diverges somewhat from these three main areas of inquiry found in the literature. For example, it is important to note that neither interval judgments nor multiple judges are required to use the method to be described herein. In other words, the method proposed here is fully compatible with the standard AHP in that no modifications of the original Saaty procedure are needed. Moreover, statistical models are used in the method presented here, but the estimation technique is Bayesian as opposed to maximum likelihood or least squares. Nonetheless, work by two sets of authors can be seen to be more related to the method proposed here. Lipovetsky and Tishler (1999) describe a more general method that corresponds with the Type D category described in Figure 1. They noted that the ratio of two normal variates has a Cauchy distribution and applied this observation to pairwise comparison ratios. Their method allows for inferential statements to be made about the priority vector, a topic to be considered in this paper, and they also specified a sampling distribution for pairwise comparison ratios. However, their method requires the use of a reciprocally nonsymmetric judgment matrix in which the elements in the upper triangle have a greater impact on the final priority vector than do the elements in the lower triangle. Additionally, their method does not address the dependence of observations in the judgment matrix, and so the results of the inferential procedure may be inaccurate. Basak (1998) provides a method for making inference on the priority vector using a Bayesian framework. In particular, the article extends a framework originally due to Gelfand, Mallick and Dey (1995) and makes it suitable for use with the AHP. The method requires several steps and two distinct parties. The first party is known as the decision maker. From him or her, it is necessary to elicit a matrix of probabilities, each element of which is the prior probability that a particular pairwise comparison will fall in a certain interval of the Saaty scale (e.g., the interval 4 to 6). There is also a second party which corresponds to a panel of experts. Each one of these experts must also supply a matrix of probabilities such as the one required of the decision maker. We can then return to the decision maker and elicit information about his or her confidence in the judgments of the experts. This is achieved by eliciting a shape parameter for a Beta distribution for each expert where larger values of this parameter correspond to smaller probability variances and therefore higher confidence levels on the part of the decision maker. The priority vector is then obtained by Bayesian Monte Carlo methods; subsequent inference on the priority vector is conducted with frequentist significance tests. In summary, the paper presents several advances but the technique described is rather involved and additionally requires both a decision maker and a separate panel of experts. Moreover, dependence between judgments is not addressed. It is worth stressing that the last of these issues is important to handle because dependency in the judgments is a fundamental characteristic of pairwise comparison matrices in the AHP. For example, in the case of perfectly consistent judgments, knowing the pairwise comparison ratios C_{ij} and C_{jk} allows us to obtain the pairwise comparison ratio C_{ik} exactly from the product $C_{ij} \times C_{jk}$. In the case of modest amounts of inconsistency, we still may obtain a degree of information about C_{ik}

from $C_{ij} \times C_{jk}$. Techniques for addressing these fundamental dependencies in pairwise comparison matrices will be discussed in the next section.

A STOCHASTIC METHOD FOR OBTAINING PRIORITIES

The Sampling Model for the Priorities

A probabilistic characterization of the process of making pairwise comparisons that can be applied to the AHP is now developed (see also Bemmaor & Wagner, 2000, for a related approach). Consider the pairwise comparison ratio C_{ij}, where *i* ¹ *j*, that has resulted from the pairwise comparison of two and only two alternatives O_i and O_j with weights w_i and w_j. For the moment, take $w_i \geq w_j$, such that $C_{ij} = \{1, 2, \ldots, 9\}$. Then C_{ij} expresses the amount by which O_i is preferred to O_j. Specifically, for every outcome of preference for O_j, there are C_{ij} outcomes of preference for O_i. We can conceptualize this as the ratio of success outcomes and failure outcomes in a binomial process. As such, the pairwise comparison ratios can be used to obtain the components of a binomial process in which w_i successes have been observed in $(w_i + w_j)$ trials subject to an unobserved preference parameter, p_i. With no loss of generality, we can divide the numerator and the denominator of C_{ij} by the sum of the weights to obtain

$$C_{ij} = \frac{w_i}{w_j} = \frac{\dfrac{w_i}{w_i + w_j}}{\dfrac{w_j}{w_i + w_j}} = \frac{p_i}{1 - p_i}. \qquad (1)$$

Here, $p_i/(1 - p_i)$ is the ratio of preferences and we see that p_i will be the stochastically derived priority. The priority p_i is such that $0 < p_i < 1$ in the present context since by definition the act of pairwise comparison requires the presence of non-zero weights w_i and w_j associated with O_i and O_j respectively. Again, we can conceptualize that w_i has a binomial distribution with parameters $w_i + w_j$ and p_i, which we write as $w_i \sim Binomial(w_i + w_j, p_i)$. Note that in the cases where $w_i < w_j$, it remains true that $w_i \sim Binomial(w_i + w_j, p_i)$.

Many times a decision maker will be faced with more than two alternatives. In this case, the underlying process is multinomial by extension. If there are K alternatives $O_1, O_2, \ldots, O_K$ with weights $w_1, w_2, \ldots, w_K$, then the i^{th} row of the pairwise comparison matrix has a multinomial distribution. That is,

$$(w_{i1}, w_{i2}, \ldots, w_{iK}) \sim Multinomial(w_{i1} + w_{i2} + \ldots + w_{iK}, \boldsymbol{p}_i),$$

where $\boldsymbol{p}_i$ is a vector of preference parameter or priorities such that $\sum_{k=1}^{K} p_{ik} = 1$. Again, since all K alternatives are present by definition, it must be true that $0 < p_{ik} < 1$. With K alternatives, the matrix of pairwise comparisons will contain K multinomial trials. Thus, the matrix of pairwise comparisons is square with K columns, each one corresponding to an alternative, and K rows, each one corresponding to a different trial. Having supplied a probabilistic characterization of the pairwise comparison process and the resulting matrix of pairwise comparisons, it is possible to specify statistical models for the prediction of outcomes. Of primary interest is $\boldsymbol{p}$, the vector of marginal priorities for the alternatives. A natural model for the problem of interest is the multinomial logit model (e.g., McFadden, 1973). Using this general model, a Bayesian perspective will be adopted for inference on $\boldsymbol{p}$, and estimation will be conducted using Markov chain Monte Carlo (MCMC) methods (for an introduction to MCMC methods in Bayesian inference, see Casella and George, 1992, Chib and Greenberg, 1995, or Gamerman, 1997).

The Multinomial Logit Model

We now specify the form of the multinomial models to be considered in greater detail. The logit link is used in all of the models considered, and the following relations hold in all of the multinomial models to be examined:

$$w_{ik} \sim Multinomial(\sum_{k=1}^{K} w_{ik}, p_{ik}), \tag{1}$$

$$p_{ik} = \frac{f_{ik}}{\sum_{k=1}^{K} f_{ik}}, \tag{2}$$

$$f_{ik} = \exp(a_k + b_{ik}). \tag{3}$$

Expression (1), which was discussed in the previous section, indicates that the weights have a multinomial distribution given the underlying priorities. Expression (2) shows that the priorities result from the normalization of the log-linear predictors, f_{ik}. The log-linear predictors are specified as $f_{ik} = \exp(a_k + b_{ik})$, as can be seen in Expression (3). Note that a set of coefficients, $a_2, \ldots, a_K$, is associated with the alternatives. The coefficient a_1 is constrained to be zero to ensure identifiability (Agresti, 1990, p. 313). A second set of coefficients, $b_{22}, \ldots, b_{2K}, b_{32} \ldots, b_{KK}$, is associated with the trials. Again, it is necessary to constrain the coefficients $b_{11}, \ldots, b_{K1}$ and $b_{12}, \ldots, b_{1K}$ to zero for the purposes of identifiability. Obtaining the final priority for each alternative involves averaging the relevant values of p_{ik} over the trials. Specifically, the final priority of the k^{th} alternative is $\frac{1}{I}\sum_{i=1}^{I} p_{ik}$ and the vector of priorities, $\boldsymbol{p}$, is the collection of these final alternative-specific priorities.

In many applications of statistical modeling, the $\boldsymbol{a}$ and $\boldsymbol{b}$ coefficients are of primary interest and other estimation tasks have limited relevance. For example, in regression the value of $\boldsymbol{b}$ in particular may shed light on an experimental hypothesis, whereas analyses regarding the predicted values, $\hat{y}$, may provide little value-add. By contrast, we can imagine some situations where the reverse is true. For example, a manager may be keenly interested in the prediction estimates and prediction intervals for her sales teams in order to determine whether her firm will be profitable. Here, $\hat{y}$ would have the focus whereas $\boldsymbol{a}$ and $\boldsymbol{b}$ would not. It is worth pausing to emphasize that in the current method the primary interest lies in obtaining $\boldsymbol{p}$. This is akin to the manager's situation above where there is a focus on outcomes as opposed to predictors. Furthermore, the $\boldsymbol{a}$ and $\boldsymbol{b}$ coefficients do not have especially relevant interpretations in the current context. The $\boldsymbol{b}$ coefficients will tend to zero as $\boldsymbol{C}$ tends toward perfect consistency, and will diverge from zero as inconsistency increases. However, there already exists an easily interpreted and commonly used measure of inconsistency for AHP, the inconsistency ratio. So the $\boldsymbol{b}$ coefficients seem to add little of managerial import. The $\boldsymbol{a}$ coefficients will tend to zero as the first baseline or reference alternative is similar to the remaining alternatives. However, they do not take the inconsistency fully into consideration. Hence, these coefficients also have limited relevance, and instead interest centers almost exclusively on $\boldsymbol{p}$.

92).

Two formulations of the multinomial logit model are considered here. Model 1 is a marginal multinomial model. Specifically, in a marginal model all of the coefficients are estimated independently of one another. As such, all of the as and bs are estimated independently in Model 1. Of course, there will be dependence across trials for several reasons. First, the AHP in its original formulation makes use of a fully specified matrix of pairwise comparisons. Therefore a certain, usually high, degree of redundancy will be present in such matrices. Indeed, information about the extent of this redundancy is used in the AHP to calculate a measure of consistency in the judgments. Second, the multinomial logit model requires the w_{ik}s, which we must obtain from $\boldsymbol{C}$. It is necessary to use the w_{ik}s implied by previous values of $\boldsymbol{C}$ to ensure that there is no missing data on any of the trials (this topic that will be addressed in the section entitled 'Obtaining w from $\boldsymbol{C}$'). Thus, Model 1 admittedly ignores a feature of the data, but is nonetheless estimated for comparative purposes. A better model would account for the dependencies among the w_{ik}s. Hence, Model 2 is a hierarchical model that allows for dependency in the weights across the different trials. In it, the bs are drawn from a common normal distribution with a mean of zero and precision parameter t. The relationship can be expressed as follows:

$$b_{22},\ldots,b_{2K},b_{32},\ldots,b_{KK} \sim Normal(0,t)\,.$$

Note that the current parameterization is one in which t is a precision parameter as opposed to a variance parameter. The precision is the reciprocal of the variance (i.e., $t = 1/s^{2}$) and is commonly used in Bayesian inference for computational reasons (see, e.g., Gill, 2002, pp. 90-92).

To summarize, Model 1 is a marginal multinomial model, whereas Model 2 is a hierarchical model. The bs are assumed to have a common normal distribution in the hierarchical models, whereas they are assumed to be independent in the marginal model. In Model 2, the common normal distribution for the bs is specified to have a mean of zero, and we then estimate the unknown precision parameter t.

Priors and MCMC Implementation

Bayesian analysis requires the specification of prior distributions for the parameters to be estimated (see for example Berger, 1985, ch. 3). Here, vague but proper priors are used for all of the parameters (excepting those which have been set to zero to ensure identifiability). Such priors are minimally informative. In Model 1, all of the as and bs are given vague but proper independent normal priors. In Model 2, t is given a vague but proper gamma prior, and the as are again given vague independent normal priors. The posterior distributions of the quantities of interest were obtained via the MCMC method of Gibbs sampling. From an MCMC computational perspective, the models considered here were well behaved. Convergence to the posterior was rapid, and the chains did not exhibit substantial autocorrelation.

Obtaining *w* from *C*

The eigenvalue decomposition makes use of a matrix of pairwise comparisons, $\boldsymbol{C}$, to obtain $\boldsymbol{p}$, whereas the multinomial logit model employs the weights, w_{ik}, from which $\boldsymbol{C}$ is composed.

Unfortunately, the w_{ik}s will typically not be known in the course of using the AHP. This is because the decision maker supplies the elements of $\boldsymbol{C}$, not the w_{ik}s, as input to the AHP. Thus, we employ the following process in every multinomial trial in order to provide a unique solution for the set of all weights, $\boldsymbol{w}$. The least preferred alternative in a particular trial is given a weight of 1. The remaining alternatives are integer multiples of the least preferred alternative in accordance with the appropriate values of $\boldsymbol{C}$. For example, if the pairwise comparison of A to B is 2:1 in the first trial, and the pairwise comparison of A to C is 4:1 in the first trial, then (w_{11}, w_{12}, w_{13}) = {4, 2, 1} in this particular trial. From an operational perspective, another way of describing this procedure is as follows. For each row of the matrix, take the reciprocal of the weights. Then multiply all the weights by a constant such that the smallest weight equals one. As an example, consider the matrix

$$\begin{pmatrix} 1 & 2 & 4 \\ 1/2 & 1 & 2 \\ 1/4 & 1/2 & 1 \end{pmatrix}.$$

Using the procedure described above, we obtain

$$\begin{pmatrix} 4 & 2 & 1 \\ 4 & 2 & 1 \\ 4 & 2 & 1 \end{pmatrix}$$

which we use as data for the multinomial logit model.

STOCHASTIC MODELS FOR OBTAINING PRIORITIES – EMPIRICAL EXAMPLES

Two stochastic models have been proposed for obtaining priorities. Model 1 was a marginal multinomial model. Model 2 was a hierarchical model in which the dependency among the *b*s (and, by extension, the w_{ik}s) was addressed by modeling them as being drawn from a common normal distribution with a mean of zero and unknown precision t. We now compare the priorities from these approaches with the priorities from AHP. We employ the mean absolute deviation (MAD) of the model priorities from the AHP priorities as our measure of a model's discrepancy. We use this measure for the following two reasons. First, the numeric value of the MAD lends itself to straightforward interpretation. A MAD of 0.001 indicates that the model's agreement with AHP extends well through the second decimal place to slight deviations at the third decimal place of accuracy, whereas a MAD of 0.01 indicates a tendency for slight deviations at the second decimal place. It turns out that other measures (such as the mean squared error) possess interpretations which may be less immediately useful. Second, linear regression will be used in a subsequent Monte Carlo study of the method's performance. The MSE criterion increases quadratically with increasing deviations and as such would unnecessarily complicate a linear regression analysis by introducing a non-linear relationship.

Priorities for Models with Three Alternatives

As a point of departure, we first examine a perfectly consistent matrix with three alternatives. Here, Matrix 1 is constructed such that C_{12} is 2:1, C_{13} is 4:1, and C_{23} is 2:1. The point estimates of the priorities obtained via the different methods appear in Table 1. These estimates were based on the results from an MCMC run of 50,000 iterations after a burn-in of 10,000 iterations. Table 1 reveals that priorities are all the same, regardless of which model is used. In general, the weights will be identical across trials for a consistent matrix, and so the multinomial models will produce the same priorities as the eigenvalue decomposition.

Table 1 – Priorities for Consistent Matrix with Three Alternatives

	Alternative 1	Alternative 2	Alternative 3
Principal Eigenvector	0.571	0.286	0.143
Model 1	0.571	0.286	0.143
Model 2	0.571	0.286	0.143

The second matrix to be considered is one in which an introduced error produces a relatively small amount of inconsistency. The elements of Matrix 2 are such that C_{12} is 2:1, C_{13} is 8:1, and C_{23} is 2:1. In words, this means that Alternative 1 is again judged to be twice as preferable as Alternative 2. However, now Alternative 1 is judged to be 8 times as preferable as Alternative 3. Alternative 2 is again twice as preferable as Alternative 3. Hence, Matrix 2 is the same as Matrix 1 except that C_{13} is twice as large in Matrix 2 as it was in Matrix 1. The AHP does not require perfect consistency among judgments in order for the calculation of priorities to proceed. If the judgments in a matrix are not perfectly consistent, information about the extent of inconsistency in the matrix can be obtained and a measure of inconsistency can be calculated (Saaty, 1977). In particular, a measure called the inconsistency ratio is often used. Heuristically, when the value of the inconsistency ratio exceeds 0.10, the inconsistency level is said to be such that the decision maker may wish to re-examine his or her judgments. The inconsistency ratio for Matrix 2 is 0.05 and so reconsideration of the judgments is not necessary.

Table 2 contains the priorities obtained via the different methods for Matrix 2. With the eigenvector method, Alternative 1 has a higher priority and Alternatives 2 and 3 have lower priorities when compared to Table 1. All of the methods again produce results that are close to the eigenvector solution. Here, Model 2 most resembled the eigenvector method (MAD = 0.0029). While both stochastic models were able to approximately replicate the priorities of the eigenvector solution, Model 1's priorities for Alternatives 1 and 3 tended to be pulled toward the central value of 0.5.

Table 2 – Priorities for Three-Alternative Matrix with Inconsistency

	Alternative 1	Alternative 2	Alternative 3	MAD
Principal Eigenvector	0.643	0.255	0.101	
Model 1	0.638	0.258	0.104	0.0035
Model 2	0.645	0.258	0.097	0.0029

Priorities for Models with Four Alternatives

The case of matrices with four alternatives is also considered. For the purposes of comparison, we construct a four-alternative matrix that is related to Matrix 1. The new matrix, Matrix 3, is designed such that C_{12} is 2:1, C_{13} is 4:1, C_{14} is 4:1, C_{23} is 2:1, C_{24} is 2:1, and C_{34} is 1:1. Thus, the new matrix is similar to Matrix 1, except a fourth alternative identical to the third is added. As was the case with Matrix 1, the different methods produce identical results when the matrix is perfectly consistent. In terms of this example, all of the methods recovered the priorities accurately as 0.5, 0.25, 0.125, and 0.125, excluding some trivial Monte Carlo error. Hence, we do not report further on this matrix. Instead, we introduce some inconsistency to it to obtain Matrix 4. In Matrix 4, C_{12} is 2:1, C_{13} is 8:1, C_{14} is 4:1, C_{23} is 1:1, C_{24} is 2:1, and C_{34} is 1:1. The inconsistency ratio of this matrix is 0.07. Table 3 contains the priorities obtained via the different methods for Matrix 4. Model 2's MAD is again the smaller of the two (MAD = 0.0072).

Table 3 – Priorities for Four-Alternative Matrix with Inconsistency

	Alternative 1	Alternative 2	Alternative 3	Alternative 4	MAD
Principal Eigenvector	0.568	0.198	0.12	0.113	
Model 1	0.551	0.208	0.126	0.115	0.0086
Model 2	0.557	0.210	0.117	0.116	0.0072

Examples of Stochastic Models for Obtaining Priorities – Summary

The models examined here replicated the eigenvector solution in the case of a perfectly consistent matrix. However, small differences in priority estimates appeared in the presence of inconsistency. These differences in priority estimates may additionally be a function of the number of alternatives under consideration. It is therefore of interest to more systematically examine the relationships between matrix inconsistency, the number of alternatives, and model performance. With this goal in mind, a Monte Carlo study was conducted to examine these possible relationships. We describe this study in the following section.

STOCHASTIC MODELS FOR OBTAINING PRIORITIES – A MONTE CARLO STUDY

To examine the performance of the models more extensively, we conducted a Monte Carlo study using an experimental design and randomly generated matrices. This approach allowed us to examine the models' performance under a variety of conditions. The main experimental factor was the *Number of Alternatives*. In particular, we considered matrices with 3, 4, 5, 6, 7, and 8 decision alternatives, and so this factor had 6 levels. We generated 20 matrices of pairwise comparisons within each level of this factor. These matrices were randomly generated to possess a Consistency Index of less than 0.10 (recall that a Consistency Index value of 0.10 is commonly used as the cutoff value for the maximum acceptable level of inconsistency). Matrices which had a greater amount of inconsistency were discarded. Thus, we obtained a randomly generated sample of matrices having differing amounts of consistency. As such, the 20 matrices allowed us to examine the performance of the models with respect to the continuous factor, or dimension, of *Matrix Inconsistency*. In total we generated 6×20 matrices for the Monte Carlo study. Estimates of the priorities under both Models 1 and 2 were calculated for each matrix. Hence, there were two levels of the factor *Model Type*, with the levels being Model 1 and Model 2. In summary, a total of $6 \times 20 \times 2$ sets of priority estimates were calculated. Because of the massive computational demands of the Monte Carlo study, each set of estimates was produced from 50,000 MCMC iterations after a 5,000-iteration burn-in period had transpired.

Figure 2 – Discrepancy from AHP Priorities: Model 1

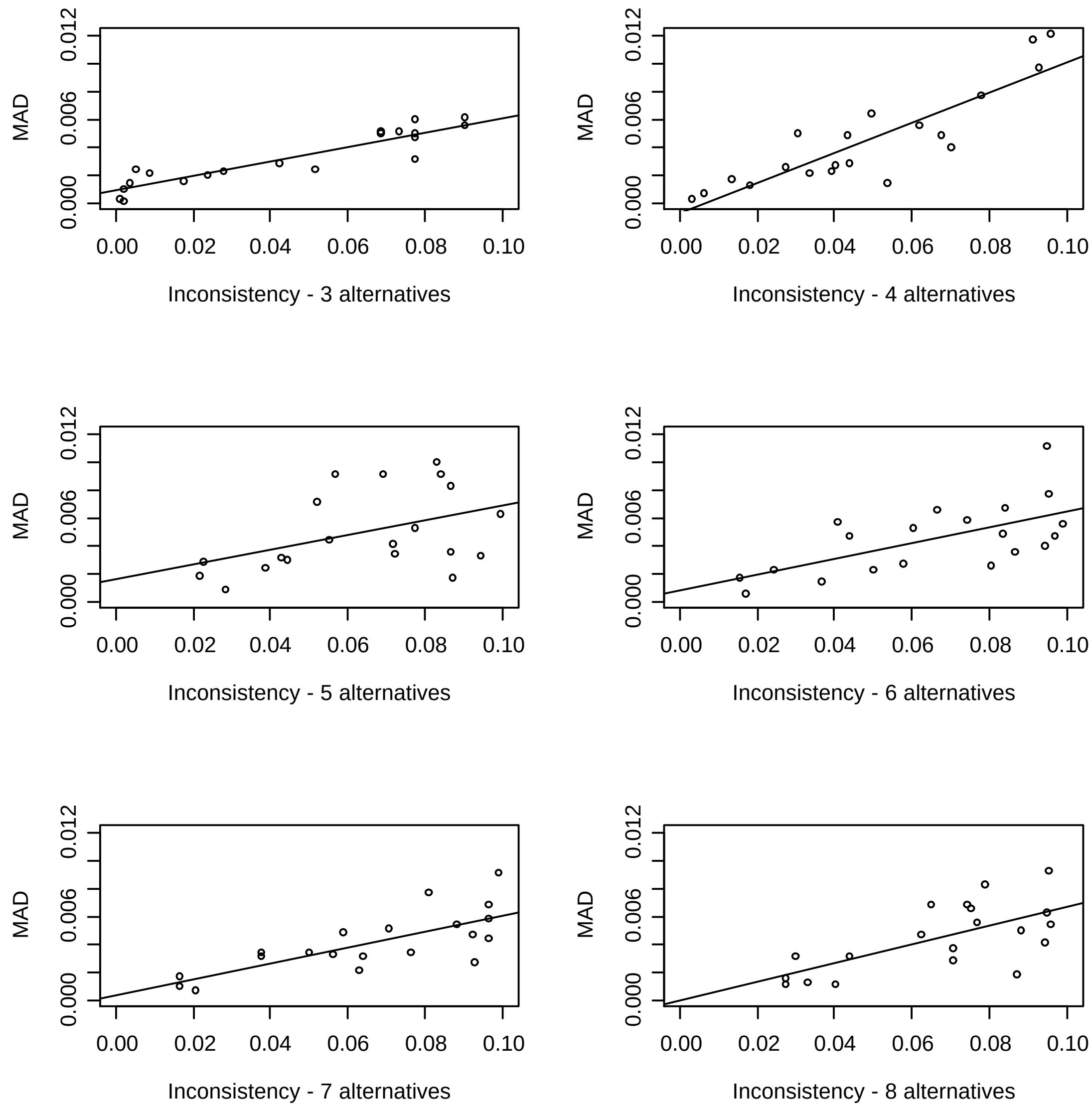


Table 4 – Summary Measures: Model 1

	Intercept	Slope	Slope S.E.	Slope/S.E.	Expected MAD at 0.10
3 Alts.	0.000848	0.0523	0.0052	10.08	0.0061
4 Alts.	-0.000788	0.1085	0.0141	7.72	0.0101
5 Alts.	0.001536	0.0528	0.0249	2.12	0.0068
6 Alts.	0.000717	0.0571	0.0162	3.53	0.0064
7 Alts.	0.000295	0.0571	0.0120	4.74	0.0060
8 Alts.	-0.000090	0.0671	0.0174	3.85	0.0066

Results – Model 1

The scatter plots of the relationships between matrix inconsistency and MAD under Model 1 with respect to the differing *Number of Alternatives* appear in Figure 2. The respective regression lines are also plotted in Figure 2. The scatter plots show that the relationship between the two variables is strongly positive. This remains true for all of the values of Number of Alternatives considered. Hence, the plots show that when inconsistency in the judgments is low, the MAD is similarly decreased. For example, in the case of 8 alternatives and an inconsistency of 0.016, the expected MAD (as indicated by the regression line) is approximately 0.001. This means that there is about a 0.1% discrepancy between the AHP priorities and the priorities from Model 1 on average when there are 8 alternatives under consideration.

The intercepts and slopes associated with the Model 1 regression lines in Figure 2 appear in the first and second columns of Table 4. Also, the third column in Table 4 gives the standard errors of these slopes. We can divide the slope by its standard error to obtain a *t* statistic, which we display for reference in the fourth column of the table. We may compare this statistic to critical values taken from a *t* distribution with 18 degrees of freedom. The latter information allows us to conclude that each particular slope is not equal to zero at the 99% probability level or better (as the Bayesian approach with flat priors yields the same results as the classical approach in simple linear regression). The final column of Table 4 displays the values of the different regression lines at an inconsistency level of 0.10. This corresponds to the expected value of the MAD at the highest level of inconsistency deemed generally acceptable. Graphically, these also correspond to the right-most points on the regression lines in Figure 2. We see that, with the exception of the case of four alternatives, the expected maximum MAD ranges from 0.006 to 0.0068. Hence, even in the case of a relatively high (yet acceptable) level of inconsistency in the judgment matrix, we would expect there to be little difference between the priorities generated by Model 1 and those generated by AHP.

Figure 3 – Discrepancy from AHP Priorities: Model 2

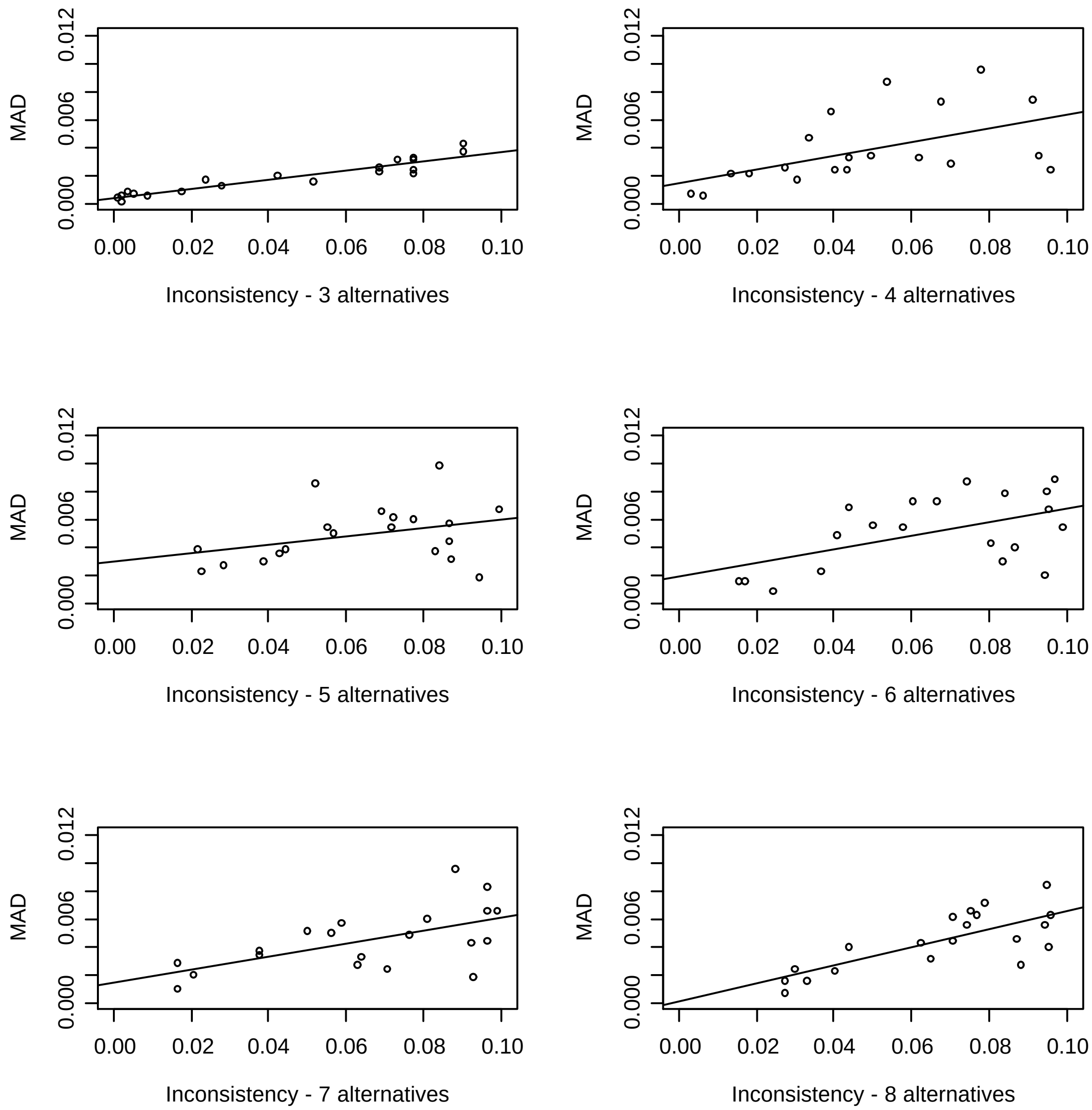


Table 5 – Summary Measures: Model 2

	Intercept	Slope	Slope S.E.	Slope/S.E.	Expected MAD at 0.10
3 Alts.	0.000324	0.0335	0.0028	11.98	0.0037
4 Alts.	0.001465	0.0490	0.0186	2.64	0.0064
5 Alts.	0.002972	0.0293	0.0187	1.57	0.0059
6 Alts.	0.001855	0.0494	0.0182	2.71	0.0068
7 Alts.	0.001395	0.0463	0.0146	3.17	0.0060
8 Alts.	0.000011	0.0649	0.0132	4.90	0.0065

Results – Model 2

The scatter plots of the relationships between matrix inconsistency and MAD under Model 2 with respect to the differing *Number of Alternatives* appear in Figure 3. It is immediately apparent that the slopes of the Model 2 regression lines in the cases of 3 and 4 alternatives are considerably less steep than their Model 1 counterparts. In the cases of 5 and greater alternatives, the differences between the Model 1 and Model 2 regression lines are considerably harder to distinguish. This is especially true in the case of 7 and 8 alternatives. One interpretation of this observation is that the impact of accounting for dependence becomes less important as the number of alternatives increases.

The intercepts and slopes associated with the Model 2 regression lines are similarly shown in Table 5. The third column in Table 5 gives the standard errors of these slopes, and we can conclude from the information in the fourth column that each particular slope is not equal to zero at the 98% probability level or beyond, excepting the case of 5 alternatives. In the case of 5 alternatives, we may only conclude that the slope is different from zero at the 86% probability level. Table 5 also contains the values of the regression lines at an inconsistency level of 0.10

under Model 2. In the case of three alternatives, the expected MAD at an inconsistency level of 0.10 under Model 2 is about 40% less than it is under Model 1. Hence, in the case of three alternatives, Model 2 seems to provide superior performance over Model 1 when there are higher levels of inconsistency. Model 2 also seems to be considerably superior in the case of 4 alternatives and higher levels of inconsistency. Model 2 remains superior, albeit at a much more modest level, with 5 alternatives. With 6 to 8 alternatives, the models tend to converge toward one another in terms of performance. Hence, in situations where the number of alternatives is from 5 to 6 or greater, there seems to be little to be gained from selecting one model over the other. However, Model 2 seems to be clearly preferable when the number of alternatives is 3 or 4. Looking at the overall performance of Model 2, we see the expected maximum MAD tends to range from 0.0037 to 0.0068. So, in the case of a high yet acceptable level of inconsistency in the judgment matrix, we would not expect there to be too great of a difference between the priorities generated by Model 2 and those generated by AHP.

Monte Carlo Study of Stochastic Models for Obtaining Priorities – Summary

The Monte Carlo study provides supporting evidence for three conclusions. First, we see that Models 1 and 2 generally approximate AHP priorities well irrespective of the number of decision alternatives under consideration. Given the similarity of the results under 6, 7, and 8 alternatives, it may be that there is some very preliminary evidence of a plateau where the models handle increasing numbers of alternatives at a particular performance level. Future research with larger numbers of alternatives is needed to establish results for those cases. Second, we find that the MAD increases with the amount of inconsistency, at least for the range of inconsistency values explored here. If a smaller MAD is felt to be desirable, it may be possible to ask the decision maker to reflect on or re-examine his or her judgments. Lastly, we see that Model 2 seems to be clearly superior to Model 1 in the case of either 3 or 4 decision alternatives. By the MAD criterion, Models 1 and 2 seem to converge for larger number of alternatives; however, recall that Model 2 is the one that addresses dependence in judgments so its use may be preferred on that criterion.

METHODS FOR INTERVAL ESTIMATION AND INFERENCE ON THE PRIORITIES

In the empirical analyses described above, different models were compared with respect to their ability to produce point estimates of the priorities that were comparable to those yielded by the eigenvector method. However, because the models examined here belong to the Type D category of multi-criteria decision methods, they also provide information about the variability associated with these estimates. This is an important advance, as this information allows us to construct confidence intervals around the estimates and thereby empirically determine whether two alternatives have different priority levels. In this section, we examine the characteristics of judgment matrices in order to determine how we may appropriately construct confidence intervals. In particular, we are concerned with determining the quantity of information in a judgment matrix. We begin by noting that some of the entries in the judgment matrix are inherently redundant, and that this artificially increases the sample size of the matrix. We then propose a method of weighting which effectively decreases the sample size of the matrix back to its appropriate size.

The existence of multiple rows in $\boldsymbol{C}$ which are generally consistent with one another of course leads to increased precision of the estimates. This increased precision is reflected in decreased values for measures of dispersion such as parameter standard deviations. However, it may be that in a given row that some of the w_{ik}s are redundant with previously supplied w_{ik}s. While it is necessary to have repetition of the w_{ik}s in order to complete $\boldsymbol{C}$, it is inappropriate for parameter standard deviations to be made unnecessarily smaller because of this repetition. One remedy for this is to use a row weight or an "information sample size" multiplier. This can be used to weight the rows downward such that the multiplicity of rows is effectively eliminated. A simple way to implement this is to weight each of the I rows by $1/I$ so that the sum of the row weights, R, is 1. This has the effect of decreasing the number of observations associated with any w_{ik} to be w_{ik}/I. This method of weighting is unduly conservative, and other methods of weighting are of course possible. We now describe one such method that has better properties.

Note that in the first row of $\boldsymbol{C}$ none of the w_{ik}s is redundant, whereas all of them are in the last row. In the middle rows of matrices having size $I > 2$, some proportion of the information will be non-redundant. Based on the quantity of non-redundant w_{ik}s, we can see that a more accurate weighting scheme would arise from weighting by the proportion of non-redundant information present in each row. Denote the sum of these proportions of non-redundant row information as R_K. Then, R_K is $1+\sum_{i=2}^{I-1}\left(1-\frac{i-1}{I}\right)$ when $I > 2$, and is 1 when $I = 2$. After obtaining R_K, we then redistribute the weight to the rows equally such that each row receives the row weight R_K/I. Note that R_K functions as the "downweighted" or corrected number of rows for the judgment matrix. So, by extension we can also think of R_K/I as sample size deflator, or alternatively as the overall proportion of non-redundant information in the matrix. We present the following theorem about R_K.

Theorem

For a square judgment matrix having I rows and columns,

$$\lim_{I \to \infty} \frac{R_K}{I} = \frac{1}{2}.$$

Moreover,

$$\lim_{I \to \infty} R_K = \frac{I}{2}$$

provided that I is finite.

Proof

A proof of the theorem appears in the Appendix. The proof also shows that we may more readily obtain R_K/I by using the simpler expression $\frac{I^2 + I - 2}{2I^2}$.

The theorem tells us that 50% of the information in a judgment matrix is non-redundant in the limit where the number of alternatives grows very large. Alternatively, we could say that the deflated sample sizes will be one half of the uncorrected sample sizes in the limit. This accords with our intuition that the half of the judgments below the diagonal (and on the entire last row) are redundant and so should not "count" toward the overall sample size. Similarly, the corrected number of rows is $I/2$ in the limit as long as I still remains finite. In conclusion, for occasions where inference on the priorities is of interest, a weighted multinomial model is appropriate. Thus, the likelihood of the models should take the form

$$w_{ik} \sim Multinomial(Q\sum_{k=1}^{K} w_{ik}, p_{ik})$$

where Q is some constant such as R_K/I.

Interval Estimation and Inference on the Priorities – Example

Consider an organizational decision regarding which of three technology products should be prototyped and eventually brought to market. The three alternatives are Product 1 (p_1 – a wrist-based personal computing device), Product 2 (p_2 – a high bandwidth cell phone), and Product 3 (p_3 – a voice-recognition personal digital assistant). It is of interest to use multi-criteria decision making methods regarding the future success of these products, but there is a non-trivial degree of uncertainty in the judgments. As such, we use the methods described in the current paper to address this problem. In particular, we concern ourselves with inference regarding these three alternatives. A weighted version of Model 2 is estimated in which the row weights are R_K/I. The decision maker is asked to make judgments, and her judgments are as follows. The wrist-based personal computing device is considered to be moderately to strongly superior to the high bandwidth cell phone from the standpoint of future success in the marketplace. The wrist-based personal computing device is judged to be extremely superior vis-à-vis the voice-recognition personal digital assistant from the standpoint of future success in the marketplace. Finally, the high bandwidth cell phone is considered to be moderately superior to the voice-recognition personal digital assistant with respect to future success in the marketplace. These judgments are translated into a 3:1 preference for Product 1 versus Product 2, a 9:1 preference for Product 1 versus Product 3, and a 2:1 preference for Product 2 versus Product 3. Hence, we have C_{12} = 3:1, C_{13} = 9:1, and C_{23} = 2:1. Via the mathematics of AHP, the priorities for Products 1 through 3 respectively are 0.705, 0.205, and 0.090. The inconsistency ratio is 0.02.

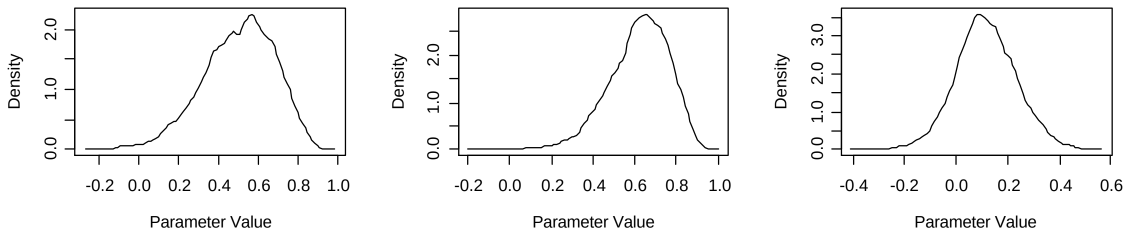
One of the benefits of an MCMC approach to Bayesian estimation is that it is straightforward to obtain the posterior distributions for arbitrary functions of the parameters once the model has been specified. Then, inference can be performed using the posterior distributions of these functions. Functions that would be of interest for inferential purposes would include the pairwise differences among the priorities. Specifically, we would like to examine the differences $p_1 - p_2$, $p_1 - p_3$, and $p_2 - p_3$. The output of the MCMC run allows us to construct credible intervals, which are Bayesian analogues of confidence intervals, for these differences. Here, we construct 95% credible intervals by identifying the values at the 2.5% and 97.5% quantiles of the posterior distribution. We may then determine whether the value of zero is included in a particular interval. If the interval does not include zero, then we may conclude that we are 95% confident that the two priorities are different from one another.

Table 6 – Summary Statistics for Posterior Distributions of Priorities and Differences in Priorities

Parameter	Mean	Standard Deviation	2.5% Quantile	Median	97.5% Quantile
p_1	0.706	0.102	0.491	0.713	0.884
p_2	0.206	0.090	0.061	0.195	0.407
p_3	0.088	0.064	0.008	0.074	0.248
$p_1 - p_2$	0.501	0.182	0.109	0.517	0.807
$p_1 - p_3$	0.618	0.144	0.298	0.633	0.855
$p_2 - p_3$	0.117	0.118	-0.114	0.114	0.357

For the analyses reported here, the values of 50,000 iterations were used for posterior estimation after a 5,000 iteration burn-in had been conducted. Table 6 displays summary statistics for the posteriors of the priorities, as well as for the differences in priorities. Additionally, Figure 4 contains plots of the posteriors for the three difference scores. The fourth row of Table 6 reveals that there is evidence to indicate that p_1 is significantly different from p_2. Additionally, there is evidence to indicate that p_1 is different from p_3. However, the 95% credible interval for the difference between p_2 and p_3 ranges from -0.114 to 0.357. Thus, the decision maker should be indifferent between Products 2 and 3 at the 95% probability level. In summary, despite uncertainty in the judgments of the decision maker, she can be at least 95% confident that the wrist-based personal computing device has a higher priority than does either of the two remaining alternatives. Furthermore, she cannot be 95% confident that the two remaining alternatives have unequal priorities.

Figure 4 – Posterior Distributions of Differences in Priorities



(a) $p_1 - p_2$ (b) $p_1 - p_3$ (c) $p_2 - p_3$

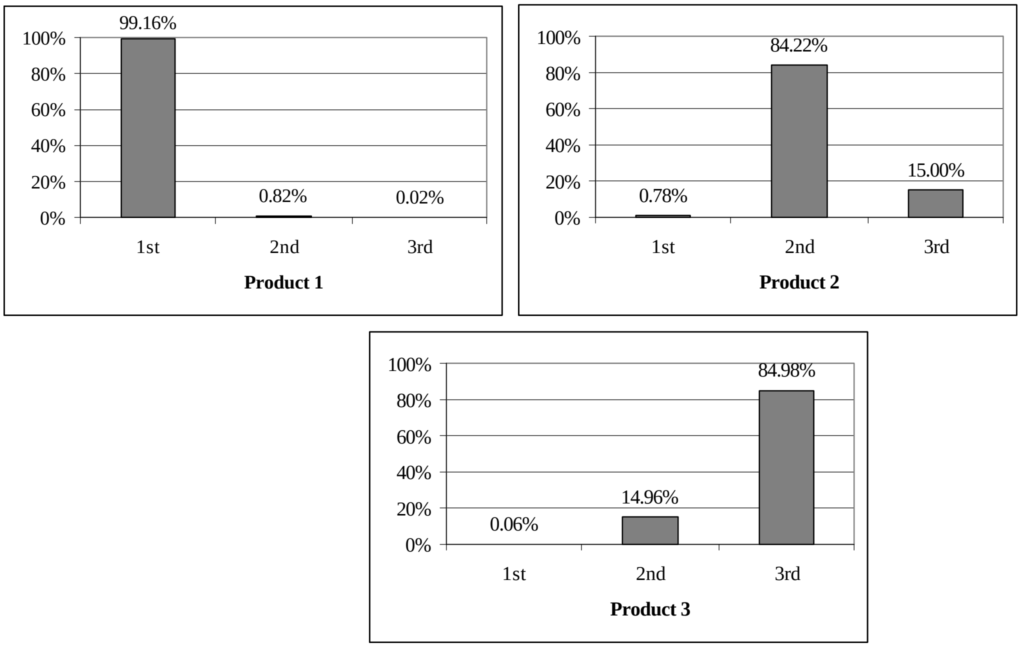
Many times a decision maker will be concerned with determining which one of several alternatives is the most attractive. In these instances, attention will center on the highest ranked alternative. Hence, some decision makers may find it preferable to make judgments using inferential information about ranks as opposed to inferential information about differences in priority. It is also possible to obtain the posterior distribution of the ranks through the use of MCMC methods. A step function is calculated in the MCMC run that computes the rank order of an alternative based on its priority. Then, the posterior distribution of this function is available for subsequent examination. Table 7 contains summary statistics for the posterior distribution of the ranks of the three products under Model 2. Figure 5 displays the posterior distributions of the ranks.

Table 7 – Summary Statistics for Posterior Distributions of Ranks

Parameter	Mean	Standard Deviation	2.5% Quantile	Median	97.5% Quantile
p_1	1.008	0.090	1	1	1
p_2	2.142	0.371	2	2	3
p_3	2.850	0.359	2	3	3

First it is worth remarking that rank is a discrete variable, and so here Figure 5 contains a posterior probability mass function. This contrasts with the difference scores examined previously which have a continuous probability density function. The credible interval for Product 1 in Table 7 reveals that the posterior probability that Product 1 is ranked first is greater than 95%. We may see in Figure 5 that the exact probability it is ranked first is 99.16%. However, there is a small, non-zero probability that Product 1 is ranked other than first, and Figure 5 shows this probability to be 0.84%. More uncertainty exists with regard to the ranks of Products 2 and 3. While the mean rank of Product 3 is greater than that of Product 2, the 95% credible interval for both extends from a rank of second to a rank of third. That is, using Figure 5 we can see there is only an 84.22% probability that Product 2 is the second most attractive alternative. In general, the conclusions reached by examining the posterior probabilities of the ranks are operationally equivalent to those reached by examining the posterior distributions of the difference scores. However, an applied decision maker may find these results are easier to comprehend and to communicate, and hence they may be perceived as being more actionable. For example, based on the results in Table 7 a decision maker could indicate that she is at least 95% confident that Product 1 is the "best" or most preferred alternative.

Figure 5 – Posterior Distributions of Product Ranks



DISCUSSION

The stochastic methods developed for obtaining priorities produced results that were generally very similar to the priorities obtained via the eigenvalue decomposition. The models' abilities to recover AHP priorities were established for varying numbers of alternatives and varying degrees of inconsistency. To the extent that judgments are more consistent, the estimates produced by the methods described here and those of AHP become more similar. Addressing dependence in the judgments seemed to be decidedly more effective when the number of alternatives is 4 or less. Thus, Model 2 should be used when the number of alternatives is 4 or less. When the number of alternatives is 5 or greater, the benefits of Model 2 become less tangible and so either model may be used. The key advantage of the stochastic approaches developed here is that they permit inferential statements about the priorities to be evaluated, which is not possible using the deterministic approach. A decision maker can now examine posterior distributions of difference scores to see if they do or do not include certain values such as zero. Alternatively, he or she can obtain the posterior probability of the rank of an alternative. For example, if a decision maker finds that a particular alternative is the best with high probability, then he or she may have more reason to feel confident that it is indeed the best alternative. Conversely, if there is considerable uncertainty about which alternative is best, the decision maker is again in possession of valuable information bearing on the ultimate decision.

The typology of multi-criteria decision making methods clearly shows that re-examining the quantification axiom has led to significant advances. Basic multi-criteria decision making methods (Type A), which capitalize on the fact that decision makers can systematically render quantitative judgments, constitute a major improvement over informal, non-systematic, and *ad hoc* approaches to decision making. Sensitivity analysis (Type B) provides a simple way of examining a decision's robustness in the case where judgments are assumed to be certain but yet uncertainty still exists. Despite its simplicity, sensitivity analysis is nonetheless a useful starting point and often a considerable improvement over no follow-up analyses at all. Yet the decision maker is left with a largely qualitative impression of the decision's robustness. By inspecting the results of the sensitivity analysis, the decision maker may be able to say that it "seems like" the decision is robust (or is not robust), but he or she may be able to say little more than this. Further, it may be difficult to compare the results of sensitivity analyses in different problems, so it may be more challenging to indicate that one decision was "more robust" than was another decision (or less robust as the case may be). With Type C methods, it becomes possible to make inferential statements regarding the decision outcomes. This is a critical advance as it now becomes possible to quantitatively evaluate the decision's robustness. Moreover, a common metric and language, that of probability, becomes available which makes it possible to compare decisions across time and space. For example, a decision in which a decision alternative was most preferred at a 95% level of probability can be compared to an entirely unrelated decision in which a decision alternative was most preferred at a 75% level of probability. The downside of this type of method is that intervals or distributions must be elicited from the decision maker. Such a task can become quite time-consuming and mentally demanding. Type D methods, by contrast, are stochastic by design. As such, decision makers can return to the much easier task of providing scalar judgments while still benefiting from the ability to quantitatively examine a decision's robustness. Type D methods for multi-criteria decision making, therefore, have much to offer decision makers.

REFERENCES

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Arbel, A. (1989). Approximate articulation of preference and priority derivation. *European Journal of Operational Research*, *43*, 317-326.

Arkes, H.R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486-498.

Basak, I. (1991). Inference in pairwise comparison experiments based on ratio scales. *Journal of Mathematical Psychology*, *35*, 80-91.

Basak, I. (1998). Probabilistic judgments specified partially in the Analytic Hierarchy Process. *European Journal of Operational Research*, *108*, 153-164.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*, (2nd ed.). New York: Springer-Verlag.

Bemmaor, A.C., & Wagner, U. (2000). A multiple-item model of paired comparisons: Separating chance from latent preference. *Journal of Marketing Research*, *37*, 514-524.

Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46**, 167-173.

Crawford, G., & Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, *29*, 387-405.

Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327-335.

de Jong, P. (1984). A statistical approach to Saaty's scaling method for priorities. *Journal of Mathematical Psychology*, *28*, 467-478.

Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. London: Chapman & Hall.

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, 38*, 477-496.

Gelfand, A.E., Mallick, B.K., & Dey, D.K. (1995). Modeling expert opinion arising as a partial probabilistic specification. *Journal of the American Statistical Association, 90*, 598-604.

Gill, J. (2002). *Bayesian methods: A social and behavioral sciences approach*. Boca Raton, FL: Chapman & Hall/CRC.

Haines, L.M. (1998). A statistical approach to the Analytic Hierarchy Process with interval judgments. (I). Distributions on feasible regions. *European Journal of Operational Research*, *110*, 112-125.

Jensen, R.E. (1984). An alternative scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *28*, 317-332.

Keeney, R.L. (1982). Decision analysis: An overview. *Operations Research, 30*, 803-838.

Lipovetsky, S., & Tishler, A. (1999). Interval estimation of priorities in the AHP. *European Journal of Operational Research*, *114*, 153-164.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics*. New York: Academic Press, 105-142.

Moskowitz, H., Tang, J. & Lam, P. (2000). Distribution of aggregate utility using stochastic elements of multiattribute utility models. *Decision Sciences, 31*, 327-360.

Pratt, J.W., Raiffa, H., & Schlaifer, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *Journal of the American Statistical Association*, *59*, 353-375.

Ramanathan, R. (1997). Stochastic decision making using multiplicative AHP. *European Journal of Operational Research*, *97*, 543-549.

Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*, 234-281.

Saaty, T.L., & Vargas, L.G. (1987). Uncertainty and rank order in the Analytic Hierarchy Process. *European Journal of Operational Research*, *32*, 107-117.

Saaty, T.L., & Vargas, L.G. (1991). *Prediction, projection and forecasting*. Boston: Kluwer Academic Publishers.

Stam, A., & Silva, A.P.D. (1997). Stochastic judgments in the AHP: The measurement of rank reversal probabilities. *Decision Sciences*, *28*, 655-688.

Triantaphyllou, E., & Sánchez, A. (1997). A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. *Decision Sciences*, *28*, 151-194.

APPENDIX

We present here a proof of the theorem regarding R_K. Recall that

$$R_K = 1 + \sum_{i=2}^{I-1}\left(1 - \frac{i-1}{I}\right).$$

So,

$$\begin{aligned}\frac{R_K}{I} &= \frac{1 + \sum_{i=2}^{I-1}\left(1 - \frac{i-1}{I}\right)}{I} \\ &= \frac{1 + \frac{I^2 - I - 2}{2I}}{I} \\ &= \frac{I^2 + I - 2}{2I^2}.\end{aligned}$$

Clearly,

$$\lim_{I \to \infty} \frac{I^2 + I - 2}{2I^2} = \frac{1}{2}$$

and so

$$\lim_{I \to \infty} \left(I \times \frac{I^2 + I - 2}{2I^2} \right) = \frac{I}{2}$$

provided that I is finite. ■

Author Biography

Dr. Eugene D. Hahn is an Assistant Professor in the Department of Information and Decision Sciences at Salisbury University. He received his Ph.D. in Information and Decision Systems from George Washington University. He also holds a masters from the University of Texas at Austin and a bachelors in the honors program at Boston College. His research interests include Bayesian methods, multi-criteria decision making, and marketing science. He has published in such journals as Organizational Behavior and Human Decision Processes and ASEAN Economic Bulletin, has published book chapters, and has presented at conferences domestically as well as in such places as Thailand, the Canary Islands, and Japan. Dr. Hahn has consulted for organizations such as the U.S. Bureau of the Census, Booz-Allen & Hamilton, and the Corporation for Public Broadcasting.