# Working Paper

## Stochastic Judgments in the AHP: The Measurement of Rank Reversal Probabilities

*Antonie Stam*
*and*
*A. Pedro Duarte Silva*

WP-94-101
September 1994

IIASA
International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 71521 ◻ Telex: 079 137 iiasa a ◻ Telefax: +43 2236 71313

# Stochastic Judgments in the AHP: The Measurement of Rank Reversal Probabilities

*Antonie Stam*
*and*
*A. Pedro Duarte Silva*

WP-94-101
September 1994



IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 71521 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

# STOCHASTIC JUDGMENTS IN THE AHP: THE MEASUREMENT OF RANK REVERSAL PROBABILITIES

Antonie Stam[1,2] and A. Pedro Duarte Silva[3]

September 28, 1994

*1*: Terry College of Business
The University of Georgia
Athens, GA 30602
U.S.A.

*2*: International Institute for Applied Systems Analysis
A-2361 Laxenburg
Austria

*3*: Universidade Católica Portuguesa
Centro Regional do Porto
Curso de Administração e Gestão de Empresas
Rua Diogo Botelho 1327
4100 Porto
Portugal

# FOREWORD

Often, decision makers have to deal with uncertain factors when faced with a decision problem. It is important that during the process of selecting the most suitable alternative, the decision maker is provided with as much information about the nature and consequences of this uncertainty. The analytic hierarchy process is a frequently used decision support tool for selecting the most preferred solution from a discrete set of alternatives. The presence of uncertainty in a decision analysis based on analytic hierarchies is that the pairwise preference ratios are interval judgments, rather than single values. One consequence of having stochastic judgment intervals is the phenomenon of rank reversal, *i.e.*, the possibility that the final ranking of the solutions in terms of their relative preference is incorrect. This paper addresses the implications of having stochastic judgment intervals, and uses multivariate statistical methods to derive point estimates and confidence intervals for the rank reversal probabilities, providing relevant information to both the decision analyst and decison maker about the robustness of the rankings of the alternatives.

# STOCHASTIC JUDGMENTS IN THE AHP: THE MEASUREMENT OF RANK REVERSAL PROBABILITIES


## ABSTRACT

Recently, the issue of rank reversal of alternatives in the Analytic Hierarchy Process (AHP) has captured the attention of a number of researchers. Most of the research on rank reversal has addressed the case where the pairwise comparisons of the alternatives are represented by single values, focusing on mathematical properties inherent to the AHP methodology that can lead to rank reversal if a new alternative is added or an existing one is deleted. A second situation, completely unrelated to the mathematical foundations of the AHP, in which rank reversal can occur is the case where the pairwise judgments are stochastic, rather than single values.

If the relative preference ratings are uncertain, one has judgment intervals, and as a consequence there is a possibility that the rankings resulting from an AHP analysis are reversed, *i.e.*, incorrect. It is important for modeler and decision maker alike to be aware of the likelihood that this situation of rank reversal will occur. In this paper, we introduce methods for assessing the relative preference of the alternatives in terms of their rankings, if the pairwise comparisons of the alternatives are stochastic.

We develop multivariate statistical techniques to obtain point estimates and confidence intervals of the rank reversal probabilities, and show how simulation experiments can be used as an effective and accurate tool for analyzing the *stability* of the preference rankings under uncertainty. This information about the extent to which the ranking of the alternatives is sensitive to the stochastic nature of the pairwise judgments should be valuable information into the decision making process, much like variability and confidence intervals are crucial tools for statistical inference. Although the focus of our analysis is on stochastic preference judgments, our sampling method for estimating rank reversal probabilities can be extended to the case of non-stochastic imprecise fuzzy judgments.

We provide simulation experiments and numerical examples comparing our method with that proposed previously by Saaty and Vargas (1987) for imprecise interval judgments.



**Keywords:** Multicriteria Decision Making, Decision Analysis, Analytic Hierarchy Process, Uncertainty, Judgments.

# STOCHASTIC JUDGMENTS IN THE AHP: THE MEASUREMENT OF RANK REVERSAL PROBABILITIES

## 1. INTRODUCTION

The Analytic Hierarchy Process (AHP), developed by Thomas Saaty (1977, 1980, 1982a), is a widely used multicriteria decision making method which is based on the decomposition of a complex decision problem into several smaller and easier to handle sub-problems. These sub-problems are organized in different levels of a hierarchy. The overall objective of the AHP is to find a relative importance (or preference) scale for the set of discrete alternatives under consideration. Using the "relative comparison" method of the AHP, pairwise comparisons are used to derive the relative attractiveness or preference of the criteria, and the degree of preference of each alternative with respect to each criterion, resulting in a set of weights representing the relative importance of each alternative.

A number of methods can be used to derive a preference ratio scale based on pairwise comparisons, for instance the geometric mean method (Barzilai, Cook and Golany 1987; Crawford and Williams 1985) or ordinary and logarithmic least squares. However, Saaty and Vargas (1987, p. 108) note that regression-based methods fail to allow for inconsistencies between the pairwise judgments (Saaty 1980, 1986), and can lead to misleading results (Saaty and Vargas 1984a, 1984b). Harker (1989) states that, even though logarithmic least squares methods have their advantages, "... the eigenvector method has the interpretation of being a simple averaging process by which the final weights **w** are taken to be the average of all possible ways of comparing alternatives. Thus, the eigenvector is a "natural" method for computing the weights. Furthermore, some theoretical evidence (Saaty 1987; Saaty and Vargas 1984) suggests that this method is the best at uncovering the true rank-order of a set of alternatives." Similarly, Saaty (1980, 1986) and Saaty and Vargas (1987) recommend using the eigenvector procedure of the AHP, as this method expressly allows for the possibility of inconsistencies between the elicited preference judgments. In the AHP, the normalized right principal eigenvector **w** of **A** is used as the vector of weights representing the importance of the alternatives (or criteria). For a detailed justification of this procedure and the mathematical concepts used in the AHP, we refer the reader to Saaty (1977, 1980, 1982a).

Although the AHP methodology has been applied successfully to many complex real-life decision problems (Zahedi 1986; Golden, Wasil and Harker 1989), the theoretical soundness of this method has recently been criticized on a number of grounds (Dyer 1990; Winkler 1990; Trout 1988; Schoner, Wedley and Choo 1992, 1993), notably the use of a ratio scale in the AHP comparisons, rather than the interval scale commonly used in Multiattribute Utility Theory (MAUT), and the rank reversal problem, *i.e.*, the phenomenon that the preference rankings produced by the AHP can change significantly by introducing a new or deleting an existing alternative.

The purpose of our research is *not* to study the appropriateness of the AHP versus competing methodologies such as MAUT, nor is it to address the issue of rank reversal under deterministic pairwise preference judgments, and we will not concern ourselves with this – admittedly interesting – phenomenon. Rather, we will focus on a second, completely unrelated, source of rank reversal, namely rank reversal in the presence of *stochastic* pairwise comparison judgments, and offers a rigorous statistical approach to analyzing outranking and rank reversal properties of the AHP methodology under these conditions. Having noted the potential drawbacks of the AHP methodology and the central role played by the phenomenon of rank reversal in the ongoing discussion of the AHP, we will take the overall AHP methodology as the basis for our analysis below, rather than focusing on the differences in viewpoint between its proponents and critics. However, our method is equally applicable if the priority weights are calculated using some variant of the original AHP. We next review the concept of judgment intervals.

## 2. JUDGMENT INTERVALS

In the original AHP methodology, the decision maker is assumed to be able to provide single values when making the pairwise preference judgments. The process of deriving the scale and corresponding weights within a particular level of an AHP hierarchy with single-valued pairwise judgments can be summarized as follows. Suppose there are $k$ decision alternatives under consideration. In the AHP, a nine point ratio scale, implying a possible range for the pairwise judgment ratios from $1/9$ to $9$, is used to pairwise compare each alternative (or criterion) with the other alternatives (or criteria), resulting in a matrix $\mathbf{A} = \{a_{ij}\}$ of preference ratios. The entry $a_{ij}$ of $\mathbf{A}$ represents the relative preference of alternative $i$ over alternative $j$ $(i < j)$ with respect to the subproblem in question. The matrix $\mathbf{A}$ is assumed to be reciprocal, so that $a_{ji} = 1/a_{ij}$, for $i, j = 1, \ldots, k$. Thus, a total of $k(k-1)/2$ pairwise judgments are needed to fully determine $\mathbf{A}$. Extensions of the AHP methodology exist for the case where not all of the pairwise judgments are available. In the case of a decision problem with multiple criterion levels, a matrix of pairwise judgments is constructed for each sub-problem, after which the overall weights are calculated by synthesizing the information of the sub-problems. Without loss of generality we will restrict our analysis to a single matrix of pairwise judgments $\mathbf{A}$.

Noting that the assumption that the decision maker is capable of providing meaningful single-valued judgments may be an over-simplification of reality, several researchers have recently extended the AHP methodology to allow for imprecise pairwise preference judgments. Some have represented this uncertainty as stochastic, using subjective probabilities (Vargas 1982), others in terms of sensitivity analysis, fuzzy sets, and interval judgments (Saaty and Vargas 1987; Arbel 1989; Boender, De Graan and Lootsma 1989; Zahir 1991; Arbel and Vargas 1992, 1993; Hämäläinen and Lauri 1993;

Salo 1993; Salo and Hämäläinen 1992, 1994; Moreno-Jimenez and Vargas 1993).

In the remainder of our paper, we distinguish between imprecise and stochastic judgments. Both cases imply judgment intervals, but whereas in the case of imprecise or fuzzy judgments the intervals reflect an inability on the part of the decision maker to express his/her relative preferences as a single value, stochastic judgments imply a probability distribution over the range of each judgment interval.

### 2.1. Imprecise Pairwise Comparisons: Judgment Intervals

Most previous attempts to incorporate imprecise judgments in the AHP were based on pairwise comparisons that are restricted to certain finite intervals, deriving intervals of variation for the components of the principal eigenvector from these intervals (Saaty and Vargas 1987: Arbel 1989; Arbel and Vargas 1991; Zahir 1991; Salo and Hämäläinen 1992, 1994). Zahir (1991, p. 207) remarks that "... in many cases absolute measurements have inherent uncertainties due to statistical errors which in turn translate into relative measurements," and that, once uncertainties affect the matrix **A** of preference ratios, there must be uncertainties in the resulting priorities of the decision elements as well. Zahir (1991, pp. 210-212) derives analytically and shows by example how uncertainty about the pairwise judgments can affect the relative rankings in the case of two and three alternatives, and presents a numerical algorithm for computing approximate lower and upper bound for the priority weights in the general case ($k \geq 2$). However, Zahir does not provide a statistical analysis of the rank reversal problem.

Arbel (1989) and Arbel and Vargas (1992, 1994) propose an optimization approach to obtain the intervals spanned by each element of the principal eigenvector, in order to determine dominance structures in the preference rankings of the alternatives. While their approach reduces to a linear program in the case of transitive and consistent judgment intervals, it requires solving a generally non-convex nonlinear programming problem in the presence of inconsistency. In a simulation study that draws on Arbel (1989) and Arbel and Vargas (1991), Moreno-Jimenez and Vargas (1993) note that if some judgments are inconsistent "... the reciprocal constraints are not convex, and, hence, the optimum obtained by traditional methods may not be the global optimum" (p. 80), and conclude that "... the more general optimization problems posed for the inconsistent case are intractable because convexity is violated." As we will focus on the general case of potentially inconsistent preference matrices, the above-mentioned linear programming representations are of limited use for our purposes.

Salo and Hämäläinen (1992) and Hämäläinen and Lauri (1993) use preference programming, a user-interactive approach to modify and fine-tune the initially specified interval judgments to a final combination of intervals for which transitivity and consistency are completely achieved. This approach is consistent with the practice in traditional decision analysis of querying the decision maker so as to minimize inconsistency and ambiguity of the preference judgments. However, while this approach may work in many decision situations, it is conceivable that the decision maker will be unable to fine-tune the judgment intervals to the point that full consistency or a fully specified dominance structure is achieved. Moreover, one of the attractive aspects of the original AHP methodology is that it allows for some (reasonable level of) inconsistency, which appears to be compatible with the way in which humans make decisions (Saaty 1980). Thus, within the AHP philosophy it is reasonable that some inconsistency remains in the final decision matrix at the conclusion of an interactive session aimed at reducing inconsistency between the judgment intervals.

While Salo and Hämäläinen (1992) seek to *reduce* the conflict between the judgment intervals, Saaty and Vargas (1987) take the jugment intervals as *given*, and use a sampling experiment to study the impact of imprecise pairwise judgments on rank reversal. As in most other previous research, the nature of the interval judgments in their approach is non-stochastic, reflecting that the decision maker is unable to select single-valued pairwise preference ratings. Arbel and Vargas (1993) study fuzzy priority derivation by simulation and preference programming.

Although the non-stochastic approach to interval judgments is interesting and reasonable in many decision problems, and provides a flexible analysis that offers valuable information to the decision maker, it has some limitations. For instance, it is difficult to use a non-stochastic approach to derive meaningful measures for interesting properties of the interval matrices, such as probabilities of rank reversal, probabilities of particular rankings, and probabilities that a given alternative will be ranked first, from the resulting sampled intervals of the principal eigenvector components. Even though Saaty and Vargas (1987) attempt to numerically approximate some of these measures through a sampling experiment, their interpretation is somewhat problematic, as they try to estimate probabilistic quantities from non-probabilistic concepts.

### 2.2. Stochastic Judgment Intervals

We believe that many decision situations exist where the nature of the judgment intervals can be considered to be stochastic, justifying a probabilistic approach that uses standard statistical methodologies to study rank reversal likelihoods. The stochastic nature of pairwise judgments can reflect either subjective probabilities that a particular alternative better achieves a given goal, or objective probabilities that reflect uncertain consequences of selecting a particular alternative.

As an example of a decision situation where stochastic judgments may be reasonable, consider the situation where the decision maker has to choose between two different investment opportunities, $I_1$ and $I_2$, that require an identical one-time investment at the beginning of the planning period. Assume that the goal of the decision problem is to maximize net present value over the planning period, and that the interest rate over the planning period is constant but unknown at the time of the investment.

decision. Of course, most realistic investment decision problems are more complex than this simple example, which merely serves as an illustration. Suppose that the break-even point of the net present value of $I_1$ and $I_2$ over the planning period occurs at an interest rate of $r^*$, and that $I_1$ will be preferred if the interest rate exceeds $r^*$, while $I_2$ will be more attractive if $r < r^*$. Thus, the probability that $I_1$ is preferred to $I_2$ equals the likelihood that the interest rate will exceed $r^*$. As the true interest rate is uncertain at the time of the investment decision, a stochastic interval for the relative attractiveness of $A_1$ and $A_2$ appears appropriate.

In our paper we choose to represent the uncertainties in the pairwise comparisons by subjective probability distributions. A rigorous analysis of stochastic judgments requires more than a sensitivity analysis, because the latter ignores important information (*e.g.*, the "confidence" attributed to each scenario) that should be taken into consideration. Although we will treat the interval judgments as stochastic, it may be possible to apply part of our methodology of determining rank reversal probabilities to the case of imprecise judgments, provided of course that the underlying assumptions and definitions of the decision process and the interpretation of the resulting preference ratings are adjusted accordingly. For the sake of a clear focus, we refrain from including such an extension in our paper, and relegate these issues to future research.

The remainder of our paper is organized as follows. In Section 3 we summarize a sampling and estimation method developed by Saaty and Vargas (1987) for the case of imprecise judgments, and discuss some properties of their estimator of rank reversal probability. In Section 4, we introduce two measures of rank reversal probability that are well-suited for the case of stochastic judgments. Section 5 offers computational examples that illustrate our proposed method, and explores several desirable properties associated with our measures of rank reversal probability. The paper concludes in Section 6 with final remarks and potential avenues of future research.

## 3. IMPRECISE JUDGMENTS IN THE AHP AND RANK REVERSAL

### 3.1. Saaty and Vargas' Method

In their 1987 article, Saaty and Vargas propose the following approach for estimating rank reversal probabilities of the alternatives in the case of imprecise preference judgments. Instead of a single judgment value when comparing two alternatives (or criteria), the decision maker is asked to specify a finite interval which covers the relevant range of values for the relative importance of the alternatives. Such interval estimates, called "interval judgments" by Saaty and Vargas (1987, p. 108), are collected for each pairwise comparison.

We will write univariate random variables in upper case italics, and realizations of random variables as well as non-stochastic variables in lower case italics. Matrices will be denoted in upper case and boldface, and vectors of random variates in upper case italics and boldface. Suppose the decision problem under consideration has a total of $k$ alternatives. Denote the pairwise comparison of alternatives $i$ and $j$ ($i, j = 1, \ldots, k$) by $m_{ij}$, and let $\mathbf{M} = \{m_{ij}\}$. If the judgments are imprecise, we will denote $m_{ij}$ in $\mathbf{M}$ by the finite range $[m_{ij}^L, m_{ij}^U]$ of its domain.

In order to calculate estimates for the rank reversal probabilities using the AHP methodology, we need information about the true principal right eigenvector $\boldsymbol{w}^T = (w_1, \ldots, w_k)$ associated with the interval judgments. Within the AHP framework, the component $w_i$ is interpreted as the relative importance weight for alternative $i$. In the presence of interval judgments, the exact nature of $\boldsymbol{w}$ is generally intractable, but approximate information can be gathered through simulation experiments.

As in Zahedi (1984), Saaty and Vargas (1987) sample pairwise judgments from a uniform distribution over the interval specified in $\mathbf{M}$, reflecting the assumption that the decision maker is unable to select a single value from the interval and considers each point within the interval equally. The purpose of their sampling experiment is to derive approximate properties of the likelihood of rank reversals and out-ranking. Saaty and Vargas (1987) show that the range of possible values for $w_i$ ($i = 1, \ldots, k$) is bounded and closed in the set of positive real numbers, since the principal eigenvector is a continuous function of the $m_{ij}$ and the judgment intervals are bounded and closed. Let the range of possible values of $w_i$ associated with $\mathbf{M}$ be defined by $[w_i^L, w_i^U]$. Realizations $a_{ij}$ are generated (simulated) for each entry of $\mathbf{M}$ above the diagonal (*i.e.*, $i < j$), after which the remaining entries are specified such that $a_{ji} = 1/a_{ij}$ for all $i$ and $j$, completing the reciprocal matrix $\mathbf{A} = \{a_{ij}\}$. As in the original AHP analysis, where $\mathbf{M}$ consists of singleton values only, inconsistencies between the pairwise comparisons are allowed. Once $\mathbf{A}$ has been computed, its principal right eigenvector $\boldsymbol{w}$ is calculated. As the sampling experiment introduces stochasticity in the measurement of the principal eigenvector, we denote the random variate representing the $i^{\text{th}}$ component of the vector that is being measured by $W_i$. Replicating the above simulation experiment $n$ times, a sample $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^n$ of principal eigenvectors is obtained. Let us denote the $i^{th}$ component of the $m^{th}$ eigenvector generated by $w_i^m$, and the standard deviation of $w_i^1, \ldots, w_i^n$ by $s_i$. The properties of this sample of eigenvectors are used to estimate the rank reversal probabilities.

It can be shown that if the pairwise comparisons are sampled from a uniform distribution over the judgment interval, the principal eigenvector components are beta distributed and can be approximated by a truncated normal distribution if the number of alternatives is sufficiently large (Saaty and Vargas 1987; Zahedi 1984). For purposes of statistical inference, Saaty and Vargas (1987) use the Kolmogorov-Smirnov test and $\chi^2$ goodness-of-fit tests to verify whether we can assume that the sample $w_i^1, \ldots, w_i^n$ for each individual component $i$ ($i = 1, \ldots, k$) of the principal eigenvector is drawn from a normal distribution. Since Saaty and Vargas analyze the $W_i$ ($i = 1, \ldots, k$) separately, without considering their interdependence, and then multiply pairwise rank reversal probabilities to calculate the overall rank reversal probability, their method implicitly assumes that these components are statistically independent. If the individual normality hypotheses cannot be rejected, a $(1-\alpha)$ level "interval of variation" ($IOV_i^\alpha$) is constructed for each component $W_i$ ($i = 1, ..., n$) of the principal eigenvector. The interval shown in (1) is centered at the maximum likelihood estimate of $W_i$, the sample mean $\overline{W}_i = \sum_r W_i^r/n$, and has a length of $2\ t_{\alpha/2,n-1}\ S_i$, where $S_i$ is the sample standard deviation of $W_i$ and $t_{\alpha/2,n-1}$ is the $\alpha/2^{\text{th}}$ percentile value for the Student $T$ distribution with $n-1$ degrees of freedom. Note that the width of this interval is determined in part by the pre-defined probability level $\alpha$.

$$IOV_i^\alpha = (\overline{W}_i \pm t_{\alpha/2,n-1}\ S_i). \tag{1}$$

The interval $IOV_i^\alpha$ can be interpreted as a central probability statement about the weight $W_i$. Define the intersection of $IOV_i^\alpha$ and $IOV_j^\alpha$ ($i \neq j$) by $IOV_{ij}^\alpha$. Saaty and Vargas determine the estimate $RR_{ij}^I$ of the "probability of rank reversal" $\Pi_{ij}$ associated with each pair of alternatives ($i$ and $j$), in the case of imprecise judgments, by (2):

$$RR_{ij}^I = \begin{cases} 0, & \text{if } IOV_{ij}^\alpha = \emptyset, \\ \mathrm{P}(W_i,\ W_j \in IOV_{ij}^\alpha), & \text{if } IOV_{ij}^\alpha \neq \emptyset;\ IOV_i^\alpha,\ IOV_j^\alpha \neq IOV_{ij}^\alpha, \\ \mathrm{P}(W_j \in IOV_i^\alpha), & \text{if } IOV_i^\alpha \subset IOV_j^\alpha, \\ \mathrm{P}(W_i \in IOV_j^\alpha), & \text{if } IOV_j^\alpha \subset IOV_j^\alpha. \end{cases} \tag{2}$$

The superscript "$I$" in $RR_{ij}^I$ indicates that this rank reversal probability is based on the assumption of imprecise (non-stochastic) pairwise judgments. Equation (2) implies that given intervals of level $(1-\alpha)$, alternatives $i$ and $j$ will never reverse ranks if $IOV_{ij}^\alpha$ is empty, whereas in the case where this interval is not empty, the rank reversal probability equals the likelihood that both $W_i$ and $W_j$ are contained in $IOV_{ij}^\alpha$. Unless there exists a potential for confusion, we will not include the $\alpha$-level in the notation of $RR_{ij}^I$.

Saaty and Vargas (1987, p. 110) note that their measure of rank reversal "... is a measure of the stability of the eigenvector components to changes in $IOV_{ij}^\alpha$. It is not a measure of the true ranks of the alternatives, because the true answer may not be known." Several approximate measures of the probability of rank reversal $\Pi_{ij}$ can be derived. The approximation $RR_{ij}^I$ in (2) selected by Saaty and Vargas defines the phenomenon of rank reversal in terms of the *stability* of the principal eigenvector components as measured by $IOV_{ij}^\alpha$, the intersection of the "intervals of variation" for principal eigenvector components $W_i$ and $W_j$. Vargas and Arbel (1992) show that the measure $RR_{ij}^I$ has a theoretical justification, as it converges to the average of the vertices of the linear program in Arbel

(1989), if the judgment intervals are fully consistent. Of course, as mentioned above, in the presence of inconsistency between the judgment intervals the problem becomes non-convex, with an untractable solution.

In addition to pairwise rank reversal probabilities $RR^I_{ij}$, Saaty and Vargas (1987) also determine expressions for the probability that at least one rank reversal occurs in the eigenvector ($\Pi$) and the likelihood that a given alternative $i$ will reverse rank with some other alternative ($\Pi_i$). Assuming independent events (as we will see below, this assumption is erroneous). Saaty and Vargas' (1987) formulas for these probabilities are given in (3) and (4).

$$\hat{\Pi} = RR^I = 1 - \prod_{1 \leq i < j \leq n} (1 - RR^I_{ij}), \tag{3}$$

$$\hat{\Pi}_i = RR^I_i = 1 - \prod_{j=1}^{n} (1 - RR^I_{ij}). \tag{4}$$

As the probabilities $\Pi$ and $\Pi_i$ are composite measures of the more detailed $\Pi_{ij}$, the rank reversal probability of specific alternatives $i$ and $j$, we will focus only on estimating $\Pi_{ij}$ in this paper.

### 3.2. Properties of $RR^I_{ij}$

In this section we will show how Saaty and Vargas' (1987) method can be improved in the case of non-stochastic imprecise judgements. In Section 4 we outline how $RR^I_{ij}$ can be extended to the case of stochastic judgments.

One issue in Saaty and Vargas' (1987) method is the way in which the "intervals of variation" ($IOV^\alpha_{ij}$) in (1) are constructed. First, the width of $IOV^\alpha_{ij}$, and as a result the estimated value $RR^I_{ij}$, depends on the particular value of $\alpha$ selected. Thus, it is always possible to *increase the estimated rank reversal probabilities* $RR^I_{ij}$ by increasing the level of $\alpha$, and to decrease them by reducing $\alpha$, so that it may be difficult to give a useful interpretation to these probabilities for a given level of $\alpha$.

Second, in the construction and use of the $IOV^\alpha_i$ we can make a more complete use of all sample information relevant to the calculation of rank reversal probabilities. For example, $RR^I_{ij}$ in (2) represents the likelihood that that $W_i$ and $W_j$ are both contained in the intersection of their "intervals of variation" $IOV^\alpha_{ij}$, without taking into consideration where (*i.e.*, how deep) in $IOV^\alpha_{ij}$ the components will be located. However, this information may be relevant for determining the *strength* of the difference in preference between alternatives $i$ and $j$, and thus of the rank reversal likelihood. It is possible, for example, to build a $(1-\alpha) = 99$ percent "interval of variation," suggesting that we are *99* percent certain that $W_i$ is included in the interval, but at the same time have a relatively high probability that a realization $w_i$ of $W_i$ within the interval will be located close to the boundary of the interval.

In addition, for each principal eigenvector component $W_i$ the interval $IOV_i^\alpha$ is computed independently of the other components, after which the rank reversal probabilities $RR_{ij}^I$ are calculated by multiplying the relevant probabilities related to $W_i$ and $W_j$ $(i \neq j)$. Thus, in the construction of the $IOV_i^\alpha$ it is implicitly assumed that the different components of $\boldsymbol{W}$ are mutually independent, and the information contained in the correlation between the components of $\boldsymbol{W}$ is ignored. Saaty and Vargas (1987, p. 108) remark that "...the eigenvector is an $n$-dimensional variable, and statistical measures can be developed for each of its components, but not for the entire vector. Thus, one must derive statistical measures to study rank reversal for single components and then use them to derive one for the entire vector." However, a simultaneous analysis of the entire vector can (and should) be conducted using multivariate statistical techniques. In Section 5 we will use simulation experiments to show that the assumption of independently distributed eigenvector components (weights) is *de facto* false, as all the weights are *simultaneously* derived from a single matrix $\mathbf{A}$ of simulated pairwise judgments. In fact, from the sampling experiments in Section 5 we will see that some of the components of $\boldsymbol{W}$ are strongly correlated.

## 4. STOCHASTIC JUDGMENTS IN THE AHP AND RANK REVERSAL

In this section, we will draw upon the conceptually appealing and interesting method developed by Saaty and Vargas (1987) (taking into account and correcting its problems as described above), extending their measures to the case of stochastic judgments. Specifically, we will introduce estimators of $\Pi_{ij}$, derive their statistical properties and build confidence intervals for $\Pi_{ij}$ to measure the stability of the preference under uncertainty.

In our stochastic approach to characterizing the nature of the judgment intervals, we ask the decision maker for information that can be used to construct a probability distribution over the range of each judgment interval. While Saaty and Vargas (1987) use uniformly distributed random variates to sample values from their non-stochastic judgment intervals, we sample from the assumed probability distributions over the interval of judgments, resulting in a stochastic estimate of the principal eigenvector, which is in turn used to estimate the true probabilities of rank reversal. Therefore, the statistical properties of our estimators of rank reversal probability are based on probabilistic concepts inherent to the nature of the judgment intervals, whereas the statistical analysis in the method of Saaty and Vargas (1987) derives only from the sampling experiment itself. The stochastic nature of the judgments themselves enables us to conduct a more rigorous statistical analysis of the rank reversal likelihoods. Moreover, since the derivation of Saaty and Vargas' "intervals of variation" and hypothesis tests are based directly on their sampling from uniformly distributed random variates, a direct application of $RR_{ij}^S$ in its original form to stochastic judgments may not be appropriate for general types of distributions over the judgment intervals. For general distributions the null hypothesis of approximately normally distributed eigenvector components may be untenable, in which case the Student $T$ distribution cannot be used to determine "intervals of variation."

The probability distributions over the judgment intervals can be assessed in numerous different ways, and any probability distribution can be used to characterize the stochastic nature of the preference judgments. Here, we describe two such ways. If a discrete distribution is appropriate, one may ask the decision maker to select a *degree of confidence* for several different discrete pairwise comparison ratio levels. These confidence levels can then be used to derive a probability distribution over the range of each judgment interval, for instance by normalizing the sum of the confidence levels to unity. Alternatively, if a continuous distribution is appropriate, one may elicit the most optimistic, most pessimistic and most likely values for the preference ratios from the decision maker, after which a (continuous) beta distribution is constructed for the ratio values, assuming that the logarithms of the preference ratings follow a beta distribution (in order to maintain a valid ratio scale). The preference elicitation procedures are not limited to uniformly distributed judgment intervals, whereas the amount of information required from the decision maker to build the probability distributions outlined above is quite modest and should not be difficult to obtain – the preference elicitation is certainly not much more involved than the effort of specifying non-stochastic interval judgments. Our replicated sampling procedure explained below uses the probability distributions over the judgment intervals to obtain a representative sample of principal eigenvectors $\boldsymbol{w}^1, \ldots, \boldsymbol{w}^n$ from their respective sampled pairwise comparison matrices, in a way similar to Saaty and Vargas (1987).

Before deriving our measures of rank reversal probabilities, we need to extend the notation to the stochastic case. Denote the random variate representing the pairwise comparison of alternatives $i$ and $j$ ($i, j = 1, \ldots, k$) by $M_{ij}$, and let $\mathbf{M} = \{M_{ij}\}$. Again, $M_{ij}$ in $\mathbf{M}$ is denoted by the finite range $[m_{ij}^L, m_{ij}^U]$ of its domain. Since the $M_{ij}$ are now stochastic, the principal eigenvector $\boldsymbol{W}$ is random as well, and our sampling experiments are designed to derive probability statements about $\boldsymbol{W}$. In contrast, in the experiments by Saaty and Vargas (1987) the only stochastic aspects derive from the sampling experiment itself.

We adopt the following definition of rank reversal. We will assume that rank reversal between two alternatives $i$ and $j$ occurs, if alternative $i$ *would* be preferred over $j$ under perfect information (*i.e.*, $i \succ j$), but is *calculated* to be less preferred based on the sample information on the interval judgments (*i.e.*, $w_i < w_j$). Let us assume the true probability that the decision maker prefers alternative $i$ over alternative $j$ is given by $\pi_{ij} = \mathrm{P}(i \succ j)$. In addition, let $\pi_{ij}^I = \mathrm{P}(W_i > W_j)$, where as before $W_i$ and $W_j$ are the stochastic weights determined using the eigenvector approach of the AHP. Then, the probability of rank reversal $\Pi_{ij}$ according to our definition is given by (5):

$$\Pi_{ij} = \pi_{ij}\,(1-\pi_{ij}^I) + (1-\pi_{ij})\,\pi_{ij}^I. \tag{5}$$

We next use (5) to develop two point estimates, $RR^S_{ij}$ and $NRR^S_{ij}$, for $\Pi_{ij}$, and use these estimates to construct confidence intervals for $\Pi_{ij}$.

### 4.1 Probability of Rank Reversal Based on Sample Frequencies of Preferences: $RR^S_{ij}$

The first estimate of $\Pi_{ij}$, $RR^S_{ij}$, is based on simple sample frequencies. The superscript "*S*" indicates that this estimate is based on stochastic judgment intervals. Neither $\pi_{ij}$ nor $\pi^I_{ij}$ in (5) is known, but if we can assume that in a given simulation trial the probability that $W_i > W_j$ is approximately equal to the probability that alternative $i$ is preferred to alternative $j$ under complete information, then $\pi^I_{ij} \doteqdot \pi_{ij}$, and both can be estimated by $\hat{\pi}^{I,RR^S}_{ij} = \hat{\pi}^{RR^S}_{ij} = P_{ij}$, the relative sample frequency of the event that $W_i$ exceeds $W_j$. Hence, assuming $\pi^I_{ij} = \pi_{ij}$, (5) implies that $\Pi_{ij}$ can be estimated by $RR^S_{ij}$ in (6):

$$\hat{\Pi}^{RR^S}_{ij} = RR^S_{ij} = 2\, P_{ij}\,(1-P_{ij}). \qquad (6)$$

From (6) we see that $RR^S_{ij}$ ranges from *0*, when one alternative is always preferred to the other, to *0.5*, when each alternative is equally likely to be preferred. Besides its simplicity and intuitive appeal, $RR^S_{ij}$ has the advantage that it is neither based on an *a priori* definition of the distribution of the principal eigenvector components, nor on an assumption that these components are independent.

An exact confidence interval for $\pi^I_{ij}$ is defined by $[p^L_{ij},\ p^U_{ij}]$ (see Cooper and Pearson 1934), where $p^L_{ij}$ and $p^U_{ij}$ are defined by (7) and (8):

$$p^L_{ij} = \frac{p_{ij}}{p_{ij} + (1-p_{ij}+n^{-1})\, F_{\alpha/2,\ 2n(1-p_{ij}+n^{-1});\ 2np_{ij}}}, \qquad (7)$$

$$p^U_{ij} = 1 - \frac{1-p_{ij}}{1-p_{ij} + (p_{ij}+n^{-1})\, F_{\alpha/2,\ 2n(p_{ij}+n^{-1});\ 2n(1-p_{ij})}}, \qquad (8)$$

$n$ is the sample size, and $F_{\alpha/2,\ n_1,\ n_2}$ is the $\frac{\alpha}{2}$ th percentile value of the *F*-distribution with $(n_1,\ n_2)$ degrees of freedom. The interval defined by (7) and (8) provides valuable information about the likely range of $P(W_i > W_j)$. The end-points $p^L_{ij}$ and $p^U_{ij}$ can be used to construct the $(1-\alpha)$ confidence interval for $\Pi_{ij}$ shown in (9):

$$\left.\begin{array}{ll} [2\, p_{ij}^L\, (1-p_{ij}^L)\ ,\ 2\, p_{ij}^U\, (1-p_{ij}^U)], & \text{if } p_{ij}^U \le 0.5, \\ \quad [2\, p_{ij}^L\, (1-p_{ij}^L)\ ,\ 0.5], & \text{if } p_{ij}^L, p_{ij} \le 0.5 \text{ and } p_{ij}^U \ge 0.5, \\ \quad [2\, p_{ij}^U\, (1-p_{ij}^U)\ ,\ 0.5], & \text{if } p_{ij}^U, p_{ij} \ge 0.5 \text{ and } p_{ij}^L \le 0.5, \\ [2\, p_{ij}^U\, (1-p_{ij}^U)\ ,\ 2\, p_{ij}^L\, (1-p_{ij}^L)], & \text{if } p_{ij}^L \ge 0.5. \end{array}\right\} \qquad (9)$$

Since $RR_{ij}^S = 2P_{ij}(1-P_{ij})$ is bounded by $0$ and $0.5$, and the relative magnitude of $p_{ij}^L(1-p_{ij}^L)$ and $p_{ij}^U(1-p_{ij}^U)$ depends on which of $p_{ij}^L$ and $p_{ij}^U$ is closer to $0.5$, we need to distinguish several different cases in (9). First, when $p_{ij}^L \le p_{ij}^U \le 0.5$, $p_{ij}^U$ is closer to $0.5$ than $p_{ij}^L$, so that $2p_{ij}^L(1-p_{ij}^L) < 2p_{ij}^U(1-p_{ij}^U) < 0.5$, whereas the case $0.5 \le p_{ij}^L \le p_{ij}^U$ implies $2p_{ij}^U(1-p_{ij}^U) < 2p_{ij}^L(1-p_{ij}^L) < 0.5$. Note that the confidence interval for $\Pi_{ij}$ is not necessarily symmetric about the point estimate $RR_{ij}^S$. In two out of the four cases in (9), the interval is enlarged to ensure that the confidence interval indeed includes $0.5$, the upper limit of $\Pi_{ij}$ (which occurs when $\pi_{ij}^1 = 0.5$).

An attractive aspect of using $RR_{ij}^S$ to approximate $\Pi_{ij}$ is that it does not require an *a priori* assumption regarding the distribution of $\boldsymbol{W}$. Moreover, $RR_{ij}^S$ implicitly takes the correlation between the components into consideration, since each $W_i$ is compared with the $W_j$ measured in the *same* replication. Its drawback, however, is that $RR_{ij}^S$ only provides an indication of how frequently in the sample alternative $i$ is preferred to alternative $j$, and ignores the ***intensity*** of the preference as reflected by the relative values of $W_i$ and $W_j$.

Next, we develop alternative rank reversal measures which take advantage of the information contained the variance-covariance structure of $W_1, \ldots, W_k$.

### 4.2. Probability of Rank Reversal Based on Magnitude of Preference Differences: $NRR_{ij}^S$

$NRR_{ij}^S$, our second approximation for $\Pi_{ij}$, explicitly takes into account the magnitude of $D_{ij} = W_i - W_j$, the difference between the relative preference weights for alternatives $i$ and $j$. Rather than estimating $\pi_{ij}$ and $\pi_{ij}^1$ from their respective sample proportions, we now derive their expressions from the distribution of $D_{ij}$. Provided that the principal eigenvector $\boldsymbol{W}$ is approximately multivariate normally distributed, $D_{ij}$ is also approximately normally distributed, with mean $\mu_{D_{ij}} = \psi_i - \psi_j$ and variance $\sigma_{D_{ij}}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$, where $\psi_i$ and $\psi_j$ are the (true) mean eigenvector components associated with attributes $i$ and $j$, $\sigma_i^2$ and $\sigma_j^2$ are the (true) variances of the weights $W_i$ and $W_j$, respectively, and $\sigma_{ij}$ is the (true) covariance of $W_i$ and $W_j$. Note that while this approach does require that the principal eigenvector is approximately multivariate normally distributed (so that it is less general than $RR_{ij}^S$ in this respect), its components are not assumed to be independent. In Section 5, our simulation experiments will show that some of the components $W_i$ are in fact strongly correlated, so that the independence assumption implicitly made in the calculation of $RR^I_{ij}$ and $RR^S_{ij}$ may lead to inferior estimation results. This remark holds true for a sampling-based analysis of imprecise judgments as well as stochastic judgments. $NRR^S_{ij}$ is now given by (10):

$$\hat{\Pi}^{NRR^S}_{ij} = NRR^S_{ij} = 2\, \mathrm{P}(D_{ij} < 0)\, \mathrm{P}(D_{ij} > 0). \tag{10}$$

The probabilities $\mathrm{P}(D_{ij} < 0)$ and $\mathrm{P}(D_{ij} > 0)$ in (10) are estimated using the maximum likelihood estimators $\overline{D}_{ij} = (\overline{W}_i - \overline{W}_j)$ and $S_{D_{ij}}$ of $\mu_{D_{ij}}$ and $\sigma_{D_{ij}}$, respectively. For example,

$$Q_{ij} = \hat{\pi}^{I,NRR^S}_{ij} = \mathrm{P}(D_{ij} > 0) = \mathrm{P}(Z > -\mu_{D_{ij}}/\sigma_{D_{ij}}) \doteq \mathrm{P}(Z > -(\overline{W}_i - \overline{W}_j)/S_{D_{ij}}), \tag{11}$$

where $Z$ is standard normally distributed. Thus, while $P_{ij}$ is the relative sample frequency of the event that $W_i$ exceeds $W_j$ for the $RR^S_{ij}$ method, $Q_{ij}$ is its counterpart for the $NRR^S_{ij}$ method. We determine a $(1-\alpha)$ level confidence interval for $\pi^I_{ij}$ by constructing a simultaneous confidence interval for $\mu_{D_{ij}}$ and $\sigma_{D_{ij}}$. Using the Bonferroni method (Neter, Wasserman and Kutner 1985, p. 582), we can derive a conservative simultaneous interval for $\pi^I_{ij}$ from the individual $(1-\alpha/2)$ level confidence intervals for $\mu_{D_{ij}}$ and $\sigma_{D_{ij}}$. These individual intervals are given in (12) and (13), respectively.

$$[\hat{\mu}^L_{D_{ij}}, \hat{\mu}^U_{D_{ij}}] = [\overline{D}_{ij} - z_{1-\alpha/4}\, S_{D_{ij}}/\sqrt{n},\ \overline{D}_{ij} + z_{1-\alpha/4}\, S_{D_{ij}}/\sqrt{n}], \tag{12}$$

$$[\hat{\sigma}^L_{D_{ij}}, \hat{\sigma}^U_{D_{ij}}] = [\sqrt{((n-1)S^2_{D_{ij}}/\chi^2_{1-\alpha/4,n-1})},\ \sqrt{((n-1)S^2_{D_{ij}}/\chi^2_{\alpha/4,n-1})}]. \tag{13}$$

Denoting $\mathrm{P}(Z < z^*)$ by $Z(z^*)$, respectively, and using (11), a $(1-\alpha)$ confidence interval for $\pi^I_{ij}$ is then defined by $[q^L_{ij}, q^U_{ij}]$, where $q^L_{ij}$ is the lowest possible value of $Z(\hat{\mu}_{D_{ij}}/\hat{\sigma}_{D_{ij}})$ for any combination of $\mu_{D_{ij}} \epsilon [\hat{\mu}^L_{D_{ij}}, \hat{\mu}^U_{D_{ij}}]$ and $\sigma_{D_{ij}} \epsilon [\hat{\sigma}^L_{D_{ij}}, \hat{\sigma}^U_{D_{ij}}]$, and similarly $q^U_{ij}$ is the highest value of $Z(\hat{\mu}_{D_{ij}}/\hat{\sigma}_{D_{ij}})$ for any values of $\mu_{D_{ij}}$ and $\sigma_{D_{ij}}$ satisfying (12) and (13). Hence, a conservative $(1-\alpha)$ confidence interval for $\pi^I_{ij}$ is as shown in (14):

$$[q^L_{ij}, q^U_{ij}] = \begin{cases} [Z(\hat{\mu}^L_{D_{ij}}/\hat{\sigma}^U_{D_{ij}}),\ Z(\hat{\mu}^U_{D_{ij}}/\hat{\sigma}^L_{D_{ij}})], & \text{if } \hat{\mu}^L_{D_{ij}} > 0, \\ [Z(\hat{\mu}^L_{D_{ij}}/\hat{\sigma}^L_{D_{ij}}),\ Z(\hat{\mu}^U_{D_{ij}}/\hat{\sigma}^L_{D_{ij}})], & \text{if } \hat{\mu}^L_{D_{ij}} < 0 < \hat{\mu}^U_{D_{ij}}, \\ [Z(\hat{\mu}^L_{D_{ij}}/\hat{\sigma}^L_{D_{ij}}),\ Z(\hat{\mu}^U_{D_{ij}}/\hat{\sigma}^U_{D_{ij}})], & \text{if } \hat{\mu}^U_{D_{ij}} < 0. \end{cases} \tag{14}$$

Similar to the case of $\hat{\Pi}^{RR^S}_{ij} = RR^S_{ij}$, assuming that $\pi^I_{ij} \doteq \pi_{ij}$ we can use (10) to calculate the point estimate $NRR^S_{ij}$ of $\Pi_{ij}$ in (15):

$$\hat{\Pi}_{ij}^{NRR^s} = NRR_{ij}^S = 2\, Q_{ij}\,(1-Q_{ij}). \tag{15}$$

A $(1-\alpha)$ confidence interval for $\Pi_{ij}$ is found in a way similar to (9), but now using $q_{ij}^L$, $Q_{ij}$ and $q_{ij}^U$ instead of $p_{ij}^L$, $P_{ij}$ and $p_{ij}^U$.

The derivation of $NRR_{ij}^S$ assumes a multivariate normal distribution for the principal eigenvectors, and takes the difference between the values of their components – and thus in a sense the strength of the relative preferences – into account. Hence, the tradeoff between $RR_{ij}^S$ and $NRR_{ij}^S$ is that the former is of a simpler nature and can be used regardless of which distribution the eigenvector components follow, whereas the latter utilizes more sample information and may therefore yield more accurate estimates provided that the eigenvector components are approximately normally distributed. However, $NRR_{ij}^S$ may not be appropriate for distributions of eigenvector components which differ substantially from the normal.

In this section, we have developed two estimates for the most detailed measure of rank reversal, $\Pi_{ij}$. The corresponding measures for $\Pi$ and $\Pi_i$ can easily be derived, analogous to Saaty and Vargas (1987). Estimates of $\Pi$ and $\Pi_i$ based on the same principles as $RR_j^S$ can easily be derived, using the frequencies of each ranking of the components of $\boldsymbol{W}$ instead of the frequency of the event that $W_i > W_j$. Estimates of $\Pi$ and $\Pi_j$ based on the same principles as $NRR_j^S$ can be derived using integrals of the multivariate normal distribution. However, their computation is not trivial.

It should be emphasized that the issue of why rank reversal in the AHP weights occurs in the case of single-valued pairwise comparisons is substantially different from that of assessing the rank reversal probabilities when the pairwise judgments *are either imprecise or stochastic*, and introducing uncertainty into the analysis does not alleviate the fundamental problems associated with the phenomenon of rank reversal in the AHP itself. In practice, it may be difficult to discern whether in a given situation rank reversal occurs due to the stochastic nature of the pairwise comparisons or as a result of the underlying mathematics of the AHP procedure itself. By combining sampling procedures for deriving point estimates of the stochastic relative importance weights with the traditional AHP methodology, our method of deriving measures for the "probability of rank reversal" may establish an interesting rationale for the occurrence of certain types of rank reversal in practice.

## 5. NUMERICAL EXAMPLES

In this section, we will use two simulation experiments to exemplify our proposed method and compare the computational results for $RR_{ij}^S$ and $NRR_{ij}^S$ with $RR_{ij}^I$. As the interval judgments in Saaty and Vargas (1987) are non-stochastic, but their sampling experiment yields a statistical analysis of rank reversal probabilities similar to ours, our comparison with their method is limited to the computational aspects only. The interpretation of both methods cannot be compared directly. Suppose that a decision maker decides to use the AHP methodology to compare four alternatives, and arrives at the following matrix of pairwise interval judgments:

$$\mathbf{M} = \begin{bmatrix} & A1 & A2 & A3 & A4 \\ & 1 & [2,4] & [3,5] & [3,5] \\ & & 1 & [1/2,1] & [2,5] \\ & & & 1 & [1/3,1] \\ & & & & 1 \end{bmatrix}. \tag{16}$$

The matrix **M** in (16) is identical to that used in the experimental study reported by Saaty and Vargas (1987) for their analysis of imprecise judgments.

## 5.1 Experimental Design

### 5.1.1. Saaty and Vargas' (1987) Experiment Evaluating $RR^I_{ij}$

First, we describe the simulation experiment conducted by Saaty and Vargas (1987) in order to illustrate the use of $RR^I_{ij}$. Saaty and Vargas simulated a total of $n = 100$ matrices **A** from uniformly distributed variates over the judgment intervals of **M** specified in (16). The respective principal right eigenvector for each **A** was computed using the procedure outlined in Saaty (1980, 1982b), yielding the estimated principal eigenvector $\boldsymbol{w}$ in Equation (17):

$$\boldsymbol{w} = \lim_{k \to \infty} (\mathbf{A}^k\mathbf{e} \,/\, \mathbf{e}^T\mathbf{A}^k\mathbf{e}), \tag{17}$$

where **e** is an appropriately dimensioned unit column vector, and $\mathbf{e}^T$ is the transpose of **e**. In their study, Saaty and Vargas verified approximate univariate normality of each component of $\boldsymbol{W}$ using the Kolmogorov-Smirnov test and the $\chi^2$ goodness-of-fit test, finding that the normality assumption could not be rejected for any $W_i$ at the $\alpha = 5$ percent significance level. Summary statistics of the sample results obtained by Saaty and Vargas (1987) are presented in Table 1.

-----------------------------

Table 1 About Here

-----------------------------

### 5.1.2. Two Experiments to Evaluate $RR^S_{ij}$ and $NRR^S_{ij}$

We perform two separate experiments to evaluate $RR^S_{ij}$ and $NRR^S_{ij}$, and computationally compare the results with $RR^I_{ij}$. In Experiment A, we generate *100* samples from a matrix **M** with uniform stochastic judgment intervals of the same form as (16). Thus, Experiment A facilitates a *direct* computational comparison between $RR^I_{ij}$ and our proposed measures for the case of stochastic judgments. In Experiment B, we again generate *100* samples from judgment intervals, but this time the variates are generated from skewed discrete distributions over three specific pairwise judgment values, with probabilities proportional to the confidence levels. In practice, this information would be provided by the decision maker. The extension to more than three levels of ratio judgments is straightforward. Matrix **C** in (18) gives the pairwise comparison ratios used in Experiment B, followed within parentheses by their associated normalized confidence levels. Note that the normalized confidence levels in **C** cover the same range of values as the matrix **M** used in both Saaty and Vargas' original study and Experiment A.

$$\mathbf{C} = \begin{array}{c} \\ A1 \\ \\ A2 \\ \\ A3 \\ \\ A4 \end{array} \begin{array}{c} \begin{array}{cccc} \quad A1 \quad & \quad A2 \quad & \quad A3 \quad & \quad A4 \quad \end{array} \\ \left[ \begin{array}{cccc} 1(1.00) & \begin{matrix} 2(0.15) \\ 3(0.60) \\ 4(0.25) \end{matrix} & \begin{matrix} 3(0.20) \\ 4(0.50) \\ 5(0.30) \end{matrix} & \begin{matrix} 3(0.20) \\ 4(0.50) \\ 5(0.30) \end{matrix} \\ & 1(1.00) & \begin{matrix} 1/2(0.15) \\ 3/4(0.60) \\ 1(0.25) \end{matrix} & \begin{matrix} 2(0.30) \\ 3(0.40) \\ 5(0.30) \end{matrix} \\ & & 1(1.00) & \begin{matrix} 1/3(0.20) \\ 3/5(0.50) \\ 1(0.30) \end{matrix} \\ & & & 1(1.00) \end{array} \right] \end{array} . \qquad (18)$$

While the range of each judgment interval in $C$ corresponds with that of **M** in (16), we constructed the problem such that those discrete values which are closer to the median of the interval are more likely than the extreme points. It appears that, in practice, distributions over the interval judgments with a mode near the mean or median value may often be more relevant than uniform distributions. Additionally, in contrast with the uniform distribution, which requires a reciprocal or a logarithmic transformation (Moreno-Jimenez and Vargas 1993), it is easy to select discrete probability values that are consistent with the use of a ratio scale. Whether it is better to represent the values within each judgment interval by a discrete or a continuous distribution will depend on the perceptions of the decision maker. We selected a discrete distribution for Experiment B because it offers an interesting comparison with the uniform intervals in Experiment A.

Using the expression in (17), we approximate $W$ in our experiments by $\mathbf{A}^{64}\mathbf{e}/\mathbf{e}^T\mathbf{A}^{64}\mathbf{e}$. Raising $\mathbf{A}$ to the power *64* proved more than sufficient for convergence. In most cases, a power of $k = 5$ was enough for convergence. For both Experiment A and Experiment B we will compute the point estimates $RR^S_{ij}$ and $NRR^S_{ij}$, and construct *99* percent confidence intervals for the probabilities of rank reversal $\Pi_{ij}$. To distinguish between the two experiments, we will denote the point estimates for Experiment A by $RR_{ij}^{A,S}$ and $NRR_{ij}^{A,S}$, and those for Experiment B by $RR_{ij}^{B,S}$ and $NRR_{ij}^{B,S}$. We also calculated $RR_{ij}^{I}$ for Experiments A and B, and denote the resulting rank reversal probability measures by $RR_{ij}^{A,I}$ and $RR_{ij}^{B,I}$. For notational convenience, the superscript indicating to which experiment (A or B) the estimate pertains is replaced by a dot (*i.e.*, $RR_{ij}^{\cdot,S}$, $NRR_{ij}^{\cdot,S}$, $RR_{ij}^{\cdot,I}$) if the discussion applies to both experiments. Whenever possible, we will compare the results from our experiments with the estimates $RR_{ij}^{I}$ obtained in Saaty and Vargas' original study (1987).

### 5.2 Results of Experiments A and B

We first determine whether it is reasonable to assume that the estimates of $W_1$, ..., $W_4$ generated in our experiments are approximately multivariate normally distributed. It is well-known that if $\boldsymbol{W}$ is normally distributed with mean vector $\boldsymbol{\psi}$ and variance-covariance matrix $\Sigma$, then $\boldsymbol{V} = (\boldsymbol{W}-\boldsymbol{\psi})^T \Sigma^{-1} (\boldsymbol{W}-\boldsymbol{\psi})$ follows a $\chi^2$ distribution with four degrees of freedom. Following Malkovich and Affifi (1973), we perform the Kolmogorov-Smirnov test for goodness-of-fit between $\boldsymbol{V}$ and the $\chi_4^2$ distribution, yielding a Kolgomorov-Smirnov statistic ($KS$) equal to *0.085* for Experiment A, with a $p$–value of *0.471*, and $KS = 0.142$, with a $p$–value of *0.035* for Experiment B. Hence, for Experiment A the null hypothesis that $\boldsymbol{W}$ is multivariate normally distributed cannot be rejected at any reasonable $\alpha$–level, so that the rank reversal measures $RR_{ij}^{A,I}$ and $NRR_{ij}^{A,S}$ (which both assume normality of $\boldsymbol{W}$) are indeed applicable. However, the validity of the normality hypothesis for Experiment B is doubtful. Recall that the estimation of $RR_{ij}^{\cdot,S}$ does not require any distributional assumption of $W_i$.

Table 1 gives the mean weight values $\overline{\boldsymbol{w}} = (\overline{w}_1, ..., \overline{w}_4)$, the standard deviations of the estimated weights $s_{w_i}$, and the minimum and maximum weight realizations, $\boldsymbol{w}_{min}$ and $\boldsymbol{w}_{max}$, for the samples of *100* principal eigenvectors generated in each of the three experiments (Experiments A and B, and the original one by Saaty and Vargas (1987). From this table we see that the summary statistics are quite similar across experiments, with slightly larger standard deviations for the weights in Experiment B than those in Experiment A, which in turn tend to be somewhat larger than the standard deviations reported in the original Saaty and Vargas (1987) experiment. Overall, the eigenvector components associated with the three sampling experiments are of statistically comparable nature, on an average basis as well as in terms of the range of the sample values. It is of interest to study how *sensitive* the three measures of rank reversal probability ($RR_{ij}^{\cdot,S}$, $NRR_{ij}^{\cdot,S}$, $RR_{ij}^{\cdot,I}$) are to the *differences in distributions* between the experiments over the judgment intervals. Presumably, an estimator gives more stable and reliable results if the point estimates and confidence intervals are similar across distributions and estimation method. The number of times $w_i$ exceeds $w_j$ ($i < j$) is given in Table 2. These figures are only available for Experiments A and B, as Saaty and Vargas (1987) do not report this information for their experiment.

---

Tables 2 and 3 About Here

---

Table 2 shows that for the vast majority of samples in Experiments A and B, the weight for alternative $i$ exceeds that for alternative $j$ $(i<j)$. For instance, $w_1$ exceeds $w_2$ for all *100* samples in both experiments, and $w_3$ is larger than $w_4$ in *94* out of *100* samples taken in Experiment A and in *69* of the *100* samples in Experiment B. The corresponding computational results of the original Saaty and Vargas (1987) experiment should closely correspond to those of Experiment A, because the sampling distributions were identical for these two experiments.

As mentioned above, Saaty and Vargas (1987) use univariate techniques to analyze the sample eigenvector components, implicitly assuming that the weight values are independently distributed. The sample correlation matrices of the weights given in Table 3 indicate that this assumption is erroneous, and several of the principal eigenvector components are indeed strongly correlated. For instance, in Experiment A weights $W_1$ and $W_2$ have a correlation of $-0.786$. Even though the correlation matrix for their original experiment is not given in Saaty and Vargas (1987), it is clear that the weights in their study must have been correlated as well, because the sampling distributions and experimental designs of their experiment and Experiment A are identical. Note that the correlation between the eigenvector components is due *only* to the sampling procedure used, and in no way depends on whether the original judgment intervals were stochastic or not.

---

Tables 4 and 5 About Here

---

We use the information in Tables 1–3 and 5 to calculate the point estimates and confidence intervals for $\pi^1_{ij}$ for Experiments A ($p_{ij}$) and B ($q_{ij}$) in Table 4. For instance, Table 3 shows that $w_3$ exceeds $w_4$ in *94* out of the *100* samples in Experiment A, so that the corresponding value of $p_{34}$ in Table 4 equals *0.94*. Table 5 presents the sample means $\bar{d}_{ij}$ and standard deviations $s_{D_{ij}}$ of the pairwise differences between the weights, and confidence intervals for the mean differences $\mu_{D_{ij}}$ and the standard deviation of the differences $\sigma_{D_{ij}}$. For instance, from this table we see that $\bar{d}_{24} = 0.0747$ and $s_{d_{24}} = 0.041$ in Experiment B, yielding $q_{24} = Z(\mu_{D_{24}}/\sigma_{D_{24}}) \doteq Z(0.0728/0.041) = Z(1.78) \doteq 0.96$ (see Table 4). The results in Table 4 show that the point estimates $p_{ij}$ and $q_{ij}$ are generally quite close to one, which is consistent with the figures in Table 2 which exhibit a near-dominance of attribute $i$ over $j$ for most $i<j$. An exception is pair $(3, 4)$ in Experiment B, in which case $p_{ij} = 0.69$ and $q_{ij} = 0.64$. Similarly, the *99* percent confidence intervals for $\pi^1_{ij}$ in Table 4 are close to one for both alternative measures. Note that these intervals are not necessarily symmetric about the point estimates.

measures. Note that these intervals are not necessarily symmetric about the point estimates.

In addition to providing information needed for for the calculation of $NRR^S_{ij}$ and the confidence intervals $[q^L_{ij}, q^U_{ij}]$ for $\pi^I_{ij}$, Table 5 also contains interesting statistics on the magnitude of the differences in the relative preferences. For example, this table shows that on average attribute *1* clearly out-ranks the other attributes, with *99.5* percent confidence intervals for $\mu_{D_{1j}}$ ranging from *0.3060* to *0.4063* in Experiment A and from *0.2911* to *0.4051* in Experiment B. The mean sample difference $\overline{d}_{ij}$ between the weights is much smaller for attributes $i, j \in \{2, 3, 4\}$. The corresponding standard deviations for these pairs of attributes, however, are somewhat smaller as well, and none of the *99.5* percent intervals in Table 5 includes $\mu_{D_{ij}} = 0$. Therefore, for each attribute pair the null hypothesis of no difference between the mean weights is rejected, so that it is highly unlikely that the mean difference is negative for any pair.

------------------------------

Table 6 About Here

------------------------------

Table 6 reports *99* percent "intervals of variation" $IOV_i^{\alpha}$ for the attribute weights $W_i$, calculated for Experiments A and B using the Saaty and Vargas (1987) method outlined in Equation (1) of Section 2. The intervals reported by Saaty and Vargas (1987, p. 113, Table 5) for their original experiment are included in Table 6 as well. From Table 6 it is seen that, as was to be expected since the interval distributions and the experimental design of Experiment A and the original experiment conducted by Saaty and Vargas (1987) are identical, the ranges of the $IOV_i^{\alpha}$ are quite similar for these two experiments. Moreover, the "intervals of variation" obtained in Experiment B are remarkably similar to Experiment A and the S&V Experiment, probably due to the fact that the finite ranges of the discrete pairwise judgment distributions of Experiment B are identical to those in the other two experiments. However, as we will see below, the estimated rank reversal probabilities for Experiment B are substantially different.

Of course, the primary purpose for the calculations underlying Tables 4 and 5 is to derive the intermediate statistics necessary for determining the point estimates $RR_{ij}^{\cdot,S} = 2P_{ij}(1-P_{ij})$ and $NRR_{ij}^{\cdot,S} = 2Q_{ij}(1-Q_{ij})$ of the unknown rank reversal probabilities $\Pi_{ij}$, and to construct $(1-\alpha) = 99$ percent confidence intervals for the rank reversal probabilities for Experiments A and B. These values are summarized in Tables 7 and 8. A direct comparison of the confidence intervals for $\Pi_{ij}$ in Table 8 with the Saaty and Vargas (1987) method is not possible, because they do not discuss or calculate these entities in their paper.

---------------------------------------

Tables 7 and 8 About Here

---------------------------------------

Table 7 summarizes the estimated rank reversal probabilities (point estimates) as calculated as our proposed measures ($RR_{ij}^{\cdot,S}$ and $NRR_{ij}^{\cdot,S}$) and the Saaty and Vargas (1987) measure ($RR_{ij}^{\cdot,I}$) for Experiments A and B. In this table, we also report the estimated probabilities obtained by Saaty and Vargas (1987) in their original study ($RR_{ij}^{I}$). Table 7 shows that in both Experiments A and B, each of the three measures ($RR^{\cdot,S}$, $NRR^{\cdot,S}$ and $RR^{\cdot,I}$) yields a zero probability of rank reversal between attribute *1* and any of the other attributes. The same is true for the original experiment by Saaty and Vargas (1987). This finding is consistent with Table 3, where attribute *1* was seen to dominates the other attributes in all *100* samples taken in Experiments A and B, with Table 5, which – as mentioned above – indicates that the confidence interval for the mean pairwise difference between attribute *1* and the other attributes is strictly positive, and with Table 6 which shows no overlap of $IOV_1^{\alpha}$ with any of the other "intervals of variation." As noted in Table 3, Saaty and Vargas do not report the corresponding information for their experiment, but based on the values of $RR_{1j}^{I}$ ($i$, $j = 2, 3, 4$) we suspect that the statistics for their study are similar. For the attribute pairs not involving attribute *1*, the $RR_{ij}^{\cdot,S}$ and $NRR_{ij}^{\cdot,S}$ estimates are similar, but differ considerably from $RR_{ij}^{\cdot,I}$. Typically, the $NNR_{ij}^{\cdot,S}$ estimates are between the corresponding $RR_{ij}^{\cdot,S}$ and $RR_{ij}^{\cdot,I}$ values, albeit much closer to $RR_{ij}^{\cdot,S}$ than to $RR_{ij}^{\cdot,I}$. In several cases, $RR_{ij}^{\cdot,I}$ differs significantly from the other two estimates. For instance, Table 7 indicates that for alternatives (*3*, *4*), $RR_{34}^{A,S} = 0.1128$, $NRR_{34}^{A,S} = 0.1486$, while $RR_{34}^{A,I}$ is much higher at *0.5434*. Interestingly, the estimates ($RR_{ij}^{I}$) reported in the original study by Saaty and Vargas (1987) differ considerably from those obtained in Experiment A ($RR_{ij}^{A,I}$), even though the underlying data conditions were identical for these experiments.

Overall, the results in Table 7 indicate that, not surprisingly, the $R_{ij}^{\cdot,I}$ estimates are less stable than $RR_{ij}^{\cdot,S}$ and $NRR_{ij}^{\cdot,S}$, since the $R_{ij}^{\cdot,I}$ figures varied considerably when the data conditions were only slightly changed, while the $RR_{ij}^{\cdot,S}$ and $NRR_{ij}^{\cdot,S}$ measures appear less sensitive to the particular sample. Again, we stress that the comparison of our stochastic measures with Saaty and Vargas' measure is limited to computational properties only, as the latter was designed for the case of imprecise (non-stochastic) pairwise judgments.

As mentioned above, an advantage of $RR_{ij}^{S}$ and $NRR_{ij}^{S}$ is that both easily facilitate the construction of tight (precise) confidence intervals for the rank reversal probability. Inspecting Table 8, we observe that at the cost of the normality assumption, the confidence intervals derived for the second measure of rank reversal ($NRR_{ij}^{\cdot,S}$) are tighter than those of the first measure ($RR_{ij}^{\cdot,S}$). Therefore, $NRR_{ij}^{S}$ is preferred to $RR_{ij}^{S}$, unless the distribution of the attribute weights is clearly non-normal.

Summarizing, it appears that the measures developed in this paper ($RR_{ij}^S$ and $NRR_{ij}^S$) yield attractive and robust estimates of the true rank reversal probabilities $\Pi_{ij}$, in the case of stochastic judgment intervals.

## 6. CONCLUDING REMARKS

In this paper, we propose two measures of rank reversal probabilities in the Analytic Hierarchy Process resulting from pairwise judgments which are stochastic in nature. These measures are based, in part, on a previously proposed measure of rank reversal probability by Saaty and Vargas (1987) in the case of imprecise judgments. We introduce straightforward yet statistically rigorous procedures for deriving both theoretically sound point estimates and tight confidence intervals for the rank reversal probabilities. One of the measures ($RR_{ij}^S$) is based on relative sample frequencies and does not require any assumption on the distribution of the attribute weights, while the other ($NRR_{ij}^S$) is based on the assumption of multivariate normality of the attribute weights.

Using two simulation experiments, we have shown that our proposed measures provide robust estimates of rank reversal probability. Specifically, as long as the normality assumption for $\boldsymbol{W}$ is reasonable, we recommend using the second measure ($NRR_{ij}^S$), which is based on a multivariate analysis, and takes advantage of the variance-covariance structure of the attribute weights and of the strength of the information of the difference in preference between the alternatives to achieve more accurate estimation results. If the normality assumption for the weights is rejected, $RR_{ij}^S$ is the method of choice. We have also shown that our approach is flexible in that it is possible to elicit the preference information in the form of discrete confidence levels associated with several values of the pairwise judgments (ratios).

Arbel and Vargas (1993) remark that exploring the sensitivity of rank order to the range of preferences and providing flexible ways for dealing with articulation of non-transitive preference structures are areas of potential avenues of future research. Their remark was made in the context of imprecise judgments. An extension of our current methodology to non-stochastic judgment intervals may prove promising, for instance modifying our rank reversal measures in order to analyze rank reversal probabilities for fuzzy (non-stochastic) judgment intervals where not all points in the interval are considered equally by the decision maker.

**TABLE 1: Descriptive Statistics for the Simulated Principal Eigenvector Components (Weights), Experiments A and B, and the Original Saaty and Vargas (1987) (S&V) Experiment**

| | Experiment A | | | | Experiment B | | | | S&V Experiment[a] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wgt | Mean | STD | Min | Max | Mean | STD | Min | Max | Mean | STD | Min | Max |
| $w_i$ | $\overline{w}_i$ | $s_{w_i}$ | $s_{min}$ | $s_{max}$ | $\overline{w}_i$ | $s_{w_i}$ | $s_{min}$ | $s_{max}$ | $\overline{w}_i$ | $s_{w_i}$ | $s_{min}$ | $s_{max}$ |
| $w_1$ | .5211 | .0305 | .4386 | .5750 | .5215 | .0378 | .4077 | .5895 | .5093 | .0273 | :4374 | .5696 |
| $w_2$ | .2005 | .0211 | .1567 | .2482 | .2041 | .0303 | .1568 | .2815 | .2131 | .0219 | .1654 | .2708 |
| $w_3$ | .1526 | .0138 | .1228 | .1848 | .1431 | .0160 | .1107 | .1901 | .1496 | .0175 | .1111 | .1971 |
| $w_4$ | .1258 | .0129 | .1007 | .1586 | .1313 | .0237 | .0890 | .1919 | .1280 | .0151 | .1011 | .1633 |

a: These values were reported in Table 2 of Saaty and Vargas (1987, p. 113).

**TABLE 2: Number of Times $w_i$ Exceeded $w_j$ Among the *100* Eigenvectors Simulated, Experiments A and B, and the Original Saaty and Vargas (1987) (S&V) Experiment**

| | | Experiment A | | | | Experiment B | | | | S&V Experiment |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *j* | | | | *j* | | | | |
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | Not Reported |
| *i* | 1 | – | 100 | 100 | 100 | – | 100 | 100 | 100 | |
| | 2 | | – | 99 | 100 | | – | 98 | 98 | |
| | 3 | | | – | 94 | | | – | 69 | |
| | 4 | | | | – | | | | – | |

**TABLE 3: Sample Correlation Matrix for the Principal Eigenvector Components (Weights), Experiments A and B, and the Original Saaty and Vargas (1987) (S&V) Experiment**

| | Experiment A | | | | | Experiment B | | | | S&V Experiment |
|---|---|---|---|---|---|---|---|---|---|---|
| | $W_1$ | $W_2$ | $W_3$ | $W_4$ | | $W_1$ | $W_2$ | $W_3$ | $W_4$ | |
| $W_1$ | 1 | −.786 | −.565 | −.472 | $W_1$ | 1 | −.752 | −.336 | −.400 | Not Reported |
| $W_2$ | | 1 | .149 | .059 | $W_2$ | | 1 | .088 | −.140 | |
| $W_3$ | | | 1 | .023 | $W_3$ | | | 1 | −.268 | |
| $W_4$ | | | | 1 | $W_4$ | | | | 1 | |

**TABLE 4: Point Estimates and *99* Percent Confidence Intervals for $\pi^{1}_{ij} = \pi_{ij}$, Experiments A and B**

**First Proposed Measure ($RR^{\cdot,S}_{ij}$)**

| | Experiment A | | Experiment B | |
|---|---|---|---|---|
| Pair $(i,j)$ | $p_{ij}$ | $[p^L_{ij},p^U_{ij}]$ | $p_{ij}$ | $[p^L_{ij},p^U_{ij}]$ |
| *(1,2)* | 1.00 | [0.948,1.0] | 1.00 | [0.948,1.0] |
| *(1,3)* | 1.00 | [0.948,1.0] | 1.00 | [0.948,1.0] |
| *(1,4)* | 1.00 | [0.948,1.0] | 1.00 | [0.948,1.0] |
| *(2,3)* | 0.99 | [0.928,1.0] | 0.98 | [0.911,0.999] |
| *(2,4)* | 1.00 | [0.948,1.0] | 0.98 | [0.911,0.999] |
| *(3,4)* | 0.94 | [0.851,0.984] | 0.69 | [0.559,0.803] |

**Second Proposed Measure ($NRR^{\cdot,S}_{ij}$)**

| | Experiment A | | Experiment B | |
|---|---|---|---|---|
| Pair $(i,j)$ | $q_{ij}$ | $[q^L_{ij},q^U_{ij}]$ | $q_{ij}$ | $[q^L_{ij},q^U_{ij}]$ |
| *(1,2)* | 1.00 | [1.0,1.0] | 1.00 | [1.0,1.0] |
| *(1,3)* | 1.00 | [1.0,1.0] | 1.00 | [1.0,1.0] |
| *(1,4)* | 1.00 | [1.0,1.0] | 1.00 | [1.0,1.0] |
| *(2,3)* | 0.97 | [0.921,0.977] | 0.97 | [0.894,0.995] |
| *(2,4)* | 1.00 | [0.988,1.0] | 0.96 | [0.903,0.996] |
| *(3,4)* | 0.94 | [0.821,0.981] | 0.64 | [0.524,0.782] |

**TABLE 5: Sample Means ($\bar{d}_{ij}$), Standard Deviations ($s_{D_{ij}}$) and *99.5* Percent Confidence Intervals (CI) for the Weight Differences and their Standard Deviations, Second Proposed Measure (*i.e.*, using $NRR^S_{ij}$), Experiments A and B**

**Experiment A**

| Pair $(i,j)$ | $\bar{d}_{ij}$ | $s_{D_{ij}}$ | CI for $\mu_{D_{ij}}$ | CI for $\sigma_{D_{ij}}$ |
|---|---|---|---|---|
| *(1,2)* | 0.3206 | 0.0489 | [0.3060,0.3346] | [0.0407,0.0608] |
| *(1,3)* | 0.3685 | 0.0400 | [0.3570,0.3800] | [0.0333,0.0493] |
| *(1,4)* | 0.3953 | 0.0383 | [0.3843,0.4063] | [0.0319,0.0476] |
| *(2,3)* | 0.0479 | 0.0235 | [0.0412,0.0546] | [0.0195,0.0292] |
| *(2,4)* | 0.0747 | 0.0241 | [0.0678,0.0816] | [0.0200,0.0300] |
| *(3,4)* | 0.0268 | 0.0186 | [0.0214,0.0322] | [0.0156,0.0232] |

**Experiment B**

| Pair $(i,j)$ | $\bar{d}_{ij}$ | $s_{D_{ij}}$ | CI for $\mu_{D_{ij}}$ | CI for $\sigma_{D_{ij}}$ |
|---|---|---|---|---|
| *(1,2)* | 0.3174 | 0.0638 | [0.2911,0.3357] | [0.0531,0.0793] |
| *(1,3)* | 0.3784 | 0.0460 | [0.3652,0.3916] | [0.0383,0.0572] |
| *(1,4)* | 0.3902 | 0.0520 | [0.3753,0.4051] | [0.0433,0.0646] |
| *(2,3)* | 0.0610 | 0.0331 | [0.0515,0.0705] | [0.0275,0.0412] |
| *(2,4)* | 0.0728 | 0.0410 | [0.0664,0.0900] | [0.0341,0.0510] |
| *(3,4)* | 0.0118 | 0.0322 | [0.0026,0.0210] | [0.0268,0.0400] |

**TABLE 6**
**Applying the Saaty and Vargas (1987) Measures to Experiments A and B, and to the Original Saaty and Vargas (1987) (S&V) Experiment: 99 Percent "Intervals of Variation" for $W_i$**

| | Experiment A | Experiment B | S&V Experiment |
|---|---|---|---|
| $W_i$ | $IOV_i^\alpha$ for $W_i$ | $IOV_i^\alpha$ for $W_i$ | $IOV_i^\alpha$ for $W_i$ |
| $W_1$ | [0.4401,0.6021] | [0.4212,0.6218] | [0.4388,0.5798] |
| $W_2$ | [0.1443,0.2567] | [0.1238,0.2845] | [0.1567,0.2695] |
| $W_3$ | [0.1160,0.1892] | [0.0996,0.1865] | [0.1043,0.1949] |
| $W_4$ | [0.0917,0.1599] | [0.0683,0.1943] | [0.0890,0.1669] |

**TABLE 7: Summary of Estimated Rank Reversal Probabilities, for the Saaty and Vargas (1987) Probability Measure and our Proposed Measures, Experiments A and B, and the Original Saaty and Vargas (1987) (S&V) Experiment**

| Pair $(i,j)$ | Experiment A | | | Experiment B | | | S&V Experiment[a] |
|---|---|---|---|---|---|---|---|
| | $RR_{ij}^{A,S}$ | $NRR_{ij}^{A,S}$ | $RR_{ij}^{A,I}$ | $RR_{ij}^{B,S}$ | $NRR_{ij}^{B,S}$ | $RR_{ij}^{B,I}$ | $RR_{ij}^{I}$ |
| (1,2) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| (1,3) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| (1,4) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| (2,3) | 0.0198 | 0.0722 | 0.2156 | 0.0396 | 0.0446 | 0.2469 | 0.0065 |
| (2,4) | 0.0 | 0.0032 | 0.0021 | 0.0396 | 0.0406 | 0.2325 | 0.0003 |
| (3,4) | 0.1128 | 0.1486 | 0.5434 | 0.4278 | 0.4240 | 0.9087 | 0.7791 |

*a*: These values are reported by Saaty and Vargas (1987), p. 114.

**TABLE 8**

**99 Percent Confidence Intervals for Probabilities of Rank Reversal,[a] Experiments A and B**

| | Experiment A | | Experiment B | | |
|---|---|---|---|---|---|
| Pair $(i,j)$ | $RR_{ij}^{A,S}$ | $NRR_{ij}^{A,S}$ | $RR_{ij}^{B,S}$ | $NRR_{ij}^{B,S}$ | $RR_{ij}^{I}$ |
| (1,2) | [0,0.0986] | [0,0] | [0,0.0986] | [0,0] | – |
| (1,3) | [0,0.0986] | [0,0] | [0,0.0986] | [0,0] | – |
| (1,4) | [0,0.0986] | [0,0] | [0,0.0986] | [0,0] | – |
| (2,3) | [0,0.1336] | [0.0060,0.1455] | [0.0020,0.1622] | [0.0100,0.1895] | – |
| (2,4) | [0,0.0986] | [0,0.0237] | [0.0020,0.1622] | [0.0080,0.1751] | – |
| (3,4) | [0.0315,0.2536] | [0.0373,0.2939] | [0.3164,0.4930] | [0.3410,0.4988] | – |

*a*: The corresponding figures have not been reported in the Saaty and Vargas (1987) study.

# REFERENCES

Arbel, A., "Approximate Articulation of Preference and Priority Derivation," *European Journal of Operational Research*, **43** (1989), 317–326.

Arbel, A. and L. G. Vargas, "The Analytic Hierarchy Process with Interval Judgments," in *Multiple Criteria Decision Making: Proceedings of the IXth International Conference: Theory and Applications in Business, Industry and Government*, A. Goicoechea, L. Duckstein and S. Zionts (Eds.), Springer-Verlag, New York, NY, 1992, 61–70.

Arbel, A. and L. G. Vargas, "Preference Simulation and Preference Programming: Robustness Issues in Priority Derivation," *European Journal of Operational Research*, **69** (1993), 200–209.

Barzilai, J., W. D. Cook and B. Golany, "Axiomatic Foundation of the Analytic Hierarchy Process," *Operations Research Letters*, **6**, 3 (1987), 131–134.

Boender, C. G. E., J. G. de Graan and F. A. Lootsma, "Multi-Criteria Decision Analysis with Fuzzy Pairwise Comparisons," *Fuzzy Sets and Systems*, **29** (1989), 133–143.

Cooper, C. J. and E. S. Pearson, "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, **26** (1934), 406–413.

Crawford, G. and C. Williams, "A Note on the Analysis of Subjective Judgement Matrices," *Journal of Mathematical Psychology*, **29** (1985), 387–405.

Dyer, J. S., "Remarks on the Analytic Hierarchy Process," *Management Science*, **36**, 3 (1990), 249–258.

Forman, E. H. and T. L. Saaty, *Expert Choice*, Version 7.0, Expert Choice, Inc., Pittsburgh, PA, 1990.

Golden, B. L., E. A. Wasil and D. E. Levy, "Applications of the Analytic Hierarchy Process: A Categorized, Annotated Bibliography," in *The Analytic Hierarchy Process: Applications and Studies*, B. L. Golden, E. A. Wasil and P. T. Harker (Eds.), Springer-Verlag, New York, NY, 1989, pp. 37–58.

Golden, B. L., E. A. Wasil and P. T. Harker (Eds.), *The Analytic Hierarchy Process: Applications and Studies*, Springer-Verlag, New York, NY, 1989.

Hämäläinen, R. P. and H. Lauri, *HIPRE $3^+$ Decision Support Software, Version 3.3*, Helsinki University of Technology, Systems Analysis Laboratory, Espoo, Finland, 1993.

Harker, P. T. "The Art and Science of Decision Making: The Analytic Hierarchy Process," in *The Analytic Hierarchy Process: Applications and Studies*, B. L. Golden, E. A. Wasil and P. T. Harker (Eds.), Springer-Verlag, New York, NY, 1989, pp. 3–36.

Harker, P. T. and L. G. Vargas, "Reply to 'Remarks on the Analytic Hierarchy Process' by J.S. Dyer," *Management Science*, **36**, 3 (1990), 269–275.

Malkovich, J. F. and A. A. Affifi, "On Tests for Multivariate Normality," *Journal of the American Statistical Association*, **69** (1973), 176–179.

Moreno-Jimenez, J. M. and L. G. Vargas, "A Probabilistic Study of the Preference Structures in the Analytic Hierarchy Process with Interval Judgments," *Mathematical and Computer Modelling*, **17** (1993), 73–81.

Morrison, D. F., *Multivariate Statistical Methods* (3rd Ed.), McGraw-Hill, New York, NY. 1990.

Neter, J., W. Wasserman and M. H. Kutner, *Applied Linear Statistical Models*, Second Edition, Irwin, New York, NY 1985.

Saaty, T. L., "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, **15** (1977), 37–57.

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

Saaty, T. L., *Decision Making for Leaders*, Wadsworth, Belmont, CA, 1982a.

Saaty, T. L., *The Logic of Priorities: Applications in Business, Energy, Health and Transportation*, Kluwer-Nijhoff, Boston, MA, 1982b.

Saaty, T. L., "Rank Generation, Preservation, and Reversal in the Analytic Hierarchy Decision Processes," *Decision Sciences*, **18** (1987), 157–177.

Saaty, T. L., "An Exposition of the AHP in Reply to the Paper 'Remarks on the Analytic Hierarchy Process,'" *Management Science*, **36**, 3 (1990), 259–268.

Saaty, T. L. and L. G. Vargas, "Inconsistency and Rank Preservation," *Journal of Mathematical Psychology*, **28** (1984), 205–214.

Saaty, T. L. and L. G. Vargas, "Uncertainty and Rank Order in the Analytic Hierarchy Process," *European Journal of Operational Research*, **32** (1987), 107–117.

Salo, A. A., "Inconsistency Analysis by Approximately Specified Priorities," *Mathematical and Computer Modelling*, **17** (1993), 123–133.

Salo, A. A. and R. P. Hämäläinen, "Processing Interval Judgments in the Analytic Hierarchy Process," in *Multiple Criteria Decision Making: Proceedings of the IXth International Conference: Theory and Applications in Business, Industry and Government*, A. Goicoechea, L. Duckstein and S. Zionts (Eds.), Springer-Verlag, New York, NY, 1992, 359–372.

Salo, A. A. and R. P. Hämäläinen, "Preference Assessment by Imprecise Ration Statements," *Operations Research*, **40** (1992), 1053–1061.

Salo, A. A. and R. P. Hämäläinen, "Preference Programming Through Approximate Ratio Comparisons," *European Journal of Operational Research*, Forthcoming, 1994.

Schoner, B., W. C. Wedley and E. U. Choo, "A Rejoinder to Forman on AHP, With Emphasis on the Requirements of Composite Ratio Scales," *Decision Sciences*, **23**, 1992, 509–517.

Schoner, B., W. C. Wedley and E. U. Choo, "A Unified Approach to AHP with Linking Pins," *European Journal of Operational Research*, **64**, 1993, 384–392.

Vargas, L. G., "Reciprocal Matrices with Random Coefficients," *Mathematical Modelling*, **3** (1982), 69–81.

Winkler, R. L., "Decision Modeling and Rational Choice: AHP and Utility Theory," *Management Science*, **36**, 3 (1990), 247–248.

Zahedi, F., "Estimation Techniques and Their Statistical Properties in the Analytic Hierarchy Process," Working Paper, University of Massachusetts, Boston, MA, December 1984.

Zahir, M. S., "Incorporating the Uncertainty of Decision Judgements in the Analytic Hierarchy Process," *European Journal of Operational Research*, **53** (1991), 206–216.