

# Link Function Selection in Stochastic Multicriteria Decision Making Models

Eugene D. Hahn  
Department of Information & Decision Sciences  
Salisbury University  
Salisbury, MD 21804  
E-mail: [edhahn@salisbury.edu](mailto:edhahn@salisbury.edu)

August 18, 2004

## Abstract

A stochastic formulation of the Analytic Hierarchy Process (AHP) using an approach based on Bayesian categorical data models has been developed. However, in categorical data models it is known that the selection of the link function may have an impact on the model estimates. In particular, the selection of the probit link implies an assumption that model error terms are normally distributed and this normality assumption is regularly utilized in other related methods such as the multiplicative AHP. We examine model performance with respect to the choice of two model link functions. With regard to point estimates, it is found that the logit formulation is better able to replicate the estimates obtained by the eigenvector decomposition associated with the original formulation of the AHP. By contrast, the probit link produces priorities which are consistently more moderate than those of the AHP. The results suggest that the logit formulation will be preferred by decision makers who wish to replicate the AHP priorities as closely as possible. The results also suggest that the unexamined use of the normality assumption in other stochastic AHP methods may have an impact on priority estimates and thus is worthy of further attention.

**Keywords** Analytic hierarchy process; Bayesian inference; Multivariate statistics; Simulation.

## 1 Introduction

The Analytic Hierarchy Process (AHP; Saaty, 1977) is a methodology useful for multicriteria decision making which has received substantial application (Vaidya and Kumar, 2004). In the AHP, a decision maker provides his or her relative preferences for the various alternatives by way of a series of pairwise comparisons. From these judgments, a matrix of pairwise comparisons is formed. Then, the priorities (or relative underlying weights) of the alternatives are obtained via a deterministic method, the eigenvalue decomposition.

As a simple motivating example, suppose a decision maker is to receive a reward, and she may choose to receive either 4 ounces, 2 ounces, or 1 ounce of chocolate as her reward. Assume that the decision maker's value or utility for the alternatives is proportional to their amounts. Let  $w_1$ ,  $w_2$ , and  $w_3$  denote the underlying weights or utilities of the alternatives. Also, let the pairwise comparison  $C_{ij}$  equal  $w_i/w_j$ . Then, assuming that the decision maker's judgments are without error, the matrix of pairwise comparisons is

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{2/1} & \mathbf{4/1} \\ 1/2 & 1 & \mathbf{2/1} \\ 1/4 & 1/2 & 1 \end{pmatrix}.$$

The quantities in boldface in the matrix are the pairwise comparisons of the decision maker. For example, four ounces is four times as preferable as one ounce, and so the top right entry where Alternatives 1 and 3 are compared is 4/1. The diagonal consists of ones as for any  $w_i \neq 0$ ,  $w_i/w_i = 1$ . The entries below the diagonal are the reciprocals of those above it. Returning to the problem at hand, the priorities can be obtained from the normalized principal eigenvector of the pairwise comparison matrix and here the priorities are 4/7, 2/7, and 1/7 respectively. Thus, the decision maker should take the 4-ounce reward as it has the highest priority.

The eigenvalue decomposition is a non-stochastic mathematical approach to deriving priorities. In adopting it, error is assumed to be non-existent or negligible. This assumption likely holds in very simple situations such as in the previous example. In many complex real world decision problems, however, the true underlying weights of the alternatives are far more difficult for a decision maker to assess. As such, we would expect the pairwise comparisons to be subject to random error. Given this state of affairs, we might be interested to know whether we can be reasonably confident that two priorities are truly different from one another, or alternatively whether any differences might be attributable to random error. We can see that a statistical approach could be helpful in answering our question. The purpose of such a stochastic multicriteria decision making method is clear: to allow one to make inferential statements on the elements of the priority vector.

As such, the development of probabilistic or statistical formulations of the AHP has attracted considerable interest (e.g., Arbel, 1989; Basak, 1991, 1998; Crawford and Williams, 1985; de Jong, 1984; Genest and Rivest, 1994; Haines, 1998; Jensen, 1984; Lipovetsky and Tishler, 1999; Ramanathan, 1997; Saaty and Vargas, 1987; Stam and Duarte Silva, 1997). Some authors have obtained stochastic judgments by eliciting intervals for pairwise comparisons as opposed to point values (e.g., Bryson and Joseph, 2000; Sugihara, Ishii, and Tanaka, 2004). While useful, these methods require additional judgments from the decision maker and thereby lack the efficiency of the original formulation of the AHP. Others have obtained a probabilistic formulation by invoking a frequentist definition of probability and thereby requiring some form of repeated sampling to occur. These methods are relevant when there exists a group of decision makers so that sampling variability can be assessed across multiple individuals. For example, Basak (2001) introduced a related method based on minimum chi-squared estimation for large group decision-making. However, when the decision maker is an individual or when the group of decision makers is small, these methods are not applicable. The remainder are not able to address the fact that there is a high degree of dependence

among relatively consistent pairwise comparisons. It is clear from the definition of cardinal consistency,  $C_{ik} = C_{ij} \times C_{jk}$ , that any one judgment is dependent upon the values of the other two. Thus, approaches based on an assumption of independence among judgments may be somewhat suspect, particularly in the context of inference.

In contrast to the above approaches, recently a stochastic multicriteria decision making method has been developed that permits inference on the elements of the priority vector without requiring interval judgments or multiple decision makers (Hahn, 2003). Moreover, the approach easily allows for dependence among pairwise comparison judgments. This stochastic judgment method utilizes the pairwise comparison matrices of the original AHP and thus problems that have been examined with AHP may be examined without modification using this method (e.g., Phillips-Wren, Hahn, and Forgionne, 2004). The value-add of such a two-phase examination is that the latter method allows individuals to make inferential statements regarding the priorities. This can be a powerful tool for decision makers. The approach assumes that the underlying weights have a multinomial sampling distribution such that the priorities may be estimated via multinomial logit models (McFadden, 1973). Inference on the priority vector is obtained by the adoption of a Bayesian approach and the use of Markov chain Monte Carlo methods.

## 2 Model Formulation

A brief review of the methodology is as follows. The pairwise comparison is stochastically conceptualized in terms of its nature as counts of relative preference outcomes. This conceptualization naturally leads us to the binomial and multinomial distributions. We can explain this with a simple example. Consider the case of a coin which has been flipped 6 times and has come up heads 5 times out of the total. We may wish to examine the propensity of this coin to exhibit the outcome of heads. A natural way of doing this is with the binomial distribution. Using the binomial distribution, we may estimate  $p$ , the propensity of the coin to exhibit a certain outcome. Here, the binomial likelihood

$$\Pr(y|p, n) = \binom{n}{y} p^y (1-p)^{n-y}$$

leads us to consider estimators of  $p$ . One estimator of  $p$  is the familiar estimator,  $y/n$ . Thus, one might use this estimator and say our best estimate of the propensity for heads given the observed pattern of results is 5/6. By extension, we may use this approach to examine an expert's propensity or preference for one of two outcomes. For example, the pairwise comparison 5/1 indicates that on a relative scale the expert gives five outcomes of preference for Alternative  $A$  for each one outcome for Alternative  $B$ . In the simple case of 2 outcomes, again this process can be described by the binomial distribution. Specifically, the weights  $w_i$  and  $w_j$  which are derivable from  $C_{ij}$  can be shown to be subject to a binomial distribution with parameter  $p$ . Here, our best estimate of the propensity for  $A$  would again be 5/6. In the more general case of  $k$  outcomes where  $k > 2$ , the multinomial distribution

$$\Pr(y_1, \dots, y_k | p_1, \dots, p_k, n) = \frac{n!}{y_1! y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k} \quad (1)$$

is the relevant distribution and  $p$  is replaced by the vector  $\mathbf{p} = (p_1, \dots, p_k)$ . In (1) note that  $n = \sum_{i=1}^k y_i$  and  $\sum_{i=1}^k p_i = 1$ .

In the method previously described, the judgment matrix may be transformed into a multinomial data matrix having  $I$  rows and  $J$  columns of data<sup>1</sup>. Here, the  $j^{\text{th}}$  column contains the weights associated with the  $j^{\text{th}}$  alternative. The  $i^{\text{th}}$  row can be thought of as the  $i^{\text{th}}$  trial. However, the multinomial logit model is not the only model that can be used for categorical data analysis. Note that in the case of binary response data, probit and logit models are commonly encountered categorical data models. We review these models briefly as they provide a useful reference point for the development later in the paper.

In binary response categorical data models, we define the linear predictor  $\eta$  as  $\eta = \mathbf{x}\beta$  where  $\beta$  is the regression coefficient vector and  $\mathbf{x}$  is the matrix of the predictor variables. Then, the probit model can be written as  $\Phi^{-1}(p) = \eta$ , which is to say that  $\eta$  is mapped to the probability of a response,  $p$ , by way of the inverse standard normal cumulative distribution function,  $\Phi^{-1}(\cdot)$ . In this situation,  $\Phi^{-1}(\cdot)$  is the link function, as it maps the linear predictor onto the probability. Hence, we may think of a link function as a mapping function between  $\eta$  and  $p$ . Similarly, the logit model is written  $\ln(p/(1-p)) = \eta$ . Here, the link function is  $\ln(p/(1-p))$ , which is the inverse of the logistic cumulative distribution function.

In the current context, it is useful to mention how the probit and the logit models can be described in terms of a latent variable formulation (for a more detailed discussion, see sections 3.3 and 7.5 of Powers and Xie, 2000). Imagine a simple choice situation between alternatives  $A$  and  $B$ . We can assume that if the total utility for  $A$  exceeds that of  $B$ , then  $A$  will be selected. Otherwise,  $B$  will be selected. Thus, in terms of our introductory example we can write a difference in choice utilities,  $v$ , as  $v = w_A - w_B$ . Of course, we do not observe this latent quantity directly, but rather we observe the preference-related outcomes. Nonetheless, it seems sensible to assume this latent quantity or propensity exists. If we could somehow observe this quantity, we could easily write a linear model for it as

$$v = \mathbf{x}\beta + \epsilon \quad (2)$$

such that the utility is equal to the linear predictor plus an error  $\epsilon$ . However, we only observe whether  $A$  or  $B$  is selected (i.e., whether  $w_A > w_B$  or  $w_A < w_B$ ), namely

$$y = \begin{cases} 1 & \text{if } v > 0 \text{ (or } w_A > w_B), \\ 0 & \text{if } v < 0 \text{ (or } w_A < w_B). \end{cases} \quad (3)$$

When we take Expressions (2) and (3) together, we have the general framework for a categorical data model (Powers and Xie, 2000). If we assume that the error term for the latent variable is normally distributed, the probit model is obtained. Conversely, the logit model is obtained when the error is taken to have a logistic distribution.

The tails of the logistic distribution are considerably heavier than that of the standard normal. It has been noted that the logistic distribution is related to a member of the family

---

<sup>1</sup>Note that the matrix is square and thus  $I = J$ , but it is useful to have a notation which distinguishes between rows and columns.

of  $t$  distributions. Specifically, Albert and Chib (1993) have indicated that the logistic is well approximated by a  $t$  distribution with 8 degrees of freedom. Given the differences between the logistic and normal distributions, we might expect that there may be instances where the choice of link function impacts the analysis. Indeed, the performance of a categorical data model is known to depend at least in part on the choice of the link function (e.g., Pregibon, 1980). Aside from the estimation considerations, the development of a probit model in the context of stochastic multicriteria decision methods would be interesting as in the literature several authors have invoked the normal distribution in the context of extensions of the AHP. For example, the method of Ramanathan (1997) is based on an assumption of normally distributed errors when there are a large number of decision makers rendering judgments. Similarly, Van den Honert (1998) described a multiplicative AHP approach in which the pairwise comparisons of multiple decision makers were assumed to be normal (see also Van den Honert, 2001). Lipovetsky and Tishler (1999) assumed that the underlying weights of the alternatives were normally distributed. They then utilized the Cauchy distribution, a distribution which is obtained as the ratio of normal variates. Laininen and Hämäläinen (2003) described a robust regression approach to priority estimation in the multiplicative AHP where the logged pairwise comparison ratio is taken to be normal. Indeed, the log-normal conceptualization of the pairwise comparison has appeared with some regularity (e.g., Crawford and Williams, 1985; Escobar and Moreno-Jiménez, 2000). Moreover, the multivariate normality of the normalized principal eigenvectors in AHP has been assessed by Stam and Duarte Silva (1997).

In the previous section, the case of binary response models was discussed. However, in the case of more than two alternatives we require a model which allows for polychotomous outcomes. Categorical data models used to examine the choices of individuals in the presence of polychotomous outcomes are sometimes termed discrete choice models (Ben-Akiva and Lerman, 1985) and one of the most commonly used of these is the multinomial logit model. One well known property of the multinomial logit model is that it implies an assumption that the error terms are independent. When it is desirable to relax this particular assumption, the multinomial probit model such as that of Hausman and Wise (1978) is sometimes used. In the general multinomial probit model, the errors are assumed to have a  $J - 1$ -variate normal distribution, which relaxes the independence assumption. Thus, the general multinomial probit model, while attractive, implies a qualitatively different error structure than does the multinomial logit model. As such, it is not suitable for a direct examination of the effect of link functions in the current context because the general multinomial probit model is different from the multinomial logit in two aspects: the link and the (correlated) error structure. However, by appropriately specifying the covariance matrix, we can obtain an independent multinomial probit model with an uncorrelated error structure which is suitable for comparison with the multinomial logit model. We describe this model in detail in Section 2.2.

## 2.1 Model 1: Logit Link

Model 1 is a multinomial logit model and can be specified in the usual way as

$$w_{ij} \sim \text{Multinomial}\left(\sum_{i=1}^I w_{ij}, p_{ij}\right), \quad (4)$$

$$p_{ij} = \frac{\phi_{ij}}{\sum_{i=1}^I \phi_{ij}}, \quad (5)$$

$$\phi_{ij} = \exp(\alpha_j + \beta_{ij}). \quad (6)$$

Here, we see in (4) that the underlying weights of the decision maker are multinomially distributed such that they depend on the priorities as well as the sum of the weights. Also, we see in (5) and (6) the inverse multinomial logistic link function which is  $\exp(\cdot)/(\sum \exp(\cdot))$ . Moreover, a simple algebraic manipulation of (6) indicates that  $\ln(\phi_{ik})$  is linearly related to the predictors  $\alpha_k$  and  $\beta_{ik}$ . Thus, we see the model is linear in the logs similar to the multiplicative AHP. We discuss this relationship further in the next section. In Model 1, the various  $\alpha$  and  $\beta$  parameters are estimated simultaneously. We have in principle  $J$   $\alpha$  coefficients to estimate, one for each of the  $J$  alternatives. For identifiability, however, it is necessary to set one of these, say  $\alpha_1$ , to zero. Similarly, there are  $(I \times J)$   $\beta$  coefficients to estimate. Again, we need to set all of the  $\beta_{1,j}$  and  $\beta_{i,1}$  coefficients to zero to ensure identifiability. Because  $\mathbf{x}$  is a design matrix of zeros and ones, it need not appear explicitly because the values of  $\alpha$  and  $\beta$  are all intercepts specified by the design matrix. In the multinomial logit model the distribution of the errors,  $\epsilon_{ij}$ , follows a type-I extreme value distribution having cumulative distribution function  $F(\epsilon_{ij} < \epsilon) = \exp[-\exp(\epsilon)]$ . This distribution implies that the random variable defined as the difference between two error terms follows a logistic distribution. Recall that the difference in utilities,  $v$ , was used in the binary model formulation in (2).

Model 1 is a marginal model and as such all of the coefficients are assumed to be independent. In the Bayesian framework to be used, we need to provide a probabilistic specification for all of the parameters to be estimated. Thus, we have

$$\Pr(\alpha_2, \dots, \alpha_J, \beta_{2,2}, \dots, \beta_{J,J}) \propto \prod_{j=2}^J \Pr(\alpha_j) \times \prod_{i=2}^I \prod_{j=2}^J \Pr(\beta_{i,j}). \quad (7)$$

Here, we can see that the joint probability density of the parameters is the product of the marginals and as such all of the parameters are independent. Hierarchical extensions of this model have been considered (Hahn, 2003) but in the current research we examine only the marginal formulation.

## 2.2 Model 2: Probit Link

We describe here a independent multinomial probit model to compare with Model 1. The case of  $J = 3$  is presented but the extension to greater values of  $J$  is straightforward.

Following Maddala (1983), with three alternatives we have

$$\begin{aligned} u_1 &= \eta_1 + \epsilon_1, \\ u_2 &= \eta_2 + \epsilon_2, \\ u_3 &= \eta_3 + \epsilon_3 \end{aligned}$$

where  $u$  is the latent utility for a particular alternative. The error terms  $(\epsilon_1, \dots, \epsilon_3)$  are assumed to have a trivariate normal distribution with mean  $\mu = (0, 0, 0)$  and variance-covariance matrix  $\Sigma$ . Here,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}.$$

The probability that, say, the first alternative is chosen can be written as  $\Pr(u_1 > u_2, u_1 > u_3)$ . The dimensionality of the problem can be reduced from  $J$  to  $J - 1$  by differencing (e.g., Alvarez and Nagler, 1995). We may therefore write the probability of the first alternative being chosen as  $\Pr(\eta_1 - \eta_2 > \epsilon_2 - \epsilon_1, \eta_1 - \eta_3 > \epsilon_3 - \epsilon_1)$ . Let  $\bar{\eta}_{1j} = \eta_1 - \eta_j$  and  $\bar{\epsilon}_{j1} = \epsilon_j - \epsilon_1$ . The joint covariance matrix for the  $\bar{\epsilon}_{j1}$  is

$$\Omega_1 = \begin{pmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} \\ \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{pmatrix}.$$

Then, the probability of alternative 1 being selected is

$$p_1 = \int_{-\infty}^{\bar{\eta}_{12}/\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} \int_{-\infty}^{\bar{\eta}_{13}/\sqrt{\sigma_1^2 + \sigma_3^2 - 2\sigma_{13}}} f(\bar{\epsilon}_{21}, \bar{\epsilon}_{31}) d\bar{\epsilon}_{21} d\bar{\epsilon}_{31},$$

where  $f(\cdot)$  is the standard bivariate normal distribution.

Analogously to Model 1, we see that Model 2 can be specified as

$$w_{ij} \sim \text{Multinomial}\left(\sum_{i=1}^I w_{ij}, p_{ij}\right), \quad (8)$$

$$p_{i1} = \Phi_2(-\eta_{i2}, -\eta_{i3}; \Omega_1), \quad (9)$$

$$p_{i2} = \Phi_2(\eta_{i2}, \eta_{i2} - \eta_{i3}; \Omega_2), \quad (10)$$

$$p_{i3} = \Phi_2(\eta_{i3}, \eta_{i3} - \eta_{i2}; \Omega_3), \quad (11)$$

where  $\Phi_2$  is the bivariate standard normal cumulative distribution function. Here, the  $\alpha$  and  $\beta$  coefficients appear in the original values of  $\eta$  prior to differencing. Again, we see the underlying weights are multinomially distributed conditional on the priorities. In (9)-(11) the differenced linear predictors are used to obtain the priorities by way of the bivariate normal link function. It can be shown (e.g., Hausman and Wise, 1978; Terza, 1998) that in an independent multinomial probit model  $\Omega$  after standardization has entries of 1 on the diagonal and 1/2 elsewhere. Note that the off-diagonal elements of 1/2 result from the differencing. Model 2 is also a marginal model and has the same general structure as in (7).

It is instructive to compare Models 1 and 2 presented here to another well-known method for estimating priorities, the multiplicative AHP. In the multiplicative AHP as well as in the current approach, linear models are used to estimate priorities by utilizing pairwise comparisons which have undergone a transformation. In the multiplicative AHP, the transformation is the natural logarithm whereas in the current models the transform is the logit transform or the inverse multivariate normal cumulative distribution function. Thus, the multiplicative AHP can also be seen to make use of a link or mapping function. The multiplicative AHP appears to have originally arisen from the line-fitting tradition of minimum least squares estimation (e.g., Lootsma, 1996), thus its stochastic characteristics were originally less explicitly presented. However, these considerations have been made more explicit over time (Van den Honert, 1998; Laininen and Hämäläinen, 2003). By contrast the method here has been explicitly derived from a stochastic characterization of the pairwise comparison as the ratio of two kinds of preference outcomes. The pairwise comparison is therefore treated like an odds ratio of preference outcomes. Among other things, in this approach the sampling distribution of the priorities is immediately obtainable and inference regarding the priorities is straightforward using MCMC. It should be noted that the models associated with the stochastic judgment method such as Models 1 and 2 may be more computationally expensive than are multiplicative AHP models. However, the statistical assumptions associated with the multinomial distribution are possibly more easily met. The statistical specification of the multiplicative AHP requires the logged pairwise comparisons to be independent and identically normally distributed with constant variance (Laininen and Hämäläinen, 2003). By contrast, in the stochastic judgment method, constancy of variance is never assumed and dependence structures are easily incorporated into the model. For example, Model 2 can easily permit dependence to be estimated from the data via estimation of the covariance parameters in  $\Omega$  whereas dependence can be introduced into Model 1 via a hierarchical formulation.

A more comprehensive understanding of the performance characteristics of the two different kinds of models can be obtained from a Monte Carlo study involving the two models. Thus, a study was conducted as follows. The main factor of interest was the effect of the link function. As such, the underlying priorities were estimated using the multinomial logit model of Section 2.1 and then were separately estimated using the independent multinomial probit model of Section 2.2. In addition to the link function, we examine the performance of the models



with respect to the number of alternatives. Here, pairwise comparison matrices having either 3 or 4 alternatives were considered. Within each of these two conditions, 20 pairwise comparison matrices were randomly generated. One important characteristic of these pairwise comparison matrices was the amount of judgmental inconsistency as measured by the consistency index ( $CI$ ; Saaty, 1977). The pairwise comparison matrices utilized here were generated to have a  $CI$  such that  $CI \leq 0.10$ . The  $CI \leq 0.10$  criterion is commonly utilized to determine whether a pairwise comparison matrix has an acceptable level of consistency. A second way in which pairwise comparison matrices may differ involves the dispersion of the priorities. If the priorities are widely dispersed, some alternatives are considerably more preferable or important than others. That is, certain alternatives will strongly dominate others. Conversely, if the dispersion is minimal, all the alternatives have relatively similar aggregate underlying weights. Here the dispersion of the priorities was measured by calculating their standard deviation. In summary, a factorial design was employed in the Monte Carlo study which had a 2 (*link function*)  $\times$  2 (*number of alternatives*) structure with 20 pairwise comparison matrices observed in each cell. The  $CI$  and dispersion of the priorities were also recorded for use as continuous covariates.

In the Bayesian framework, it is necessary to place priors on model parameters. Here, very flat normal priors were used for all  $\alpha$  and  $\beta$  regression coefficients as non-informative priors. Specifically, the normal priors utilized had mean zero and variance 1,000,000. Thus, the priors were flat across all reasonable values of the model parameters. For each model, a burn-in period of 1,000 iterations of the Markov chain was used. This burn-in period allows the Markov chain to move from initial parameter values to its final distribution. Here, the initial values were set to be zero. Due to the nature of both multinomial logit and probit models in the context of stochastic multicriteria decision making problems, parameter posterior distributions regularly either include zero or are very close to zero. Therefore, a burn-in period of 1,000 iterations was conservative in that it was more than sufficient to allow the Markov chain to reach its final steady state. The Markov chain was then monitored for an additional 50,000 iterations. These iterations were used for inference. In general, estimation of these models is straightforward and as a consequence the Markov chains were well-behaved.

In the current work we focus on the point estimation problem as interval estimation has been addressed in previous research. We compare the model priority estimates to the AHP priorities and quantify the differences between them using the mean absolute deviation (MAD). For each judgment matrix, the discrepancy between the estimated priorities and the priorities obtained from AHP was summarized by using the MAD between the two sets of priorities. Figure 1 contains plots of the models' discrepancies as a function of the  $CI$ . The plots reveal clearly that as inconsistency increases, the discrepancy between the estimated priorities and the AHP priorities likewise increases. Still, we see that there are some differences. For example, the first row of Figure 1 shows that the discrepancy level between Model 1 and AHP is comparatively lower whereas the second row shows that the discrepancy level between Model 2 and AHP is comparatively higher. More precise conclusions may be gained by examining the data using a regression approach. Thus, Table 1 contains summary statistics from regressions of  $CI$  on MAD for the four models<sup>2</sup>. In addition

<sup>2</sup>Note that Bayesian and OLS regression coefficients are identical under flat priors.

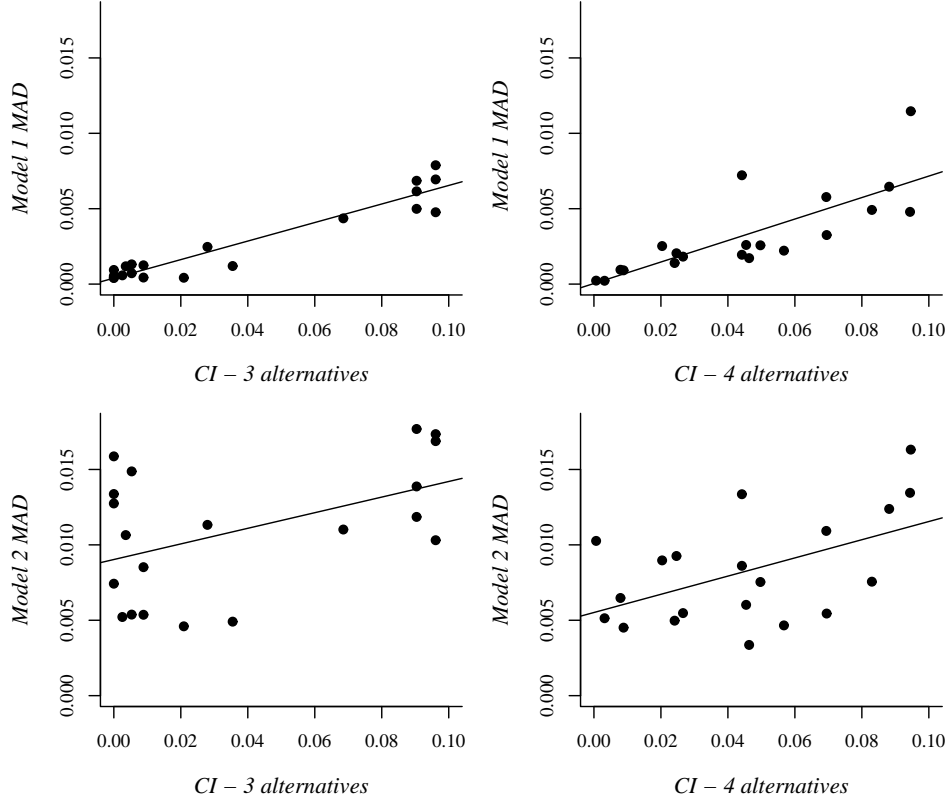
to the coefficients, we also present in Table 1 the standardized slopes. These quantities provide a common basis for discerning whether a slope is large relative to its standard error and moreover are the  $t$  statistics used in classical hypothesis testing. Table 1 shows a relatively prominent difference in the intercepts. The Model 1 intercepts are approximately zero whereas the Model 2 intercepts are considerably larger (.0055 and .009). This indicates that under the logit link with perfect or nearly perfect consistency the model estimated priorities recovered the AHP priorities almost exactly. However, under the probit link we obtain estimates that are considerably different from the priorities from AHP, even when judgments are perfectly consistent. This evidence suggests that an assumption of normality of errors in stochastic multicriteria decision models may lead to estimates that are somewhat divergent from those of AHP. This seems to suggest that an assumption of normality of errors is not consonant with the priorities of AHP, or conversely that the priorities of AHP are not consonant with an assumption of normality of errors. Table 1 also shows that values of the slopes are roughly comparable across the different models and different numbers of alternatives. For a .01 increase in  $CI$  the MAD would be expected to increase by between .00052 to .00071. This indicates that as inconsistency increases, the discrepancy between the model estimated priorities and the AHP priorities increases at an attenuated rate. Finally, Table 1 shows the values of  $MAD_{0.1}$ , the predicted MAD at the maximum  $CI$  of 0.1. We see from the values of  $MAD_{0.1}$  that the discrepancy between the AHP priorities and the estimated priorities is approximately twice as large under Model 2 as it is under Model 1. Phrased differently, in a worst case scenario of high inconsistency, Model 1 is predicted to be half as discrepant from AHP on average as Model 2 under the MAD criterion.

Table 1: Regression Coefficients for Discrepancy versus  $CI$

	Alternatives	Intercept	Slope	Slope/S.E.	$MAD_{0.1}$
Model 1	3	0.0004	0.0615	13.39	0.0065
	4	0.0000	0.0712	5.32	0.0072
Model 2	3	0.0090	0.0517	2.32	0.0142
	4	0.0055	0.0603	2.55	0.0115

We now examine model performance with respect to dispersion of the priorities. Recall that if the priorities are widely dispersed, there are some alternatives that are judged to be particularly preferable and others that are judged to be particularly undesirable. Conversely, if the priorities are all very similar to each other in magnitude, then the alternatives are judged to be approximately equally preferable. Dispersion is quantified here by taking the standard deviation of the priorities generated by AHP as the dispersion metric. Figure 2 displays plots of the MADs as a function of the dispersion metric. The figure clearly shows that the slopes for Model 2 are more pronounced than those of Model 1. Incidentally, Figure 2 also shows that as the number of alternatives goes up, the standard deviation of the priorities tends to go down. This is due to the fact that the priorities must sum to one. For example, consider the case where one alternative is strongly preferred to the remainder. As

Figure 1: Discrepancy and Judgment Matrix Consistency

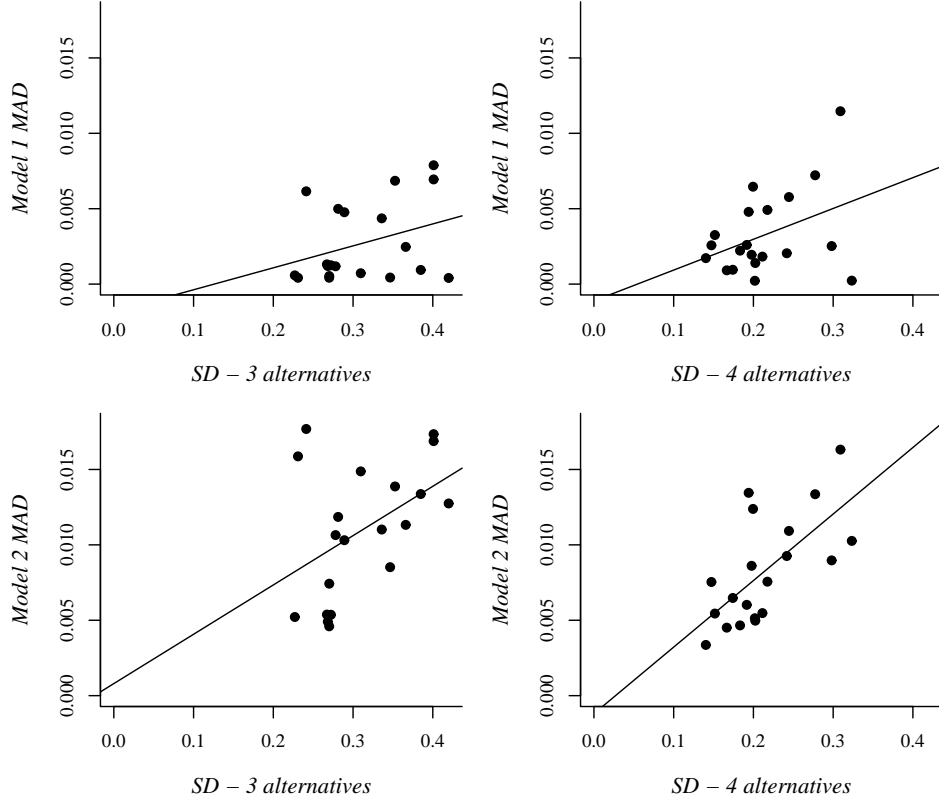


the number of alternatives  $J$  increases, the  $J - 1$  alternatives which are not preferred cluster near zero thereby shrinking the standard deviation of the priorities.

Table 2 contains summary statistics from regressions of priority standard deviation on MAD for the four models. The table shows some variability in the intercepts. The Model 1 intercepts are negative while the Model 2 intercepts have both signs. The intercepts theoretically have a useful interpretation as the expected MAD when the priorities are equal (giving a standard deviation of zero). However, the intercepts are more difficult to interpret reliably here as the bulk of the data lies away from the origin. Table 2 confirms that the slopes are more pronounced in Model 2 than in Model 1. This indicates that when the priority standard deviation decreases, the Model 2 priorities approach the AHP priorities at a faster rate than the Model 1 priorities. Conversely, as standard deviations increase, Model 2 priorities diverge from those of AHP faster than do the Model 1 priorities. Because none of the randomly generated judgment matrices had standard deviations close to zero, there is a risk of overinterpreting the results. However, the results suggest that the assumption of normal errors may yield estimates consonant with AHP when all alternatives are approximately equally preferable. Conversely, AHP may generate priorities consonant with a normality of errors assumption when all alternatives are essentially equally preferable.

The MAD is useful for examining the discrepancies between the priorities of AHP and

Figure 2: Discrepancy and Priority Standard Deviation



those of Models 1 and 2. However, a more detailed examination of the nature of the discrepancies is possible using an appropriate metric. As a point of departure, we note that Stam and Duarte Silva (2003) found the multiplicative AHP had a moderating effect on priorities compared to the original AHP. That is, for a given set of judgments the priorities produced by multiplicative AHP tended to be more distant from the extremes of 0 and 1 than were the priorities of the original AHP. We develop here an comparative measure of extremity and use it to examine the priorities from Models 1 and 2. Denote the AHP priorities as  $\tilde{p}$  and the priorities of Models 1 and 2 as  $\hat{p}$ . Then a mean extremity index can be defined as

$$\kappa = \frac{1}{J} \sum_{j=1}^J ((\tilde{p}_j - \hat{p}_j) \times \text{sgn}(\tilde{p}_j - \bar{\tilde{p}}))$$

where  $\bar{\tilde{p}}$  is the mean of the AHP priorities and  $\text{sgn}(\cdot)$  denotes the signum or sign function. Here,  $\bar{\tilde{p}}$  is used as a measure of central tendency by which we may categorize priorities as being above the center or below the center. If, for a given pairwise comparison matrix, the AHP priorities are more extreme on average than those of a particular model, the index will be positive. Conversely, if the priorities of a model are more extreme than those of AHP, the index will be negative.

Figure 3 contains the values of  $\kappa$  for the two models under the conditions of both 3 and 4 alternatives. The actual values of  $\kappa$  for the 20 judgment matrices appear as short

Table 2: Regression Coefficients for Discrepancy versus Priority Standard Deviation

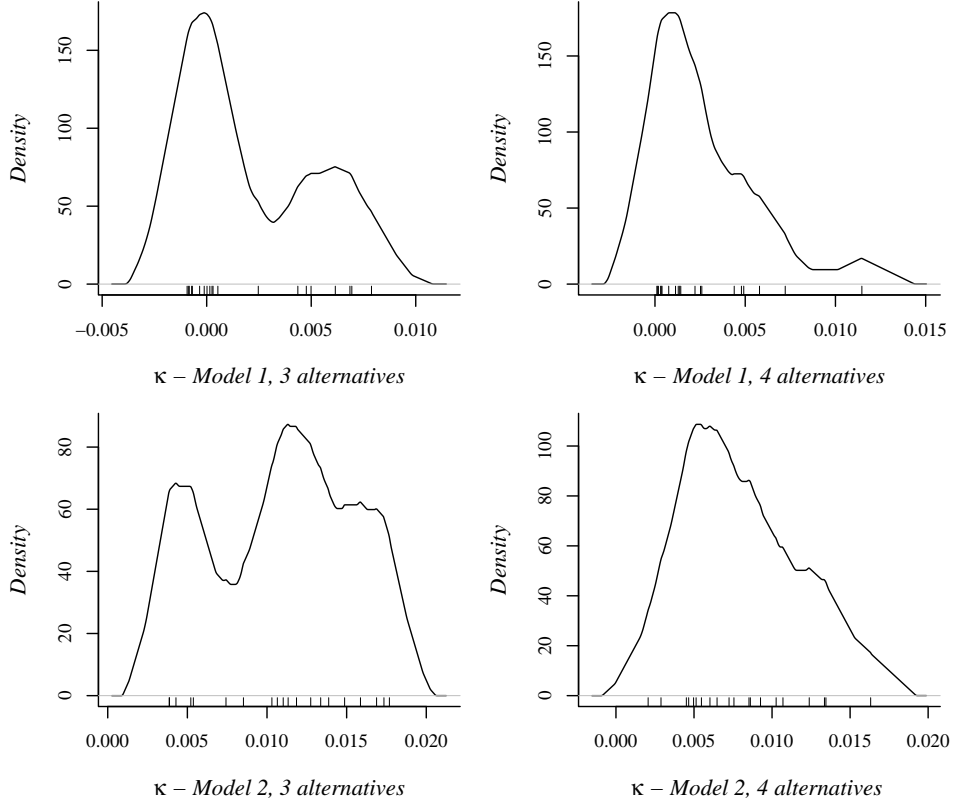
	Alternatives	Intercept	Slope	Slope/S.E.
Model 1	3	-0.0018	0.0146	1.50
	4	-0.0011	0.0204	1.80
Model 2	3	0.0008	0.0327	2.13
	4	-0.0012	0.0441	3.69

vertical lines above the  $x$ -axis. Above the values of  $\kappa$  is a nonparametric kernel density estimate of the distribution of  $\kappa$  for that condition. For example, the top left portion of Figure 3 pertains to  $\kappa$  under Model 1 for the 20 3-alternative judgment matrices. We can see there is a cluster of values very close to zero where essentially neither set of estimates is appreciably more extreme than the other. We then see another cluster of values in the positive portion of the  $x$ -axis in the vicinity of 0.005. These values indicate the presence of a number of judgment matrices for which each AHP priority is on average more extreme than its Model 1 counterpart by roughly 0.005. The top right portion of the figure displays a similar pattern of a cluster near zero accompanied by a number of increasingly positive values of  $\kappa$ . This indicates that the priorities of Model 1 may be as extreme as those of AHP in some circumstances, but in other circumstances the AHP priorities may be more extreme. The lower two panels of Figure 3 provide insight as to why Model 2's priorities were found to be consistently discrepant from the AHP priorities. We see that the AHP priorities are consistently more extreme than those of Model 2. Indeed, many of the values of  $\kappa$  are at 0.01 or greater. Such values indicate that in those circumstances the priorities from the two approaches would not agree even at the second decimal place because on average the AHP priority would be 0.01 more extreme than Model 2's priority. Viewed from another perspective, Model 2 and its assumption of normally distributed errors can be seen to have a moderating effect of the priorities *vis-à-vis* AHP.

## 4 Example

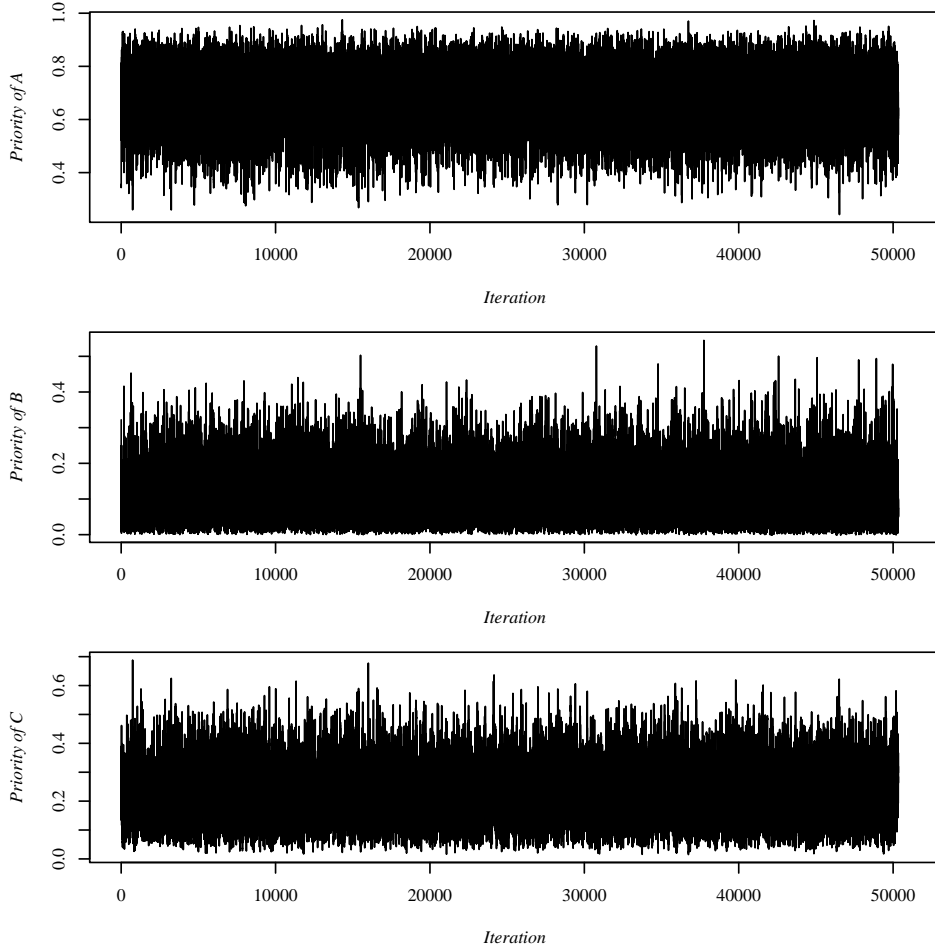
As an example, we consider a decision regarding which of three features to provide as an option on a luxury automobile. Alternative  $A$  is a rear-view wide-angle camera with a dashboard display, Alternative  $B$  is an anti-rollover system, and Alternative  $C$  is a drowsy-driver warning system. Judgments regarding the alternatives were assessed on the AHP scale and  $A$  was very strongly preferred to  $B$ ,  $A$  was moderately preferred to  $C$ , and the preference for  $C$  over  $B$  was between moderate and equal. In terms of numeric representations, these judgments corresponded to a 7/1 pairwise comparison for  $A$  versus  $B$ , a 3/1 pairwise comparison for  $A$  versus  $C$ , and a 1/2 pairwise comparison for  $B$  versus  $C$ . These judgments result in AHP priorities of 0.6817, 0.1025, and 0.2158 respectively for  $A$ ,  $B$  and  $C$ . The  $CI$  is quite low at 0.003 indicating a high degree of consistency.

The estimation of Models 1 and 2 was conducted using the same conditions as were used

Figure 3: Kernel Densities of  $\kappa$ 


in the Monte Carlo study in Section 3. Specifically, flat normal priors having mean zero and variance of 1,000,000 were utilized for all regression coefficients. The initial 1,000 iterations of the Markov chain were discarded as a burn-in period and inference was conducted using a subsequent 50,000 iterations. A trace plot of the Markov chain for the priorities,  $\mathbf{p}$ , of Model 1 appears in Figure 4. It is evident from Figure 4 that the Markov chains move to the stationary distributions immediately, well before the 1,000-iteration burn-in period has elapsed. The point estimates of the priorities under Model 1 are 0.6809, 0.1025, and 0.2166. The corresponding MAD is 0.0005 which is small. Thus, Model 1 recovers the AHP point estimates well. We might expect the priorities of Model 2 to be somewhat different from the AHP priorities, particularly since the dispersion of the priorities is not low ( $sd = 0.307$ ). This is confirmed with respective priority estimates of 0.6546, 0.1193 and 0.2260 under Model 2. The MAD here is a somewhat appreciable 0.018.

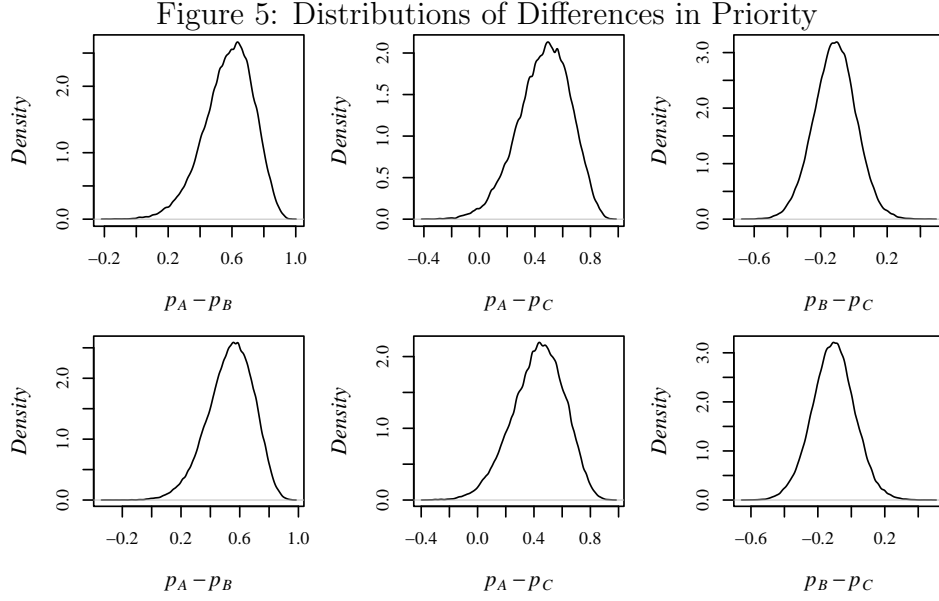
Inference regarding the priorities can be conducted by constructing the posterior distributions of  $p_A - p_B$ ,  $p_A - p_C$ , and  $p_B - p_C$  from the MCMC run. Plots of these distributions appear in Figure 5. Table 3 contains descriptive statistics for these distributions. We may conduct inference by looking at the 95% posterior probability intervals for the differences in the priorities. For example, consider the distributions of  $p_A - p_B$  at the left of Figure 5. We can see that the distributions are essentially positive, with negative values occurring rarely. Thus, our intuition is that  $p_A > p_B$  since  $p_A - p_B > 0$ . Table 3 permits a more formal

Figure 4: Traces of  $\mathbf{p}$ 

inferential treatment of this issue. From Table 3, we see that the 95% interval for  $p_A - p_B$  under Model 1 is 0.240 – 0.838. Since this interval does not include the value of zero, we may conclude that  $p_A > p_B$  at the 95% probability level. The same conclusion is reached under Model 2. Inspection of Table 3 also indicates we may conclude that  $p_A > p_C$  at the 95% probability level under both models. However, we may not conclude that  $p_C > p_B$  at the 95% probability level under either model since the relevant intervals contain the value zero.

## 5 Discussion

In general, Model 1 which utilized the logit link yielded point estimates of stochastic priorities that were relatively consonant with the priorities of AHP. This finding suggests that of the two stochastic judgment method approaches described here, Model 1 may be preferred by decision makers who are seeking to replicate AHP results as closely as possible. One interpretation of the findings is that AHP produces priorities that are similar to a categorical



data model in which logistic errors are assumed. This stands in contrast with many of the stochastic AHP approaches described previously which are predicated on underlying normality assumptions. In the models examined here, such normality assumptions lead to priority estimates which were consistently less toward extreme values than were those of AHP. Previous research on the multiplicative AHP has shown it to produce priority estimates that are also more moderate than those of AHP. The results also suggest that the frequently-invoked assumption of normal errors in other stochastic approaches such as the multiplicative AHP may not necessarily be the assumption that produces results which best replicate those of AHP. The effects of alternative assumptions such as assuming a heavy-tailed distribution for the errors would provide greater insights into the original formulations of these approaches.

As mentioned in Section 1, there are a number of stochastic multicriteria decision making methods that are related to AHP. Some require interval judgments and others require multiple decision makers. Others make assumptions about independence of judgments, and still others (such as the ones described here) do not require such assumptions. It is not uncommon in this literature for authors to compare the results of a particular method to that of the original AHP. However, there appears to be a relative dearth of empirical research comparing the various stochastic multicriteria decision making methods to one another. Thus, many questions remain unanswered regarding the relationships among these stochastic multicriteria decision making methods. In the current paper, we attempt to outline some similarities and some contrasts between the models described herein and an alternative approach, the multiplicative AHP. Future research should attempt to explicitly answer questions about the circumstances under which certain stochastic methods become preferred over others as such research will help unify and integrate this growing field of inquiry.



Table 3: Descriptive Statistics for Posterior Distributions of Differences in Priorities

	Mean	Standard Deviation	2.5% Quantile	97.5% Quantile
Model 1				
$p_A - p_B$	0.578	0.153	0.240	0.838
$p_A - p_C$	0.464	0.187	0.067	0.789
$p_B - p_C$	-0.114	0.126	-0.365	0.132
Model 2				
$p_A - p_B$	0.535	0.155	0.202	0.801
$p_A - p_C$	0.429	0.183	0.042	0.750
$p_B - p_C$	-0.107	0.127	-0.359	0.144

## References

- Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669–679.
- Alvarez, R.M., Nagler, J., 1995. Economics, issues and the Perot candidacy: Voter choice in the 1992 presidential election. *American Journal of Political Science* 39(3), 714–744.
- Arbel, A., 1989. Approximate articulation of preference and priority derivation. *European Journal of Operational Research* 43(3), 317–326.
- Basak, I., 1991. Inference in pairwise comparison experiments based on ratio scales. *Journal of Mathematical Psychology* 35(1), 80–91.
- Basak, I., 1998. Probabilistic judgments specified partially in the Analytic Hierarchy Process. *European Journal of Operational Research* 108(1), 153–164.
- Basak, I., 2001. The categorical data analysis approach for ratio model of pairwise comparisons. *European Journal of Operational Research* 128(3), 532–544.
- Ben-Akiva, M., Lerman, S.R., 1985. *Discrete Choice Analysis*. Cambridge, MA, MIT Press.
- Bryson, N., Joseph, A., 2000. Generating consensus priority interval vectors for group decision-making in the AHP. *Journal of Multicriteria Decision Analysis* 9(4), 127–137.
- Crawford, G., Williams, C., 1985. A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology* 29(4), 387–405.
- de Jong, P., 1984. A statistical approach to Saaty’s scaling method for priorities. *Journal of Mathematical Psychology* 28(4), 467–478.
- Escobar, M.T., Moreno-Jiménez, J.M., 2000. Reciprocal distributions in the Analytic Hierarchy Process. *European Journal of Operational Research* 123(1), 154–174.

- 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(4), 477–496.
- Hahn, E.D., 2003. Decision making with uncertain judgments: A stochastic formulation of the Analytic Hierarchy Process. *Decision Sciences* 34(3), 443–466.
- 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(1), 112–125.
- Hausman, J.A., Wise, D.A., 1978. A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica* 46(2), 403–426.
- Imai, K., van Dyk, D.A., 2004. A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, forthcoming .
- Jensen, R.E., 1984. An alternative scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 28(3), 317–332.
- Laininen, P., Hämäläinen, R.P., 2003. Analyzing AHP-matrices by regression. *European Journal of Operational Research* 148(3), 514–524.
- Lipovetsky, S., Tishler, A., 1999. Interval estimation of priorities in the AHP. *European Journal of Operational Research* 114(1), 153–164.
- Lootsma, F.A., 1996. A model for the relative importance of the criteria in the multiplicative AHP and SMART. *European Journal of Operational Research* 94(3), 467–476.
- Maddala, G.S., 1983. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge, Cambridge University Press.
- McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), *Frontiers of Econometrics*. Academic Press, New York, pp. 105–142.
- Phillips-Wren, G.E., Hahn, E.D., Forgionne, G.A., 2004. A multiple-criteria framework for evaluation of decision support systems. *Omega* 32(4), 323–332.
- Powers, D.A., Xie, Y., 2000. *Statistical Methods for Categorical Data Analysis*. San Diego, Academic Press.
- Pregibon, D., 1980. Goodness of link tests for generalized linear models. *Applied Statistics* 29(1), 15–24.
- Ramanathan, R., 1997. Stochastic decision making using multiplicative AHP. *European Journal of Operational Research* 97(3), 543–549.
- Saaty, T.L., 1977. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15(3), 234–281.

- Saaty, T.L., Vargas, L.G., 1987. Uncertainty and rank order in the Analytic Hierarchy Process. *European Journal of Operational Research* 32(1), 107–117.
- Stam, A., Duarte Silva, A.P., 1997. Stochastic judgments in the AHP: The measurement of rank reversal probabilities. *Decision Sciences* 28(3), 655–688.
- Stam, A., Duarte Silva, A.P., 2003. On multiplicative priority rating methods for the AHP. *European Journal of Operational Research* 145(1), 92–108.
- Sugihara, K., Ishii, H., Tanaka, H., 2004. Interval priorities in AHP by interval regression analysis. *European Journal of Operational Research* 158(3), 745–754.
- Terza, J.V., 1998. Independence in multinomial choice models: The Horowitz test revisited. Working paper, Penn State University. Available at <http://grizzly.la.psu.edu/~terza/>. Retrieved Jul. 21, 2004.
- Vaidya, O.S., Kumar, S., 2004. Analytic Hierarchy Process: An overview of applications. *European Journal of Operational Research*, forthcoming .
- Van den Honert, R.C., 1998. Stochastic group preference modelling in the multiplicative AHP: A model of group consensus. *European Journal of Operational Research* 110(3), 99–111.
- Van den Honert, R.C., 2001. Decisional power in group decision making: A note on the allocation of group members' weights in the multiplicative AHP and SMART. *Group Decision and Negotiation* 10(3), 275–286.