SPC

An application of two sided power distribution in Bayesian analysis of paired comparison of relative importance of predictors in linear regression models

Xiaoyin Wang

Department of Mathematics, Towson University, Towson, MD 21252*, USA*
Email: xwang@towson.edu

Abstract

The purpose of determining the relative importance of predictors is to expose the extent of the individual contribution of a predictor in the presence of other predictors within a selected model. The goal of this article is to expand the current research practice by developing a statistical paired comparison model with Two Sided Power (TSP) link function in the Bayesian framework to evaluate the relative importance of each predictor in a multiple regression model. Results from simulation studies and empirical example reveal that the proposed Two Sided Power link function provides similar conclusions as the commonly used logit link function, but has more advantages from both practical and computational perspectives.

Keywords: *Bayesian Estimate; Markov Chain Monte Carlo Method(MCMC); Paired Comparison; Relative Importance of Predictor; Two Sided Power Distribution*

1. Introduction

Finding out the relative importance of predictors in linear regressions has been a long debated issue, with literature being developed in various subjects such as, statistics, psychology, political economy, organizational research, and medicine [1]. The purpose of determining the predictor importance is to uncover the individual contribution of a predictor relative to each other within a selected model. Here, the question of interest becomes understanding the extent to which each variable drives the prediction.

Over the years, researchers have suggested various methods for measuring the relative importance of predictors; for a comprehensive review see [1]. However, there has been a lack of agreement on the definition of the relative importance of a predictor in the multiple regression context. Past research has documented how indices commonly produced by multiple regression analyses fail to appropriately answer this question when predictors are correlated [2]. Later, an alternative approaches, the Dominance Analysis [3], has been produced, which allows for more accurate variance partitioning among correlated predictors. The Dominance Analysis approaches the problem of relative importance by examining the change in R^2 resulting from adding a predictor to all possible subset regression models. The novelty of this approach is that predictors are compared in a pairwise fashion based on a common

reference model across all possible subset models, and a hierarchy of levels of dominance is established, *i.e.*, complete dominance, conditional dominance and general dominance.

The Dominance Analysis is unique in that it measures the relative importance of predictors in a pairwise fashion using the average increase in R^2 associated with a variable across all possible subset models, and two predictors are compared in the context of all possible models that contain the same subsets of other predictors. Nevertheless, the general dominance weight, which is calculated as the average of conditional dominance weights over difference sizes of subset models, introduces bias in the aggregation because the additional contribution of a predictor is more substantial in subset models with fewer number of predictors, and as the number of predictors increases, the additional contribution of the predictor gets less. The value of the general order dominance index is influenced to a greater extent by the conditional dominance weights of the subset model with a smaller number of predictors, see [4] for detailed illustration.

Moreover, the Dominance Analysis method is established based on observational statistics, and the estimated relative importance may differ from the dominance order in the population. The reproducibility process used to derive the probability of pairwise dominance without any particular model assumptions is based on a non-parametric bootstrapping technique that yields similar results as the Bayesian approach only under certain conditions [5, 6]. Lahiri (2003) [5] provides a critical review of various modifications of Efrons original bootstrap to handle complex issues in survey sampling. For example, it has been shown that Efrons basic bootstrap procedure fails when the observed sample points are not independent [5].

Suppressor variables are well known in the context of multiple regression analysis. A suppressor variable is a predictor that is uncorrelated with the response variable but whose presence improves prediction because of its correlation with other predictors. Wang and Yao (2004) [4] states that, in certain circumstance, when the suppressor is highly correlated with the other predictors, the Dominance Analysis method is not able to correctly identify the suppressor variable as the least important variable. Johnson (2000) [7] explicitly stated that relative importance analysis should not replace the model selection in multiple regression analysis. On the other side, researchers in the applied area might want to retain all variables in their predictive models to study the evolutions of these parameters. Thus, they are not seeking variable selection modelisation schemes. Furthermore, a promising measure of relative importance should correctly identify the suppressor, if it exists, as the least important variable.

With regards to these crises, Wang et. al. (2013) [8] proposed an inferential approach to discover the relative importance of predictors in a Bayesian statistical framework using the Bradley-Terry paired comparison model. In most of the ordinary cases, conclusions from the Bayesian approach are consistent with those from the Dominance Analysis methods. However, the basic Bradley-Terry model is not very effective when there exists suppressor variables. As a remedy, Wang and Yao (2014) [4] proposed a weighted Bradley-Terry model in the Bayesian framework. Simulation results show that this weighted model is more efficient than the basic unweighted Bradley-Terry model and the Dominance Analysis. The dominance index resulting from the Bradley-Terry models varies from $-\infty$ to ∞. Researchers in the applied area have a preference to an index that is on the scale of $0-1$. It is more comprehensible to practitioners when one says that the closer an index is to 1.0, the more important the predictor is.

In this paper, we propose to apply the cumulative distribution function of the two-sided power (TSP) distribution in the paired comparison model to ensure the resulting dominance index vary between 0 and 1. In section 2, we formulate paired comparison models applying the logit and the TSP link functions with weighted and un-weighted likelihood functions. In section 3, we introduce the Bayesian inference and the Markov chain Monte Carlo (MCMC) computational procedures. In section 4, we perform simulation studies to compare the outcomes of different models using different sets population correlation matrices. In section 5, we report an empirical example of a math proficiency research study. Finally, in section 6, we discuss potential future research avenues regarding the relative importance of predictors.

2.   Paired Comparison Models

Statistical methods of determining relative dominance abilities based on paired comparisons have long history dated back to Bradley and Terry (1952) [9] in the context of chess tournaments, and have been broadly applied in many fields such as statistics, psychometrics, marketing research, preference measurement, sports competition, behavioural study, education, machine learning, and many others. In this paper, we propose to determine the relative importance of predictors from a behavioural study point of view by applying the paired comparisons based on the additional increase of R^2 amongst all possible subset models.

2.1. Likelihood Function

Suppose there are p predictors in a multiple regression model, $X_1, \cdots, X_p$. Let X^* be a subset of $(X_1, \cdots, X_p)$ and denote $R^2_{Y \cdot X^*}$ as the coefficient of determination of Y and X^*. Let $\Delta R^2_{Y \cdot X_i | X^*} = R^2_{Y \cdot (X_i, X^*)} - R^2_{Y \cdot X^*}$ be the increase in R^2 by adding X_i to the subset model with X^*. Here, the model with X^* is considered as the baseline reference model. The predictor X_i is said to dominate or "win" X_j if $\Delta R^2_{Y \cdot X_i | X^*} > \Delta R^2_{Y \cdot X_j | X^*}$ because adding X_i to the model leads to a greater increase in R^2 than would be obtained by adding X_j to the model with the same subsets of other variables. In a regression model with p predictors, X_i and X_j are compared amongst 2^{p-2} baseline reference models. Let W_{ij} be the number of times the predictor X_i dominating X_j, and θ_{ij} be the probability of the predictor X_i dominating X_j. Here, W_{ij} has a binomial distribution with parameters $(2^{p-2}, \theta_{ij})$ and the likelihood function as follows.

$$\prod_{i<j} C^{w_{ij}}_{2^{p-2}} \theta_{ij}^{w_{ij}} (1 - \theta_{ij})^{2^{p-2} - w_{ij}}.$$

Wang and Yao (2014) [4] points out that the number of times that the predictors X_i and X_j encounter each other depends on the size of the baseline reference model. For a regression model with p predictors, based on the baseline reference model with k predictors, the number of times that the predictors X_i and X_j encounter is C^k_{p-2}. For example, when $p = 4$ and $k = 0$, X_1 and X_2 only meet once, that is $R^2_{Y \cdot X_1}$ and $R^2_{Y \cdot X_2}$; when $k = 1$, X_1 and X_2 are compared under two baseline reference models, which are $\Delta R^2_{Y \cdot X_1 | X_3}$ vs $\Delta R^2_{Y \cdot X_2 | X_3}$, and $\Delta R^2_{Y \cdot X_1 | X_4}$ vs $\Delta R^2_{Y \cdot X_2 | X_4}$; and when $k = 2$, X_1 and X_2 only meet once under the same baseline reference model with two predictors X_3 and X_4. Taking this fact into consideration, a weighted Bradley-Terry model is proposed with $C^k_{p-2}/2^{p-2}$ as the weight. Let $w_{ij \cdot k}$ be the number of times that the predictor X_i outweighs X_j when the baseline reference model has k predictors. The likelihood function of the weighted Bradley-Terry model is

$$\sum_{k=0}^{p-2} \frac{C^k_{p-2}}{2^{p-2}} \prod_{i<j} \begin{pmatrix} C^k_{p-2} \\ w_{ij \cdot k} \end{pmatrix} \theta_{ij}^{w_{ij \cdot k}} (1 - \theta_{ij})^{C^k_{p-2} - w_{ij \cdot k}}.$$

2.2. Link Functions

The original Bradley-Terry model assumes that $\theta_{ij} = \xi_i / (\xi_i + \xi_j)$, where ξ_i represents the relative dominance ability of predictor X_i. Let $d_i = ln(\xi_i)$, then d_i can be interpreted as the relative dominance ability of predictor X_i on the logarithm scale. Based upon this reparameterization, we have $logit(\theta_{ij}) = d_i - d_j$, or $\theta_{ij} = \exp(d_i - d_j)/(1 + \exp(d_i - d_j))$, which means that the dominance probability depends only on the difference of dominance indices. Therefore, the Bradley-Terry model assumes the probability that one predictor prevails over another is a logit function (the link function) of the difference in dominance indices between these two predictors.

A more general approach is to assume that $\theta_{ij} = H(d_i - d_j)$, where H is a link function, which maps the difference in dominance indices to a probability that lies between 0 and 1, inclusively. Also, the link function should have the following characteristics: (1) the larger the difference, the larger the value of θ_{ij}; that is, the more likely one predictor will dominate the other; (2) when two dominance indices are equal, θ_{ij} equals 0.5; that is, each predictor has a 50% chance of prevailing when dominance indices of two predictors are the same.

The paired comparisons of predictors are based upon the additional R^2, which ranges between 0 and 1. If the returning value of d_i is also within 0 and 1, then, to a extend, it reflects the additional R^2 associated X_i in the presence of other predictors. In practice, a index on the 0-1 scale is more favourable by the practitioners. Recently, vanDorp and Kotz (2002) [10] provide a four-parameter two-sided power (TSP) distribution with the cumulative distribution function that satisfies those characters. This link function is sufficiently rich from the mathematical perspective and also allows efficient implementations in practice. The TSP distribution is described by parameters denoting the minimum value, a, the maximum value, b, the mode or most likely value, c, and a fourth parameter, η, describing the curvature of the distribution. The parameter η requires expert evidence for the relative importance of the most likely value relative to distribution bounds a and b. That is, if $\eta = 2$, c is equally important as a and b; if $\eta > 2$, c has more weight than the bounds; if $\eta < 2$, c is given less emphasis relative to the bounds. The probability density function of $TSP(a, b, c, \eta)$ is presented in the equation (1).

$$f(x) = \begin{cases} \frac{\eta}{b-a} \left(\frac{x-a}{c-a} \right)^{\eta-1} & \text{if } a < x \le c \\ \frac{\eta}{b-a} \left(\frac{b-x}{b-c} \right)^{\eta-1} & \text{if } c < x \le b \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The expected value of a TSP distribution is

$$E(X) = \frac{a + (\eta - 1)c + b}{\eta + 1}$$

The cumulative distribution function follows the expression in the equation (2)

$$F(x) = \begin{cases} \frac{c-a}{b-a}\left(\frac{x-a}{c-a}\right)^{\eta} & \text{if } a < x \leq c \\ 1 - \frac{b-c}{b-a}\left(\frac{b-x}{b-c}\right)^{\eta} & \text{if } c < x \leq b \end{cases} \tag{2}$$

When values of d_i's are assumed falling between 0 and 1, values of $d_i - d_j$ vary between -1 and $+1$. Therefore, it is reasonable to choose $a = -1$ and $b = 1$. Because that we do not have any preference regrading the possible value of d_i's, we use $c = (a + b)/2 = 0$ and $\eta = 2$. As a results, the probability density function of this TSP distribution has a triangular shape with expected value of 0.

The logit and the TSP link function have very similar shapes of distribution, see Figure 1. In this paper, we focus on comparing the performance between the logit and the TSP link functions with weighted and unweighted likelihood functions, i.e., Model I: Logit link with unweighted likelihood; Model II: Logit link with weighted likelihood; Model III: TSP link with unweighted likelihood; and Model IV: TSP link with weighted likelihood.

3.   Bayesian Dominance Inference

Algorithms used to produce maximum likelihood estimates of dominance abilities under the Bradley-Terry model fail to converge to finite values, and so cannot be used for many data sets with zero counts. Davidson and Solomon (1973) [11] and Leonard (1977) [12] describe the Bayesian version of the methods of paired comparisons. The major benefit of the Bayesian approach is that prior information can be incorporated into the analysis so that the resulting estimates of dominance abilities are always finite[12].

3.1.   Prior Distribution

The prior distribution plays a key role of the Bayesian inference. It represents the information about an uncertain parameter, and is combined with the probability distribution of observed data to yield the posterior distribution, which in turn is used for future inferences and decisions. In practice, often there is no or very limited prior information about the dominance ability of predictors. Hierarchical (multilevel) and objective prior models are central to the modern Bayesian statistics for both conceptual and practical reasons. On the theoretical side, hierarchical models allow a more "objective" approach rather than requiring them to be specified using subjective information [13, 14, 15]. At a practical level, hierarchical models are flexible tools for aggregating information and partial pooling of inferences [16, 17, 18, 19, 20].

When the parameter varies from $-\infty$ to $+\infty$, a normal distribution is the most commonly used prior distribution because it corresponds to the prior belief that modest values of the parameter are nearly equally likely, and that very large values are somewhat less probable. For the logit link, values of d_i's range from $-\infty$ to $+\infty$. It is conventional to assume that there is the same symmetrical prior probability distribution for each predictor's dominance index. Therefore, without loss of generality, we assume that $d_i \sim N(\mu_i, \sigma^2)$.

The second stage of the hierarchical prior assumes that another normal distribution applies to the mean of the prior normal distribution, i.e., $\mu_i \sim N(\nu, \tau^2)$, and an Inverse-Gamma distribution is applied to the variance of the prior normal distribution, i.e., $\sigma^2 \sim IG(a, a)$. Here, ν, τ and a are predetermined hyper-parameters, which are chosen using vague information as follows. When d_i's vary between $-l$ and l, by applying the 3σ or empirical rule, the expected value of the standard deviation of d_i's is around $l/3$, namely the expected value of σ^2 is about $l^2/9$. On the other hand, σ^2 is assumed to have an $IG(a, a)$, which has a mean value of $a/(a-1)$. By letting $a/(a-1) = l^2/9$ and solving for a, we can get the hyper-parameter $a = l^2/(l^2 - 9)$. When values of d_i are in the range of $(-4, 4)$, values of $d_i - d_j$ are in the range of $(-8, 8)$. As a result, values of $\theta_{ij} = \exp(d_i - d_j)/(1 + \exp(d_i - d_j))$ vary from 0.0003 to 0.9997, which is sufficiently wide for practical purposes. Therefore, it is adequate to presume that d_i's range between ± 4. Thus, the value of a is about 2. In addition, it is also reasonable to assume the mean μ_i has a standard normal distribution with $\nu = 0$ and $\tau = 1$.

For the TSP link function, we presume that d_i's range from 0 to 1, and we do not have any preference towards any particular value. It is sufficient enough to apply the Uniform distribution on $(0, 1)$ as the prior distribution to each d_i.

3.2. Bayesian Computation

Under the square error loss function, the posterior mean is the Bayesian estimator. However, the closed form expressions of the posterior distribution of d_i's are not easy to achieve. With the advent of Markov chain Monte Carlo Method (MCMC), it has become possible to do the calculations on these much more complex models to obtain numerical results. In order to implement the MCMC procedure, it is necessary to have the full conditional posterior distributions, which are the conditional distributions of one parameter given all the other unknown parameters and data.

let $\mathbf{d} = (d_1, \cdots, d_p)$, $\mathbf{d}[-j] = (d_1, \cdots, d_{j-1}, d_{j+1}, \cdots, d_p)$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_p)$, $\boldsymbol{\mu}[-j] = (\mu_1, \cdots, \mu_{j-1}, \mu_{j+1}, \cdots, \mu_p)$, $\phi(\cdot|\mu, \sigma^2)$ be the probability density function of $N(\mu, \sigma^2)$, $H_1(\cdot)$ be the logit link function, and $H_2(\cdot)$ be the TSP link function. Also, let $(\cdot| \ )$ denote the conditional distribution and $[\cdot| \ ]$ denote the conditional density function.

Proposition 1 The full conditional posterior distributions of $(d_1, \cdots d_p, \mu_1, \cdots, \mu_p, \sigma^2)$ given data $\mathbf{X}$ under the Model I with unweighted likelihood and the logit link function are as follows.

- For $i = 1, \cdots, p$,

$$[d_i \mid \mathbf{d}_{[-i]}, \boldsymbol{\mu}, \sigma^2, \mathbf{X}] \propto \prod_{i<j} C_{2^{p-2}}^{w_{ij}} H_1(d_i - d_j)^{w_{ij}} (1 - H_1(d_i - d_j))^{2^{p-2} - w_{ij}} \phi(d_i|\mu_i, \sigma^2)$$

- For $i = 1, \cdots, p$, $(\mu_i \mid \mathbf{d}, \boldsymbol{\mu}[-i], \sigma^2, \mathbf{X}) \sim N\left(\frac{d_i \tau^2 + \sigma^2 \nu}{\tau^2 + \sigma^2}, \ \frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2}\right)$

- $(\sigma^2 \mid \mathbf{d}, \boldsymbol{\mu}, \mathbf{X}) \sim IG\left(\frac{p}{2} + a, \ \sum_{i=1}^k \frac{(d_i - \mu_i)^2}{2} + a\right)$,

Proposition 2 The full conditional posterior distributions of $(d_1, \cdots d_p, \mu_1, \cdots, \mu_p, \sigma^2)$ given data $\mathbf{X}$, under the Model II with weighted likelihood and the logit link function are as follows.

- For $i = 1, \cdots, p$,

$$[d_i \mid \mathbf{d}_{[-i]}, \boldsymbol{\mu}, \sigma^2, \mathbf{X}] \propto \sum_{k=0}^{p-2} \frac{C_{p-2}^k}{2^{p-2}} \prod_{i \neq j} \binom{C_{p-2}^k}{w_{ij \cdot k}} H_1(d_i - d_j)^{w_{ij \cdot k}} (1 - H_1(d_i - d_j))^{C_{p-2}^k - w_{ij \cdot k}} \phi(d_i|\mu_i, \sigma^2)$$

- For $i = 1, \cdots, p$, $(\mu_i \mid \mathbf{d}, \boldsymbol{\mu}[-i], \sigma^2, \mathbf{X}) \sim N\left(\frac{d_i \tau^2 + \sigma^2 \nu}{\tau^2 + \sigma^2}, \ \frac{\sigma^2 \tau^2}{\tau^2 + \sigma^2}\right)$

- $(\sigma^2 \mid \mathbf{d}, \boldsymbol{\mu}, \mathbf{X}) \sim IG\left(\frac{p}{2} + a, \ \sum_{i=1}^k \frac{(d_i - \mu_i)^2}{2} + a\right)$,

Proposition 3 The full conditional posterior distributions of $(d_1, \cdots d_p,)$ given data $\mathbf{X}$ under the Model III with unweighted likelihood and the TSP link function are as follows.

$$[d_i \mid \mathbf{d}_{[-i]}, \boldsymbol{\mu}, \sigma^2, \mathbf{X}] \propto \prod_{i<j} C_{2^{p-2}}^{w_{ij}} H_2(d_i - d_j)^{w_{ij}} (1 - H_2(d_i - d_j))^{2^{p-2} - w_{ij}},$$

for $i = 1, \cdots, p$,

Proposition 4 The full conditional posterior distributions of $(d_1, \cdots d_p)$ given data $\mathbf{X}$ under the Model IV with weighted likelihood and the TSP link function are as follows.

$$[d_i \mid \mathbf{d}_{[-i]}, \boldsymbol{\mu}, \sigma^2, \mathbf{X}] \propto \sum_{k=0}^{p-2} \frac{C_{p-2}^k}{2^{p-2}} \prod_{i \neq j} \binom{C_{p-2}^k}{w_{ij \cdot k}} H_2(d_i - d_j)^{w_{ij \cdot k}} (1 - H_2(d_i - d_j))^{C_{p-2}^k - w_{ij \cdot k}},$$

for $i = 1, \cdots, p$,

Comparing full conditional distributions, one can observe that full conditional distributions with the TSP link are in a more simpler form than those with the logit link. Therefore, the TSP link function simplifies the computation procedure. Vihola (2012) [21] introduces a new robust adaptive Metropolis algorithm estimating the shape of the target distribution and simultaneously concerning the acceptance rate. In this paper, the R package, adapMCMC which is based on Vihola (2012) [21], is utilised to implement the Bayesian computation.

The Bayesian dominance method provides prosperous information about the relative importance of predictors via posterior distributions of dominance indices d_i's. First of all, the posterior mean of the dominance index d_i can be applied as an overall measure of relative importance of the predictor X_i. Secondly, the posterior distribution of dominance probability θ_{ij} can be obtained by applying the corresponding link function to the MCMC chain, and the posterior mean of θ_{ij} can be applied as an estimate of the dominance probability to reveal the conclusion of paired comparisons among predictors. Moreover, the lower and upper 2.5^{th} percentiles of the posterior distributions of d_i and θ_{ij} can be used to construct 95% confidence intervals of d_i and θ_{ij}, respectively. Last but not least, the posterior probability of one particular order of dominance can be estimated by the proportion of steps occurred in the MCMC chain that is in the same order.

4. Simulations

To better compare the proposed Bayesian approaches in multiple regression settings, we conduct two sets of simulation studies based on different population correlation matrices. One hypothetical example with four predictors, which consists possible strong collinearity between X_1 and X_2 with the pairwise correlation coefficient of 0.80, and is used by Azen and Budescu (2003) [22]; the other example includes three predictors with a suppressor variable, X_2. The population correlation matrices are listed in Table 1, and a data of 1000 observations are generated in each set of simulations.

4.1. Hypothetical Example With Four Predictors

In this setting of correlation, the Dominance Analysis method concludes that the dominance order is $X_1 > X_4 > X_3 > X_2$. Applying the proposed Bayesian approach, a paired comparison of predictors based on the additional increase in R^2 is conducted among $2^{4-2} = 4$ possible subset regression models. Numbers of times that the variable X_i prevails over X_j under each and all possible size of subset models are shown in Table 2. Under baseline reference models with $k = 0$ and 2, each pair of predictors is only compared once; while under baseline reference models with $k = 1$, each pair of variables is compared twice. The predictor X_1 outweighs other predictors in all possible subset models. The predictor X_3 and X_4 outweigh X_2 in all possible subset models. Comparing X_3 and X_4, we find that X_4 outweighs X_3 when $k = 0$ and $k = 1$; X_3 outweighs X_4 when $k = 2$. Altogether, X_4 outweighs X_3 three out of four times. These results indicate that the order of dominance is $X_1 > X_4 > X_3 > X_2$.

The Bayesian point estimates and 95% confidence intervals of dominance indices d_i's in all four models are listed in Table 3. In all four models, the order of the dominance index is $d_1 > d_4 > d_3 > d_2$. The 95% confidence intervals of dominance indices overlap each other, which makes it challenging to conclude a definite dominance order of the predictors.

Estimates of dominance probabilities, θ_{ij}'s, are listed in Table 4, with $i < j$. For $i > j$, $\theta_{ij} = 1 - \theta_{ji}$. There exists consistency in all four models, such that θ_{12}, θ_{13} and θ_{14} are all greater than 0.5, which implies that X_1 is more likely to dominate the other predictors. Furthermore, the results that $\theta_{12} > \theta_{13} > \theta_{14}$ indicate that X_1 has more chance to dominate X_2 than X_3, and more chance to dominate X_3 than X_4. This finding, as well, suggests that the dominance order between X_2, X_3, and X_4 is $X_4 > X_3 > X_2$. Likewise, the results $0.5 > \theta_{24} > \theta_{23}$ and $\theta_{34} < 0.5$ support the same order of dominance as previous.

The posterior probability of a particular order of the dominance ranking is estimated by the proportion that the order is encountered in the MCMC steps. Altogether there are 24 possible ranking orders. The results show that, under all four models, the ranking order of $X_1 > X_4 > X_3 > X_2$ receives the largest posterior probability.

As a summary, results of this simulation study show that the proposed Bayesian approach determines the predictor with possible multicollinearity as the least important variable, and produce the same order of dominance as the Dominance Analysis.

4.2. Example with Classical Suppressor

In this setting of correlation, the predictor X_2 acts as a classical suppressor. However, the Dominance Analysis does not categorize X_2 as the least important variable, and claim that the dominance order is $X_1 > X_2 > X_3$. Paired

comparisons of predictors based on the additional increase in R^2 are conducted among $2^{3-2} = 2$ possible subset regression models. As displayed in Table 2, the predictor X_1 outweighs other predictors under all possible subset models; the predictor X_3 outweighs X_2 when $k = 0$; while X_2 outweighs X_3 when $k = 1$. The dominance order is not easily recognizable from the paired comparison results.

The Bayesian point estimates and 95% confidence intervals of dominance indices d_i's under all four models are listed in Table 3. In all four models, the resulting order of dominance is $d_1 > d_3 > d_2$. In this case, the proposed Bayesian approach successfully identifies the suppressor as the least important variable. However, the 95% confidence intervals of dominance indices overlap each other. It is challenging to uncover which predictor definitely dominates the other predictors.

In Table 4, estimates of dominance probabilities show that $\theta_{12} > \theta_{13} > 0.5$, which implies that X_1 is more likely to dominate the other predictors in all four models. Moreover, this result also indicates that X_1 has more chance to dominate X_2 and X_3. The result $\theta_{23} < 0.5$ reveals that X_2 is less important than X_3. The posterior probability of a particular order of the dominance ranking is estimated by the proportion that the order is encountered in the MCMC steps. The ordering rank of $X_1 > X_3 > X_2$ has the largest posterior probability in all four models.

As a summary, results of this simulation indicate that the proposed Bayesian approach has the better capability to categorize the suppressor as the least important variable than the Dominance Analysis methods.

5. Empirical Example

Neter et.al (2004) [23] provides an interesting data set about a national study of the educational achievements of eighth-grade students in mathematics. The response variable Y is the average math proficiency score from the 1990 National Assessment of Educational Progress for 37 states, the District of Columbia and the Virgin Islands. There are five predictor variables: the percentage of students living with both parents (X_1); percentage of students having three or more types of reading material at home (X_2); percentage of students reading more than 10 pages a day (X_3); percentage of students watching TV for six or more hours (X_4); percentage of students absented three days or more in the previous month (X_5).

The sample correlation matrix of the variables, displayed in Table 5, shows that the average math proficiency score is linearly correlated with all the predictors. The sample correlations of the predictors and the response variable have the order, from largest to smallest, as $X_4 > X_2 > X_1 > X_3 > X_5$. The Dominance Analysis also discovers the order of dominance as $X_4 > X_2 > X_1 > X_3 > X_5$. However, there exists possible multicollinearity among the predictors in that the sample correlation between X_1 and X_4 is -0.831, and the sample correlation between X_3 and X_4 is -0.79. The results with X_4 being the most important predictor is somewhat questionable. Paired comparisons of predictors based on the additional increase in R^2 are conducted among $2^{5-2} = 8$ possible subset regression models. The total numbers of times that the variable X_i prevails over X_j in all possible subset reference models are presented in Table 6. The predictors X_2 and X_4 completely dominate the other predictors in all possible subset models; and X_2 outweighs X_4 six out of eight times. The predictor X_1 and X_3 both outweigh X_5 six out eight times; the predictor X_1 and X_3 are even. These results suggest a different order of dominance as $X_2 > X_4 > X_1 > X_3 > X_5$.

The Bayesian point estimates and 95% confidence intervals of dominance indices d_i's under four different models are listed in Table 7. The resulting order rank of dominance indices from all four models provide the same dominance order as $X_2 > X_4 > X_1 > X_3 > X_5$. The lower and upper 2.5^{th} percentiles of the MCMC chain are applied to construct the 95% confidence intervals as shown in Table 7. Because the 95% confidence intervals of dominance indices overlap each other, it is challenging to determine which predictor definitely dominates the other predictors.

Bayesian approach provides a probabilistic solution to this difficulty. By applying the link functions to the dominance indices within the MCMC, we achieve the estimate of dominance probabilities, θ_{ij}'s, which are listed in Table 8, with $i < j$. While, for $i > j$, $\theta_{ij} = 1 - \theta_{ji}$. The results also support the order of dominance in Table 7. For example, θ_{23}, θ_{25}, and θ_{45} are very large, which implies that X_2 and X_4 have large chance dominating X_3 and X_5. The results that θ_{12}, θ_{14} and θ_{34} are very small also suggest that X_2 and X_4 have large chance dominating X_3. Furthermore, $\theta_{24} > 0.5$ implies that X_2 has more change to dominate X_4. Comparing the probability θ_{13}, θ_{15}, and θ_{35}, we can also discover the dominance order of X_1, X_3 and X_5 is $X_1 > X_3 > X_5$.

The posterior probability of a particular order of the dominance ranking is estimated by the proportion of this order occurring in the MCMC steps. Among the 120 possible of ranking orders, the order ranking $X_2 > X_4 > X_1 > X_3 > X_5$ receives the highest probability under all four models. Once again, the results support previous findings in dominance indexes and probabilities.

As a summary, both sample correlation between the predictors and response variables and the Dominance Analysis discover the order of dominance as $X_4 > X_2 > X_1 > X_3 > X_5$ while the proposed Bayesian approaches

consistently identify the dominance order as $X_2 > X_4 > X_1 > X_3 > X_5$ under both weighted and unweighted likelihood functions with both logit and TSP link functions. We have more confidence on the latter conclusion because it is well supported by the sample correlation matrix, the dominance probability, and the posterior order of dominance. Moreover, from the practice perspective, the percentage of students having three or more types of reading material at home (X_2) is more academic relates, hence, should have more influence on the math achievement score (Y) than the percentage of students watching TV for six or more hours (X_4).

6.   Summary and Concluding Remarks

Relative importance analyses permit a greater understanding of the particular role played by variables in a multiple regression equation. Crucially, these analyses can reveal the underlying impact of a particular predictor more accurately than standardized regression coefficients or simple correlations. This paper extends the current research practices of measuring the relative importance of predictors in linear regression models by introducing an original method built on weighted and un-weighted paired comparison models with two different link functions in the Bayesian framework.

The advantage of the Bayesian approach is that it allows the use of genuine prior information in addition to the information that is available in the observed data to produce better results. In general, Bayesian methods provide a better approximation to the level of uncertainty than other approaches which use only information provided by the model and the observed data. In addition to providing useful statistics, such as, the mean and percentiles of the posterior distribution of the unknown parameters, Bayesian methods give more reliable results for small samples [24, 25, 26].

More specifically, the Bayesian approach presented in this paper offers several advantages over the current methods in determining the relative importance of predictors in a regression model. First, the general dominance weight of the Dominance Analysis is biased towards the results of subset models with small size. The proposed approach resolves this problem by modelling the number of times a predictor outranking the others with the paired comparison model. This probabilistic model based approach provides more comprehensive inference about the population relative dominance ability of predictors. Secondly, the Bayesian approach provides more information about the relative importance of the predictor by making straightforward statements about the dominance ability of the predictors, the dominance probability of each possible pair of predictors, and the probability of each possible order of dominance.

The advantages of applying the TSP link function over the logit link function come from both practical aspect and computational aspect. The TSP link function is more flexible in modelling the results of paired comparisons using indices on the $0 - 1$ scale. In managerial terms saying that one variable dominates another on a scale of 0 to 1 communicates more to the practitioners than those range from $-\infty$ to $+\infty$. From computational point of view, the TSP link function simplifies the computational procedure with a simple uniform prior distribution. The MCMC chain produced by the TSP link function converges faster and is less correlated than those by the logit link function. Figure 2 and Figure 3, which present the trace plots of the MCMC chains of the four models using data in the empirical example, show that the MCMC chains produced by the TSP link function mixed better and converge faster than those by the logit link function. Moreover, autocorrelation functions of the four models in the empirical example, as presenting in Figure 4 and Figure 5, show that the MCMC chains produced by the TSP link function die down faster than those from the logit link function, which means the former ones are less correlated and more stationary than the latter ones.

Relative importance analyses are also applicable to situations commonly confronted by organizational scholars where the criteria may not meet the distributional assumptions of ordinary least squares (OLS) regression, such as predicting binary criteria like turnover, promotion decisions, or training success. Although both Dominance Analysis and Bayesian approach are developed for use with OLS regression, Azen and Traxel (2009) [27] presented modifications of these respective analyses to handle categorical criterion variables that would typically be analysed using logistic regression. Tonidandel and LeBreton (2010) [28] developed an application of relative weight analysis to logistic regression. Thus, we will continue our work about questions of the relative contribution of each of the variables in terms of predicting the categorical criterion, and how the proposed method can be examined in this context as well.

One of the central questions in a multivariate analysis of variance (MANOVA) considers identifying the dependent variables that are driving the significant multivariate F-test. Unfortunately, the correlations among the various dependent variables often make it difficult to accurately identify the role being played by the various dependent variables. The Bayesian approach of relative importance analyses may also be a useful supplement to analyse other than multiple regression, such as MANOVA as well.

Table 1: Population Correlation Matrix for Simulation

Example with Four Predictors

	Y	X_1	X_2	X_3	X_4
Y	1.0	0.6	0.3	0.4	0.5
X_1	0.6	1.0	0.8	0.1	0.3
X_2	0.3	0.8	1.0	0.1	0.1
X_3	0.4	0.1	0.1	1.0	0.2
X_4	0.5	0.3	0.1	0.2	1.0

Example with Classical Suppressor

	Y	X_1	X_2	X_3
Y	1.0	0.3	0.0	0.25
X_1	0.3	1.0	0.75	0.55
X_2	0.0	0.75	1.0	0.15
X_3	0.25	0.55	0.15	1.0

Table 2: Paired Comparisons of Predictors in Simulation Studies

Example With Four Predictors

Null k=0

	X_1	X_2	X_3	X_4
X_1	0	1	1	1
X_2	0	0	0	0
X_3	0	1	0	0
X_4	0	1	1	0

k=1

	X_1	X_2	X_3	X_4
X_1	0	2	2	2
X_2	0	0	0	0
X_3	0	2	0	0
X_4	0	2	2	0

k=2

	X_1	X_2	X_3	X_4
X_1	0	1	1	1
X_2	0	0	0	0
X_3	0	1	0	1
X_4	0	1	0	0

Total

	X_1	X_2	X_3	X_4
X_1	0	4	4	4
X_2	0	0	0	0
X_3	0	4	0	1
X_4	0	4	3	0

Example with Classical Suppressor

Null k=0

	X_1	X_2	X_3
X_1	0	1	1
X_2	0	0	0
X_3	0	1	0

k=1

	X_1	X_2	X_3
X_1	0	1	1
X_2	0	0	1
X_3	0	0	0

Total

	X_1	X_2	X_3
X_1	0	2	2
X_2	0	0	1
X_3	0	1	0

Table 3: Summary of Posterior Statistics of Dominance Index of Predictors in Simulations

	Bayesian Analysis			
	Logit Link		TSP Link	
	Un-weighted Model I	Weighted Model II	Un-weighted Model III	Weighted Model IV
	Example with Four Predictors			
d_1	2.8928 (0.7293 , 6.1262)	1.7645 (-0.6591 , 5.2106)	0.8885 (0.6351 , 0.9959)	0.7608 (0.2835 , 0.9929)
d_2	-2.5517 (-6.7476 , -0.5002)	-1.7865 (-5.2678 , 0.7165)	0.1112 (0.0041 , 0.3593)	0.2325 (0.0091 , 0.6967)
d_3	-0.2688 (-2.6886 , 1.8047)	-0.1801 (-3.0025 , 2.6071)	0.4136 (0.0711 , 0.7944)	0.4838 (0.0335 , 0.9590)
d_4	0.5763 (-1.5670 , 2.7098)	0.0534 (-2.7240 , 2.8409)	0.5825 (0.1834 , 0.9471)	0.5241 (0.0422 , 0.9709)
	Example with Classical Suppressor			
d_1	1.3835 (-0.8060 , 4.0337)	1.1571 (-1.2259 , 4.0560)	0.7640 (0.3048 , 0.9911)	0.6924 (0.1271 , 0.9909)
d_2	-0.8441 (-3.3536 , 1.4278)	-0.5876 (-3.1505 , 1.9929)	0.3192 (0.0157 , 0.8109)	0.4056 (0.0113 , 0.9522)
d_3	-0.8383 (-3.2240 , 1.4271)	-0.5417 (-3.3472 , 2.1919)	0.3342 (0.0152 , 0.8021)	0.4222 (0.0159 , 0.9502)

Table 4: Summary of Posterior Statistics of Dominance Probability of Predictors in Simulation

	Logit Link		TSP Link	
	Un-weighted	Weighted	Un-weighted	Weighted
	Model I	Model II	Model III	Model IV
Example With Four Predictors				
$\theta_{1,2}$	0.9864	0.9166	0.9670	0.8566
$\theta_{1,3}$	0.9271	0.7818	0.8429	0.7006
$\theta_{1,4}$	0.8616	0.7642	0.7398	0.6727
$\theta_{2,3}$	0.1457	0.2533	0.2615	0.3161
$\theta_{2,4}$	0.0782	0.2279	0.1597	0.2895
$\theta_{3,4}$	0.3291	0.4623	0.3624	0.4694
Example with Classical Suppressor				
$\theta_{1,2}$	0.8433	0.7614	0.8158	0.7039
$\theta_{1,3}$	0.8409	0.7515	0.8071	0.6912
$\theta_{2,3}$	0.4983	0.4902	0.4873	0.4901

Table 5: Sample Correlation Matrix in Example

	Y	X_1	X_2	X_3	X_4	X_5
Y	1.000	0.740	0.744	0.714	-0.872	-0.524
X_1	0.740	1.000	0.392	0.691	-0.831	-0.599
X_2	0.744	0.392	1.000	0.374	-0.592	-0.471
X_3	0.714	0.691	0.374	1.000	-0.790	-0.390
X_4	-0.872	-0.831	-0.592	-0.790	1.000	0.556
X_5	-0.524	-0.599	-0.471	-0.390	0.556	1.000

Table 6: Total of Paired Comparisons of Predictors in Empirical Example

	X_1	X_2	X_3	X_4	X_5
X_1	0	0	4	0	6
X_2	8	0	8	6	8
X_3	4	0	0	0	6
X_4	8	2	8	0	8
X_5	2	0	2	0	0

Table 7: Summary of Posterior Statistics of Dominance Index of Predictors in Example

	Logit Link		TSP Link	
	Un-weighted Model I	Weighted Model II	Un-weighted Model III	Weighted Model IV
d_1	-1.2226 (-2.8667 , 0.4566)	-0.0075 (-2.2691 , 2.2144)	0.2000 (0.0040 , 0.4663)	0.4769 (0.0385 , 0.9551)
d_2	2.7012 (0.9236 , 4.6888)	2.0872 (0.2512 , 4.5215)	0.9280 (0.7565 , 0.9984)	0.7482 (0.2388 , 0.9919)
d_3	-1.2699 (-3.0816 , 0.2826)	-0.5380 (-2.4782 , 1.3904)	0.1915 (0.0055 , 0.4487)	0.3090 (0.0120 , 0.8292)
d_4	1.7785 (0.1793 , 3.7052)	1.0884 (-0.8847 , 3.4413)	0.8445 (0.5796 , 0.9939)	0.7375 (0.2603 , 0.9952)
d_5	-2.2999 (-4.1318 , -0.7340)	-1.6131 (-4.1911 , 0.5371)	0.0729 (0.0016 , 0.2443)	0.1959 (0.0065 , 0.5936)

Table 8: Summary of Posterior Statistics of Dominance Probability of Predictors in Example

	Logit Link		TSP Link	
	Un-weighted	Weighted	Un-weighted	Weighted
	Model I	Model II	Model III	Model IV
$\theta_{1,2}$	0.0286	0.1663	0.0452	0.3028
$\theta_{1,3}$	0.5112	0.5985	0.5077	0.6254
$\theta_{1,4}$	0.0630	0.3162	0.0742	0.3095
$\theta_{1,5}$	0.7296	0.7663	0.6127	0.7072
$\theta_{2,3}$	0.9719	0.8721	0.9571	0.8067
$\theta_{2,4}$	0.6942	0.6817	0.5747	0.5092
$\theta_{2,5}$	0.9890	0.9448	0.9857	0.8686
$\theta_{3,4}$	0.0616	0.2233	0.0702	0.1980
$\theta_{3,5}$	0.7217	0.6930	0.6053	0.5878
$\theta_{4,5}$	0.9755	0.8835	0.9664	0.8672
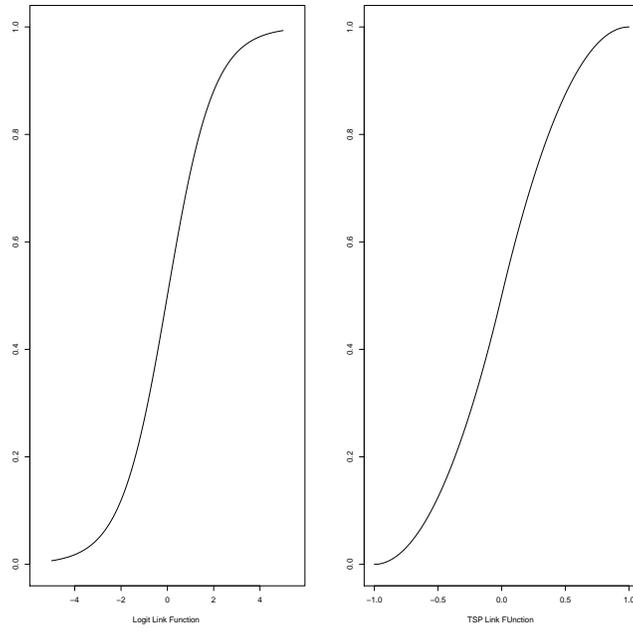
Figure 1: Logit and TSP Link Functions

(a) with Unweighted Likelihood

(b) with Weighted Likelihood
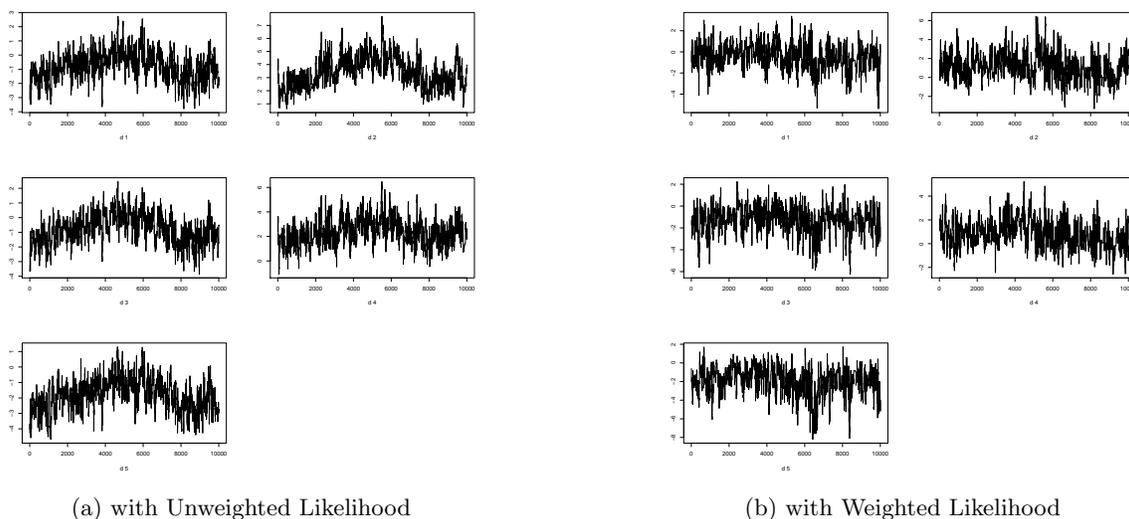
Figure 2: The Trace-Plot of Dominance Index with Logit Link Functions

(a) with Unweighted Likelihood

(b) with Weighted Likelihood
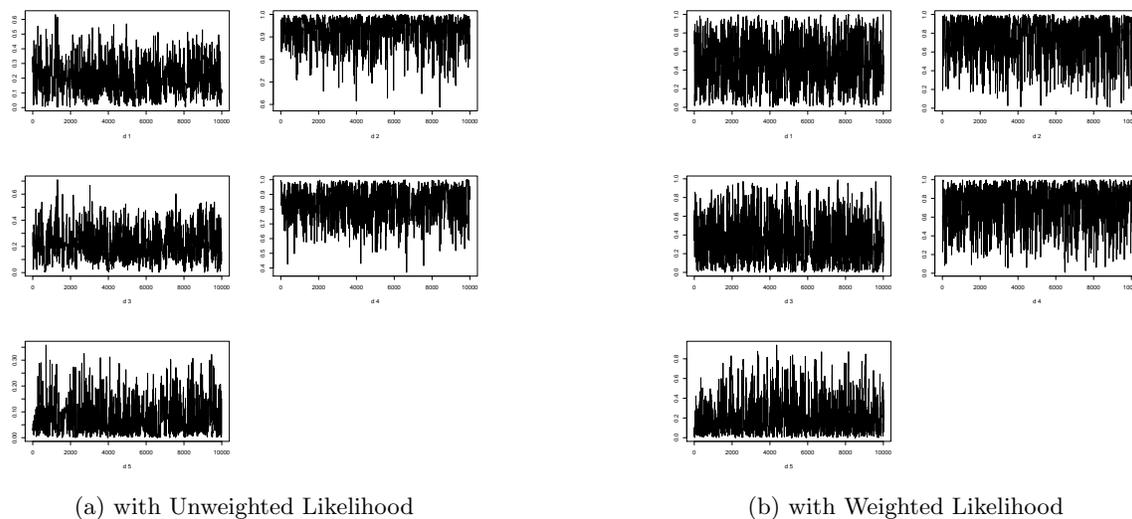
Figure 3: The Trace-Plot of Dominance Index with TSP Link Functions
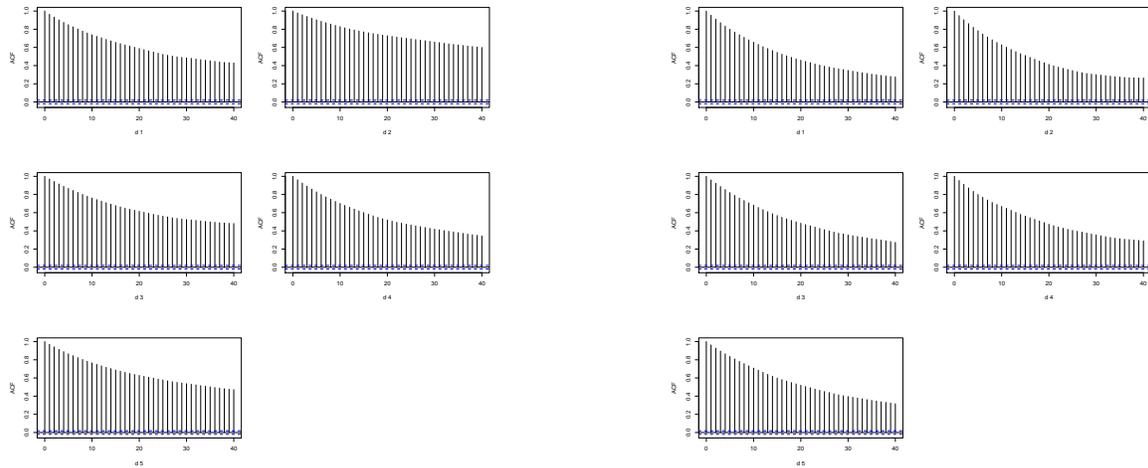
(a) with Unweighted Likelihood   (b) with Weighted Likelihood

Figure 4: The Auto Correlation Functions of Dominance Index with Logit Link Functions
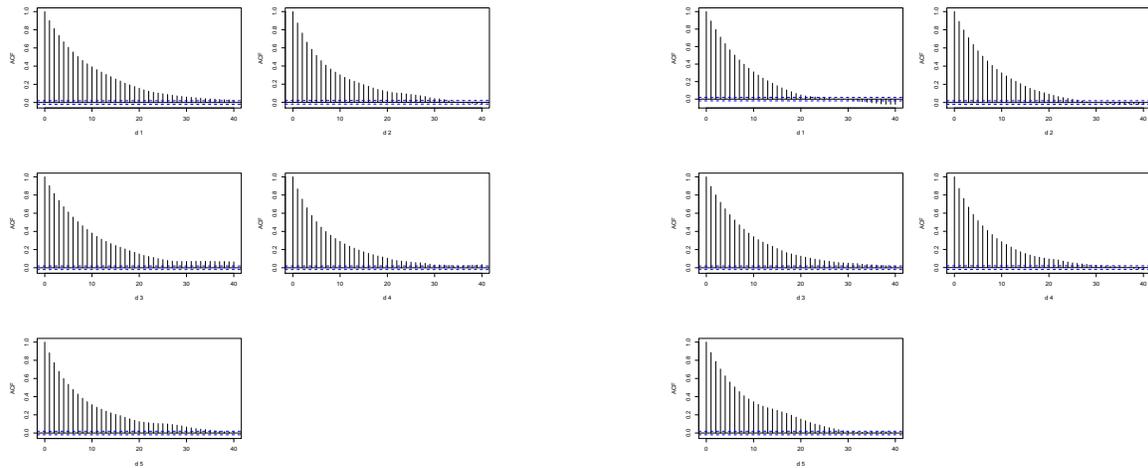
(a) with Unweighted Likelihood   (b) with Weighted Likelihood

Figure 5: The Auto Correlation Functions of Dominance Index with TSP Link Functions

References

[1] Jeff W Johnson and James M LeBreton. History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7(3):238–257, 2004.

[2] Richard B Darlington. Multiple regression in psychological research and practice. *Psychological bulletin*, 69(3):161, 1968.

[3] David V Budescu. Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114(3):542, 1993.

[4] Xiaoyin Wang and Dongqing Yao. Bayesian inference of predictor's relative importance in multiple regression. *Advances and Applications in Statistics*, 41(1):69, 2014.

[5] Pierre Lahiri et al. On the impact of bootstrap in survey sampling and small-area estimation. *Statistical Science*, 18(2):199–210, 2003.

[6] Razia Azen, David V Budescu, and Benjamin Reiser. Criticality of predictors in multiple regression. *British Journal of Mathematical and Statistical Psychology*, 54(2):201–226, 2001.

[7] Jeff W Johnson. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1):1–19, 2000.

[8] X Wang, P Duverger, and HS Bansal. Bayesian inference of predictors relative importance in linear regression model using dominance hierarchies. *International Journal of Pure and Applied Mathematics*, 88(3):321–339, 2013.

[9] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs the method of paired comparisons. *Biometrika*, 39(3-4):324–345, 1952.

[10] J Rene Van Dorp and Samuel Kotz. The standard two-sided power distribution and its properties: with applications in financial engineering. *The American Statistician*, 56(2):90–99, 2002.

[11] Roger R Davidson and Daniel L Solomon. A bayesian approach to paired comparison experimentation. *Biometrika*, 60(3):477–487, 1973.

[12] Tom Leonard. An alternative bayesian approach to the bradley-terry model for paired comparisons. *Biometrics*, pages 121–132, 1977.

[13] William James and Charles Stein. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 361–379, 1961.

[14] Bradley Efron and Carl Morris. Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319, 1975.

[15] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, 1983.

[16] Ita GG Kreft, Ita Kreft, and Jan de Leeuw. *Introducing multilevel modeling*. Sage, 1998.

[17] Roel Bosker and Tom Snijders. Multilevel analysis: An introduction to basic and advanced multilevel modeling. *New York*, 1999.

[18] Bradley P Carlin and Thomas A Louis. *Bayes and empirical Bayes methods for data analysis*. Chapman and Hall/CRC, 2 edition, 2001.

[19] Stephen W Raudenbush and Anthony S Bryk. *Hierarchical linear models: Applications and data analysis methods*, volume 1. Sage, 2002.

[20] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. 2 edition, 2003.

[21] Matti Vihola. Robust adaptive metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, 22(5):997–1008, 2012.

[22] Razia Azen and David V Budescu. The dominance analysis approach for comparing predictors in multiple regression. *Psychological methods*, 8(2):129, 2003.

[23] Michael H Kutner, Chris Nachtsheim, and John Neter. *Applied linear regression models*. McGraw-Hill/Irwin, 2004.

[24] David B Dunson. Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, pages 355–366, 2000.

[25] Sik-Yum Lee and Xin-Yuan Song. Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal categorical data. *British Journal of Mathematical and Statistical Psychology*, 57(1):131–150, 2004.

[26] Richard Scheines, Herbert Hoijtink, and Anne Boomsma. Bayesian estimation and testing of structural equation models. *Psychometrika*, 64(1):37–52, 1999.

[27] Razia Azen and Nicole Traxel. Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34(3):319–347, 2009.

[28] Scott Tonidandel and James M LeBreton. Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods*, 13(4):767–781, 2010.