

BAYESIAN ESTIMATION OF CIRCUMPLEX MODELS SUBJECT TO PRIOR THEORY CONSTRAINTS AND SCALE-USAGE BIAS

PETER LENK, MICHEL WEDEL

UNIVERSITY OF MICHIGAN

ULF BÖCKENHOLT

MCGILL UNIVERSITY

This paper presents a hierarchical Bayes circumplex model for ordinal ratings data. The circumplex model was proposed to represent the circular ordering of items in psychological testing by imposing inequalities on the correlations of the items. We provide a specification of the circumplex, propose identifying constraints and conjugate priors for the angular parameters, and accommodate theory-driven constraints in the form of inequalities. We investigate the performance of the proposed MCMC algorithm and apply the model to the analysis of value priorities data obtained from a representative sample of Dutch citizens.

Key words: Bayesian inference, circumplex correlations, inequality constraints, latent variables, ordinal scales.

1. Introduction

A classical finding in psychometrics is that similarity judgments of different colors can be represented in a two-dimensional space in the form of Newton's color circle (Shepard, 1962a, 1962b). Based on this work, similar circular representations proved useful for describing variations among experiences or judgments in a wide range of psychological and related disciplines. For example, affective states are commonly depicted by a circular structure based on the dimensions of valence and arousal (Russell & Carroll, 1999). Numerous other applications can be found in personality and social psychology (Lippe, 1995; Plutchik & Conte, 1997). The circular ordering of the responses implies that the elements of the corresponding correlation matrix follow a so-called circumplex structure with correlations first decreasing but then increasing as one moves from the main diagonal. Guttman (1954) and Anderson (1960) suggested stochastic processes on the perimeter of the circle that produce positive correlations obeying the circumplex structure with, respectively, moving average and Markov properties. Models that allow for negative correlations were developed by Cudeck (1986) and by Wiggins, Steiger, and Gaelick (1981). Browne (1992) proposed an extension of Anderson's (1960) model that allows for negative correlations.

We extend the work by Browne (1992) as follows. First, we introduce a Bayesian specification of the circumplex for ratings data and present identifying constraints and conjugate priors for the angular parameters. Second, we specify inequality constraints on blocks of variables in the circumplex as defined by psychological theories. Third, we accommodate idiosyncratic response-scale usage by persons (see, e.g., Rossi, Gilula, & Allenby, 2001) that, if not accounted for, may substantially distort the derived circumplex. Since surveys are often burdened with item nonresponse, we also capitalize on the Markov Chain Monte Carlo (MCMC) estimation algorithm to impute missing values. As a result, our approach facilitates powerful tests of

We wish to thank Michael Browne and two anonymous reviewers for their comments. The data for this study were collected as part of the project AIR2-CT94-1066, sponsored by the European Commission.

Requests for reprints should be sent to Peter Lenk, The University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234, USA. E-mail: plenk@umich.edu

psychological theories based on a circumplex structure and controls for a number of nuisance effects.

The next section describes the proposed model and discusses estimation and inference issues. In section 3, the model is fitted to value ratings collected in The Netherlands, and it is investigated whether the underlying data structure is consistent with the prominent circumplex value theory of Schwartz and Bilsky (1987, 1990). Section 4 summarizes the paper.

2. Model

Subject i responds $W_{i,j}$ to item j on an rating scale with H ordered categories. Such scales are very common in psychology and other social science applications. We assume that the observed response is driven by a latent variable $Y_{i,j}$ falling between two cutpoints:

$$w_{i,j} = k \quad \text{iff} \quad c_{i,k-1} < y_{i,j} < c_{i,k} \quad \text{for} \quad k = 1, \dots, H,$$

where the cutpoints $\{c_{i,k}\}$ may vary from person to person. We consider it important to estimate person-specific cutpoints, since response scale bias has been reported to be highly idiosyncratic (Rossi et al., 2001). The cutpoints are ordered: $c_{i,k-1} < c_{i,k}$, and the first two and last two cutpoints are fixed without loss of generality: $c_{i,0} = -\infty, c_{i,1} = -1, c_{i,H-1} = 1, c_{i,H} = \infty$. The probability of the ordinal response is

$$Pr(W_{i,j} = k) = \int_{c_{i,k-1}}^{c_{i,k}} f(y_{i,j}) dy \quad \text{for} \quad k = 1, \dots, H, \quad (1)$$

where f is the density of $Y_{i,j}$.

Browne (1992) proposed using trigonometric series to model the circumplex correlations between items and developed a corresponding factor analytical model. We similarly specify a random effects model for the latent response variable $Y_{i,j}$ to describe individual differences for person i and item j :

$$Y_{i,j} = \mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j) + \epsilon_{i,j} \\ \text{for } i = 1, \dots, n \quad \text{and} \quad j = 1, \dots, J; \quad \theta_1 = 0; \quad \text{and} \quad 0 \leq \theta_j < 2\pi. \quad (2)$$

The mean, latent response for item j is μ_j ; ϕ_i is a subject-specific random effect that captures scale-usage effects, and $\alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j)$ is a subject by item interaction term that provides circumplex correlations. The error terms, $\{\epsilon_{i,j}\}$, are mutually independent, normally distributed random error terms with zero mean, and item-specific variances: $\text{var}(\epsilon_{i,j}) = \sigma_j^2$.

The model can be viewed as a three-factor model where $\phi_i, \alpha_i,$ and β_i are subject-specific factor scores. The first factor score ϕ_i is a random effect that takes into account subject-specific scale-usage effects. These effects are artifacts of the measurement system and usually do not have substantive, field-dependent implications: they reflect that subjects use systematically different parts of the ordinal measurement scale. Respondents with a positive ϕ_i tend to use the upper end of the rating scale, and respondents with a negative ϕ_i tend to use the lower end. We will see in the application that ignoring scale-usage effects can severely distort the estimated circumplex.

The other two factor scores α_i and β_i are individual-level random coefficients that have substantive meaning for the psychological phenomenon under investigation: they represent bipolar latent constructs. Their item-specific loadings, $\sin(\theta_j)$ and $\cos(\theta_j)$, are constrained to the unit circle; thus, they are expressed in polar coordinates. With the appropriate assumptions about these random effects and the constrained loadings, interitem correlations, after adjusting for scale-usage bias, have a circumplex structure.

We assume that the random effects, $(\phi_i, \alpha_i, \beta_i)$, are mutually independent and normally distributed with zero means and the following variances: $\text{var}(\phi_i) = \lambda^2$; $\text{var}(\alpha_i) = \tau_a^2$; and $\text{var}(\beta_i) = \tau_b^2$. Circumplex correlations are obtained when $\tau_a = \tau_b = \tau$. If these variances are unequal, then one can reparametrize the subject by item interactions as $\alpha_i \tau_a \sin(\theta_j) + \beta_i \tau_b \cos(\theta_j)$ were α_i and β_j are factor scores with mean 0 and variance 1. Then the loadings $\tau_a \sin(\theta_j)$ and $\tau_b \cos(\theta_j)$ are constrained to the ellipse. In the empirical application, we will compare the circumplex model to this more general one.

After integrating out the random effects, the variance and covariances of the latent variables for circumplex correlations ($\tau_a = \tau_b = \tau$) for the items conditional on the angles are

$$\begin{aligned} \text{var}(Y_{i,j}) &= \lambda^2 + \tau^2[\sin(\theta_j)^2 + \cos(\theta_j)^2] + \sigma_j^2 \\ &= \lambda^2 + \tau^2 + \sigma_j^2, \end{aligned} \quad (3)$$

$$\begin{aligned} \text{cov}(Y_{i,j}, Y_{i,k}) &= \lambda^2 + \tau^2[\sin(\theta_j) \sin(\theta_k) + \cos(\theta_j) \cos(\theta_k)] \\ &= \lambda^2 + \tau^2 \cos(\theta_j - \theta_k). \end{aligned} \quad (4)$$

This covariance is a special case of Browne's (1992) approximation using first-order, trigonometric polynomials. In the classical analysis of random-effects models, the variances and covariances in equations (3) and (4) determine the error covariance matrix in the log-likelihood function. In Bayesian inference, the random effects $(\phi_i, \alpha_i, \beta_i)$ are frequently treated as unknown parameters that are estimable: they are not just nuisance parameters.

2.1. Identifiability

The part of the circumplex correlation function that depends on item angles, $\tau^2 \cos(\theta_j - \theta_k)$ from equation (4), depends only on the differences in the angles so that the origin is arbitrary. Thus, we fix θ_1 to 0, but this alone does not identify the model, which can be seen as follows. Define another set of angles as $\psi_1 = 0$ and $\psi_j = 2\pi - \theta_j$ for $j \geq 2$. Because $\sin(2\pi - \theta) = -\sin(\theta)$ and $\cos(2\pi - \theta) = \cos(\theta)$, the likelihoods $L[\alpha, \beta, \theta]$ and $L[-\alpha, \beta, \psi]$ are equal. Consequently, in addition to establishing the origin with $\theta_1 = 0$, we also need to establish the ‘‘positive’’ directions for the angles.

The positive direction is implicitly identified when using block constraints, as defined in the next section, where blocks of angles are consecutively ordered. Without block constraints, we identify the positive directions by imposing the condition that one of the remaining angles, say θ_2 , is between 0 and π . In theory, one can choose any angle, other than θ_1 , to constrain the model. In practice, if the item selected, say item 2, has an angle close to zero (high positive correlation with item 1), then the model is ‘‘nearly’’ unidentified, which can be seen by defining $\psi_j = 2\pi - \theta_j + 2\theta_2$ for $j > 1$. Then $\cos(\psi_j - \psi_k) = \cos(\theta_j - \theta_k)$ for j and $k > 1$, and $\cos(\psi_j - 0) = \cos(\theta_j - 2\theta_2) \approx \cos(\theta_j)$ if θ_2 is close to zero. Thus, we recommend selecting an item for the 0 to π constraint that does not have a large positive correlation with item 1.

We selected the cosine function in the covariance terms to represent the circumplex, but other functions (see Browne, 1992) could be used as well. A function satisfies the circumplex properties for correlations if it is even, continuous, monotonically decreasing on $(0, \pi)$, monotonically increasing on $(\pi, 2\pi)$ with maxima of 1 at 0 and 2π , and a minimum of -1 at π . We choose the cosine without loss of generality, however, since the angles and trigonometric function are simultaneously unidentified for a finite set of items. If another function f has the circumplex properties, then it is possible to define a new set of angles ϕ such that (ϕ, f) and (θ, \cos) result in the same covariances and likelihoods for a finite set of items. However, strictly speaking, the invariance only holds with respect to the likelihood. For the posterior distributions of the

parameters for the two models defined through (ϕ, f) and (θ, \cos) to be equivalent, the prior for θ would have to be transformed into an equivalent prior for ϕ . In practice, since the prior specification is often chosen as a compromise between realism and convenience, the transformed prior for ϕ would rarely match a preferred direct specification of it. For example, an uninformative prior for θ may not result in an uninformative prior for ϕ .¹

2.2. Block Constraints on the Angles

Substantive theory often postulates that subsets of angles in the circumplex, characterizing the items in a certain domain, are less than or greater than other subsets of angles in other domains, thus imposing blocks of constraints on the directions $\{\theta_j\}$. In psychological theories on personality and value priorities, such domains, consisting of groups of substantively homogeneous items, are often distinguished. Suppose that there are K blocks of constraints, and let \mathcal{B}_k be the set of indices for the k th block. We are interested in specifying prior constraints on the order of these blocks of angles. Without loss of generality we assume the blocks are ordered from $1, \dots, K$ with $\theta_1 = 0$ belonging to the first block.

Expressing the block constraints is fairly straightforward, except for the fact that angles in the first block can be on both sides of the origin. For blocks 2 to K , define the minimum and maximum angles:

$$\underline{\mathcal{B}}_k = \min\{\theta_j : j \in \mathcal{B}_k\} \quad \text{and} \quad \overline{\mathcal{B}}_k = \max\{\theta_j : j \in \mathcal{B}_k\} \quad \text{for } k = 2, \dots, K.$$

The “minimum” and “maximum” angles for the first block require some care because angles in the first block can be on both sides of the zero value:

$$\underline{\mathcal{B}}_1 = \begin{cases} 2\pi & \text{if } \underline{\mathcal{B}}_K > \max\{\theta_j : j \in \mathcal{B}_1\}, \\ \min\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j > \overline{\mathcal{B}}_K\} & \text{if } \underline{\mathcal{B}}_K < \max\{\theta_j : j \in \mathcal{B}_1\}, \end{cases}$$

$$\overline{\mathcal{B}}_1 = \max\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j < \underline{\mathcal{B}}_2\}.$$

That is, $\underline{\mathcal{B}}_1$ is the smallest angle in the first block that is larger than the angles in the K th block. If no angle satisfies this requirement, it is defined to be 2π . $\overline{\mathcal{B}}_1$ is the largest angle in the first block that is smaller than the angles in the second block. With this nonstandard definition of “minimum” and “maximum” for the first block, we obtain the ordering

$$0 \leq \overline{\mathcal{B}}_1 < \underline{\mathcal{B}}_2 < \overline{\mathcal{B}}_2 < \dots < \underline{\mathcal{B}}_K < \overline{\mathcal{B}}_K < \underline{\mathcal{B}}_1 \leq 2\pi.$$

These constraints are additional to $\theta_1 = 0$. In the presence of block constraints, one does not need to constrain the angle of an item to be between 0 and π in order to establish the positive direction: it is implicit in the blocking structure.

2.3. Distributions

Introducing matrix notation simplifies defining the circumplex model in equation (2) and its distributions. The $n \times J$ matrix for the latent scores $\{y_{i,j}\}$ (subject i and item j) from the cutpoint model is

$$Y = \begin{bmatrix} y_{1,1} & \cdots & y_{1,J} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,J} \end{bmatrix} = \begin{bmatrix} \underline{Y}'_{1,\bullet} \\ \vdots \\ \underline{Y}'_{n,\bullet} \end{bmatrix} = [\underline{Y}_{\bullet,1} \cdots \underline{Y}_{\bullet,J}],$$

¹We thank one anonymous reviewer for pointing this out to us.

where

$\underline{Y}_{i,\bullet} = J$ vector of latent item scores for subject i ,

$\underline{Y}_{\bullet,j} = n$ vector of latent subject scores for item j .

The $n \times J$ matrix of error terms is

$$E = \begin{bmatrix} \epsilon_{1,1} & \cdots & \epsilon_{1,J} \\ \vdots & \ddots & \vdots \\ \epsilon_{n,1} & \cdots & \epsilon_{n,J} \end{bmatrix} = \begin{bmatrix} \underline{\epsilon}'_1 \\ \vdots \\ \underline{\epsilon}'_n \end{bmatrix}.$$

The item specific J -vectors for means and angles are

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_J \end{bmatrix}, \quad \underline{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_J \end{bmatrix}, \quad \underline{X}_S = \begin{bmatrix} \sin(\theta_1) \\ \vdots \\ \sin(\theta_J) \end{bmatrix}, \quad \text{and} \quad \underline{X}_C = \begin{bmatrix} \cos(\theta_1) \\ \vdots \\ \cos(\theta_J) \end{bmatrix}.$$

The $J \times 2$ factor loading matrix is represented as

$$X = [\underline{X}_S \quad \underline{X}_C].$$

The $J \times J$ diagonal matrix of error variances is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_J^2 \end{bmatrix}.$$

Factor scores and scale-usage effects are collected into n -vectors:

$$\underline{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \text{and} \quad \underline{\phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_n \end{bmatrix}.$$

With these definitions, equation (2) becomes

$$Y = \underline{\phi}'_J + \underline{1}_n \underline{\mu}' + \underline{\alpha} \underline{X}'_S + \underline{\beta} \underline{X}'_C + E,$$

where $\underline{1}_K$ is a K vector of ones.

We use the bracket notation “[•]” of Gelfand and Smith (1990) to designate a distribution or density for a random variable. The argument in the brackets identifies the distribution; for example, $[X]$ and $[Y|X]$ are the distributions of X and Y given X , respectively. The model and analysis require four distributions: uniform, normal, inverted gamma, and univariate extended Von Mises. The densities for the first three, standard distributions are displayed below to establish notation:

$$[v|a, b] = U(v|a, b) = (b - a)^{-1} \quad \text{for } a \leq v \leq b,$$

$$[\underline{x}|\underline{\mu}, \Sigma] = N_m(\underline{x}|\underline{\mu}, \Sigma) = (2\pi)^{-1/2} |\Sigma|^{-1/2} \exp[-1/2(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})],$$

$$[y|a, b] = IG(y|a, b) = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp(-b/y) \quad \text{for } y > 0.$$

The density for the extended Von Mises distribution is

$$[\theta|\underline{d}, \underline{Q}, \underline{C}] = VM(\theta|\underline{d}, \underline{Q}, \underline{C}) \propto \exp\left\{-\frac{1}{2}[\underline{\xi}(\theta) - \underline{d}]' \underline{Q} [\underline{\xi}(\theta) - \underline{d}]\right\} \chi(\theta \in \underline{C}),$$

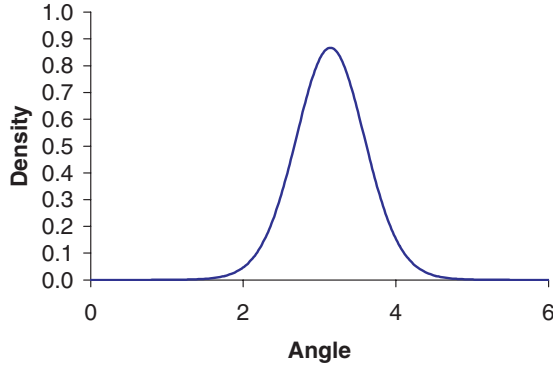


FIGURE 1.
Extended Von Mises distribution.

where $\xi(\theta) = [\sin(\theta), \cos(\theta)]'$, \underline{d} is a two-dimensional vector, Q is a 2×2 matrix, $\chi(\bullet)$ is the indicator function, and \mathcal{C} is a subset of $[0, 2\pi)$. Q does not need to be symmetric or positive definite because the range of ξ is finite. Figure 1 graphs the extended Von Mises distribution for $\underline{d} = (\sin(\pi), \cos(\pi))'$ and $Q = 5I_2$ when $\mathcal{C} = [0, 2\pi]$. If Q is a matrix of zeros the density is uniform. If $\underline{d} = (\sin(\tilde{\theta}), \cos(\tilde{\theta}))'$, then the mode of the distribution is $\tilde{\theta}$. We will restrict our attention to quadratic forms because the likelihood function for the angles takes this expression, and the extended Von Mises is the natural conjugate prior distribution for the angles.

The error terms of the latent variables are mutually independent and normally distributed:

$$[\underline{\epsilon}_i | \Sigma] = N_J(\underline{\epsilon}_i | \underline{0}_J, \Sigma) \quad \text{where } \underline{0}_J \text{ is a } J \text{ vector of zeros.}$$

The random effects are also mutually independent and normally distributed:

$$[\underline{\phi} | \lambda^2] = N_n(\underline{\phi} | \underline{0}_n, \lambda^2 I_n), \quad [\underline{\alpha} | \tau^2] = N_n(\underline{\alpha} | \underline{0}_n, \tau^2 I_n), \quad \text{and} \quad [\underline{\beta} | \tau^2] = N_n(\underline{\beta} | \underline{0}_n, \tau^2 I_n),$$

where I_n is the $n \times n$ identity matrix. Given the latent variables and cutpoints, the distribution of the response for subject i is

$$\begin{aligned} &Pr(w_{i,1}, \dots, w_{i,J} | c_i, \mu, \phi_i, \alpha_i, \beta_i, \theta, \Sigma) \\ &= \prod_{j=1}^J \left[\int_{c_{i,w_{i,j-1}}}^{c_{i,w_{i,j}}} N_1[y_{i,j} | \mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2] dy_{i,j} \right]. \end{aligned}$$

The prior distribution for the cutpoints is conditionally uniform:

$$[c_i] \propto \chi(-1 < c_{i,2} < \dots < c_{i,H-2} < 1).$$

That is, given $c_{i,k-1}$ and $c_{i,k+1}$, the conditional distribution of $c_{i,k}$ is uniform:

$$[c_{i,k} | c_{i,k-1}, c_{i,k+1}] = U(c_{i,k} | c_{i,k-1}, c_{i,k+1}) \quad \text{for } k = 2, \dots, H - 2.$$

The mean latent scores have a normal prior:

$$[\underline{\mu}] = N_J(\underline{\mu} | \underline{m}_0, V_0),$$

and the error and random effects variances have inverted gamma distributions:

$$[\sigma_j^2] = IG\left(\sigma_j^2 \mid \frac{r_0}{2}, \frac{s_0}{2}\right), \quad [\lambda^2] = IG\left(\lambda^2 \mid \frac{u_{0,1}}{2}, \frac{v_{0,1}}{2}\right), \quad \text{and} \quad [\tau^2] = IG\left(\tau^2 \mid \frac{u_{0,2}}{2}, \frac{v_{0,2}}{2}\right).$$

The prior distribution for the angles is extended Von Mises:

$$[\theta_j] = \begin{cases} VM(\theta_j | \underline{d}_0, Q_0, [0, \pi]) & \text{for } j = 2, \\ VM(\theta_j | \underline{d}_0, Q_0, [0, 2\pi)) & \text{for } j > 2, \end{cases}$$

with $\underline{d}_0 = (0, 1)'$, $Q_0 = 0.2I_2$ where I_2 is the 2×2 identity matrix. This prior distribution is fairly flat on $[0, 2\pi)$.

2.4. Markov Chain Monte Carlo Estimation

A primary goal of Bayesian inference is to compute the posterior distribution of the unknown parameters given the data. The posterior distribution quantifies the uncertainty about unknown parameters after observing the data. The posterior mean is the Bayes estimator under squared-error loss, and the posterior standard deviation is a measure of uncertainty about the parameter. For sufficiently large samples and well-behaved models, posterior distributions are approximately normal, and there is approximately 95% probability that the true parameter is within \pm two posterior standard deviations of the posterior mean. For non-Bayesians, the posterior mean is the point estimator, and the posterior standard deviation roughly resembles the standard error for the point estimator. Estimation of the model is accomplished via MCMC (see Gelfand & Smith, 1990). The Appendix gives details for the application to the circumplex model. After an initial transition period, the random deviates from MCMC can be treated as random draws from the posterior distribution and used to numerically approximate posterior statistics of the parameters. For example, the posterior mean is approximated by the average of the random draws.

The accuracy of these numerical approximations can be ascertained by the root mean squared simulation error (RMSSE). The RMSSE is the standard deviation of the MCMC approximation to the posterior mean and accounts for the autocorrelation in the Markov chain. The RMSSE tends to decrease as one uses more iterations in MCMC. In comparison, the theoretical posterior standard deviation does not depend on the estimation algorithm, and it tends to decrease as sample sizes increase. It is important to differentiate between the posterior standard deviation and the RMSSE. The first quantifies the posterior uncertainty about a parameter, while the latter quantifies the accuracy of the numerical algorithm in approximating the posterior mean. We will report the RMSSE to give an indication of the accuracy of the numerical approximations from the MCMC algorithm.

2.5. Brier Scores

Brier (1950) proposed a squared error loss statistic that compares predictive probabilities and random outcomes (see Gordon & Lenk, 1991, 1992). Let $\{z_i\}$ be n uncertain events where $z_i = 1$ if the event occurs and $z_i = 0$ otherwise. The Brier score is $BS = n^{-1} \sum_{i=1}^n (z_i - p_i)^2$ where p_i is the predictive probability for z_i . Suppose that one uses m different predictive probabilities $\{q_j\}$. The Brier score can be decomposed into two components, called “calibration” and “refinement”:

$$\begin{aligned} BS &= \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^n \chi(p_i = q_j) (z_i - q_j)^2 \\ &= \frac{1}{n} \left[\sum_{j=1}^m n_j (q_j - \hat{q}_j)^2 + \sum_{j=1}^m \sum_{i=1}^n \chi(p_i = q_j) (z_i - \hat{q}_j)^2 \right], \\ n_j &= \sum_{i=1}^n \chi(p_i = q_j) \quad \text{and} \quad \hat{q}_j = n_j^{-1} \sum_{i=1}^n \chi(p_i = q_j) z_i, \end{aligned}$$

where n_j is the number of times that q_j is used, and \hat{q}_j is the relative frequency of event j given that one predicted it would happen with probability q_j .

The first term of the decomposition is calibration and is related to bias. The calibration measure is zero when the predictive probability and conditional relative frequencies are equal. Clearly, calibration alone does not imply an accurate forecasting system. For instance, a system is well-calibrated if it always reports the base rates for events; however, base rates may not be very informative. The second measure, refinement, is similar to variance, and measures the propensity of the prediction system to use values close to zero or one: in a well-calibrated system, forecasts closer to zero or one are more useful than forecasts in the middle of the unit interval. DeGroot and Feinberg (1982) showed that if two systems are well-calibrated and if system A is more refined than system B , then B 's forecasts are equivalent to passing A 's forecasts through a noisy filter.

Our fit measure is based on a modified Brier score. Instead of using the predictive probabilities given the data in the computation, we compute the predictive probabilities given the parameters Ω and the data, and use these to compute a Brier score on each iteration of the Markov chain:

$$BS^{(m)} = \frac{1}{NH} \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^H \delta_{i,j} [z_{i,j,k} - P(W_{i,j} = k | \Omega^{(m)})]^2$$

where N is the total number of observations; $z_{i,j,k} = 1$ if person i responded k to variable j , and 0 otherwise; and $\delta_{i,j} = 1$ if the variable is observed and 0 if it is missing. That is, missing observations are excluded from the Brier score. The $\{BS^{(m)}\}$ are then used in computing posterior means and standard deviations. This approach extends the Brier score to include calibration, refinement, and uncertainty in the predictive probabilities.

2.6. Model Test on Synthetic Data

Before presenting the results of an empirical application of the model, we discuss the results of a synthetic data analysis mimicking 16 variables measured on a seven point ordinal scale in a sample of 100 persons. The angles θ_j were randomly generated under the constraints of four blocks with four angles in each block. Each item independently had a 7% probability of deletion. We estimated circumplex models with the correct constraints on the angles, with unconstrained angles, and with incorrect constraints, assigning variable 3 to block 3 and variable 11 to block 1. The incorrect constraints constitute a mild violation of the true model.

For each set of constraints, we ran 5, independent MCMC chains where each chain had 20,000 iterations. The initial burn-in period for each chain consisted of 10,000 iterations, and every tenth, subsequent iteration was used for estimation resulting in 1000 iterations per chain for estimation. The main reason for thinning the chains was to conserve memory. Starting values for the parameters were generated from the prior distributions, which were fairly noninformative. Convergence was checked with the potential scale reduction factor (PSRF) (Gelman & Rubin, 1992). The maximum univariate PSRFs for the angles were 1.006 without constraints, 1.003 with correct constraints, and 1.007 with incorrect constraints. The multivariate PSRFs (Brooks & Gelman, 1998) were 1.017 without constraints, 1.011 with correct constraints, and 1.021 with incorrect constraints. Brooks and Gelman (1998) conclude that PSRFs close to one indicate that the MCMC simulations are close to the posterior distribution. The average and maximum autocorrelations of lag one for the angles were 0.30 and 0.44 without constraints, 0.29 and 0.43 with correct constraints, and 0.31 and 0.48 with incorrect constraints. Autocorrelations of higher lags rapidly decreased. These statistics along with plots of the random draws versus iterations indicated convergence of the chain.

Although the posterior mean of the log-likelihood is a flawed measure for Bayesian model selection, it is sometimes computed because of its similarity to the log-likelihood statistic in

TABLE 1.

Estimated directions for the synthetic data. The constraints given to the program were either: no constraints (“None”), the right constraints (“Good”), or the wrong constraints (“Bad”). The true constraints sequentially assign four items to four blocks. The “Bad” constraints incorrectly assume that item 3 belongs to block 3 and item 11 to block 1.

Item	Block	TRUE	Posterior mean			Posterior Std. Dev.		
			None	Good	Bad	None	Good	Bad
1	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1	1.112	1.063	1.059	1.046	0.093	0.090	0.092
3	1	1.103	1.041	1.038	2.340	0.098	0.093	0.106
4	1	0.815	0.771	0.766	0.758	0.084	0.082	0.082
5	2	2.213	2.144	2.137	2.142	0.096	0.094	0.099
6	2	1.781	1.725	1.717	1.721	0.112	0.111	0.116
7	2	2.112	2.060	2.052	2.062	0.100	0.098	0.103
8	2	2.381	2.393	2.387	2.312	0.096	0.093	0.100
9	3	3.706	3.591	3.583	3.542	0.104	0.102	0.105
10	3	4.599	4.549	4.540	4.540	0.099	0.095	0.099
11	3	4.209	4.141	4.132	5.971	0.106	0.103	0.091
12	3	4.034	3.859	3.850	3.818	0.109	0.103	0.109
13	4	5.731	5.688	5.684	5.704	0.089	0.085	0.089
14	4	5.916	5.897	5.892	5.879	0.083	0.081	0.079
15	4	6.132	6.036	6.028	5.934	0.090	0.086	0.085
16	4	5.408	5.305	5.299	5.334	0.095	0.092	0.093

maximum likelihood estimation. For the simulation, the posterior means of the log-likelihood for the model without constraints and correct constraints were nearly identical: -715.84 and -715.50 . For the incorrectly constrained model, it was much smaller: -956.39 . Moreover, the Brier scores were also nearly the same for the unconstrained and correctly constrained models: 0.01618 and 0.01617 , and much larger, 0.02065 , for the incorrectly constrained model.

The algorithm was able to recover accurately the grand means μ_j , error standard deviations σ_j , and random effects standard deviations λ and τ , and all of the angles, except for items 3 and 11 with incorrectly specified blocks, as judged by their posterior means and posterior standard deviations. The RMSSEs for the posterior means of the angles ranged between 0.00133 and 0.00209 for the unconstrained model, 0.00126 and 0.00182 for the correct constraints, and 0.00133 and 0.00206 for the incorrect constraints.

With incorrectly specified constraints, item 3 was incorrectly assigned to block 3 instead of block 1. Its estimated angle, 2.340 , is the minimum angle in block 3 (between 2.312 for item 8 and 3.542 for item 9), which results from trying to recover the observed covariance structure. Similarly, item 11 was incorrectly assigned to block 1. Its estimated angle, 5.971 , is the largest estimated angle and close to 0 on the circle. It is slightly larger than the estimated angle for item 15, which has the maximum true angle 6.132 .

The differences between the models with various sets of constraints are reflected in the posterior means and standard deviations of the angles in Table 1. The posterior standard deviations with correct constraints tends to be slightly smaller than the posterior standard deviations without constraints. The information in the constraints leads to sharper estimators. This effect is stronger with smaller sample sizes (indicated by simulations not reported here). Generally, the posterior standard deviation with incorrect constraints are larger than the other two conditions, though not uniformly. For example, the posterior standard deviation for the incorrectly assigned item 11 is

PSYCHOMETRIKA

smaller than that without constraints and with the correct constraints. Apparently, the algorithm is trying its best to keep item 11 as close as possible to block 3 while still maintaining the incorrect assignment to block 1. Overall, the correct constraints reduce the posterior uncertainty about the angles. As one would anticipate, other simulations indicated that the amount of reduction depends on the number of observations with the greatest impact for very small samples. With large samples, the difference between the unconstrained and correctly constrained estimates are very small. Not surprisingly, incorrect constraints can result in inconsistent estimators.

3. Application to Schwartz Value Ratings

3.1. Schwartz Value Theory

Building on work by Rokeach (1973), Schwartz and Bilsky (1987, 1990) provided a detailed psychological theory of value content and structure. Values are defined as beliefs that pertain

TABLE 2.
Dimensions, domains, and values according to Schwartz's theory.

Openness-to-change		Conservation	
Domain	Values	Domain	Values
Self direction	Creativity	Security	Family security
	Freedom		National security
	Independent		Social order
	Curious		Cleanliness
Stimulation	Choosing own goals	Conformity	Reciprocation of favors
	Daring		Politeness
	Varied life		Obedient
	Exciting life		Self-discipline
		Tradition	Humble
			Accepting fate
			Religious
			Respect for tradition
			Moderate
Self-transcendence		Self-enhancement	
Domain	Values	Domain	Values
Benevolence	Helpful	Hedonism	Pleasure
	Honest	Achievement	Enjoying life
	Forgiving		Successful
	Loyal		Capable
Universalism	Responsible	Power	Ambitious
	Broad minded		Influential
	Wisdom		Social power
	Social justice		Authority
	Equality		Wealth
	World peace		
	World beauty		
	Unity with nature		
Environment			

to desirable states or behaviors, transcend specific situations, guide selection or evaluation of behavior, and are ordered by relative importance. They have been widely used in the social sciences to explain, for example, voting behavior (Rokeach, 1973), mass media usage (Rokeach & Ball-Rokeach, 1989), charity contributions (Manner & Miller, 1978), socially responsible behavior (Anderson & Cunningham, 1972), ecological behavior (Ellen, 1994), and innovativeness (Steenkamp, Ter Hofstede, & Wedel, 1999).

Schwartz and his collaborators postulate a comprehensive typology of the content and structure of domains of values. They distinguish ten value domains, summarized in Table 2, that are organized along two bipolar dimensions. The first dimension is defined as “openness to change versus conservation,” and opposes values of self-direction and stimulation to those of security, conformity, and tradition. The second dimension is called “self-enhancement versus self-transcendence” and opposes values of universalism and benevolence to those of hedonism, power, and achievement. The universal structure of values was investigated by Schwartz in a number of studies conducted in different countries. Smallest space analysis (Guttman, 1968) of the correlation matrices provided qualitative support for the postulates of the theory (Schwartz & Sagiv, 1995). From those studies, it is apparent that the theory borrows strength from ideas of circumplex representations. However, the circumplex model, although popular in psychology as a model for the representation of attitudes (e.g., Plutchik & Conte, 1997), has not been directly applied to the analysis of values, nor has Schwartz theory been subjected to statistical testing. Here we set out to examine more rigorously the validity of Schwartz’s value system by investigating the constraints that the theory imposes on the hypothesized circumplex structure.

3.2. *Sample and Data*

The data used here are part of a larger data set that was collected for the European Commission. A sample was drawn randomly from the household consumer panel of a market research agency in The Netherlands. This panel is representative of the Dutch population with respect to a large number of socio-demographic characteristics. For data collection, mail questionnaires were sent out to households in The Netherlands. The questionnaires included the Schwartz values measurement instrument, assessing 44 value priorities on 9-point scales. Before collecting the data extensive pretests were conducted. After sending reminders, the overall response was around 70% (for more details on data collection, see Ter Hofstede, Steenkamp, & Wedel, 1999). The sample size was 157 for a total of 6698 observations with 3% missing.

3.3. *Results*

We analyzed the Dutch value priority data with seven models derived from equation (2). The first model has random scale-usage effects only and zeros-out the circumplex structure ($\alpha_i = \beta_i = 0$). We estimated four different circumplex models ($\tau_a = \tau_b = \tau$). Model 2 is an unconstrained circumplex without scale-usage effects ($\phi_i = 0$). Models 3 to 5 are circumplex models with scale-usage effects. Model 3 has value domain constraints, and Model 4 has value dimension constraints (see Table 2). The latter is a weaker set of constraints as compared to the former. Model 5 is an unconstrained circumplex model. Model 6 is a two-factor, elliptical model ($\tau_a \neq \tau_b$) that has a more general correlation structure than the circumplex. It also has random scale-usage effects. Model 7 allows for an unconstrained error covariance ($\text{cov}(\epsilon_{i,j}, \epsilon_{i,j'}) = \sigma_{j,j'}$) among items and removes the random scale-usage effects and circumplex ($\phi_i = \alpha_i = \beta_i = 0$). The likelihood function of a model with full error covariance and random scale-usage or circumplex factors is not identified.

All models were estimated using 200,000 draws, with a burn-in of 100,000. After burn-in, every 10th iteration was used in the computations for a total of 10,000 draws. We ran five independent chains to evaluate convergence, resulting in a total of one million iterations. The

PSYCHOMETRIKA

TABLE 3.
Fit statistics for the Schwartz value data.

	Number [†] of parameters	Log-likelihood	Brier score
Model 1	89 ^a	-10271	0.0162
Model 2	133 ^b	-9910	0.0158
Model 3	134 ^c	-9769	0.0157
Model 4	134 ^c	-9745	0.0157
Model 5	134 ^c	-9702	0.0156
Model 6	135 ^d	-9698	0.0156
Model 7	1034 ^e	-11743	0.0880

Model 1: Random scale-usage effects and no circumplex ($\alpha_i = \beta_i = 0$)

Model 2: Circumplex ($\tau_a = \tau_b$), no scale-usage effects ($\phi_i = 0$), no constraints

Model 3: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, domain constraints

Model 4: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, dimension constraints

Model 5: Circumplex ($\tau_a = \tau_b$), random scale-usage effects, no constraints

Model 6: Elliptical model ($\tau_a \neq \tau_b$), random scale-usage effects, no constraints

Model 7: Full error covariance

[†]Counting the number of parameters in Bayesian, random effects models is not straightforward.

We do not include the latent variables $Y_{i,j}$, the individual-level cutpoints for the ordinal model, the random effects ($\phi_i, \alpha_i, \beta_i$), nor the prior parameters.

^a44 means μ_j , 44 error variances σ_j^2 , and random effects variance λ^2 .

^b44 means μ_j , 44 error variances σ_j^2 , 44 angles θ_j , and random coefficient variance τ^2 .

^cSame as ^b plus random effects variance λ^2 .

^dSame as ^c plus unique τ_a and τ_b instead of common τ .

^e44 means μ_j and $44(44 + 1)/2$ error variance and covariance terms.

multivariate PSRFs for the angles ranged between 1.002 for the model with dominant constraints to 1.004 for the model without scale-usage effects. The first autocorrelation coefficients for all angles for all models were less than 0.6, with most of the angles having much lower values. These statistics, along with plots of the iterations, indicated that the chains had converged and mixed fairly well.

Table 3 displays the fit statistics. The fit statistics indicate that the circumplex models (Models 3, 4, and 5) with scale-usage effects fit better than the circumplex model without scale-usage effects (Model 2) and the model with scale-usage and without the circumplex structure (Model 1). Model 6 relaxes the constraint that the standard deviations for α and β are equal, thus adding one parameter. Its Brier score is the same as that for circumplex Model 5, but its mean likelihood is slightly better than the other circumplex models. The two random effects standard deviations for Model 6 are $\tau_\alpha = 0.22$ and $\tau_\beta = 0.29$. The difference, 0.07, has a posterior standard deviation of 0.03. We take this as support, though not conclusive, that the circumplex model with one common factor variance holds. Model 7 is the most general model, but did not perform best: it has a very large number of parameters relative to the sample size. These fits provide evidence that Schwartz's value scales conform to the circumplex once scale-usage effects are properly handled. Apparently, the proposed approach of dealing with the response scale bias is both effective and important. For the circumplex models with different constraints the Brier scores are comparatively close. However, the fit deteriorates slightly as domain constraints are imposed. Thus, the value priority data seem to violate Schwartz's theory of value domains to a certain extent, although judged by the differences in fit between those models, the violations appear to be minor.

Table 4 reports the posterior means and posterior standard deviations of the estimated circumplex angles for Models 3, 4, and 5, the circumplex models with scale-usage random

TABLE 4.

Value angles for Netherlands data using three sets of constraints. “Full” is ordering of angles based on value domains; “Partial” is ordering of angles based on bipolar dimensions; “None” is unrestricted mode.

Dimension <i>Domain</i>	Posterior mean			Posterior Std. Dev.		
	Full	Partial	None	Full	Partial	None
Values						
Self-transcendence						
<i>Benevolence</i>						
Helpful	0	0	0	0	0	0
Honest	0.123	0.142	0.570	0.146	0.141	0.299
Forgiving	0.296	0.387	0.846	0.168	0.231	0.349
Loyal	0.330	0.351	0.828	0.156	0.176	0.293
Responsible	0.235	0.283	0.729	0.167	0.190	0.307
<i>Universalism</i>						
Broad-minded	1.458	1.288	1.951	0.193	0.175	0.316
Wisdom	0.940	0.823	1.331	0.263	0.244	0.330
Social justice	0.532	0.255	0.694	0.166	0.183	0.320
Equality	1.005	0.840	1.265	0.260	0.253	0.348
World peace	0.579	0.394	0.896	0.174	0.191	0.303
World beauty	0.763	0.612	1.072	0.215	0.220	0.320
Unity with nature	0.643	0.436	0.931	0.192	0.211	0.319
Environment	0.976	0.838	1.304	0.219	0.210	0.316
Openness-to-change						
<i>Self-direction</i>						
Creativity	2.052	1.957	2.853	0.213	0.209	0.382
Freedom	1.592	1.407	1.564	0.206	0.184	0.314
Independent	1.845	1.577	1.697	0.278	0.257	0.362
Curious	1.970	1.855	2.395	0.218	0.217	0.355
Choosing own goals	1.906	1.735	2.124	0.216	0.208	0.315
<i>Stimulation</i>						
Daring	2.314	2.074	3.477	0.188	0.181	0.324
Varied life	2.263	1.975	2.559	0.185	0.186	0.312
Exciting life	2.252	1.908	2.417	0.187	0.194	0.310
Self-enhancement						
<i>Hedonism</i>						
Pleasure	2.394	2.181	2.069	0.191	0.187	0.328
Enjoyment	2.398	2.193	2.289	0.191	0.187	0.329
<i>Achievement</i>						
Successful	3.118	2.988	3.629	0.233	0.220	0.324
Capable	2.510	2.185	1.649	0.217	0.201	0.327
Ambitious	3.479	3.659	4.321	0.237	0.285	0.363
Influential	3.506	3.607	4.239	0.218	0.242	0.343
<i>Power</i>						
Social power	3.935	3.784	4.324	0.266	0.263	0.337
Authority	3.954	3.828	4.390	0.240	0.235	0.333
Wealth	3.672	3.176	3.720	0.207	0.240	0.330

PSYCHOMETRIKA

TABLE 4.
cont'd

Dimension <i>Domain</i>	Posterior mean			Posterior Std. Dev.		
	Full	Partial	None	Full	Partial	None
Values						
Conservation						
<i>Security</i>						
Family security	5.572	5.831	0.175	0.184	0.221	0.437
National security	5.540	5.841	0.236	0.206	0.242	0.485
Social order	5.640	6.075	0.472	0.163	0.150	0.325
Cleanliness	5.007	4.932	5.471	0.228	0.228	0.321
Reciprocation of favors	4.917	4.926	5.711	0.318	0.346	0.513
<i>Conformity</i>						
Politeness	5.818	5.919	0.182	0.139	0.196	0.390
Obedient	5.800	5.671	6.167	0.137	0.211	0.396
Self-discipline	5.849	6.142	0.800	0.136	0.123	0.326
Honoring elders	5.810	5.664	6.126	0.137	0.215	0.364
<i>Tradition</i>						
Humble	6.041	5.328	5.614	0.135	0.339	0.386
Accepting fate	6.142	6.071	0.721	0.108	0.190	0.368
Devout	5.984	4.734	5.242	0.138	0.287	0.358
Respect for tradition	6.005	5.483	6.040	0.138	0.271	0.490
Moderate	6.111	6.019	0.427	0.118	0.178	0.323

effects and domain constraints (Full), dimension constraints (Partial), and no constraints (None), respectively. The RMSEs are small in all cases: they were less than 0.002 for the “Full” and “Partial” and less than 0.003 for “None” models. Some of the posterior distributions for the angles are bimodal because the support of the distribution spans zero. For example, if there are constraints, angles in the first block can be less than zero. In these cases, we compute the posterior means and standard deviations by “unrolling the circle” in post-processing the MCMC draws. If the posterior distribution of θ_j is bimodal and if more than half of the posterior distribution is between 0 and π , we recode MCMC draws $\theta_j^{(g)}$ for the g th iteration that are between π and 2π as $\theta_j^{(g)} - 2\pi$. Similarly, if more than half of the distribution is between π and 2π , we recode draws that are between 0 and π as $2\pi + \theta_j^{(g)}$. This recoding does not change the circumplex variances and covariances and is only used in approximating the posterior means and standard deviations. Ignoring the bimodal distributions results in nonsensical posterior means and standard deviations: if the posterior distribution is concentrated on both sides of 0, then the posterior mean will be around π , a region of zero probability.

When comparing the models with the domain (Full) and the dimension (Partial) constraints, it is apparent that only a few value angles, using the more general dimension constraints, differ from the more exact ordering, using domain constraints. The violations to the constraints misplace the angles in neighboring value domains within the value dimensions (Table 4). Most of the violations using no constraints, compared to the domain and dimension constraints, occur for angles near zero or 2π .

The posterior standard deviations indicate the uncertainty about the angles. Based on their posterior means and standard deviations, the posterior distribution of the angles from the three models are similar, with the exceptions of angles for conservation. Even here, though, the

differences are more apparent than real if one keeps in mind that 0.1 radian is very close to 2π on the circle. In comparing posterior standard deviations, estimators using domain constraints are more precise than using dimension constraints, which are, in turn, more precise than those without constraints.

Figure 2 provides a graphical display of the posterior means for each of the three circumplex models with random intercepts as well as of the model with fixed intercept. The radii of the vectors for angles in the four dimensions have been jittered so that the points do not overlap. Figure 2(a) shows significant distortions of the circumplex structure if scale usage is not taken into account: the value angles are almost entirely confined to the positive quadrant. Figure 2(b) graphs the angles for the unconstrained circumplex model with random effects, and Figures 2(c) and 2(d) graph the angles with bipolar dimension and value domain constraints.

Inspection of Figure 2 reveals that, although the unconstrained circumplex does fit the data marginally better, the differences in the locations of the values on the circumplex are minor. To reveal their correspondence, Figure 3 presents scatter plots of the angles for the four circumplex models. Figure 3(a) plots the angles from the unconstrained circumplex models with and without scale-usage effects and reinforces the observation that scale-usage effects are needed. Figures 3(b) and 3(c) plot the constrained models against the unconstrained one with scale-usage effects. The points in the top left-hand corners are due to values in the conservation dimension overlapping with those of self-transcendence in the unconstrained model. Finally, Figure 3(d) plots the domain and dimension constrained models. These plots indicate that the circumplex structures are quite similar: the estimated angles are virtually on a straight line. Because of the few and minor violations of the dimension and domain constraints, we are inclined to conclude that the Schwartz value theory holds fairly well in The Netherlands sample, even though the model fit criterion points toward the unconstrained model.

Individual differences in the value judgments are depicted in Figure 4. This figure is based on the domain constraints and displays average interaction effects. The averages are over the cosines or sines of the angles in the same value domain for fixed values of α and β :

$$\frac{\alpha}{\text{card}(\mathcal{B}_k)} \sum_{j \in \mathcal{B}_k} \sin(\theta_j) + \frac{\beta}{\text{card}(\mathcal{B}_k)} \sum_{j \in \mathcal{B}_k} \cos(\theta_j), \quad (5)$$

where $\text{card}(\mathcal{B}_k)$ is the cardinality of \mathcal{B}_k . Figure 4 contains four curves with α and β equal to ± 1 . The figure illustrates that a person who has high values for self-transcendence has low values for self-enhancement and moderate values for openness-to-change and conservation. Likewise, a person with high values for openness-to-change has low values for conservation. Similar patterns can be observed for the other two value dimensions.

4. Conclusion

Models for covariance structures are popular in the social sciences for assessing latent psychological constructs from proxy variables that are intended to represent the psychological domains in question. Whereas the exploratory factor analysis model has been used frequently since 1960, confirmatory factor models (Jöreskog, 1974) became popular in the 1970s for applications where prior theory guided the identification of the underlying latent variable structure. However, because of their linear form, these broad modeling frameworks can include only a subset of relevant models for covariance structures. One of the significant exceptions, which is not included in the confirmatory factor modeling framework, is the class of circumplex models (Guttman, 1954; Browne, 1992) which imposes nonlinear constraints on the correlation matrix. These constraints are derived from the ordering of the proxy variables on the circumplex and, thus, avoid the need to achieve simple structure through either rotation or identifying constraints.

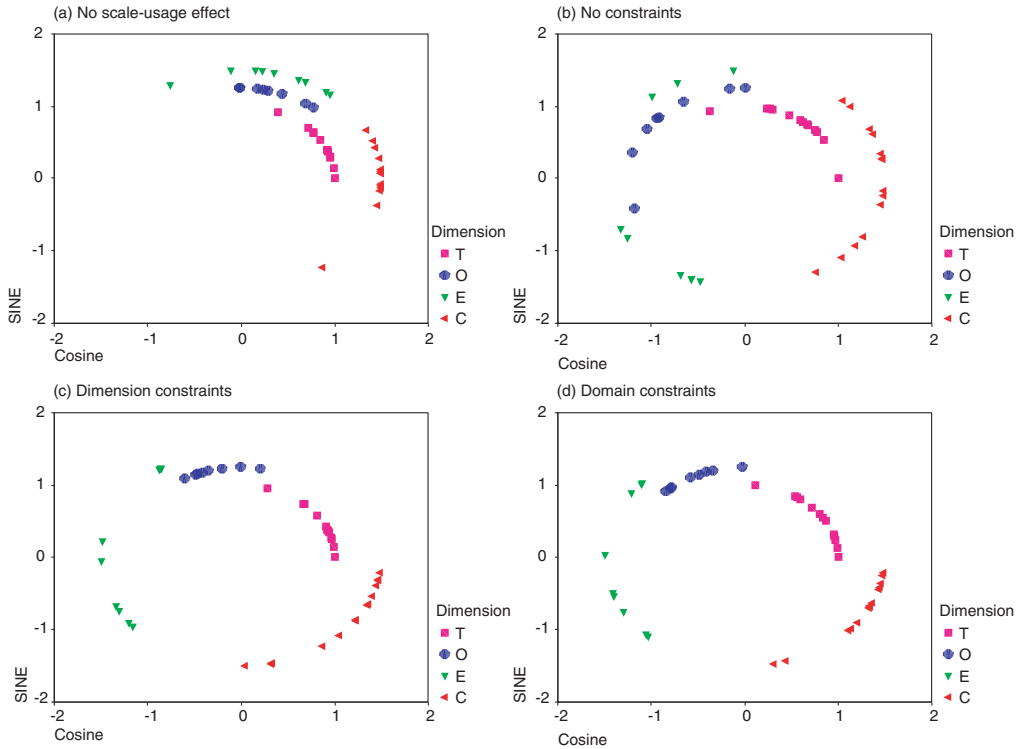


FIGURE 2.

Sine versus cosine of posterior means of item angles for Schwartz value data. (a) No constraints and fixed intercepts; (b) no constraints and random intercepts; (c) bipolar constraints and random intercepts; and (d) value domain constraints and random intercepts. Angles are identified by their bipolar dimensions: T = self-transcendence, O = openness-to-change, E = self-enhancement, and C = conservation.

In our Bayesian formulation of circumplex models for rating scales, we explicitly account for idiosyncratic response scale-usage by using an individual level cutpoint approach that assumes that respondents map an underlying latent trait onto the response scale and by a random effects specification that allows for differential scale-usage tendencies. In the synthetic data application, we demonstrated that the individual-level cutpoints can be recovered well even when the sample size is small, while the empirical application showed that failure to accommodate response scale usage seriously distorts the recovered circumplex structure. A potential drawback of our Bayesian approach, however, is that as yet standard software is not available and that it requires more computer time than maximum likelihood methods.

The circumplex model has been of much appeal to social science researchers because of its implied properties for the correlation structure of the measured items. Our approach yields a tractable representation that deals with different sources of person-specific heterogeneity. Moreover, the Bayesian formulation of the model and the proposed MCMC algorithm allow us to impose inequality constraints on the circumplex that are derived from substantive theory. In the synthetic data analysis and empirical application we showed how to investigate the validity of these constraints. We believe that these contributions will facilitate rigorous tests and further increase the popularity of circumplex models for the analysis of psychological constructs in the social sciences.

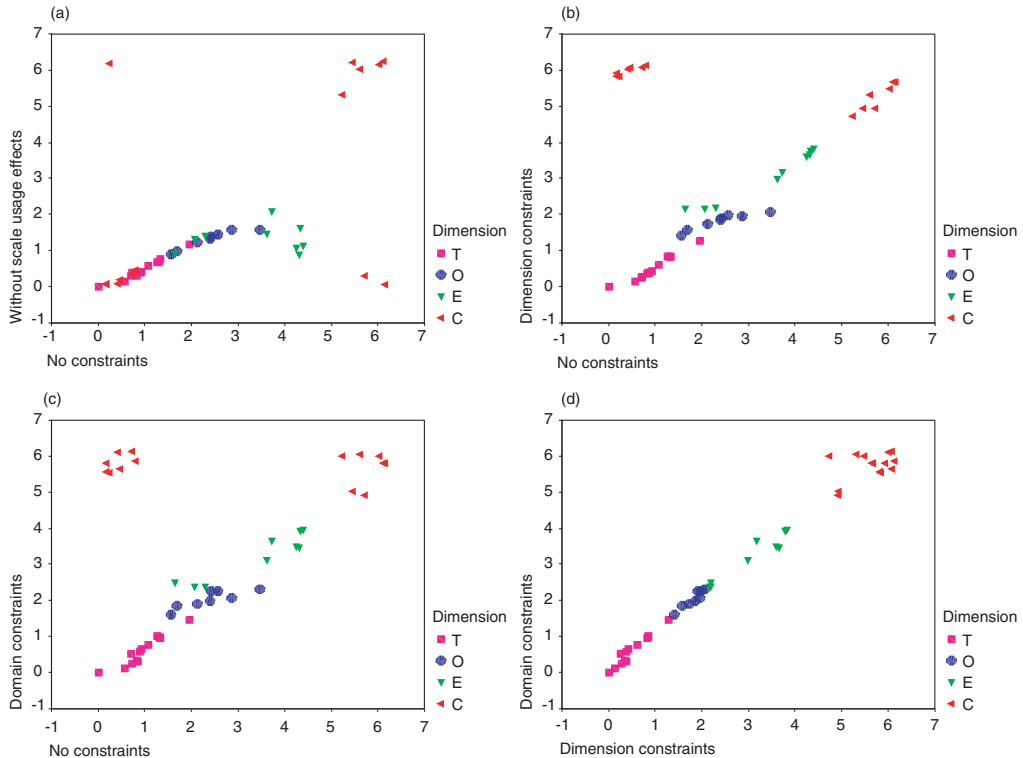


FIGURE 3.

Plots of angles from different circumplex models. (a) No scale-usage effects versus scale-usage effects without constraints; (b) dimension constraints versus no constraints with scale-usage effects; (c) domain constraints versus no constraints with scale-usage effects; and (d) domain constraints versus dimension constraints. T = self-transcendence, O = openness-to-change, E = self-enhancement, and C = conservation.

Appendix

A. MCMC

All of the full conditionals, except those for the angles, are standard distributions. The MCMC algorithm proceeds by drawing recursively from the full conditional distributions of the parameters, as provided below. Each of those full conditional distributions takes a standard form, with the exception of the full conditional for the angles, θ . We will use the matrix notation and distributions in section 2.3. The algorithm was implemented in the GAUSS language, and the code can be obtained from the first author.

A.1. Full Conditional for $Y_{i,j}$ for Observed $W_{i,j}$

$$[y_{i,j}|\text{Rest}] \propto N_1[y_{i,j}|\mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2] \chi(c_{i,w_{i,j}-1} < y_{i,j} \leq c_{i,w_{i,j}}),$$

where $\chi(\bullet)$ is the indicator function. The full conditional distribution is a truncated normal where the truncation depends on the cutpoints and the observed ordinal response. We use the inverse cumulative distribution function transform to generate truncated normal random variables.

PSYCHOMETRIKA

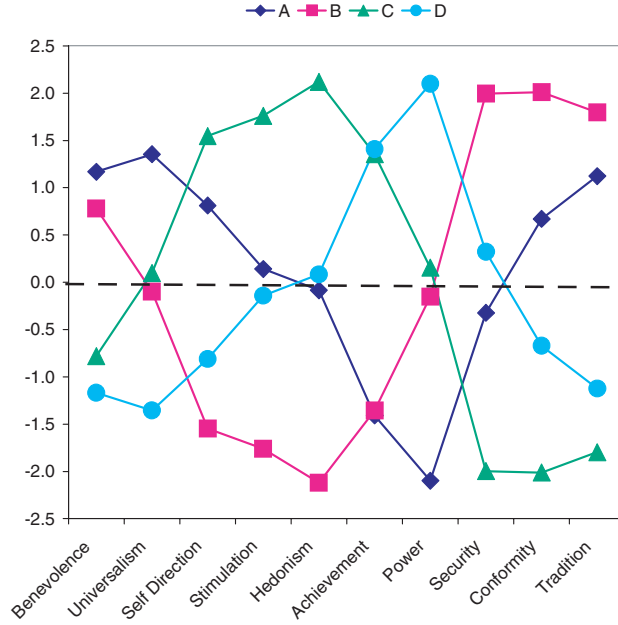


FIGURE 4.

Plot of average interactions for selected values of the random coefficients for the Schwartz value model with domain constraints. The four curves are averaged interaction effects where the averaging is within value domains for different combinations of $\alpha = \pm 1$ and $\beta = \pm 1$ in equation (5). Line A has $\alpha = 1$ and $\beta = 1$. Line B has $\alpha = 1$ and $\beta = -1$. Line C has $\alpha = -1$ and $\beta = 1$. Line D has $\alpha = -1$ and $\beta = -1$.

A.2. Full Conditional for $Y_{i,j}$ for Missing $W_{i,j}$

The model easily accommodates missing data assuming missing at random (MAR). If $W_{i,j}$ is missing, then $Y_{i,j}$ is normal:

$$[y_{i,j}|\text{Rest}] = N_1[y_{i,j}|\mu_j + \phi_i + \alpha_i \sin(\theta_j) + \beta_i \cos(\theta_j), \sigma_j^2].$$

That is, one does not know which cutpoints $y_{i,j}$ would have fallen between.

A.3. Full Conditional for Cutpoints c_i

Given $c_{i,k-1}$, $c_{i,k+1}$, and the latent variables Y_i , the cutpoint $c_{i,k}$ is uniformly distributed. Define $\bar{y}_k = \max\{y_{i,j} : w_{i,j} = k\}$ and $\underline{y}_{k+1} = \min\{y_{i,j} : w_{i,j} = k + 1\}$. Then

$$[c_{i,k}|\text{Rest}] = U(c_{i,k} | \max(\bar{y}_k, c_{i,k-1}), \min(\underline{y}_{k+1}, c_{i,k+1}) \quad \text{for } k = 2, \dots, H - 2.$$

A.4. Full Conditional for $\underline{\mu}$

$$[\underline{\mu}|\text{Rest}] = N_J(\underline{\mu}|\underline{m}_n, V_n),$$

$$V_n = (n\Sigma^{-1} + V_0^{-1})^{-1},$$

$$\underline{m}_n = V_n \left(\Sigma^{-1}(Y - \underline{\phi}1'_J - \underline{\alpha}X'_S - \underline{\beta}X'_C)'1_n + V_0^{-1}\underline{m}_0 \right).$$

A.5. Full Conditional for $\underline{\phi}$

$$\begin{aligned} [\underline{\phi}|\text{Rest}] &= N_n(\underline{\phi}|\underline{m}_\phi, v_\phi I_n), \\ v_\phi &= (\underline{1}'_J \Sigma^{-1} \underline{1}_J + \lambda^{-2})^{-1}, \\ \underline{m}_\phi &= v_\phi (Y - \underline{1}_n \underline{\mu}' - \underline{\alpha} X'_S - \underline{\beta} X'_C) \Sigma^{-1} \underline{1}_J. \end{aligned}$$

 A.6. Full Conditional for $\underline{\alpha}$ and $\underline{\beta}$

$$\begin{aligned} [\alpha_i, \beta_i|\text{Rest}] &= N_2[(\alpha_i, \beta_i)'|\underline{m}_{\alpha_i, \beta_i}, V_{\alpha_i, \beta_i}], \\ V_{\alpha_i, \beta_i} &= (X' \Sigma^{-1} X + \tau^{-2} I_2)^{-1}, \\ \underline{m}_{\alpha_i, \beta_i} &= V_{\alpha_i, \beta_i} X' \Sigma^{-1} (\underline{Y}_{i, \bullet} - \underline{\mu} - \phi_i \underline{1}_J). \end{aligned}$$

 A.7. Full Conditional for σ_j^2

$$\begin{aligned} [\sigma_j^2|\text{Rest}] &= IG\left(\sigma_j^2 \left| \frac{r_n}{2}, \frac{s_n}{2}\right.\right), \\ r_n &= r_0 + n, \\ s_n &= s_0 + \sum_{i=1}^n [y_{i,j} - \mu_j - \phi_i - \alpha_i \sin(\theta_j) - \beta_i \cos(\theta_j)]^2. \end{aligned}$$

 A.8. Full Conditional for λ^2

$$\begin{aligned} [\lambda^2|\text{Rest}] &= IG\left(\lambda^2 \left| \frac{u_{n,1}}{2}, \frac{v_{n,1}}{2}\right.\right), \\ u_{n,1} &= u_{0,1} + n, \\ v_{n,1} &= v_{0,1} + \underline{\phi}' \underline{\phi}. \end{aligned}$$

 A.9. Full Conditional for τ^2

$$\begin{aligned} [\tau^2|\text{Rest}] &= IG\left(\tau^2 \left| \frac{u_{n,2}}{2}, \frac{v_{n,2}}{2}\right.\right), \\ u_{n,2} &= u_{0,2} + 2n, \\ v_{n,2} &= v_{0,2} + \underline{\alpha}' \underline{\alpha} + \underline{\beta}' \underline{\beta}. \end{aligned}$$

 A.10. Full Conditional for $\underline{\theta}$

$$\begin{aligned} [\underline{\theta}|\text{Rest}] &\propto \exp \left\{ - \sum_{i=1}^n \sum_{j=1}^J \frac{1}{2\sigma_j^2} [y_{i,j} - \mu_j - \phi_i - \alpha_i \sin(\theta_j) - \beta_i \cos(\theta_j)]^2 \right\} \chi(\underline{\theta} \in \mathcal{C}) \\ &\times \exp \left\{ - \frac{1}{2} \sum_{j=2}^J [\underline{\xi}(\theta_j) - \underline{d}_0]' \mathcal{Q}_0 [\underline{\xi}(\theta_j) - \underline{d}_0] \right\} \end{aligned}$$

$$\begin{aligned} &\propto \prod_{j=2}^J VM(\theta_j | \underline{d}_j, Q_j, C_j), \\ Q_j &= \sigma_j^{-2} \begin{bmatrix} \underline{\alpha}'\underline{\alpha} & \underline{\alpha}'\underline{\beta} \\ \underline{\beta}'\underline{\alpha} & \underline{\beta}'\underline{\beta} \end{bmatrix} + Q_0, \\ \underline{d}_j &= Q_j^{-1} \left\{ \begin{bmatrix} \underline{\alpha}' \\ \underline{\beta}' \end{bmatrix} (\underline{Y}_{\bullet,j} - \mu_j \underline{1}_n - \underline{\phi}) + Q_0 \underline{d}_0 \right\}. \end{aligned}$$

where C_j is the constraint set for θ_j , and $C = \bigcup_{j=2}^J C_j$. Because the prior distributions for the angles are independent, the full conditionals depend on each other only through the constraints. Thus, generating from the full conditionals involves generating from univariate distributions, which we do by using the inverse cumulative distribution transform of a uniform random deviate.

Our experience has been that once the random coefficients and angles are in the vicinity of their true values, generating directly from the full conditionals works very well. However, the algorithm can get “stuck,” because the random coefficients (X matrix in the posterior density) limit the range of values attainable for the angles and visa versa. Thus, given one set of parameters, draws from the other set may not visit high probability areas of the parameter space.

Therefore, we pursue a hybrid sampling strategy and, in addition to generating the angles from their full conditional distributions directly, we include several methods for generating the candidate angles and random coefficients in a Metropolis step at every iteration of the sampler. First, we use a random walk to generate the angles where the jump distribution is a finite mixture of L uniform distributions where the endpoints depend on the current angle and the constraints. Figure 5(a) graphs the density of a typical mixture of uniforms, centered at zero, for the error distribution. It puts large mass around zero, which implies that most candidates’ values, ψ_j , will be close to the current value, θ_j , and it allows large jumps with relatively low probability. Figure 5(b) graphs the jump distribution given the current θ_j is 2.5 and the block constraints imply that $0.5 < \psi_j < 2.8$. With probability p_l , the candidate value, ψ_j , for θ_j is generated from a uniform that is proportional to $\chi(\theta_j \in C_j)U(\psi_j | \theta_j - u_l, \theta_j + u_l)$ where u_l is a prespecified positive constant, and C_j are the constraints on θ_j given the other values of θ . Once we have candidate values for the angles, we generate candidate values for the random coefficients α and β . The candidate values for the angles and random coefficients are jointly accepted or rejected. Because the acceptance rate, when all of the angles are simultaneously generated, is low, we also generate and test each angle separately.

The constraints on the angles result in relatively complex expressions, though easy to compute, for the acceptance probabilities. Without loss of generality, suppose $\theta_1 = 0$ is in the first block \mathcal{B}_1 , and there are K blocks. The indices b_1, \dots, b_K will give the last angle that belongs to the blocks:

$$1, \dots, b_1 \in \mathcal{B}_1 \quad \text{and} \quad b_{k-1} + 1, \dots, b_k \in \mathcal{B}_k \quad \text{for} \quad k = 2, \dots, K.$$

The blocks follow the order in section 2.2.

Candidate values ψ_2, \dots, ψ_M for the angles are generated sequentially. We will use the definitions of the minimum and maximum angles, $(\underline{\mathcal{B}}_k, \overline{\mathcal{B}}_k)$, from section 2.2. where it is to be understood that these minimum and maximum angles change as current values of θ_m are replaced by the candidates ψ_m as the candidates are generated. Define “ \vee ” as the maximum operator, and “ \wedge ” as the minimum operator. The random walk is a mixture of L uniform distributions: in Figure 5(a) the endpoints for component l are $\pm u_l$ with mixture probability p_l . For angles in the

first block, the jump distributions are

$$g_1[\psi_m | \psi_2, \dots, \psi_{m-1}, \theta_m, \dots, \theta_J] = \sum_{l=1}^L p_l \frac{\chi(a_{l,m,1} < \psi_m < b_{l,m,1})}{b_{l,m,1} - a_{l,m,1}},$$

$$a_{l,m,1} = (\overline{\mathcal{B}}_K - 2\pi) \vee (\theta_m - u_l),$$

$$b_{l,m,1} = \underline{\mathcal{B}}_2 \wedge (\theta_m + u_l) \quad \text{for } m = 2, \dots, b_1.$$

If the candidate value ψ_m in the first block is negative, then it is recoded as $2\pi + \psi_m$.

For angles in blocks $k = 2$ to K , the uniform random walk has density:

$$g_k[\psi_m | \psi_2, \dots, \psi_{m-1}, \theta_m, \dots, \theta_J] = \sum_{l=1}^L p_l \frac{\chi(a_{l,m,k} < \psi_m < b_{l,m,k})}{b_{l,m,k} - a_{l,m,k}},$$

$$a_{l,m,k} = \underline{\mathcal{B}}_{k-1} \vee (\theta_m - u_l),$$

$$b_{l,m,k} = \overline{\mathcal{B}}_{k+1} \wedge (\theta_m + u_l) \quad \text{for } m = b_{k-1} + 1, \dots, b_k.$$

where $\overline{\mathcal{B}}_{K+1} = \underline{\mathcal{B}}_1$.

After generating the candidate angles $\{\psi_m\}$, candidate values of the random coefficients $\{\alpha^c, \beta^c\}$ are generated from normal distributions in section A.6. The acceptance probability for the candidates are

$$\min \left\{ 1, \frac{[\psi, \alpha^c, \beta^c | Y] \prod_{k=1}^K \prod_{m=b_{k-1}+1}^{b_k} g_k[\theta_m | \theta_2, \dots, \theta_{m-1}, \psi_m, \dots, \psi_J][\alpha, \beta | \theta]}{[\theta, \alpha, \beta | Y] \prod_{k=1}^K \prod_{m=b_{k-1}+1}^{b_k} g_k[\psi_m | \psi_2, \dots, \psi_{m-1}, \theta_m, \dots, \theta_J][\alpha^c, \beta^c | \psi]} \right\},$$

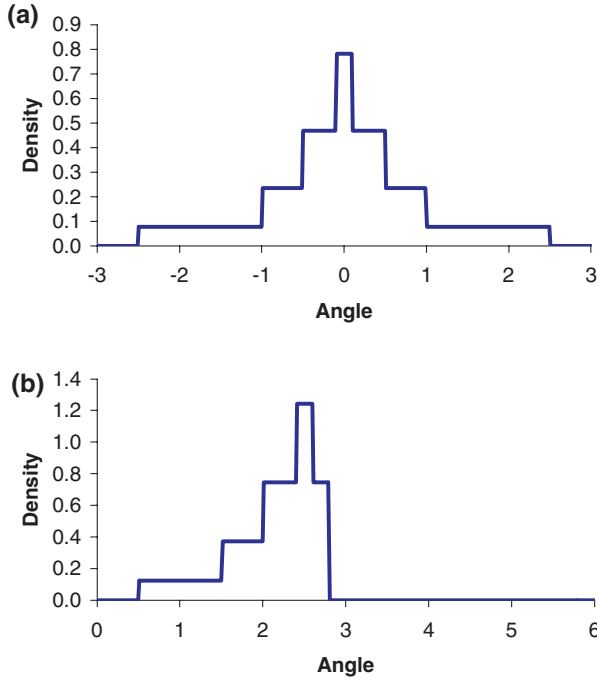


FIGURE 5.

Random walk jump distribution for angles. (a) Mixture of four uniforms. (b) Random walk based on mixture of uniforms assuming that the current value of the angle is 2.5 radians and the block constraints imply that the angle is restricted between 0.5 and 2.8 radians.

where $b_0 + 1 = 2$, and $[\theta, \alpha, \beta|Y] \propto [Y|\mu, \phi, \alpha, \beta, \theta, \sigma][\theta][\alpha][\beta]$. In addition to generating all angles at once according to the random walk Metropolis, we generated them one by one.

In the model without block constraints, mixing is improved if one of the uniform distributions in the mixture allows for reflections such as $\chi(C_j)U(\psi_j|\pi - \theta_j - u_l, \pi + \theta_j + u_l)$ for $j = 2$ and $\chi(C_j)U(\psi_j|2\pi - \theta_j - u_l, 2\pi + \theta_j + u_l)$ for $j > 2$. The rationale for the reflection is that the random walk chain has to progress from a region of high probability, through a region of low probability, to arrive at another area of high probability. For example, suppose that the current value of the angle is 0.2, so $\cos(.2)$ is close to one and $\sin(.2)$ is close to zero. Angles close to $2\pi - .2$ result in similar values for the sine, cosine, and covariances among Y variables. However, for a random walk to reach $2\pi - .2$, it has to pass through regions around π where the sines, cosines, and covariances are very different. If .2 is a highly probable value for the angle, the random walk will not reach the other side of the circle because of the low probability region that intervenes. In the same vein, we included another Metropolis step where the possible candidate values θ^N are determined from $\cos(\theta^N) = \pm \cos(\theta)$ for the current θ . Lastly, we included random phase shifts in the algorithm, by adding a small random amount to each angle, which slightly rotates the entire configuration and helps to escape from regions of low probability.

Our experience has been that these additional random walk Metropolis steps explore the parameter space more rapidly than generating angles from their full conditional distributions only. However, once the chain is in a high probability region, the additional Metropolis steps have very low acceptance rates, especially those that are designed to jump to other sectors of the circle. Then generating from the full conditionals is more efficient because none of the random deviates are rejected. Although the Metropolis steps have very low acceptance rates, they seem to be very important in mixing the chain: these large jumps are very seldom needed, but are essential in probing different regions of the circumplex and in establishing in the burn-in period the global configuration of angles that are consistent with the observed covariances when the model does not have order constraints.

Another concern is using the iterations to estimate the posterior mean and standard deviation of the angles. If an angle is close to 0 or 2π , then its posterior distribution will often be bimodal, and the posterior mean, which will be around π , is not a valid estimator. In this situation, we recode the iterations. If the majority of iterations are between 0 and π , then we recode θ between π and 2π as $2\pi - \theta$. Similarly, if the majority of iterations are between π and 2π , then we recode θ between 0 and π as $2\pi + \theta$. This recoding also needs to be performed across chains when multiple chains are run to verify convergence. For example, if the true angle is 0 (or 2π), then some of the chains may have more iterations between 0 and π , while other chains will have iterations between π and 2π . If the recoding was performed within each chain without regard for the other chains, the PSRF convergence statistics would incorrectly signal that the chains had not converged because some of posterior means would be close to zero while others would be close to 2π . This false negative is a consequence of the support being a circle. Consistently recoding the iterations across chains is required for the convergence statistics to give a true reading.

References

- Anderson, T.W. (1960). Some stochastic process models for intelligence test scores. In K.J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences* (pp. 205–220). Stanford, CA: Stanford, University Press.
- Anderson, T.W., & Cunningham, W.H. (1972). Socially conscious consumer behavior. *Journal of Marketing*, 36, 22–31.
- Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 1–3.
- Brooks, S.P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7, 434–455.
- Browne, M. (1992). Circumplex models for correlation matrices. *Psychometrika*, 57, 469–497.
- Cudeck, R. (1986). A note on structural models on the circumplex. *Psychometrika*, 51, 143–147.
- Ellen, P.S. (1994). Do we know what we need to know? Objective and subjective knowledge effects on pro-ecological behaviors. *Journal of Business Research*, 30, 43–52.

PSYCHOMETRIKA

- Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.
- Gelman, A., & Rubin, D. (1992). Inference from iterative simulations using multiple sequences. *Statistical Science*, 4, 457–472.
- Gordon, M.D., & Lenk, P. (1991). A utility theoretic examination of the probability ranking principle in information retrieval. *Journal of the American Society for Information Science*, 42, 703–714.
- Gordon, M.D., & Lenk, P. (1992). When is the probability ranking principle suboptimal? *Journal of the American Society for Information Science*, 43, 1–14.
- Guttman, L. (1954). A new approach to factor analysis: The radex. In P.F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). New York: Columbia University Press.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469–506.
- Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D.H. Krantz, R.C. Atkinson, R.D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*. (Vol. 2, pp. 1–56). San Francisco: Freeman.
- Lippe, R. (1995). Gender-related individual differences and psychological adjustment in terms of the big five and circumplex models. *Journal of Personality and Social Psychology*, 69, 1184–1202.
- Manner, L., & Miller, S.J. (1978). An examination of the value-attitude structure in the study of donor behavior. In *Proceedings of the American Institute of Decision Sciences*, St. Louis, 12, 532–538.
- Plutchik, R., & Conte, H.R. (1997). *Circumplex models of personality and emotions*. Washington, DC: American Psychological Association.
- Rokeach, M.J. (1973). *The nature of human values*. New York: The Free Press.
- Rokeach, M.J., & Ball-Rokeach, S.J. (1989). Stability and change in American value priorities, 1968–1981. *American Psychologist*, 44, 775–784.
- Rossi, P.E., Gilula, Z., & Allenby, G.M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association*, 96, 20–31.
- Russell, J.A., & Carroll, J.M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, 125, 3–30.
- Schwartz, S.H., & Bilsky, W. (1987). Toward a psychological structure of human values. *Journal of Personality and Social Psychology*, 53, 550–562.
- Schwartz S.H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology*, 58, 878–891.
- Schwartz, S.H., & Sagiv, L. (1995). Identifying culture specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, 26, 92–116.
- Shepard, R.N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part I. *Psychometrika*, 27(3), 125–140.
- Shepard, R.N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part II. *Psychometrika*, 27(3), 219–246.
- Steenkamp, J.E.B.M., Hofstede, F. Ter, & Wedel, M. (1999). A cross-national investigation into the individual and cultural antecedents of consumer innovativeness. *Journal of Marketing*, 63, 55–69.
- Ter Hofstede, F., Steenkamp, J.E.B.M., & Wedel, M. (1999). International market segmentation based on consumer-product relations. *Journal of Marketing Research*, 36, 1–17.
- Wiggins, J.S., Steiger, J.H., & Gaelick, L. (1981). Evaluating circumplexity models in personality data. *Multivariate Behavioral Research*, 16, 263–289.

Manuscript received 14 DEC 2001

Final version received 17 SEP 2004

Published Online Date: 16 MAR 2006