Research Master's programme: METHODOLOGY AND STATISTICS OF BEHAVIORAL AND SOCIAL SCIENCES
Utrecht University, the Netherlands

MSc Thesis *Jolien Cremers (3463230)*
TITLE: Assessing a Bayesian Embedding Approach to Circular Regression Models
May 2014

Supervisor:
Dr. *I.G. Klugkist*

Preferred journal of publication: JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS
Word count: 5436

Assessing a Bayesian Embedding Approach to Circular Regression Models

Jolien Cremers and Irene Klugkist

Department of Methodology and Statistics, Utrecht University

Abstract

Circular data is different from linear data and its analysis therefore also requires methods that are different from the conventional methods. In this paper the Bayesian embedding approach to estimating circular regression models as presented in Nuñez-Antonio, Gutiérrez-Peña, and Escarela (2011) will be investigated, by means of five simulation studies, in terms of performance, efficiency and flexibility. In addition an empirical example of a regression model predicting teachers' scores on the interpersonal circumplex will be used throughout the paper. From the simulation studies we may conclude that performance is reasonable in most situations and the method is deemed efficient and very flexible. Researchers should however take care when using this method on extreme data and with the interpretation of their results.

Keywords: circular data, regression, Bayesian methods, embedding approach

1 Introduction

Circular data is different from linear data in the sense that it contains information about directions or angles. One may come across circular data in many fields of research. Examples of circular variables include orientations of rock formations, migratory patterns of birds, eye movement patterns and clock times. Within the field of social science, data measured by circumplex measuring instruments is a type of circular data.

The analysis of circular data requires methods that are different from the conventional methods for linear data. As an illustration of circular data and their need for different methods we consider a dataset collected for the research of Mainhard, Brekelmans, Brok, and Wubbels (2011). This research is conducted in the field of educational science and includes the scores of teachers from 48 classes on the interpersonal circumplex as assessed by their students in the first week of the schoolyear (ACS). The interpersonal circumplex consists of the two axes Agency and Communion. Agency summarizes the aspects of status, power, dominance and control and Communion summarizes the aspects of solidarity, friendliness, warmth, and love (Horowitz & Strack, 2011). Figure 1 is a graphical representation of the interpersonal circumplex showing the two orthogonal axes Agency and Communion and the scores of four teachers from our dataset on this circumplex. The two axes split the depicted circle into four quadrants: the first is the upper right quadrant which ranges from $0°$ to $90°$, the second ranges from $90°$ to $180°$,

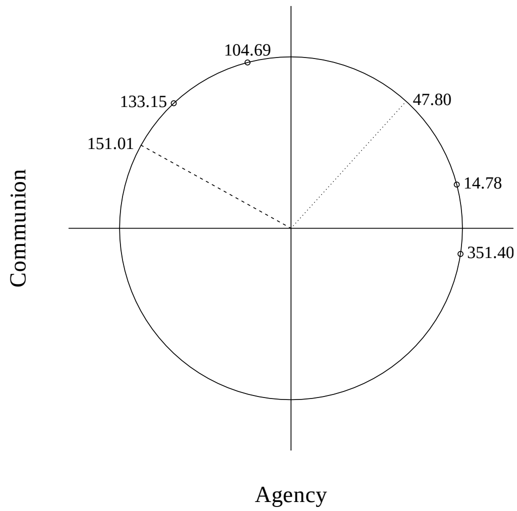


Figure 1: Plot showing the interpersonal circumplex with axes Agency and Communion, where the points indicate the score of a teacher on the circumplex measured in degrees, the dashed line represents the direction of the linear mean and the dotted line represents the direction of the circular mean.

etc. The dataset also includes, amongst others, the teachers' self-perception of their score (ACSSP), their experience measured in years (TEX) and a measure of extraversion (EV). Table 1 shows scores of four teachers from the dataset on these variables. Note that a circular outcome can be measured in degrees on a scale from $0°$ to $360°$, as in Table 1, or in radians on a scale from 0 to 2π radians. Radians can easily be transformed into degrees and vice versa.

The need for different methods can be illustrated by looking at the linear and circular mean of the scores of the first four teachers on the interpersonal circumplex (ACS, see Table 1). The linear mean of these scores is computed by

$$(351.40° + 14.78° + 133.15° + 104.69°)/4 = 151.01°.$$

However, as can be seen from Figure 1, 151.01° is not the average direction of the scores of the 4 teachers. The correct circular mean of the values takes into account the directional nature of the data and is in this specific case computed by

$$\tan^{-1}\left(\frac{\sin(351.40°) + \sin(14.78°) + \sin(133.15°) + \sin(104.69°)}{\cos(351.40°) + \cos(14.78°) + \cos(133.15°) + \cos(104.69°)}\right) = 47.80°$$

(Fisher, 1995).

In this paper regression models for circular data will be considered. In terms of the example data outlined before this would imply predicting the score of a teacher on the interpersonal circumplex by one or more linear or circular predictors (e.g. ACSSP, TEX, EV). As for the computation of the mean of circular data, regression models with a circular outcome need different estimation methods (Fisher, 1995). Mardia and Jupp (1999) and Fisher (1995) provide an overview of frequentist methods to estimate circular regression models. More specifically, Fisher and Lee (1994) and Presnell, Morrison, and Littell (1998) use Expectation Maximization algorithm techniques to obtain Maximum Likelihood estimates for multiple regression models with a circular outcome. The

Table 1: Scores of four teachers on the interpersonal circumplex as assessed by their students and themselves (measured in degrees), as well as the variables teaching experience (centered) and extraversion (centered)

	ACS	ACSSP	TEX	EV
1	351.40°	23.25°	-9.68	-1.02
2	14.78°	246.56°	-4.43	-1.86
3	133.15°	176.78°	-4.43	0.81
4	104.69°	121.31°	5.57	-0.19

downsides of frequentist methods are that they are more complex (Ravindran & Ghosh, 2011) and less flexible than Bayesian (sampling based) estimation methods for circular regression models. Additional advantages of Bayesian methods are that no asymptotic assumptions are required (Gill, 2008; Lynch, 2007), knowledge from previous research can be included in the specification of priors, and that it is not possible to obtain out of range parameter estimates (e.g. negative variance estimates).

Only a few Bayesian methods for estimating parametric circular regression models are available in the literature. Gill and Hangartner (2010) and Rodrigues, Galvao Leite, and Milan (2000) provide Markov chain Monte Carlo (MCMC) methods for circular regression based on the von Mises distribution, which is directly defined on the circle. In the literature, methods using distributions directly defined on the circle are referred to as having an 'intrinsic' approach. Other approaches are the 'wrapping' and the 'embedding' approach which make use of wrapped distributions and projected distributions respectively. Ravindran and Ghosh (2011) and Ravindran (2002) provide an MCMC method for wrapped distributions. Nuñez-Antonio et al. (2011), Wang and Gelfand (2013) and Wang (2013) provide models for a circular response based on the projected normal and general projected normal distribution.

In this paper we investigate the Bayesian embedding approach as presented in Nuñez-Antonio et al. (2011) in terms of performance, efficiency and flexibility by means of simulation studies. Although Nuñez-Antonio et al. (2011) introduce the mathematical model and develop a Bayesian sampler to estimate circular regression models, no systematic simulations to assess the quality of the method have been reported so far.

In Section 2, the model for the Bayesian embedding approach to circular regression will be introduced. This will be done with reference to the Agency-Communion example data introduced before. Sections 3 and 4 contain the methodology and results of the simulation studies for one and multiple predictors. These are followed by Section 5 which presents the interpretation and results from the model fit on the Agency-Communion data. The paper will be concluded with a discussion of the performance, efficiency and flexibility of the approach in Section 6.

2 The Model

In this section the Bayesian embedding approach will be introduced. First, the idea underlying the embedding approach will be explained. Second, the model as used in this paper will be introduced. Third, the Bayesian estimation method as used in the sampler by Nuñez-Antonio et al. (2011) will be explained. Fourth and last, a regression model to be fit onto the Agency-Communion data is introduced.

2.1 The Embedding Approach

As used in this paper, the embedding approach assumes that the circular outcome variable Θ has a projected bivariate normal distribution $PN(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^2$ is a mean vector and $\boldsymbol{\Sigma}$ is a variance-covariance matrix. Throughout this paper we will only consider projected normal distributions in which $\boldsymbol{\Sigma}$ is equal to the identity matrix, $\boldsymbol{I}$. To be able to understand how projecting a distribution works, imagine that we have one bivariate normal outcome variable, $Y \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Conceptually, we can imagine the projection of this bivariate normal outcome variable on the unit circle as shown in Figure 2. The three plots in Figure 2 show datapoints from three bivariate normal distributions with different mean vectors and an identity variance-covariance matrix, $N_2(\boldsymbol{\mu}, \boldsymbol{I})$. Lines are drawn from each datapoint to the origin $(0, 0)$ of the plot. The length of these lines is referred to with r_i for one datapoint θ_i, where $i = 1, \ldots, N$ and N is the sample size, and $\mathbf{r}$ for the vector of all r_i for one dataset. The intersections of these lines with the unit circle can be interpreted as the datapoints of the circular outcome vector $\boldsymbol{\theta}$. In Figure 2 we observe that the datapoints of the circular outcome are closer together, and thus the spread is smaller, for projected normal distributions with means further from the origin.

Projecting bivariate normal data on a circle is relatively easy and produces a circular outcome vector $\boldsymbol{\theta}$ and a vector with distances to the origin $\mathbf{r}$. However, when we start with circular data the process is reversed. Imagine one circular outcome variable measured in angles, the vector $\boldsymbol{\theta}$. Since the outcome variable is angular we might decompose each datapoint θ_i into its sine component, $\sin(\theta_i)$, and its cosine component, $\cos(\theta_i)$, and see those as a bivariate normal outcome. Figure 3 shows a unit circle with one angle, θ_i, decomposed into a sine and cosine component. The distance of this point to the origin is always 1 in a unit circle. However, the datapoints from the underlying bivariate normal outcome can theoretically be located at any distance from the origin and we can thus not easily obtain the r for a datapoint. In fact, we need a Bayesian method that treats this distance as a latent variable. This method is introduced in Section 2.3.

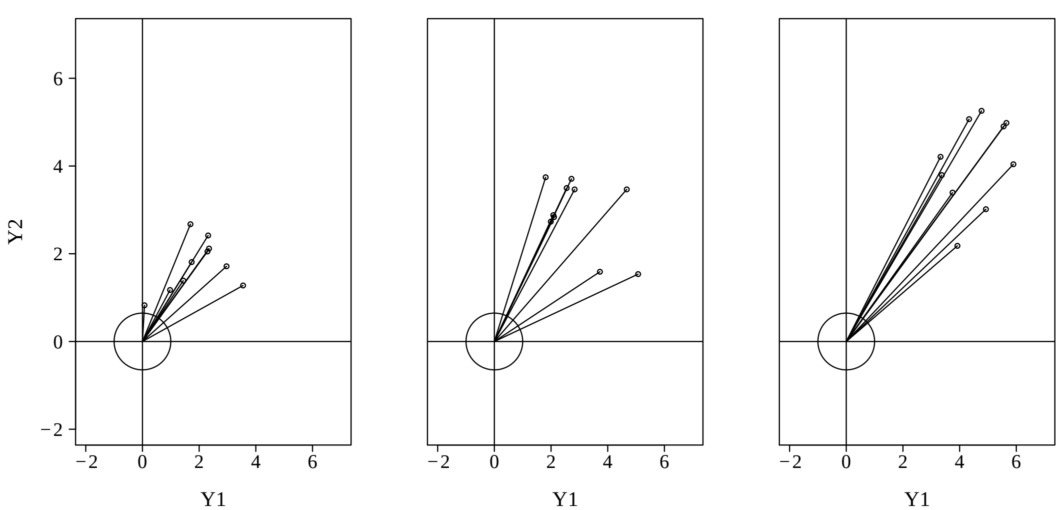


Figure 2: Three sets of bivariate normal data projected on the unit circle. From left to right the respective mean vectors, $\boldsymbol{\mu}$, are: $(1.5, 1.5)$, $(3, 3)$, $(4.5, 4.5)$. The lengths of the lines drawn from each datapoint to the origin represent r and the intersection of the lines with the circle corresponds to the individual values of the circular outcome vector, $\boldsymbol{\theta}$.

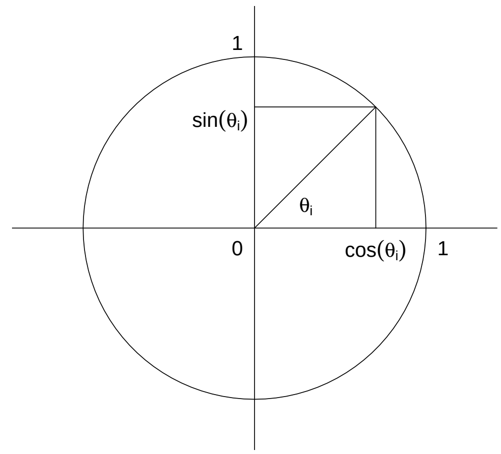


Figure 3: An angular measurement on the unit circle decomposed into its sine and cosine component.

2.2 Circular Regression

In regression models, the projected bivariate normal distribution has the following density (Nuñez-Antonio et al., 2011):

$$PN(\theta|\boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} e^{-\frac{1}{2}||\boldsymbol{\mu}||^2} \left[1 + \frac{\boldsymbol{u}^t\boldsymbol{\mu}\Phi(\boldsymbol{u}^t\boldsymbol{\mu})}{\phi(\boldsymbol{u}^t\boldsymbol{\mu})}\right], \tag{1}$$

where $0 < \theta \leq 2\pi$, $\boldsymbol{\mu}$ is the mean of the projected bivariate normal distribution with identity variance-covariance matrix $\boldsymbol{I}$. Furthermore, $\mathbf{u}$ is the vector $(\cos\,\theta, \sin\,\theta)^t$, and $\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}$ where $\boldsymbol{x}$ is a matrix with predictor variables and $\boldsymbol{B} = [\boldsymbol{\beta^1}, \boldsymbol{\beta^2}]$. The two components of $\boldsymbol{B}$, $\boldsymbol{\beta^1}$ and $\boldsymbol{\beta^2}$, are vectors with regression coefficients. The fact that $\boldsymbol{B}$ has two components means that the sine and the cosine of the outcome variable can be predicted using different regression equations. Lastly, $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function and the probability density function of the standard normal distribution. The relation between the circular and bivariate normal outcomes is defined as $\boldsymbol{U} = \boldsymbol{Y}/R$ where $\boldsymbol{U}$ is a random direction, $\boldsymbol{Y}$ is a random bivariate normal vector and $R = ||\boldsymbol{Y}||$ is the euclidean norm.

If we apply this to the Agency-Communion scores from the example dataset, and assume these to come from a projected bivariate normal distribution, we may then interpret the cosine of the combined score (the first component of $\boldsymbol{u}$) as a score just on the Communion axis and the sine of the combined score (the second component of $\boldsymbol{u}$) as a score only on the Agency axis. We can then compose separate regression equations for the Communion and the Agency scores. The vectors with regression coefficients for these two scores are referred to with $\boldsymbol{\beta^1}$ and $\boldsymbol{\beta^2}$ respectively.

2.3 Bayesian Estimation

In Bayesian analyses, prior distributions have to be specified for all the model parameters. In the circular regression model, a normal prior is specified for the two components of the matrix $\boldsymbol{B}$:

$$N(\boldsymbol{\beta}^j|\boldsymbol{\beta}_0^j, \boldsymbol{\Lambda}_0^j)\ \forall\, j = 1, 2, \tag{2}$$

where $\boldsymbol{\beta}_0^j$ are prior values for the regression coefficients and intercept and $\boldsymbol{\Lambda}_0^j$ is the prior precision matrix of component j. In this paper an uniformative prior was selected in which $\boldsymbol{\beta}_0^j = (0, 0)$ and $\boldsymbol{\Lambda}_0^j = (0.0001, 0.0001)$. Since r is not observed, we consider a latent variable R defined on $(0, \infty)$. This latent variable, together with the prior in (2) and the likelihood of the data results in the following posterior:

$$f(\theta, r|\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}) = N_2(r\boldsymbol{u}|\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}, \boldsymbol{I})|J|, \tag{3}$$

where $|J| = r$ is the Jacobian of the transformation $\boldsymbol{y} \mapsto (\theta, r)$, where $\boldsymbol{y}$ is bivariate normal, and $\mathbf{u} = (\cos\,\theta, \sin\,\theta)^t$. The sampler that can be used to obtain estimates from the resulting posterior was developed by Nuñez-Antonio et al. (2011). It contains the following steps:

1. Starting values for the r_i are chosen. In this paper they are set to 1.

2. The two components of $\boldsymbol{B}$ are sampled from their conditional posterior

$$f(\boldsymbol{\beta}^j|\theta_1, \ldots, \theta_n, \boldsymbol{r}) = N(\cdot\,|\boldsymbol{\mu}_F^j, \boldsymbol{\Lambda}_F^j)\ \forall\, j = 1, 2, \tag{4}$$

   where $\boldsymbol{\mu}_F^j = (\boldsymbol{\Lambda}_F^j)^{-1}(\boldsymbol{\Lambda}_0^j\boldsymbol{\beta}_0^j + (\boldsymbol{X}^j)^t\boldsymbol{y}^j)$, $\boldsymbol{\Lambda}_F^j = \boldsymbol{\Lambda}_0^j + (\boldsymbol{X}^j)^t\boldsymbol{X}^j$ and $\boldsymbol{X}^j$ is a design matrix.

3. Using the estimates for the two components of $\mathbf{B}$ new r_i are generated in a Metropolis-Hastings step from

$$f(r_i|\theta_i, \boldsymbol{\mu}_i = \boldsymbol{B}^t\boldsymbol{x}_i) \propto r_i \exp(-0.5r_i^2 + b_i r_i), \tag{5}$$

   where $0 < r_i < \infty$ and $b_i = \boldsymbol{u}_i^t\boldsymbol{\mu}_i$.

4. Steps 2 and 3 are repeated in MH-within-Gibbs sampler for a specified amount of iterations. After the iterations are completed one should check whether convergence has been reached. If there are problems with convergence one should run additional iterations.

More detail on the method of sampling can be found in Nuñez-Antonio et al. (2011).

2.4 A Regression Model for the Agency-Communion Data

In this section a regression model will be specified for the Agency-Communion data from the introduction. As outcome variable we choose the score of the teachers on the interpersonal circumplex as assesed by their students in the first week of the schoolyear (ACS). The regression equations for the two components of this outcome (Agency and Communion) are specified using the teachers' self assessed score on the interpersonal circumplex (ACSSP), the variable teacher experience (TEX) and the extraversion measure (EV). Since ACSSP is a circular variable we may split it up into its sine and cosine components and use these as two separate predictors in the equation (Fisher, 1995). Experience of a teacher is said to have an influence on the difference between students' and the teacher's own perception of the Agency-Communion score (Wubbels, Brekelmans, den Brok, & van Tartwijk, 2006). Extraversion is a Big Five personality trait that is related to the interpersonal circumplex (DeYoung, Weisberg, Quilty, & Peterson, 2013). It is hypothesized that scores of extravert teachers are located in the first quadrant of the circle. For illustrative purposes, to create different regression equations for the two components, EV will only be used in the regression equation for the Agency component. The resulting regression equations are as follows:

$$\mu_1 = \beta_0^1 + \beta_1^1 \cos(\text{ACSSP}) + \beta_2^1 \sin(\text{ACSSP}) + \beta_3^1 \text{TEX}$$
$$\mu_2 = \beta_0^2 + \beta_1^2 \cos(\text{ACSSP}) + \beta_2^2 \sin(\text{ACSSP}) + \beta_3^2 \text{TEX} + \beta_4^2 \text{EV}$$

where μ_1 is the predicted value for the Communion axis and μ_2 is the predicted value for the Agency axis. The results from the analysis of the data using this regression model are reported in Section 5.

3 Simulations for One Predictor Models

In order to assess the performance, efficiency and flexibility of the method for estimating circular regression models illustrated above, several simulation studies were conducted. This section outlines the methodology and results of the first two studies in which models with one predictor were investigated. For both studies 500 simulated datasets were analyzed. The precise methods for simulation are outlined in Sections 3.1.1 and 3.2.1, for one linear and one circular predictor respectively.

In both studies the convergence of the parameter estimates in the separate designs was checked before conducting simulations. Because of time considerations, designs in which convergence was not reached at 5000 iterations were not included in the simulation study. This is indicated with 'Slow Convergence' in Tables 2 and 3. One should note that although some of the datasets in these designs with slow convergence did converge within 5000 iterations, the majority of datasets did not do so. For the designs with slow convergence, a couple of pilot simulations showed that most of them did converge after 10,000 iterations. In practice this means that data with the characteristics of the slowly converging designs can be analyzed by using more iterations. In general, all designs in which convergence was reached within 5000 iterations of the sampler did so after 500 to 750 iterations. Because convergence in these designs was reached well within 5000 iterations, it was decided to run fewer iterations in the simulation studies. All results are based on 3000 iterations of the sampler of which a burn-in of 750 iterations is not taken into account for the final estimates.

The bias, coverage and mean computation time (MCT) are reported for the designs of both studies. Bias is the deviation of the posterior mean of the parameters from the population value ($Population - Estimated$) averaged over the simulated datasets. Coverage of the credible interval is the percentage of simulated datasets in which the population value lies within the 95% credible interval of the posterior distribution. Computation time is defined as the time it takes to analyze one sample from the population, that is the time it takes to estimate the parameters for one simulated dataset in seconds. These computation times per dataset are averaged over all simulated datasets from the population to obtain the MCT. For the MCT one should note that this is not a very precise measure because simulations were done on a laptop that sometimes also ran other programmes.

3.1 One Linear Predictor

3.1.1 Design

The first simulation study encompasses the situation in which one circular outcome is predicted by one linear variable. The circular outcome vector ($\boldsymbol{\theta}$) was generated by sampling N bivariate normal outcomes $Y \sim N_2(\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}, \boldsymbol{I})$ and subsequently projecting these bivariate outcomes on the circle by using the arctangent (tan^{-1}). The vector composing the predictor matrix ($\boldsymbol{x}$) was sampled from a normal distribution $N(\mu_x, \sigma_x)$.

The parameters that were varied are: the sample size (N), the slope (β_1^1) for the regression equation predicting the cosine component of the outcome, the slope (β_1^2) for the regression equation predicting the sine component of the outcome, and the mean (μ_x) of the predictor variable. Table 2 shows the chosen values. Each row in this table represents one design. The mean of the linear predictor was varied to be able to investigate the influence of the variance of the circular outcome on the performance of the sampler. As was shown in Figure 2, the mean of the circular predictor influences the variance of the circular outcome. Increasing the values in the mean vector of the bivariate normal outcome decreases the spread of the circular outcome variable. The values for the intercepts (β_0^1 and β_0^2) and variance of the linear predictor (σ_x) were not varied and set to 0 and 1 respectively.

3.1.2 Results

The results for the simulation study with one linear predictor are shown in Table 2. We see that a part of the estimates for the intercepts and the regression coefficients is biased. This bias is highest (2.41) in one of the designs with a small sample size and high predictor mean. The results show that the sample size, population mean of the predictor and population β value seem to influence the bias. First of all, the bias is highest for lower sample sizes. Where changing from a sample size of 10 to 50, the bias decreases a lot, it does not do so when changing from 50 to 100. In some of the estimates for the intercepts for these higher sample sizes the bias disappears completely. The bias also decreases with population β values and predictor means that are closer to 0.

In terms of coverage we see that it does not reach the desired 95% level in any of the designs. Coverages lie between 76.4% and 94% and in general the coverage is better for higher sample sizes, population β values that lie closer to 0, data with predictor means closer to 0 and estimates with a lower bias.

In the last column of Table 2 the MCT is reported. It may be observed that the computation time is higher for larger samples but does not seem to vary with other parameters.

3.2 One Circular Predictor

3.2.1 Design

The second simulation study encompasses the situation in which one circular outcome is predicted by one circular variable. When using a circular predictor, we take the centered sine and cosine of this predictor as two separate variables in the regression equation. The circular outcome vector ($\boldsymbol{\theta}$) was generated by sampling N bivariate normal outcomes $Y \sim N_2(\boldsymbol{\mu} = \boldsymbol{B}^t\boldsymbol{x}, \boldsymbol{I})$ and subsequently projecting these bivariate outcomes on

Table 2: Bias and Coverage of the intercepts and coefficients, and Mean Computation Time (MCT) for various simulation designs for the study with one linear predictor

Population values				Bias				Coverage				MCT (s)
β_1^1	β_1^2	μ_x	N	β_0^1	β_1^1	β_0^2	β_1^2	β_0^1	β_1^1	β_0^2	β_1^2	
			10	0.02	-0.29	0.05	-0.33	89.8	81.4	89.6	82.8	14.36
		0	50	-0.01	-0.07	0.01	-0.08	92.0	88.6	90.6	87.6	59.10
			100	0.00	-0.06	-0.00	-0.05	91.6	91.4	92.6	91.0	111.20
			10	-0.04	-0.24	-0.20	-0.29	89.4	89.4	87.8	87.4	15.28
0.5	0.5	-4	50	0.00	-0.07	-0.02	-0.08	87.4	86.8	90.8	90.4	55.15
			100	0.04	-0.05	0.05	-0.05	88.6	87.0	90.2	87.6	108.51
			10	Slow convergence								
		10	50	Slow convergence								
			100	Slow convergence								
			10	-0.01	0.08	0.02	0.03	89.4	84.4	89.8	86.0	21.46
		0	50	-0.01	0.04	0.01	0.03	91.6	90.0	91.8	91.2	56.15
			100	0.00	0.03	0.00	0.02	93.4	90.2	93.0	93.2	116.01
			10	0.20	0.12	-0.00	0.06	86.0	83.2	88.4	87.0	14.40
-0.2	-0.2	-4	50	0.07	0.05	0.04	0.03	89.8	89.0	88.0	87.8	57.27
			100	0.01	0.02	0.02	0.02	91.8	90.6	94.0	93.4	111.59
			10	1.73	-0.08	2.41	-0.14	90.4	90.6	87.2	88.2	15.41
		10	50	-0.03	0.03	0.06	0.03	88.8	88.8	93.0	92.2	55.42
			100	-0.06	0.03	-0.09	0.04	89.8	88.6	90.6	90.6	109.03
			10	-0.02	-1.01	0.01	-1.03	87.8	76.4	86.2	77.4	20.31
		0	50	-0.00	-0.30	-0.00	-0.29	90.8	82.6	91.6	85.0	53.63
			100	-0.00	-0.25	-0.01	-0.24	91.2	77.4	90.4	77.8	112.68
			10	Slow convergence								
2	2	-4	50	Slow convergence								
			100	Slow convergence								
			10	Slow convergence								
		10	50	Slow convergence								
			100	Slow convergence								

Table 3: Bias and Coverage of the intercepts and coefficients, and Mean Computation Time (MCT) for various simulation designs for the study with one circular predictor

Population values						Bias						Coverage						MCT (s)
β_1^1	β_1^2	β_2^1	β_2^2	κ_x	N	β_0^1	β_1^1	β_2^1	β_0^2	β_1^2	β_2^2	β_0^1	β_1^1	β_2^1	β_0^2	β_1^2	β_2^2	
					10	0.14	-0.91	-0.50	0.10	-0.98	-0.40	82.0	70.0	74.8	82.2	73.4	77.8	13.08
				2	50	0.01	-0.12	-0.09	-0.00	-0.14	-0.06	90.0	88.8	90.0	93.2	86.4	90.4	62.96
					100	0.00	-0.08	-0.06	0.00	-0.07	-0.04	92.0	91.6	92.2	91.6	88.4	90.4	107.20
					10	-0.01	0.18	-0.47	0.00	0.30	-0.69	77.2	72.2	75.0	77.6	69.2	74.8	17.83
0.5	0.5	0.5	0.5	10	50	0.00	-0.22	-0.13	-0.00	-0.06	-0.11	93.6	82.0	89.6	89.4	85.4	91.4	57.32
					100	-0.00	-0.22	-0.05	0.00	-0.11	-0.05	93.8	90.0	91.0	92.0	89.4	90.4	118.61
					10	0.01	-0.48	-0.10	0.05	3.18	-0.47	84.6	80.8	78.0	81.4	84.0	80.0	14.66
				60	50	-0.00	0.38	-0.15	0.00	-1.89	-0.12	90.2	89.0	88.2	91.8	87.8	90.4	57.38
					100	-0.00	0.35	-0.04	0.00	-0.27	-0.06	90.8	90.6	91.2	91.8	92.0	91.8	119.30
					10	Slow convergence												
				2	50	-0.01	0.07	0.02	-0.02	0.04	0.05	90.0	88.4	91.4	91.4	90.4	90.6	58.06
					100	-0.01	0.03	0.02	-0.03	0.04	0.03	92.6	92.2	92.0	92.0	91.2	92.4	121.85
					10	-0.02	0.96	0.24	0.02	-0.38	0.15	77.0	72.8	71.6	79.6	69.8	73.6	14.49
-0.2	-0.2	-0.2	-0.2	10	50	-0.00	0.04	0.01	-0.01	0.18	0.05	91.6	84.6	89.6	91.6	86.4	91.4	56.79
					100	-0.01	-0.06	0.04	0.00	0.04	0.05	94.0	91.4	92.0	91.8	91.0	90.8	109.92
					10	0.02	-1.11	0.53	0.04	3.50	0.19	84.4	81.8	78.6	82.6	83.8	79.4	14.66
				60	50	-0.00	0.92	0.05	0.00	-1.60	0.07	89.8	88.6	89.8	92.2	87.6	90.6	53.52
					100	-0.00	0.33	0.06	0.00	-0.10	0.04	91.0	89.2	91.2	90.6	91.2	93.0	123.96
					10	Slow convergence												
				2	50	0.02	-0.36	-0.36	0.01	-0.37	-0.33	91.4	88.0	80.6	91.2	86.0	83.0	59.86
					100	0.00	-0.28	-0.26	0.00	-0.26	-0.25	89.8	87.4	78.6	89.6	87.8	82.6	118.35
					10	Slow convergence												
2	2	2	2	10	50	0.01	-0.34	-0.42	0.00	-0.16	-0.42	93.4	87.8	86.0	89.4	87.2	83.8	59.43
					100	-0.00	-0.31	-0.29	0.00	-0.26	-0.27	91.2	91.2	86.2	90.2	90.8	86.4	112.09
					10	0.02	-0.72	-1.48	0.05	2.74	-1.79	83.4	82.6	81.2	81.0	84.4	78.8	14.67
				60	50	0.00	-0.11	-0.50	0.00	-2.12	-0.51	90.2	87.2	87.0	91.0	88.0	90.0	50.94
					100	-0.00	0.26	-0.26	0.00	-0.67	-0.28	91.0	92.4	89.4	90.8	91.2	90.4	95.38

the circle by using the arctangent ($\tan^{-1}$). The two vectors composing the predictor matrix ($\boldsymbol{x}$) are the centered cosine and sine components of a vector sampled from VM($\mu_{circ,x}, \kappa_x$); a von Mises distribution with circular mean $\mu_{circ,x}$ and concentration parameter κ_x. This concentration parameter is analogous to a variance (Fisher, 1995).

The parameters that were varied are: the sample size (N), the slopes of the centered cosine (β_1^1) and sine (β_2^1) component of the circular predictor for the regression equation predicting the cosine component of the outcome, the slopes (β_1^2, β_2^2) for the regression equation predicting the sine component of the outcome, and a value for the concentration parameter (κ_x) for the circular predictor. Table 3 shows the chosen values, with each row representing one design. The concentration parameter of the circular predictor was varied to be able to investigate the influence of the variance of the circular outcome on the performance of the sampler. The values for the intercepts (β_0^1 and β_0^2) and the circular mean ($\mu_{circ,x}$) were not varied and set to 0 in all designs.

3.2.2 Results

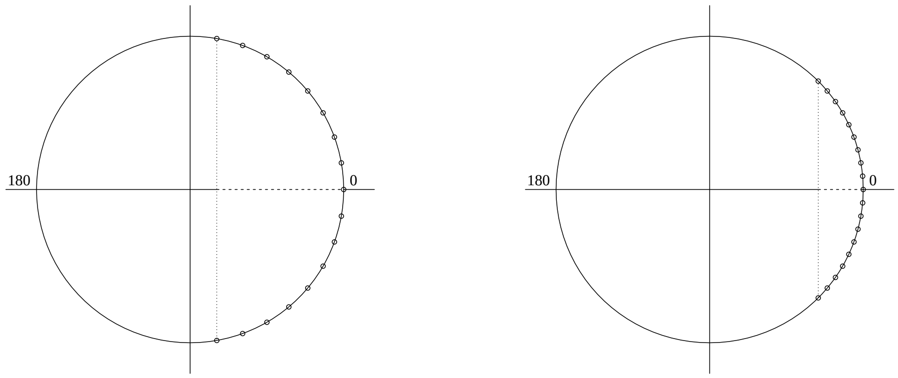
The results for the simulation study with one circular predictor are shown in Table 3. We see that also in this study a part of the estimates for the intercepts and the regression coefficients is biased. This bias is highest (3.50) in one of the designs with a high concentration of the circular predictor and low sample size. The results show that the sample size, concentration of the predictor and population β value influence the bias. The bias is highest for lower sample sizes. When changing from a sample size of 10 to 50, the bias decreases a lot while most of the time it does not do so when changing from 50 to 100. In the designs with a higher sample size the bias of the intercepts often disappears. Furthermore, the bias decreases with population β values closer to 0 and lower concentration. The bias of β_1^1 and β_1^2 is bigger than the bias of β_2^1 and β_2^2 for most designs. This is likely to be caused by the fact that β_1^1 and β_1^2 are the centered cosine components of the circular predictor while β_2^1 and β_2^2 are the centered sine components of the circular predictor. Figure 4 shows that when using a circular mean of 0 for a circular variable, the cosine component of the variable necessarily has less variation than the sine component. Furthermore, with increasing concentration, the spread of the cosine component decreases more rapidly than the spread of the sine component of the circular variable. This may influence the bias and coverage of these parameters.

In terms of coverage we see that the lowest coverage reported is one of 69.2% and the highest is one of 94.8%. The coverage is generally closer to 95% for designs with higher sample sizes and for the sine component of the predictor. An improvement of coverage with population β values closer to 0 or lower concentration is not apparent. We also see that estimates with a lower bias generally show a better coverage.

The MCT is also reported for this study. Again, it may be observed that the computation time is higher for larger samples but does not seem to vary with other parameters.

3.3 Conclusions

Results from the previous two simulation studies have shown that there is a bias in some of the parameter estimates. For the regression coeffiecients this bias is at best 10% of the value of the population β. In most designs the bias and coverage are acceptable. However, we have seen that the bias increases and the coverage is worse for designs with



(a) (b)

Figure 4: Two circles with data that have a lower (a) and higher (b) concentration on a circle, the dotted and dashed line indicate the spread of the sine and cosine of the data.

a high concentration of the circular predictor and a value for the predictor mean further away from 0. Since these parameters influence the concentration of the outcome variable this means that in data with a highly concentrated variable the estimates produced by the embedding approach and their credible intervals may deviate from the truth. Furthermore the bias is bigger for smaller sample sizes, so also in small samples results produced by the embedding aproach may be wrong.

4 Simulations for Multiple Predictor Models

In this section, the methodology and results for three simulation studies with multiple predictors are presented. The first of these is a study with two linear predictors, the second a study with one circular and one linear predictor and the third a study with different regression equations for the two components of the outcome. The parameter values were chosen such that the combinations of parameters that gave relatively good results in Section 3 would return in the studies with multiple predictors. This way it could be investigated whether these combinations still gave good results when changing the amount, type and combination of predictor variables.

Before simulations were run the convergence was checked for each design, as explained in Section 3. All results are based on 3000 iterations of the sampler of which a burn-in of 750 iterations was not taken into account for the final estimates.

4.1 Designs

The simulation methods for the studies with multiple predictors are combinations of the methods for the one predictor models described in Sections 3.1.1 and 3.2.1. The parameters that were varied are the sample size (N) and the slopes. In all designs the mean of the linear predictors (μ_x), the mean of the circular predictos ($\mu_{circ,x}$) and the va-

Table 4: Bias and Coverage of the intercepts and coefficients, and Mean Computation Time (MCT) for various simulation designs for the study with two linear predictors

Population values					Bias						Coverage						MCT (s)
β_1^1	β_1^2	β_2^1	β_2^2	N	β_0^1	β_1^1	β_2^1	β_0^2	β_1^2	β_2^2	β_0^1	β_1^1	β_2^1	β_0^2	β_1^2	β_2^2	
-0.2	-0.2	-0.2	-0.2	50	0.00	0.04	0.03	0.00	0.03	0.04	92.0	90.6	89.8	93.0	89.8	90.0	54.67
				100	-0.02	0.03	0.03	-0.00	0.02	0.02	92.2	92.2	90.8	92.4	93.2	93.2	99.25
0.5	-0.2	0.5	-0.2	50	-0.00	-0.09	0.03	0.00	-0.09	0.04	91.8	89.2	89.8	90.6	87.4	90.2	57.20
				100	-0.01	-0.06	0.03	0.00	-0.06	0.02	93.0	88.6	91.0	92.0	89.6	93.2	102.67
0.5	0.5	0.5	0.5	50	0.00	-0.08	-0.11	0.00	-0.09	-0.09	92.2	91.0	86.8	91.4	87.2	87.8	64.05
				100	-0.01	-0.06	-0.06	0.01	-0.07	-0.07	92.0	90.6	88.4	93.0	87.6	89.6	117.76
-0.2	2	-0.2	2	50	0.00	0.03	-0.35	0.00	0.02	-0.34	91.4	92.8	76.6	91.4	88.0	76.4	61.10
				100	-0.01	0.03	-0.26	-0.00	0.02	-0.26	91.8	89.8	73.2	91.2	94.2	72.4	103.09
0.5	2	0.5	2	50	Slow convergence												
				100	Slow convergence												

Table 5: Bias and Coverage of the regression coefficients, and Mean Computation Time (MCT) for various simulation designs for the study with one linear and one circular predictor

Population values						Bias						Coverage						MCT (s)
β_1^1	β_2^1	β_3^1	β_1^2	β_2^2	β_3^2	β_1^1	β_2^1	β_3^1	β_1^2	β_2^2	β_3^2	β_1^1	β_2^1	β_3^1	β_1^2	β_2^2	β_3^2	
-0.2	-0.2	0.5	-0.2	-0.2	0.5	0.05	0.02	-0.11	0.06	0.02	-0.10	89.0	90.6	85.6	88.8	91.4	88.2	82.81
-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0.04	0.03	0.03	0.06	0.03	0.03	88.4	91.0	89.8	89.4	91.2	91.6	58.97
0.5	0.5	-0.2	0.5	0.5	-0.2	-0.16	-0.11	0.04	-0.13	-0.10	0.03	89.4	89.2	89.0	90.8	91.4	92.2	59.39
0.5	0.5	0.5	0.5	0.5	0.5	-0.16	-0.09	-0.11	-0.11	-0.09	-0.11	85.8	90.0	86.0	88.4	91.6	85.4	58.59

Table 6: Bias and Coverage of the intercepts and coefficients, and Mean Computation Time (MCT) for various simulation designs for the study with different regression equations for the sine and cosine component of the outcome

Population values										Bias									MCT (s)
β_0^1	β_1^1	β_2^1	β_3^1	β_0^2	β_1^2	β_2^2	β_3^2	β_4^2	N	β_0^1	β_1^1	β_2^1	β_3^1	β_0^2	β_1^2	β_2^2	β_3^2	β_4^2	
0	2	-0.002	0.5	0	-0.4	1.5	0.5	-2.5	50	0.02	-0.46	-0.01	-0.11	-0.01	0.09	-0.36	-0.12	0.51	101.26
									100	0.00	-0.29	0.02	-0.07	-0.01	0.05	-0.23	-0.07	0.37	120.57
0	2	2	0.5	0	-0.4	-0.4	0.5	-2.5	50	0.02	-0.43	-0.41	-0.11	-0.01	0.08	0.05	-0.12	0.50	78.03
									100	0.00	-0.29	-0.29	-0.07	-0.01	0.06	0.07	-0.07	0.37	146.20
0	-0.1	1.5	0.5	0	-0.1	1.5	0.5	-2.5	50	-0.00	0.06	-0.31	-0.11	-0.01	0.01	-0.34	-0.12	0.50	77.52
									100	-0.00	0.06	-0.21	-0.07	-0.01	0.01	-0.23	-0.06	0.37	145.84

Population values										Coverage									MCT (s)
β_0^1	β_1^1	β_2^1	β_3^1	β_0^2	β_1^2	β_2^2	β_3^2	β_4^2	N	β_0^1	β_1^1	β_2^1	β_3^1	β_0^2	β_1^2	β_2^2	β_3^2	β_4^2	
0	2	-0.002	0.5	0	-0.4	1.5	0.5	-2.5	50	88.8	85.4	90.0	86.2	91.8	87.8	80.2	88.2	68.2	101.26
									100	91.4	83.8	89.8	89.2	92.0	91.4	82.4	90.2	67.8	120.57
0	2	2	0.5	0	-0.4	-0.4	0.5	-2.5	50	90.0	85.8	75.8	88.4	90.8	87.0	87.6	87.6	68.6	78.03
									100	91.8	85.0	77.4	89.2	91.8	91.8	90.4	89.0	63.4	146.20
0	-0.1	1.5	0.5	0	-0.1	1.5	0.5	-2.5	50	90.4	90.2	80.2	86.8	91.8	89.2	82.8	88.8	69.6	77.52
									100	92.0	88.6	82.4	88.4	89.4	91.6	81.8	90.0	63.8	145.84

lues for the intercepts (β_0^1 and β_0^2) were set to 0. The concentration of the circular predictors (κ_x) and the variance of the linear predictor (σ_x) were set to 2 and 1 respectively. In the study with two linear preditors, the slopes (β_1^2, β_2^2) for the regression equation predicting the sine component of the outcome were varied (see Table 4). In the study with one linear and one circular predictor, the slopes of the centered cosine component of the circular predictor (β_1^1), centered sine component of the circular predictor (β_2^1) and linear predictor(β_3^1) for the regression equation predicting the cosine component of the outcome and the slopes (β_1^2, β_2^2, β_3^2) for the regression equation predicting the sine component of the outcome were varied (see Table 5). The sample size in this study was not varied and is equal to 50 for each design. Lastly, in the study with different regression equations for the two components of the outcome the slopes of the centered cosine component of the circular predictor (β_1^1), centered sine component of the circular predictor (β_2^1) and linear predictors (β_3^1, β_4^1) for the regression equation predicting the cosine component of the outcome and the slopes (β_1^2, β_2^2, β_3^2, β_4^2) for the regression equation predicting the sine component of the outcome were varied (see Table 6).

4.2 Results

The results of the studies with multiple predictors are shown in Tables 4, 5 and 6 and show the same patterns for the bias as observed in the studies with one predictor. The bias is highest for lower sample sizes and it increases with population β values further away from 0. The MCT for the studies with multiple predictors shows the same patterns as in the studies with one predictor, the computation time is higher for larger samples.

In the study with two linear predictors (Table 4) the coverage does not reach the desired 95% level in any of the designs. The highest coverage reported is one of 94.2% and the lowest one of 72.4%. In general the coverage is highest in a sample size of 100. The coverage decreases with population βs further from 0 when holding the sample size constant. In the study with one circular and one linear predictor (Table 5) the results show the same patterns in coverage as the previous studies. The coverage is mostly smaller for values of the population β further from 0. We additionally see that the coverage for the cosine components of the circular predictor (β_1^1 and β_1^2) is in almost all of the cases slightly lower than the coverage of the sine components of the circular predictor (β_2^1 and β_2^2). In the study with different regression equations for the two components of the outcome (Table 6) the results show similar patterns in coverage as the previous studies. The coverage is smaller for values of the population β further from 0 in most cases.

4.3 Conclusions

Results from the previous studies have shown that there is no notable difference in patterns of bias and coverage between models with multiple predictors and models with one predictor. The bias and coverage are acceptable in designs with a value for the population regression coefficient closer to 0 and do not worsen if more and different combinations of predictors are included in the model. This is useful to know as in practice the models investigated in this section show more resemblance, in terms of the amount of predictor variables, to the models estimated to answer empirical research questions.

5 Estimation for the Agency-Communion data

In this section the results of the analysis in which the regression model from Section 2.4 was fit on the Agency-Communion data are presented and interpreted. Teachers with missing values on any of the variables were removed resulting in a sample size of 43. Convergence was reached within 750 iterations. After substracting a burn-in of 750 from a total of 5000 iterations the results in Table 7 were obtained.

The posterior means of the coefficients from the third column of Table 7 tell us more about the linear relations between the predictors and the two outcome components. Figures 5a, 5b, 5c and 5d show the relation by plotting the predicted linear outcome components Agency and Communion against various values of one of the predictor variables. The other predictors are kept constant at their means in the data. The coefficients for the two linear components can be interpreted as usual. For extraversion this interpretation is: 'An increase of 1 unit on the variable Extraversion leads to a 0.34 increase in predicted score on the Agency component'.

The last two columns of Table 7 show the lower and upper bounds for the 95% credible intervals. The interpretation of these is that the probability that the regression coefficient lies within these intervals is 95%. Only for the coefficients Communion Self Perception for the Communion component and Teacher Experience for the Agency component the credible intervals do not include zero and indicate that there is an effect.

To interpret the coefficients in a circular context and thus combine the coefficients of both components is less straightforward. To visualize what happens if we combine the outcome components Figures 5e, 5f, 5g and 5h were constructed. Here the predicted circular outcome is plotted against various values of one of the predictor variables. The other predictors are kept constant at their means in the data. The circular outcome was obtained by using the arctangent of the ratio of the two linear outcomes and transforming this to a measure in radians. For the circular relation plots of Figure 5 the slope is not constant indicating that the relation between outcome and predictor is not linear. This means that a change between two values of the predictor variable may induce a bigger or smaller change in the outcome than a change between two other values of the predictor variable. For a more detailed interpretation we may take a look at Figure 6. This figure is equivalent to Figure 5e but has a smaller range on the predictor variable axis. We see that the difference between a score of 1 and a score of 2 on Communion Self Perception in terms of the predicted outcome is larger than the difference between a score of 99 and 100.

Table 7: Mean and the lower and upper bounds for the 95% credible interval (CI) of the posterior distributions of the intercept and coefficients for the Agency-Communion data

	Parameter	Posterior Mean	Lower Bound CI	Upper Bound CI
Communion	Intercept	0.37	-0.03	0.77
	Communion Self Perception	0.54	0.06	1.01
	Agency Self Perception	-0.13	-0.68	0.42
	Teacher Experience	-0.02	-0.06	0.02
Agency	Intercept	1.66	1.16	2.20
	Communion Self Perception	0.28	-0.36	0.89
	Agency Self Perception	0.50	-0.15	1.19
	Teacher Experience	0.10	0.04	0.15
	Extraversion	0.34	-0.01	0.70

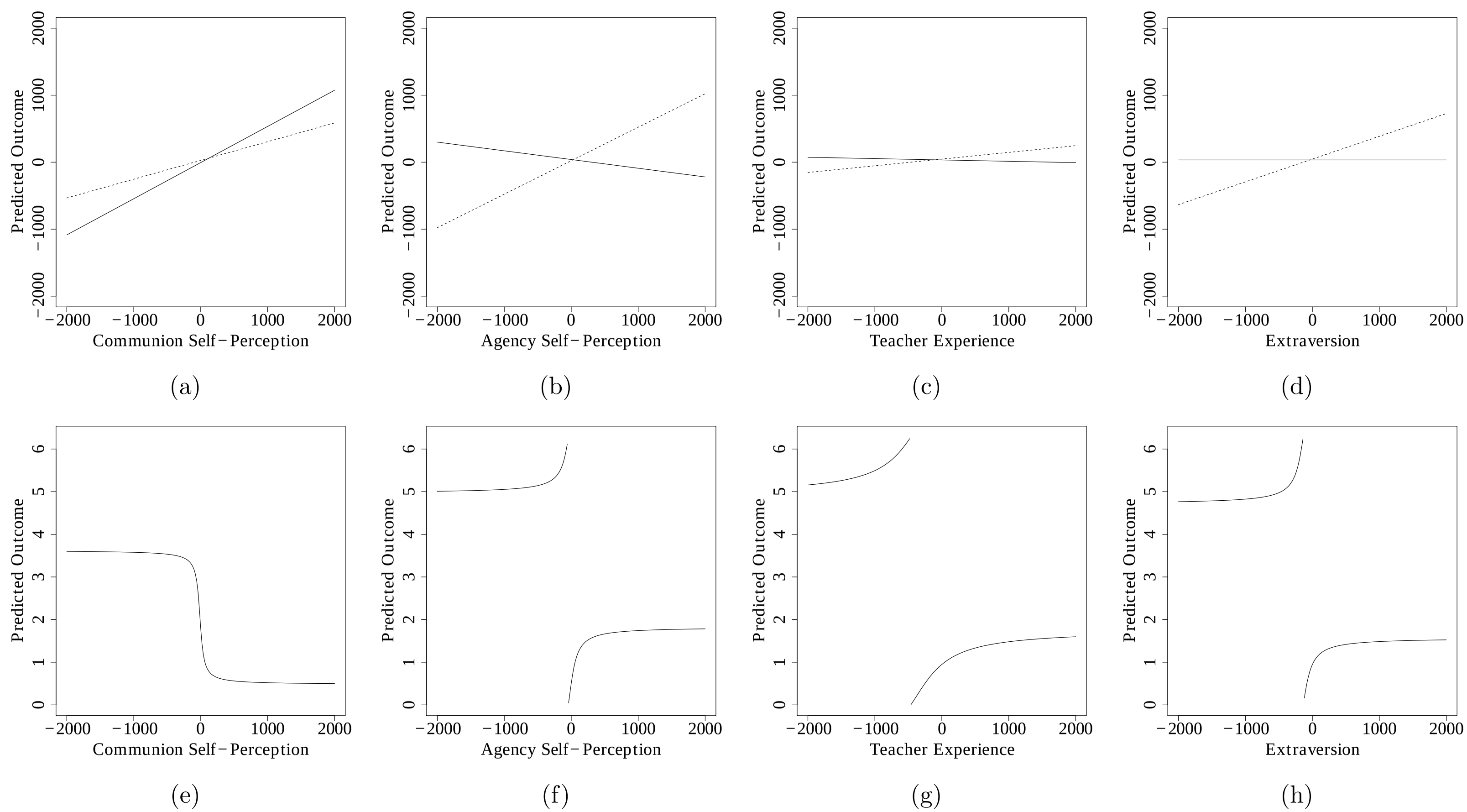


Figure 5: Predicted outcome for changes in 1 predictor. Upper tables are linear outcomes for the separate components (dotted=Agency, continuous=Communion), Lower tables are circular outcomes (measured in radians).

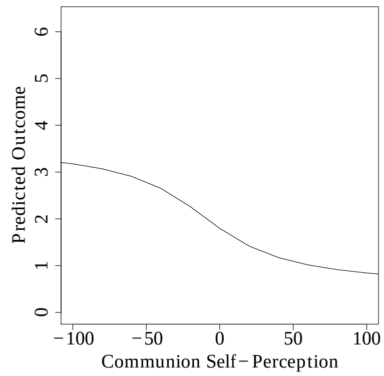


Figure 6: Predicted circular outcome in radians plotted against different values of the predictor variable Communion Self Perception.

6 Discussion

In the current paper the Bayesian embedding approach to circular regression as introduced by Nuñez-Antonio et al. (2011) has been investigated by means of simulation studies. In this section the performance, efficiency and flexibility of the approach will be discussed and the consequences this has for researchers employing the embedding approach will be described.

Concerning the flexibility of the Bayesian embedding approach we may look at several aspects. With regards to the types of regression models it can estimate the approach is very flexible. Both circular and linear predictor variables may be included in the model. Although not discussed in this paper, categorical predictors may also be included by means of creating dummy variables. The effects may then be interpreted by comparing predicted outcomes for persons with and without a score of 1 on these dummies. Furthermore, the two components of the outcome may be predicted by different combinations of variables. Applied researchers who have theories that indicate that one component of a circular outcome may be predicted by different variables than the other component can thus use this method. In terms of interpretation however, the embedding approach may have limited flexibility. Whereas it is straightforward to estimate regression coefficients for two components separately, combining them leads to more complex interpretations. How useful this approach is to applied researchers depends on their specific theoretical interests. When interested in the separate effects on the predictors on the two components of the circular outcome this approach will meet their needs. When interested in a circular interpretation of the effects however, the dependence of its size on the predictor values will complicate interpretation and researchers will have to take care to interpret effects correctly.

From the results of the simulation studies, we may reach several conclusions regarding performance. All of the designs investigated in the simulation studies show bias in the estimated parameters. This bias is dependent on the size of the parameters to be estimated, the sample size, the concentration of the circular predictor, and the mean of the linear predictor. The coverage of the 95% credible interval of the posterior distribution of the estimated parameters shows the same, but less strong, pattern as the bias. Because bias and coverage depend on what kind of data is investigated, researchers should inspect their data carefully before using the embedding approach as described in this paper. If data is too highly concentrated on the circle or has a too small sample size the embedding approach will produce biased estimates that have a low coverage.

Regarding efficiency, for most of the designs, the Bayesian sampler that was used converged well within 750 iterations. The designs that took longer than 5000 iterations to converge were those with a high concentration of the circular predictor, with means of the linear predictor further from 0 or with small sample size. Regarding computation time, the time it takes to analyze one dataset is reasonably low and is higher for designs that have larger sample sizes.

In conclusion, the Bayesian embedding approach to circular regression as introduced by Nuñez-Antonio et al. (2011) has a reasonable performance in sample sizes from about 50. Researchers should take care when analyzing data with very high concentrations of the circular outcome and small sample sizes since in these cases using the approach may result in estimates that contain a sizeable bias. Regarding efficiency and flexibility this approach performs rather well. Estimation time and convergence is fast and with regard to types and amount of predictors that can be included in the regression equations the method is as flexible as possible. A situation in which this approach is not that flexible is when circular effects are interpreted. These effects cannot be computed and interpreted as straightforwardly as the linear effects for the two separate components of the circular outcome.

References

DeYoung, C. G., Weisberg, Y. J., Quilty, L. C., & Peterson, J. B. (2013). Unifying the aspects of the big five, the interpersonal circumplex, and trait affiliation. *Journal of personality*, *81*(5), 465–475.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Fisher, N. I., & Lee, A. J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*(3), 327–339.

Gill, J. (2008). *Bayesian methods: A social and bebehavior sscience approach* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Gill, J., & Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, *18*(3), 316–336.

Horowitz, L. M., & Strack, S. (2011). *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*. Hoboken, NJ: John Wiley & Sons.

Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Mainhard, M. T., Brekelmans, M., Brok, P. d., & Wubbels, T. (2011). The development of the classroom social climate during the first months of the school year. *Contemporary Educational Psychology*, *36*(3), 190-200.

Mardia, K. V., & Jupp, P. E. (1999). *Directional statistics.* New York: Wiley.

Nuñez-Antonio, G., Gutiérrez-Peña, E., & Escarela, G. (2011). A bayesian regression model for circular data based on the projected normal distribution. *Statistical Modelling*, *11*(3), 185–201.

Presnell, B., Morrison, S. P., & Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, *93*(443), 1068–1077.

Ravindran, P. (2002). *Bayesian analysis of circular data using wrapped distributions* (Unpublished doctoral dissertation). North Carolina State University.

Ravindran, P., & Ghosh, S. K. (2011). Bayesian analysis of circular data using wrapped distributions. *Journal of Statistical Theory and Practice*, *5*(4), 547–561.

Rodrigues, J., Galvao Leite, J., & Milan, L. A. (2000). Theory & methods: An empirical Bayes inference for the von mises distribution. *Australian & New Zealand Journal of Statistics*, *42*(4), 433–440.

Wang, F. (2013). *Space and space-time modelling of directional data* (Unpublished doctoral dissertation). Duke University.

Wang, F., & Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. *Statistical Methodology*, *10*(1), 113-127.

Wubbels, T., Brekelmans, M., den Brok, P., & van Tartwijk, J. (2006). An interpersonal perspective on classroom management in secondary classrooms in the netherlands. In J. Fagerberg, D. Mowery, & R. Nelson (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1161–1191). Malwah, NJ: Lawrence Erlbaum Associates.