Advance Access publication June 18, 2010

Political Analysis (2010) 18:316–336
doi:10.1093/pan/mpq009

Circular Data in Political Science and How to Handle It

Jeff Gill
Center for Applied Statistics, Department of Political Science, Washington University, One Brookings Dr., Seigle Hall, LL-085, Box 1203, St. Louis, MO 63130-4899

Dominik Hangartner
Institute for Political Science, University of Bern, Lerchenweg 36, Bern, BE 3012 Switzerland and Center for Applied Statistics, Washington University, One Brookings Dr., Seigle Hall, LL-085, Box 1203, St. Louis, MO 63130-4899

There has been no attention to circular (purely cyclical) data in political science research. We show that such data exist and are mishandled by models that do not take into account the inherently recycling nature of some phenomenon. Clock and calendar effects are the obvious cases, but directional data are observed as well. We describe a standard maximum likelihood regression modeling framework based on the von Mises distribution, then develop a general Bayesian regression procedure for the first time, providing an easy-to-use Metropolis-Hastings sampler for this approach. Applications include a chronographic analysis of U.S. domestic terrorism and directional party preferences in a two-dimensional ideological space for German Bundestag elections. The results demonstrate the importance of circular models to handle periodic and directional data in political science.

1 Overview

There has been no attention to the problem of circular data in political science: those measured angularly to reflect automatic return on the measurement scale. Yet, governments routinely perform tasks based on specific dates, some countries have fixed election periods, legislatures are very cyclical in their activities, migration patterns follow strong directional patterns around the world, and spatial political models are often concerned with the direction of effects. Such iterative data are viewed as sitting on the circumference of a circle and come from measurement on: the 24-hour clock, various calendar types, or compass direction. Because these data recycle through their measure, failing to account for the circular nature would introduce a misleading break in the support of the data.

Why is it deleterious to ignore circular effects and simply use a bounded interval measure to model the data? Statistical summaries and models will produce results that mislead or misinform readers. Here, we develop a circular regression model for domestic terrorism events that finds within-year cycles tied to particular attacking groups and particular target groups. The key is that the political nature of these entities is an important determinant of the cyclical timing of attacks, controlling for other covariates, and this would be missed with conventional approaches. In our second study, we look at German Bundestag elections and find that in a specific two-dimensional ideological space, the parties make direction

© The Author 2010. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

decisions *relative to each other* and in response to social and political pressure from the electorate. The developed Bayesian circular regression model captures these relationships in a way that would be missed otherwise since it provides a direct modeling link between the covariates and the circular direction changes.

Consider the analysis of crime data in an urban environment. We know that crime rates at 11:15 p.m. are related to crime rates at 10:45 p.m., which is logical since such activities may be continuing, the same individuals are likely to be present, and the recorded time of the crime may be imprecise. However, treating clock time as the bounded interval 0–24 for hours means that events occurring at 12:15 a.m. appear to be temporally unrelated to events at 11:45 p.m. This is clearly an unrealistic assumption and is not reflective of the problem being studied. It is functionally equivalent to the notion that social actors ''reset'' at midnight and start some cycle anew. Such cycling is the so-called wraparound effect, also known as the crossover problem, which exists because there is no minimum or maximum on the circle. Consequently, using standard summary statistics, such as the mean or variance, produces nonsensical results.

It is important to distinguish between cases when time is considered as a continuing linear variable, which is best modeled in a time series context, and cases in which interest lies in analyzing the cyclical repeating effects. Circular data analysis is analogous to, but different than, an *autoregressive integrated moving average* process (Hamilton 1994, 437) with the assumption that cyclical effects are constant and not trending given the modeled factors. Time series models are able to estimate the seasonal pattern of the data, that is, the marginal unconditional seasonality. However, only the circular regression model is able to model the conditional expectation of the seasonal pattern as a function of one or more explanatory variables in an *intuitive* regression context on the circular metric of the data.

Circular data analysis, and more generally spherical data analysis, has been practiced in astronomy, demography, geology, geography, meteorology, earth sciences, oceanography, and biology in particular. Perhaps the most common application is in ornithology where the nest orientation (Bergin 1991; Squires and Ruggiero 1996), migration direction (Beason 1980), or general flight pattern (Schmidt-Koenig 1963; Matthews 1974; Bryan and Coulter 1987) of birds is studied with respect to compass direction. Bartels (1984) does a similar analysis of a much slower animal (the slug). In demography, circular data analysis has been used to look at geographic marital patterns (Coleman and Haskey 1986), occupational relocation in the same city (Clark and Burt 1980), and settlement trends (Upton 1986b). The seminal citation for circular data analysis in biology (and beyond) is Batschelet (1981), and another standard citation is Upton and Fingleton (1989). General interest in the statistical analysis of circular data was invigorated by the publication of Mardia's (1972, 1975b) very thorough reviews, which led to the primary era of theory development in the 1970s and 1980s.

2 Graphical Analysis and Summary Statistics

Failing to account for the reiterative nature in graphics deceives viewers because it appears that there are explicit endpoints. For example, consider data on gun crimes committed in Pittsburgh, Pennsylvania, over the period 1987–98. These data are recorded at the hourly level on the 24-h clock (Cohen and Gorr, ICPSR Study Number 2895) and certainly affect local political decision making about law enforcement priorities. The left panel of Fig. 1 shows these events in a *linear* histogram. It appears that the events start at a high level, decrease, and then return to a high level at the ends. What this histogram approach misses is the connection of the endpoints: gun crimes occur in much higher levels between 10 p.m.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

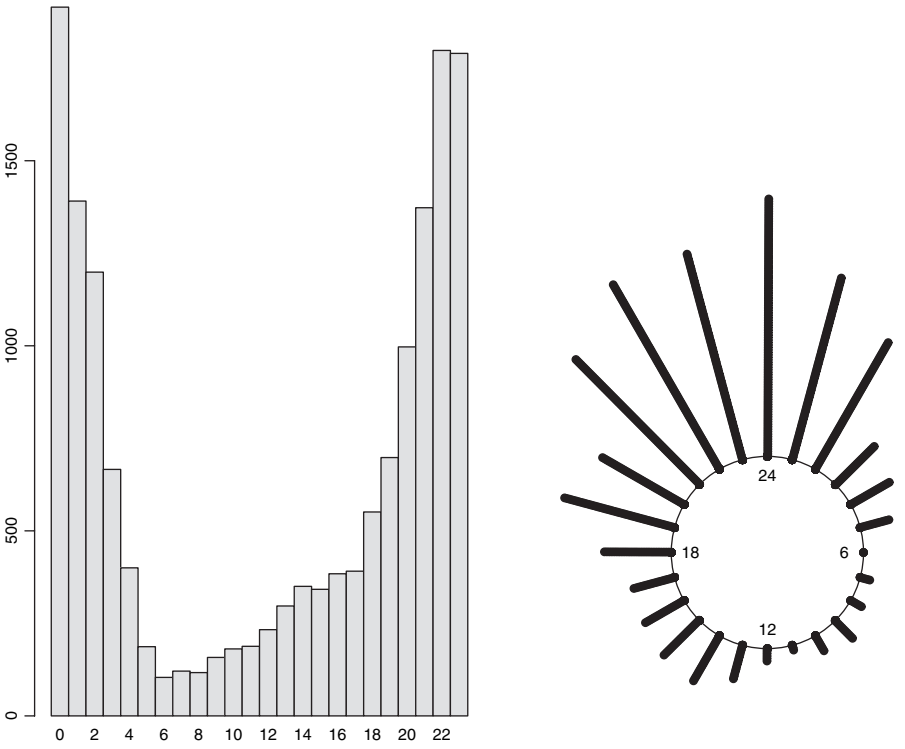


Fig. 1 Two displays of gun crime data.

and 2 a.m. Conversely, the *angular* histogram in the right panel of Fig. 1 accurately depicts the late-night association that occurs across the midnight boundary.

A related figure is the *rose diagram*, which is an angular histogram formed such that the bins are projected as connected triangles emanating from the origin. The radius of each bin is proportional to the square root of the relative frequency of the group depicted in that bin. This means that the area of the displayed triangle is proportional to that group's fraction of the total. The greater the number of bins, the smoother in general a rose diagram appears. In Fig. 2, we show a rose diagram again for the Pittsburgh gun crime data, with jittered hourly values around the enclosing circle, which shows in dramatic fashion the upturn centered at midnight.

We can also use circular nonparametric smoothing for the gun violence data. Figure 3 highlights smooth continuity in the high-event areas across the p.m. to a.m. boundary. Notice also the slight asymmetry favoring the left side of the circle. Although there are fewer gun crimes in the afternoon and mornings relative to late at night, there are slightly more in the afternoon than in the morning, which makes intuitive sense. The density estimation differs from the right panel of Fig. 1 in that it provides a smoothing of the period between hourly measurements. The angular histogram in Fig. 1 implies events directly on the hours since that is how the police recorded the event. We know, however, that these gun crimes are not actually scheduled as such, and therefore, the density estimation in Fig. 3 is more realistic.

In addition to graphical approaches, the initial exploration of data includes summary statistics. As already mentioned, naively applying linear statistics leads to misleading

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

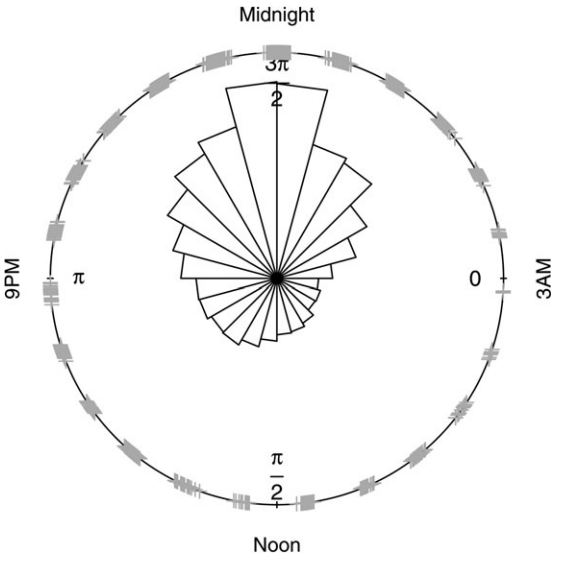


Fig. 2 Rose diagram of gun crime data.

summaries: the linear mean for the Pittsburgh data in radians is 3.41, which corresponds to a little after noon, and the linear variance for these data, 6.15 in radians or 10 hours and 30 min, makes even less sense where dispersion from a low point gives great weight in the "tails." As seen in the left panel of Fig. 1, the two modes at either end of the graph roughly balance out, with a slightly greater mass later in the day, thus explaining this mean. We certainly know already that this is misleading for circular data. The circular data mean, also called the *mean direction*, is necessarily more complicated. Start with defining the following two mean quantities from the sample of angles, hereafter denoted by **Z**:

$$C = \frac{1}{n}\sum_{i=1}^{n}\cos(z_i) \quad S = \frac{1}{n}\sum_{i=1}^{n}\sin(z_i). \tag{1}$$

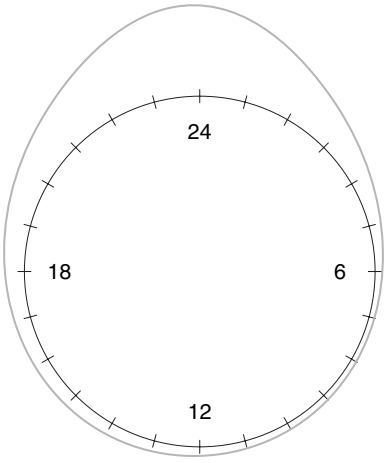


Fig. 3 Density estimate for gun crime data.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

The sample mean direction then accounts for the wrapping:

$$\bar{z} = \tan^{-1}(S/C) + I(C<0)\pi + I(S<0, C>0)2\pi, \tag{2}$$

where $I()$ is 1 if the enclosed condition or conditions are met and 0 otherwise. The circular mean for these data is -0.21 radians, which corresponds to the time of 10:15 p.m.

The circular variance is based on tying the tails of the circle together. Using the definitions from above, the circular variance is defined as follows:

$$\text{Var}(z) = 1 - \frac{C^2 + S^2}{n}, \tag{3}$$

which is bounded by [0:1]. The interpretation is the same as the linear variance, subject to the limits, where values close to 0 indicate values concentrated near the mean and values close to 1 indicate highly dispersed values around the circle. Unfortunately, $\text{Var}(z) = 1$ does not mean that the values are uniformly dispersed around the circle. In addition, due to the form of the trigonometric moments (Fisher 1993), the standard deviation is a little more complicated than the square root of the variance, $\text{SD}(z) = \sqrt{-2\log(1-\text{Var}(z))}$. The circular variance and standard deviation for the Pittsburgh data are $\text{Var}(z) = 0.42$ (about 1.5 h) and $\text{SD}(z) = 1.04$ (almost 4 hours). Unlike the circular variance, the circular standard deviation has the same support as the linear variance or standard error (note the logarithm in the formula) and is therefore preferred by some researchers to the variance.

3 The von Mises Distribution for Circular Data

There are a number of ways to specify distributions for circular data. Obviously, any effective choice must have the property that the support of the data resets as it increases. Specifically, consider a continuous form, $f(z)$, with the requirements that

$$0 \leqslant z < 2\pi, \quad \text{and} \quad \int_0^{2\pi} f(z)\mathrm{d}z = 1, \tag{4}$$

such that z only takes on values of $\mathscr{R}$ in the interval $[0:2\pi)$. Any constant works here instead of 2π, but this maximum value makes the use of trigonometric functions particularly easy since z now specifies the direction of cosines in radians. We could use degrees (0–360) instead, but radian measurement is vastly more convenient. Normally, the circle is specified also with unity radius, without any loss of generality since the angle is the real measure of interest. Noncircular manifolds are mathematically feasible, but they make the intuition and calculations unnecessarily complicated for statistical purposes. With the additional stipulation that

$$f(z+2\pi) = f(z), \tag{5}$$

we have a generic form that meets the circular requirement, which is referred to as *rotational* because of this last property above. We now need two additional rules that are not technically tied to the definition of a probability density function (PDF) but are key conveniences for modeling circular data. First, it is important to pick a starting point in $[0:2\pi)$. Typically, the north or east points are the zero level (equivalent to 12 o′clock and 3 o′clock points). Any point is acceptable, but other choices may be more difficult to explain to readers. Second, the direction needs to be imposed: clockwise or counterclockwise. Usually, the trigonometric convention is applied, and z values increase in the clockwise direction. There is rarely a motivation to change these rules.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

The distribution introduced by von Mises (1918) dominates in the recent analysis of circular data because it is flexible with regard to the effect of parameters and easy to interpret. This distribution plays the same central role in the analysis of circular data as does the normal distribution for linear data, and in some ways, it resembles the *wrapped* normal (Downs and Gould 1967; Collett and Lewis 1981). The von Mises PDF for random variable z with mean direction μ and concentration parameter κ is given by

$$f(z) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(z-\mu)], \quad 0 \leqslant z, \quad \mu < 2\pi, \quad \kappa > 0, \tag{6}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind (order 0):

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp[\kappa \cos(\phi-\mu)] \mathrm{d}\phi. \tag{7}$$

As $\kappa \to \infty$ or as the data become more tightly clustered, this distribution converges to a wrapped normal with mean μ and variance $1/\kappa$. The support of z is sometimes specified as $[-\pi{:}\pi]$, but this is obviously an arbitrary decision. Theoretical descriptions of the von Mises distribution are given by many authors: Bagchi and Guttman (1988), Cotterill and Csorgo (1982), de Waal (1979), Gruet (2000), Lwin (1975), Gordon and Hudson (1977), and Mardia (1975a). The Bessel function is a major inconvenience here and a number of works have worried about parameter estimation for the von Mises (Bingham and Mardia 1975; Hill 1976; Best and Fisher 1981; Abeysasekara and Collett 1982; Bartels 1984; Chan and He 1993).

From a random sample $z_1, z_2, \ldots, z_n$, we can calculate C and S according to equation (1) so that the maximum likelihood estimate (MLE) of κ is given (Dobson 1978, Downs and Mardia 2002) by the solution from

$$A(\hat{\kappa}) = \left(C^2 + S^2\right)^{\frac{1}{2}}, \tag{8}$$

where the function $A(x) = I_1(x)/I_0(x)$ is defined with $I_1(x)$, the modified Bessel function of the first kind (order 1; see Abramowitz and Stegun 1972). Meaning that we need

$$\hat{\kappa} = A^{-1}\left(\left(C^2 + S^2\right)^{\frac{1}{2}}\right). \tag{9}$$

This is sufficiently unwieldy that tabular lookup was used historically and we are now advantaged by software. Unfortunately, the MLE here is also biased for finite samples (Upton 1986a). Defining

$$R^2 = [nI_1(x)/I_0(x)]^2. \tag{10}$$

Schou (1978) and Batschelet (1981) tabulate unbiased values of $\tilde{\kappa}$, which is the solution for $A(\kappa) = RA(R\kappa)/n$ when $R^2 \geqslant n$. These authors recommend that $\tilde{\kappa} = 0$ when $R^2 < n$. A very useful set of approximations is given by Fisher (1993, 88), according to

$$\hat{\kappa} = \begin{cases} 2(R/n) + (R/n)^3 + 5(R/n)/6 & \text{for} \quad R < 0.53n, \\ -0.4 + 1.39(R/n) + 0.43(1-R/n) & \text{for} \quad 0.53n \leqslant R < 0.85n, \\ 1/((R/n)^3 - 4(R/n) + 3(R/n)) & \text{for} \quad R \geqslant 0.85n, \end{cases} \tag{11}$$

which needs to be adjusted for small sample size and small R/n:

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

$$\hat{\kappa}_{\text{adj}} = \begin{cases} \max\left(\hat{\kappa} - \frac{2}{n\hat{\kappa}}\right) & \text{for} \quad \hat{\kappa} < 2, \\ \frac{(n-1)^3 \hat{\kappa}}{n^3 + n} & \text{for} \quad \hat{\kappa} \geqslant 2. \end{cases} \tag{12}$$

The MLE for the mean direction is the value $\hat{\mu}$ that satisfies the equations

$$\hat{\mu} = C/R \quad \hat{\mu} = S/R, \tag{13}$$

with C, S, and R defined as above. Upton (1973) gives a likelihood ratio test statistic for $H_0{:}\mu = \mu_0$ versus $H_1{:}\mu \neq \mu_0$, provided that κ is not large, that rejects H_0 if

$$R^2 > X^2 + \frac{\left(2n^2 - X^2\right) Z_\alpha}{4n}, \quad X^2 = \left[R\cos\left(\mu_0 - \hat{\mu}\right)\right]^2 \tag{14}$$

for an α significance level. See also the likelihood ratio test for von Mises mixtures given by Grimshaw, Whiting, and Morris (2001), as well as multisample tests from Stephens (1972) and Upton (1976). Alternately, Damien and Walker (1999), Mardia and El-Atoum (1976), and Bagchi and Kadane (1991) provide (limited) Bayesian treatments to inference and model comparison with circular data. Upton (1986a) extends his hypothesis test to derive two confidence intervals for μ:

$$\begin{aligned} &\hat{\mu} \pm \cos^{-1}\left[\left(\frac{4nR^2 - 4n^2 Z_\alpha}{4nR^2 - R^2 Z_\alpha}\right)^{\frac{1}{2}}\right] && \text{for} \quad R \leqslant 0.9n, \\ &\hat{\mu} \pm \cos^{-1}\left[\left(\frac{n^2 - \left(n^2 - R^2\right)\exp\left(Z_\alpha/n\right)}{R^2}\right)^{\frac{1}{2}}\right] && \text{for} \quad R > 0.9n \end{aligned} \tag{15}$$

Under the normal approximation, this reduces to the simpler form

$$\mu \pm \left(Z_\alpha / \left(R\hat{\kappa}\right)\right)^{\frac{1}{2}}. \tag{16}$$

Finally, note that the R package `circular` (Ulric Lund and Claudio Agostinelli) provides many of these calculations using the von Mises distribution.

4 A Model for Circular Regression

So far we have only discussed how to graph and analyze univariate circular data. However, political scientists are really interested in connecting a set of explanatory variables to some outcome of interest with regression-style models. To specify a linear-additive right-hand side structure and its estimated effect on some circular outcome variable, we need to develop the appropriate generalized linear model (GLM) structure in the conventional McCullagh and Nelder (1989) sense. Since GLMs connect the nonlinear outcome variable to the linear-additive contribution from the explanatory variables through a *mean function*, the mistake of calculating linear means from circular data (section 3) applies to regression settings as well.

The existing literature on circular statistics provides little guidance for building standard regression models with circular outcomes, and we may want to reparameterize the von Mises distribution in terms of a set of possible explanatory variables: (1) the mean direction μ in terms of a vector of covariates, (2) the dispersion κ in terms of covariates, or (3) model

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

both the mean direction and the dispersion simultaneously. We focus here on the most important model, (1), thereby making the testable assumption that all concentration parameters are homoscedastic, that is, $\kappa_i = \kappa$, $\forall i$. Following Fisher and Lee (1992), our model assumes that the individual mean directions μ_is are related to the covariates by a link function $g^{-1}()$ such that

$$\mu_i = \mu + g^{-1}(\mathbf{x}_i\boldsymbol{\beta}), \tag{17}$$

where $\mathbf{x}_i$ is the ith row vector of the explanatory variable matrix, $\mathbf{X}$, corresponding to the ith observation, and $\boldsymbol{\beta}$ is a vector of regression coefficients. The function $g^{-1}()$ maps the real line to the scale of the outcome variable, which is a circle or sphere. We consider only monotone link functions having the following two properties: (1) as $\mathbf{x}_i\boldsymbol{\beta}$ ranges from $-\infty$ to $+\infty$ and $g^{-1}(\mathbf{x}_i\boldsymbol{\beta})$ ranges from $-\pi$ to $+\pi$ and (2) $g^{-1}(0) = 0$, thereby preserving the interpretation of the constant μ as the origin. Functions having these two properties are a generalization of the Johnson and Wehrly (1978) model and are referred to as ''angular monotonic'' (Fisher and Lee 1992).

Possible choices for $g^{-1}()$, which satisfy the restrictions (1) and (2), are discussed in Jammalamadaka and Sengupta (2001) and Fisher and Lee (1992). Here, we focus on a specific choice, namely

$$g^{-1}(z) = 2\arctan(z). \tag{18}$$

Using this link function and assuming our sample of size n is i.i.d., we can write the log-likelihood function for the homoscedastic von Mises regression model according to

$$\ell(\boldsymbol{\beta}|\mathbf{X}) = -n\log(2\pi I_0(\kappa)) + \kappa\sum_{i=1}^{N}\cos(z_i - \mu - 2\arctan(\mathbf{x}_i\boldsymbol{\beta})). \tag{19}$$

As we will show in simulations, this log-likelihood function is not globally concave. However, given suitable starting values, the iterative reweighted least squares algorithm (IRLS) of Green (1984) can successfully be applied to this optimization problem. To estimate the unknown parameters μ, $\boldsymbol{\beta}$, and κ, define first the following intermediate quantities (Fisher and Lee 1992; Fisher 1993):

$$\mathbf{u}' = (u_1, \ldots, u_N) \text{ with elements } u_i = \sin(z_i - \mu - g^{-1}(\mathbf{x}_i\boldsymbol{\beta})),$$

$$\mathbf{X} = [\mathbf{x}_i, \ldots, \mathbf{x}_n]',$$

$$\mathbf{G} = \operatorname{diag}(g^{-1}(\mathbf{x}_i\boldsymbol{\beta}), \ldots, g^{-1}(\mathbf{x}_n\boldsymbol{\beta})). \tag{20}$$

Similar to the previous section, we (re-)define

$$S = \frac{1}{n}\sum_{i=1}^{n}\sin(z_i - g(\mathbf{x}_i\boldsymbol{\beta})) \quad C = \frac{1}{n}\sum_{i=1}^{n}\cos(z_i - g(\mathbf{x}_i\boldsymbol{\beta})), \tag{21}$$

and relabel the quantity from the right-hand side of equation (8) as $R = (S^2 + C^2)^{\frac{1}{2}}$. Now, the MLEs are solutions to the equations

$$\mathbf{X}'\mathbf{G}\mathbf{u} = 0, \quad R\sin(\hat{\mu}) = S, \quad A(\hat{\kappa}) = R, \quad R\cos(\hat{\mu}) = C, \tag{22}$$

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

where the last equation comes from equation (21). Fisher and Lee (1992) propose to solve this system of equations with the following iterative method:

1. Start with initial values for $\hat{\boldsymbol{\beta}}$; calculate S, C, R, $\hat{\mu}$, and $\hat{\kappa}$.
2. Insert these values into equation $\mathbf{X}'\mathbf{G}\mathbf{u} = 0$ to produce $\hat{\boldsymbol{\beta}}^{+}$ as an update of $\hat{\boldsymbol{\beta}}$.
3. Solve $\mathbf{X}'\mathbf{G}\mathbf{u} = 0$ iteratively using the updating equations:

$$\mathbf{X}'\mathbf{G}^2\mathbf{X}\left(\hat{\boldsymbol{\beta}}^{+} - \hat{\boldsymbol{\beta}}\right) = \mathbf{X}'\mathbf{G}^2\mathbf{r}, \tag{23}$$

where $\mathbf{r}$ is a vector of length n with elements $r_i = u_i / \left\{A(\hat{\kappa}) g'\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)\right\}$.

4. Update $\hat{\mu}$ and $\hat{\kappa}$ from $R\sin(\hat{\mu}) = S$ and $R\cos(\hat{\mu}) = C$.

This process is then repeated with the new values until convergence is obtained. Fisher and Lee (1992) also derive the large-sample variance of these MLEs as follows:

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{\kappa A(\kappa)}\left(\left(\mathbf{X}'\mathbf{G}^2\mathbf{X}\right)^{-1} + \frac{\left(\mathbf{X}'\mathbf{G}^2\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{g}\boldsymbol{g}'\mathbf{X}\left(\mathbf{X}'\mathbf{G}^2\mathbf{X}\right)^{-1}}{n - \mathbf{g}'\mathbf{X}\left(\mathbf{X}'\mathbf{G}^2\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{g}}\right), \tag{24}$$

where $\mathbf{g}$ is a vector whose elements are the diagonal elements of $\mathbf{G}$. The variance of $\hat{\kappa}$ is given by

$$\mathrm{Var}(\hat{\kappa}) = \frac{1}{nA'\left(\hat{\kappa}\right)}, \quad \text{where } A'(\kappa) = \frac{\partial A(\kappa)}{\partial \kappa} = 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa). \tag{25}$$

This algorithm is implemented in our R package to be hosted at the Comprehensive R Archive Network (along with our new Bayesian approach below).

5 Monte Carlo Simulation Study

To build intuition, we illustrate the likelihood version of the homoscedastic von Mises regression model using a small Monte Carlo simulation. We draw three samples, each of size $n = 1000$ from the von Mises PDF: $z_i \sim \mathrm{VM}(\mu_i, \kappa)$, with $\mu_i = \mu + \arctan(\mathbf{x}_i\boldsymbol{\beta})$ and x values uniform on $(-1, 1)$. For three different contrived but realistic parameter groups, Fig. 4 shows the scatter plots of the raw data (first row), kernel densities on the circle (second row), and the corresponding log-likelihood functions as a function of the estimated $\boldsymbol{\beta}$ coefficients where the $\hat{\boldsymbol{\beta}}$ coefficients are estimated using the IRLS algorithm (third row). The first row visualizes possible patterns in the plot of (z, x) and the additional points $(z + \pi, x)$ in Cartesian coordinates. For the two strongly nonuniform cases, we see the effect of ''rolling'' past the zero point as the mode determines a well-defined point cloud. The second row shows that these are not strongly modal forms despite the patterns in the first row. The third row is the most informative because it demonstrates the difficulty in naively applying a mode-finding algorithm, and this is why we recommend always using multiple starting points to find the global maxima (dashed vertical line).

The first case (column) of Fig. 4 shows that if the true $\boldsymbol{\beta}$ is equal to or near zero, the log-likelihood function not only has a peak near $\hat{\boldsymbol{\beta}} = 0$ but also asymptotes out to $\pm\infty$ as $\boldsymbol{\beta}$ gets big in absolute value. In such a case, the only practical estimate of the mode is the peak near zero, which is the solution produced by equation (4). Despite this seemingly arbitrary choice, the parameter $\hat{\boldsymbol{\beta}}(\mathrm{mod}2\pi)$ is fully identified in the mathematical sense,

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

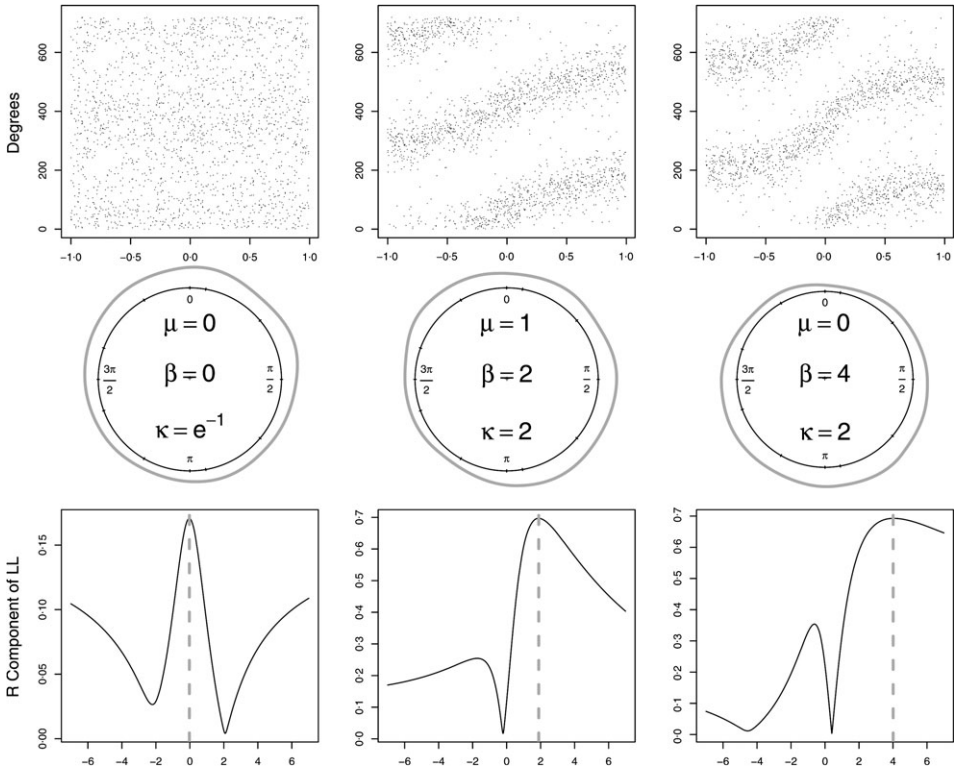


Fig. 4 Simulation summaries for the von Mises distribution.

as opposed to the econometric sense. This difference in identification definitions is a direct result of wrapping around the circle.

If the true $|\boldsymbol{\beta}| \gg 0$, as given in the second and third cases (columns) of Fig. 4, then the R component of the likelihood function is not globally concave and local maxima can exist quite close to the global maxima. Inspection of R as a function of $\boldsymbol{\beta}$ usually produces good starting values for maximization process. In many applications, the coefficients of a simple linear regression give surprisingly close starting values for this process. If computationally feasible—which critically depends on the dimension of $\mathbf{x}_i$—we recommend using a grid of starting values for subsequent runs of the IRLS algorithm. Given the lack of global log concavity, a careful exploration of the likelihood surface is always highly advisable when using the likelihood version of the circular regression model.

6 A Bayesian Circular Data Regression Model

We now derive an easy-to-use Bayesian version of the circular regression model. To date, efforts to produce a general Bayesian framework for the regression analysis of circular data have not made much progress. There are considerable difficulties in producing marginal posterior distributions from the joint posterior due to the mathematical forms of the modified Bessel function of the first kind and the arctan link function. Some work has been done to deal with these challenges, but there is no general approach to producing an analytical or Markov chain Monte Carlo (MCMC) solution. Here, we address this deficiency.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Nearly, all Bayesian approaches to circular data modeling do not include covariates. Tover and Ruiz-Rivas (1986) produce a Bayesian construct for the mean direction only with the goal of handling outliers, but outliers are not as important in circular data as in linear data since they wraparound to the other side of the origin if large enough, thus losing their outlier status. In another early piece without covariates, Guttorp and Lockhart (1988) use the Bayesian circular model as a way to use prior distributions in radio direction finding. Another paper demonstrates how Fourier transforms can help with the estimation process for the mean direction (Healy and Kim 1996). Ghosh, Jammalamadaka, and Vasudaven (1999) focus on estimating change points. More recently, Morelande (2008) applies the mean-only Bayesian approach to an engineering setting. The closest paper to producing a general Bayesian regression approach is that by Damien and Walker (1999). They introduce latent variables and avoid all major restrictions in the process of getting a workable Gibbs sampler, which unfortunately is setup only for μ and κ (our notation). In fact, these authors state: "But, to the best of our knowledge, there is no paper in the literature that provides a full Monte Carlo solution to the problem, despite the low-dimensional nature of the model." We solve this problem by introducing a hybrid MCMC algorithm that contains subcomponents that deal with the two difficult parameters. The Metropolis-within-Gibbs sampler (Robert and Casella 2004) on each cycle first draws (easily) a conditional Von Mises for the mean vector, μ, a conditional normal vector for the regression coefficient parameters, $\boldsymbol{\beta}$, and then calculates κ using the most recent draws for μ and $\boldsymbol{\beta}$. We then accept or decline this collection of values as a single unit in a Metropolis acceptance step. Since the process is really a Metropolis-Hastings algorithm where the joint candidate distribution is produced by an unchanging probabilistic process, we know that the full chain remains ergodic (Gill 2007, chap. 11).

More specifically, the algorithm is given by the following steps:

1. **Initialization:** Using the IRLS algorithm from the previous section, compute the MLEs $\hat{\mu}$ and $\hat{\kappa}$, then set $\mu^{(0)} = \hat{\mu}$. Compute now the MLE $\hat{\boldsymbol{\beta}}$ and the matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ corresponding to the asymptotic covariance of $\hat{\boldsymbol{\beta}}$ and set $\boldsymbol{\beta}^{(0)} = \hat{\boldsymbol{\beta}}$.
2. **Iterate for *M* steps:**

   (a) Generate $(\tilde{\mu} \sim \mathcal{VM}(\mu^{(t-1)}, \tau_{\mu}^2 \hat{\kappa}))$, where τ_{μ} is a tuning parameter to enhance efficient proposals.

   (b) Generate $\left(\tilde{\boldsymbol{\beta}} \sim \mathcal{N}_k(\boldsymbol{\beta}^{(t-1)}, \tau^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^2)\right)^2$, where τ^2 is another tuning parameter, usually different from τ_{μ}.

   (c) Calculate $\tilde{\kappa} = A^{-1}\left(\left(C^2 + S^2\right)^{\frac{1}{2}}\right)$, where both $C\left(\tilde{\mu}, \tilde{\boldsymbol{\beta}}\right)$ and $S\left(\tilde{\mu}, \tilde{\boldsymbol{\beta}}\right)$ are defined in the previous section about the IRLS algorithm.

   (d) Compute $\rho(\{\tilde{\mu}, \tilde{\boldsymbol{\beta}}, \tilde{\kappa}\}, \{\mu^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \kappa^{(t-1)}\}) = \min\left(1, \frac{\pi\left(\{\tilde{\mu}, \tilde{\boldsymbol{\beta}}, \tilde{\kappa}\} | \mathbf{y}\right)}{\pi\left(\{\mu^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \kappa^{(t-1)}\} | \mathbf{y}\right)}\right)$.

   (e) With probability $\rho(\{\tilde{\mu}, \tilde{\boldsymbol{\beta}}, \tilde{\kappa}\}, \{\mu^{(t-1)}, \boldsymbol{\beta}^{(t-1)}, \kappa^{(t-1)}\})$, take $\mu^{(t)} = \tilde{\mu}$, $\boldsymbol{\beta}(t) = \tilde{\boldsymbol{\beta}}$, and $\kappa^{(t)} = \tilde{\kappa}$; otherwise, take $\mu^{(t)} = \mu^{(t-1)}$, $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)}$ and $\kappa^{(t)} = \kappa^{(t-1)}$.

We observe that this MCMC algorithm mixes slowly through the sample space in part due to the small steps generated by the heavily scaled-down covariance matrix built into the proposal distribution to accomplish a "reasonable" acceptance rate (about 22% in the application in Section 8). However, this is simply a computational issue, and it means that users have to make sure to run the algorithm long enough to ensure that all subspaces of the posterior distribution are visited. Since the algorithm is very fast (5 million iterations of

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

our algorithm took less than 1 h on a standard laptop computer), this is only a minor inconvenience.

Finally, we recommend use of the following forms of priors to complete the Bayesian von Mises regression specification:

$$p(\mu) \sim \mathcal{VM}(\hat{\mu}, 5) \qquad p(\beta_j) \sim \mathcal{N}(0,10) \qquad p(\kappa) \sim \mathcal{U}([0:+\infty]) \ ,$$

although the specification is not particularly sensitive to the choice of hyperparameter values. Here, we emphasize uninformative priors for all parameters except for μ, although this could easily be changed in presence of reliable prior knowledge. Another uninformed prior for μ is the uniform distribution over $[0{:}2\pi)$.

7 Domestic Terrorism in the United States

In this section, we demonstrate circular regression using ML estimation only. The statistical analysis of terrorism is problematic because data consist of either selected publicly observed events or classified data at government agencies, existing tools for filling in missing information are difficult to apply, qualitative and technical experts have not traditionally cooperated, and it can even be physically dangerous to study. Yet, terrorism is an important *political* problem because it affects internal government policy, public perception, and relations between states. Statistical modeling has yielded some limited insights into the determinants and timing of terrorist incidents (see, e.g., Enders and Sandler 1995; Li and Schaub 2004; Enders 2007).

The data set *Political Violence in the United States, 1819–1968* (Levy 1991) provides an almost unique look at incidents of political violence resulting in injury or death over a very long period. The data were produced from coding articles in randomly selected issues of the New York Times and the Washington National Intelligence (now defunct) covering 150 years. Potential explanatory variables include the nature of the target, the number of attackers, the level of violence inflicted, the motivation for the attack, a count of deaths and injuries to targeted individuals and to attackers, the type of attacker, any property damage, and the number of pages in the newspaper describing the event. The critical category is whether the attacker or the targets are considered to be *political* groups, and we will use this as the baseline for attackers and victims in the model. We have also collected groups that are not well defined into an ''other'' category. These two variables are plotted against each other in Fig. 5 where darker cells denote higher frequencies. Although racial group and labor attackers are dominant (especially with regard to racial group targets in the case of racial group attackers), the political group row and column show important interactions.

In addition to our concern about groups, we include the variable Year to test for an underlying trend across time (calendar year minus 1819 scaled by 100), a dichotomous measure for resulting deaths or not (Some Killed), and the variable Large Attack Group, which is 1 for more than 10 attacking individuals and 0 otherwise. Lone attackers are excluded from the sample so that attacker group identification is meaningful (reducing the sample size to 746). Our central research question is whether there is a cyclical pattern within the years. Therefore, the outcome variable is the day of the year, which is obviously defined on the circular metric.

Table 1 gives the output from the circular regression model for these data, which is fit using our version of the Fisher and Lee (1992) algorithm described on the previous page. To ensure that the IRLS algorithm converged to the global maxima, the algorithm in Section 4 was run from several different starting values since the likelihood function is not

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

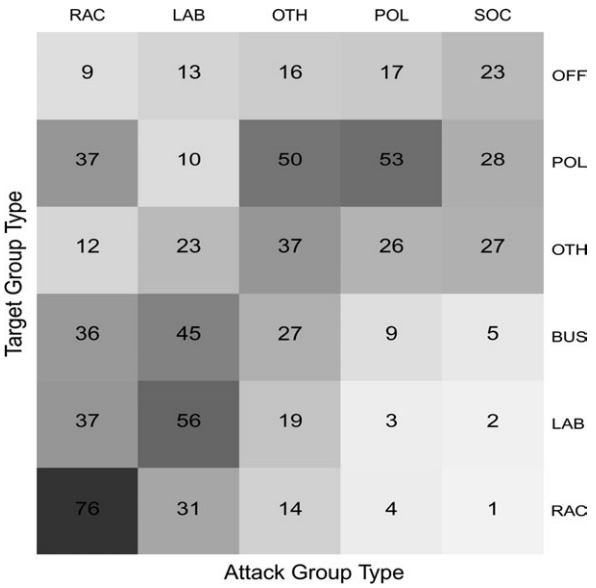


Fig. 5 Attackers versus targets, domestic terrorism. *Note.* RAC, racial group; LAB, labor group; BUS, business group; OTH, other group; POL, political group; SOC, social group; OFF, public officials.

guaranteed to be unimodal, making the model somewhat fragile as seen in Fig. 4. We also treated the missing data with multiple imputation (Little and Rubin 1983; Rubin 1987), so that the results in Table 1 are averaged across five replicate filled-in data sets. The fitted model provides a substantial improvement over the null (μ only) model since the difference in deviance (46) is far into the tail of a chi-squared distribution with 14 df.

The most important general finding is that there is evidence of a *within-year* cyclical trend as indicated by the proportion of statistically reliable explanatory variable coefficient estimates (10 of 12), and there is some evidence of an *across-year* trend from the results for Year. The model summary therefore supports the use of a circular data model on the annual scale. The negative sign on Year indicates that over the 150-year period attacks are increasingly occurring earlier in the year and we can expect that trend to continue. Conversely, ''successful'' terrorist attacks in terms of fatalities are more likely later in the year, perhaps due to the corresponding concentration of electoral and holiday gatherings. In terms of the attacking group, all types of groups are more likely to initiate violence earlier in the year than political groups. Since the U.S. electoral calendar is heavily weighted toward the latter part of the year, this makes sense. However, the target group coefficients tell the opposite story. These groups are all more likely to be attacked earlier in the year relative to political groups. Together the two blocks of group coefficient estimates tell a story: attacking political groups are motivated toward violence against *other* groups later in the calendar year, but not typically toward other political groups, except for the political groups in office (Public Officials).

8 Party Preferences in the German Bundestag

In this section, we turn to our fully Bayesian regression setup with another example. Two classes of spatial models dominate in political science: proximity models and directional

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Table 1 ML results for domestic terrorism

Explanatory variable		*Coefficient*	*SE*	$\lvert z \rvert$	*p*
Effect of attack type (relative to political)	$\hat{\mu}$	−2.462	0.282	−8.725	.001
	Year	−0.319	0.183	−1.746	.090
	Some Killed	0.517	0.208	2.493	.037
	Large Attack Group	0.190	0.270	0.704	.513
	Racial Group	−1.139	0.435	−2.616	.037
	Labor Group	−0.737	0.441	−1.672	.149
	Social Protest Group	−0.949	0.606	−1.565	.177
	Other Groups	−0.684	0.520	−1.315	.250
Effect of target type (relative to political)	Racial Group	1.135	0.428	2.649	.035
	Labor Group	0.694	0.404	1.717	.132
	Business Group	0.246	0.561	0.439	.678
	Public Officials	1.661	0.408	4.067	.000
	Other Groups	0.445	0.328	1.355	.212
	$\hat{\kappa}$	0.453	0.057	7.986	.000

Note. Null deviance = 2714.47, on 745 df; $N = 746$. Residual deviance = 2668.18, on 731 df; *Akaike's information criterion* = 2692.18.

models. Although proximity models follow a distance logic and are based on Euclidean spaces, directional models assume radial measures to be more relevant and are based on vector spaces. This distinction affects many interests in political science such as voting theory, coalition theory, and legislative politics. Purely directional models, our focus here, concentrate not on the *positions* of political actors, but rather the *direction* in which these actors try to shift policies. Linhart and Shikano (2009) categorize directional models into three types. The first two models are due to Grofman (1985) and Rabinowitz and MacDonald (1989) and assume that both direction and intensity define the ideal vector of a voter. Although these two models have many things in common, they differ in some important aspects. The third group of directional models, which is due to Matthews (1979) and Schofield (1985), is *purely* directional. Only the linear path of an actor's ideal vector, that is, the vector beginning in some starting point (e.g., the status quo) and ending in the actor's ideal point, is relevant. The length of the ideal vector, that is, the intensity, plays no role in these models. With every policy change, a new interim status quo is reached and the ideal direction is updated. Actors do not have to follow their ideal vectors exactly but can negotiate and change directions along the way. Hence, for the actor's utility, only directions matter, and vector lengths are irrelevant (Linhart and Shikano 2009, 4). Schofield (1985, 1993) and Linhart and Shikano (2009) provide the basic tools for the formal analysis of this purely directional approach. We combine their approach with the circular regression model to analyze party preferences over policy issues for all Bundestag elections in post–World War (WW) II Germany.

Measurement of the parties' preferences are based on the Comparative Manifesto Project (CMP; see Budge et al. 2001). The CMP uses party manifestos to code statements of left or right preferences with regard to various policies. We use CMP data rather than expert surveys because CMP data cover the whole period of interest: all 16 elections from the first post-WWII contest in 1949 to the recent 2005 one that produced the grand coalition. Parties are included if they won seats in multiple elections (Social Democratic

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Party [SPD] from 1949 to 2005, Christian Democratic Union/Christian Social Union [CDU/CSU] 1949–2005, Free Democratic Party [FDP] 1949–2005, Alliance '90/The Greens [B90/Greens] 1983–2005, and Party of Democratic Socialism [PDS] 1990–2005), producing 60 observations. These data include coding of both valence issues and positional issues that enables estimation of both parties' ideal positions and preferred directions. Based on the CMP data, Linhart and Shikano (2007) estimated the German parties' ideal directions for all Bundestag elections between 1949 and 2005. Consistent with this literature, we assume a two-dimensional policy space for Germany, where the first (''Dimension 1'') dimension captures economic issues and the second (''Dimension 2'') captures mostly social or cultural issues (Linhart and Shikano 2009, 13). Although parties in the lower left quadrant prefer policies more to the economic left and culturally more liberal, parties in the higher right quadrant prefer economically and culturally more conservative policies. The assumption of two dimensions is key for the applicability of the more basic circular regression model since a three-dimensional policy space would require spherical regression models. As with every other directional model, we have to make an assumption about the origin of the two-dimensional policy space. Does this origin reflect some sort of neutral point or rather the current status quo? The literature has not fully sorted out the defining criteria here, and so following the CMP group and Linhart and Shikano (2009), we simply assume that parties have the same origin point in mind when writing their manifestos.

Figure 6 displays nonparametric Gaussian kernel density estimates of the preferred directions of the ideal vectors for the five parties in the 16 Bundestag elections between 1949 and 2005. The ideal vectors of the leftist parties, SPD, Greens, and PDS, all point in the lower left direction. Although the Greens and the PDS express almost exactly equally strong preferences for more leftist policies in the first and second dimension, the SPD favored in some years more leftist policies along the economic dimension while being satisfied with the current status quo/neutral point of the second dimension. In contrast, the FDP consistently favored more liberal policies along the second dimension while favoring more right-leaning policies along the economic dimension, especially during the more recent elections. The CDU/CSU have a less consistent pattern of ideal directions but prefer in general more conservative policies along the second dimension while being satisfied with the status quo/neutral point of the first dimension.

The explanatory variables include the unemployment rate, which is assumed to influence primarily Dimension 1, as well as the proportion of out of wedlock births (in Western Germany), which is assumed to influence primarily Dimension 2. The remaining covariates are more directly political in nature. We use the party dummies (with the FDP as reference category) to contrast party direction. As a control for changes in the status quo between elections by including the *year* of the election and its square, making party movement relative to each other and therefore internally consistent. To reflect the great political upheaval of German reunification, a binary variable is specified that is 0 until the year 1990 and 1 afterward.

As a quality control measure for our MCMC sampler (fully describing the dominant mode), and a convenient means of comparing Bayesian and non-Bayesian estimation for this example, we run both the circular ML IRLS algorithm from Section 4 and the circular Bayesian estimation process described in Section 6 from four different starting values. The first estimation process conveniently specifies an asymptotic covariance matrix from the IRLS process that we can use as part of our proposal distribution for the Metropolis-Hastings algorithm. For the MCMC process, we performed 5,000,000 iterations, throwing away the first 1,000,000 of the chain as a burnin period, and thinning by keeping every 4000th

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

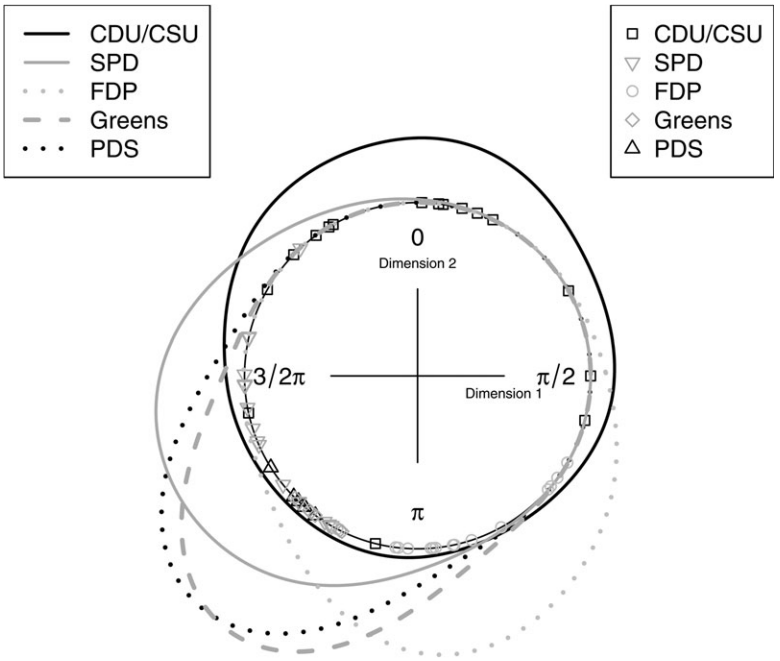


Fig. 6 Preferred directions of German parties' ideal vector.

iteration. All the conventional diagnostics pointed toward convergence (see Gill [2008] for relevant issues), including those of Geweke (1992), Heidelberger and Welch (1981a, 1981b), and Gelman and Rubin (1992), and standard graphical analysis. Interested readers can download the full set of MCMC sample values, and the diagnostic reports, from our replication page for this project. `http : //jgill.wustl.edu/replication.html`.

The ML and Bayesian approaches in Table 2 yield very similar results, which is not surprising given our use of uninformative priors. The ML results also show a good fit to the data in terms of deviance improvement since the difference in deviance between the fitted and the null models (90) is well past any arbitrary significance threshold for a chi-squared distribution with 10 df. The seemingly huge difference in the mean parameter μ between the two models is in fact negligible since $-0.81 + 2\pi = 5.47$, which is close to the mean of the Bayesian posterior of 5.89. Notice that all the Bayesian model posterior means for the explanatory variable coefficients are slightly bigger in absolute value compared to the ML point estimates. This is because the Bayesian summary takes into account the full form of the posterior when calculating this mean, as opposed to finding the dominant mode and then measuring curvature around it.

The estimated party dummies confirm the visual impression of Fig. 6, specifically that relative to the FDP, all other parties' ideal directions have larger radians. It is important to keep in mind that although the baseline party is an arbitrary choice, we picked the FDP *because doing so highlights the most interesting political differences between the parties*. This choice means that the average angle between some selected party listed in Table 2 and the FDP is the marginal effect (first difference) of switching on and off the relevant coefficient. For example, in the case of the Greens, it is the expected difference between the two scenarios run through the link function,

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Table 2 Results for Bundestag party directional preferences

	MLE			*Marginal posteriors*			
	Coefficient	*SE*	$\|z\|$	*Mean*	*SD*	*95% Highest Posterior Density*	
$\hat{\mu}$	−0.81	0.08	9.85	5.68	0.38	5.26	6.13
Unemployment rate	−0.08	0.03	3.04	−0.11	0.05	−0.23	−0.02
Out of wedlock	−0.34	0.08	4.15	−0.49	0.17	−0.86	−0.25
Reunification	−0.50	0.31	1.64	−0.70	0.43	−1.56	0.08
SPD	2.68	0.76	3.53	3.91	1.37	1.89	6.84
CDU/CSU	3.59	0.77	4.66	4.96	1.42	2.78	8.09
Greens	2.02	0.80	2.53	2.72	1.39	0.39	5.38
PDS	2.11	0.81	2.63	2.89	1.40	0.59	5.38
Year	−0.17	0.04	4.32	−0.25	0.08	−0.41	−0.12
Year^2/100	0.51	0.11	4.66	0.73	0.23	0.39	1.22
$\hat{\kappa}$	3.45	0.56	6.21	3.01	0.18	2.60	3.30

Note. Null deviance = 197.30, on 59 df; Deviance information criterion = 246.95. Residual deviance = 107.75, on 49 df; *Akaike's information criterion* = 125.75, on 49 df. $N = 60$ more than 16 elections, FDP as party reference category.

$$\begin{aligned} 2 \cdot \text{atan}[&+2.02 \cdot \text{Greens} - 0.08 \cdot \text{mean}(\text{Unemployment rate}) \\ &- 0.34 \cdot \text{mean}(\text{Out of wedlock}) - 0.50 \cdot \text{mean}(\text{Reunification}) \\ &- 0.17 \cdot \text{mean}(\text{Year}) + 0.51 \cdot \text{mean}(\text{Year squared})], \end{aligned}$$

and

$$\begin{aligned} 2 \cdot \text{atan}[&0.00 \cdot \text{Greens} - 0.08 \cdot \text{mean}(\text{Unemployment rate}) \\ &- 0.34 \cdot \text{mean}(\text{Out of wedlock}) - 0.50 \cdot \text{mean}(\text{Reunification}) \\ &- 0.17 \cdot \text{mean}(\text{Year}) + 0.51 \cdot \text{mean}(\text{Year squared})], \end{aligned}$$

which gives the average angle of the bargaining set as $\measuredangle$(FDP,Greens) = 0.86 radians. Over time this turns out to be the smallest difference between any party and the FDP, making the Greens a high-probability coalition partner. Interestingly, though, the FDP and the Greens have never jointly accumulated enough seats to form a minimal winning coalition. However, a potential coalition of CDU/CSU, FDP, and Greens (called the ''Jamaica-coalition'' because the party colors together resemble the Jamaican flag) was heavily discussed in the aftermath of the important 2005 elections. We can also calculate other parties' progressively large differences relative to the FDP: the left party $\measuredangle$(FDP,PDS) = 0.96 radians, $\measuredangle$(FDP,SPD) = 1.96 radians, and $\measuredangle$(FDP,CDU/CSU) = 2.88 radians. All these differences between the party dummies and the FDP's preferred direction are statistically reliable in the sense that they are estimated with high precision relative to effect size.

On average, both unemployment rates and proportion of out of wedlock births move all parties substantially counterclockwise, where the effect of out of wedlock births is almost four times larger than unemployment rates. The binary indicator for reunification has no reliable effect on the parties' ideal directions. We include the linear and squared term of years since the first election in 1949 into the arctan function to control, at least partially, for a change in the implicit status quo between election years. Nevertheless, these coefficients have a substantive interpretation as well. The negative effect of the linear time variable

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

indicates that in more recent elections, parties move in clockwise directions, although this effect is decreasing over time as shown by the estimated positive effect of its squared term.

The average direction for the FDP is 2.95 radians, which is consistent with Fig. 6, calculated by

$$\begin{aligned}-0.81+2\cdot\operatorname{atan}[&-0.08\cdot\operatorname{mean}(\text{Unemployment rate})\\&-0.34\cdot\operatorname{mean}(\text{Out of wedlock})-0.50\cdot\operatorname{mean}(\text{Reunification})\\&-0.17\cdot\operatorname{mean}(\text{Year})+0.51\cdot\operatorname{mean}(\text{Year squared})]\quad \operatorname{mod}2\pi,\end{aligned}$$

giving 2.95 from the wrapping property. This, and similar, calculations show that even though this model was more challenging to estimate, the marginal effects of interest are produced with the same level of effort as any standard GLM. Finally, the estimated $\hat{\kappa} = 3.45$ reflects the rather high clustering of the preferred directions, apparently mostly toward the lower left quadrant. This pervasive trend over time is consistent with most studies of German political parties.

9 Conclusions

Our purpose here is two-fold: to point out that there exist circular data problems in political science and to provide a means of building regression-style models that account for such purely cyclical outcomes. This work builds on standard von Mises approaches but is tailored for the type of data analysis that political scientists routinely perform. We demonstrate that ML estimation works, provided that one is careful about the application of the IRLSs algorithm. Furthermore, we provide an easy-to-use Metropolis-Hastings sampler, which is less prone to get stuck in a local maxima and offers all the advantages of a Bayesian model. Our code for these approaches is freely available to interested researchers.

Terrorism data are notoriously tricky to model in a regression context. Poor measurement, surreptitious subjects, and an inability to collect data firsthand make this a difficult research area in political science. When the data are measured in a circular fashion, the modeling exercise becomes even more challenging. We demonstrate here, however, that it is possible to pull findings out of such material. We learn from the results that there is a within-year cycle, implying stability of effects. Political groups, whether they are attackers or victims of domestic terrorism, appear to have a different profile than other defined groups in the data. In particular, political targets are more likely to be attacked later in the year closer to the election cycle.

In our second application, we looked at an important approach to spatial analysis of German Bundestag elections. The directional analysis of parties' policy positions and coalition preferences in a two-dimensional ideological space offers new insights to common political science data and questions. Our focus here on circular effects highlights the substantive interplay of the two dimensions, which are clearly not independent thus making a circular treatment appealing. We found that parties were remarkably consistent over time in their preferences relative to each other. The Liberals (FDP) as a baseline provide a steady scale of differences relative to other mainstream political groups. Interestingly, unemployment and out of wedlock births were shown to be reliable influences on party behavior, but reunification was not. This may be because the East Germans did not have the same diverse party history as citizens in the West, and therefore, they readily adopted the structures and cleavages of the more mature democratic partner.

Other research interests in political science generate circular data as well. Legislatures, Congress in particular, operate on a reasonably fixed schedule such that deviations often

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

point toward critical political events. Migration and immigration, with strong political consequences, often follow well-defined directional patterns. This is notably true in Europe and North America. There is a large literature in formal models that evaluate directional effects well beyond the Matthews-Schofield structure discussed here. Scholars in political economy encounter regularly measured calendar variables since governments often release data and indicators at fixed times. Finally, we also see rich potential for circular data analysis in the increasing use of geographic information systems data in political science, particularly with the release of the 2010 U.S. Census containing geocoding information.

References

Abeysasekara, S., and D. Collett. 1982. On the estimation of the parameters of the von Mises distribution. *Communications in Statistics–Theory and Methods* 11:2083–90.

Abramowitz, M., and I. A. Stegun. 1972. Modified Bessel functions I and K, definitions and properties. Section 9.6. In *Handbook of mathematical functions with formulas, graphs, and mathematical tables,* 9th printing, 374–7. New York: Dover.

Bagchi, P., and I. Guttman. 1988. Theoretical considerations of the multivariate von Mises-Fisher distribution. *Journal of Applied Statistics* 15:149–69.

Bagchi, P., and J. B. Kadane. 1991. Laplace approximations to posterior moments and marginal distributions on circles, spheres, and cylinders. *Canadian Journal of Statistics* 19:67–77.

Bartels, Robert. 1984. Estimation in a bidirectional mixture of von Mises distributions. *Biometrics* 40:777–84.

Batschelet, E. 1981. *Circular statistics in biology.* London: Academic Press.

Beason, Robert C. 1980. Orientation of waterfowl migration in the southwestern United States. *Journal of Wildlife Management* 44:447–55.

Bergin, Timothy M. 1991. A comparison of goodness-of-fit tests for analysis of nest orientation in western kingbirds (Tyrannus Verticalis). *Condor* 93:164–171.

Best, D. J., and N. I. Fisher. 1981. The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communications in Statistics–Theory and Methods* 10:493–502.

Bingham, M. S., and K. V. Mardia. 1975. Maximum likelihood characterization of the von Mises distribution. In *Statistical distributions in scientific work*. 3rd ed. ed. G. P. Patil, S. Kotz, and J. K. Ord, 387–98. Dordrecht, The Netherlands: Reidel.

Bryan, A. Lawrence Jr., and Malcolm C. Coulter. 1987. Foraging flight characteristics of wood storks in East-Central Georgia, U.S.A. *Colonial Waterbirds* 10:157–61.

Budge, I., H.-D. Klingemann, A. Volkens, and J. Bara. 2001. *Mapping policy preferences: Estimates for parties, electors, and governments 1945-1998.* Oxford: Oxford University Press.

Chan, Y. M., and Xuming He. 1993. On median-type estimators of direction for the von Mises-Fisher distribution. *Biometrika* 80:869–75.

Clark, W. A. V., and James E. Burt. 1980. The impact of workplace on residential relocation. *Annals of the Association of American Geographers* 70:59–67.

Coleman, D. A., and J. C. Haskey. 1986. Marital distance and its geographical orientation in England and Wales, 1979. *Transactions of the Institute of British Geographers, New Series* 11:337–55.

Collett, D., and T. Lewis. 1981. Discriminating between the von Mises and wrapped normal distributions. *Australian & New Zealand Journal of Statistics* 23:73–9.

Cotterill, Derek S., and Miklos Csorgo. 1982. On the limiting distribution of and critical values for the multivariate Cramer-von Mises statistic. *Annals of Statistics* 10:233–44.

Damien, Paul, and Stephen Walker. 1999. A full Bayesian analysis of circular data using the von Mises distribution. *Canadian Journal of Statistics* 27:291–8.

de Waal, D. J. 1979. On the normalizing constant for the Bingham-von Mises-Fisher matrix distribution. *South African Journal of Statistics* 13:103–12.

Dobson, Annette J. 1978. Simple approximations for the von Mises concentration statistic. *Applied Statistics* 27:345–7.

Downs, Thomas D., and Lawrence A. Gould. 1967. Some relationships between the normal and von Mises distributions. *Biometrika* 54:684–7.

Downs, Thomas D., and K. V. Mardia. 2002. Circular regression. *Biometrika* 89:683–98.

Enders, Walter. 2007. Terrorism: An empirical analysis. In *Handbook of defense economics*. Vol. 2, ed. T. Sandler and K. Hartley, 815–66. Amsterdam, The Netherlands: Elsevier.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Enders, Walter, and Todd Sandler. 1995. Terrorism: Theory and applications. In *Handbook of defense economics*. Vol. 1, ed. K. Hartley and T. Sandler, 213–49. Amsterdam, The Netherlands: Elsevier.
Fisher, N. I. 1993. *Statistical analysis of circular data*. Cambridge: Cambridge University Press.
Fisher, N. I., and A. J. Lee. 1992. Regression models for an angular response. *Biometrics* 48:665–77.
Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–511.
Geweke, John. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics*. Vol. 4. ed. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger, 169–73. Oxford: Oxford University Press.
Ghosh, Kaushik, Rao Jammalamadaka, and Mangalam Vasudaven. 1999. Change-point problems for the von Mises distribution. *Journal of Applied Statistics* 26:423–34.
Gill, Jeff. 2007. Bayesian methods for the social and behavioral sciences. 2nd ed. New York: Chapman and Hall.
———. 2008. Is partial-dimension convergence a problem for inferences from MCMC algorithms? *Political Analysis* 16:153–78.
Gordon, Louis, and Malcolm Hudson. 1977. A characterization of the von Mises distribution. *Annals of Statistics* 5:813–14.
Green, P. J. 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society Series B* 46:149–92.
Grimshaw, Scott D., David G. Whiting, and Thomas H. Morris. 2001. Likelihood ratio tests for a mixture of two von Mises distributions. *Biometrics* 57:260–5.
Grofman, Bernard. 1985. The neglected role of the status quo in models of issue voting. *Journal of Politics* 47:230–7.
Gruet, Jean-Claude. 2000. A note on hyperbolic von Mises distributions. *Bernoulli* 6:1007–20.
Guttorp, Peter, and Richard A. Lockhart. 1988. Finding the location of a signal: A Bayesian analysis. *Journal of the American Statistical Association* 83:322–30.
Hamilton, James D. 1994. *Time series analysis.*. Princeton, NJ: Princeton University Press.
Healy, Dennis M. Jr., and Peter T. Kim. 1996. An empirical Bayes approach to directional data and efficient computation on the sphere. *Annals of Statistics* 24:232–54.
Heidelberger, P., and P. D. Welch. 1981a. Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development* 25:860–76.
———. 1981b. A spectral method for confidence interval generation and run length control in simulations. *Communications of the Association for Computing Machinery* 24:233–45.
Hill, G. W. 1976. New approximations to the von Mises distribution. *Biometrika* 63:673–6.
Jammalamadaka, S. R., and A. Sengupta. 2001. *Topics in circular statistics.*. Singapore: World Scientific Press.
Johnson, Richard A., and Thomas E. Wehrly. 1978. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association* 73:602–6.
Levy, Sheldon G. 1991. *Political Violence in the United States, 1819–1968* [computer file], ed., ICPSR. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
Li, Quan, and Drew Schaub. 2004. Economic globalization and transnational terrorist incidents. *Journal of Conflict Resolution* 48:230–58.
Linhart, Eric, and Susumu Shikano. 2007. *Die Generierung von Parteipositionen aus vorverschluesselten Wahlprogrammen für die Bundesrepublik Deutschland,* MZES working paper.
———. 2009. A basic tool set for a generalized directional model. *Public Choice* 140(1/2):85–104.
Little, Roderick J. A., and Donald B. Rubin. 1983. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *American Statistician* 37:218–20.
Lwin, T. 1975. On von Mises directions. *Annals of the Institute of Statistical Mathematics* 27:79–85.
Mardia, K. V. 1972. *Statistics of directional data.*. London: Academic Press.
———. 1975a. Distribution theory for the von Mises-Fisher distribution and its application. In *Statistical distributions in scientific work*. Vol. 1, ed. G. P. Patil, S. Kotz, and J. K. Ord, 113–30. Dordrecht, The Netherlands: Reidel.
———. 1975b. Statistics of directional data. *Journal of the Royal Statistical Society Series B* 37:349–93.
Mardia, K. V., and S. A. M. El-Atoum. 1976. Bayesian inference for the von Mises-Fisher distribution. *Biometrika* 63:203–5.
Matthews, G. V. T. 1974. On bird navigation, with some statistical undertones. *Journal of the Royal Statistical Society Series B* 36:349–64.
Matthews, S. A. 1979. A simple direction model of electoral competition. *Public Choice* 34:141–56.
McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models*. 2nd ed. New York: Chapman and Hall.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press

Morelande, Mark R. 2008. *Circular regression using Bayesian unwrapping.* IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008 (March 31 2008–April 4 2008), Las Vegas, NV, 3441–4.

Rabinowitz, G., and S. E. MacDonald. 1989. A directional theory of issue voting. *American Political Science Review* 83(1):93–121.

Robert, C. P., and G. Casella. 2004. *Monte Carlo statistical methods*. 2nd ed. New York: Springer.

Rubin, Donald. 1987. *Multiple imputation for nonresponse in surveys*. New York: John Wiley and Sons.

Schmidt-Koenig, K. 1963. On the role of the loft, the distance and site of release in pigeon homing (the 'cross-loft experiment'). *Biological Bulletin* 125:154–64.

Schofield, Norman J. 1985. *Social choice and democracy.* New York: Springer.

———. 1993. Party competition in a spatial model of coalition formation. *Political economy: Institutions, competition, and representation*, ed. W. A. Barnett, M. J. Hinich, and N. J. Schofield, 135–74. Cambridge: Cambridge University Press.

Schou, G. 1978. Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika* 65:369–77.

Squires, John R., and Leonard F. Ruggiero. 1996. Nest-site preference of northern goshawks in Southcentral Wyoming. *Journal of Wildlife Management* 60:170–7.

Stephens, M. A. 1972. Multisample tests for the von Mises distribution. *Journal of the American Statistical Association* 67:456–61.

Tover, Arnaiz G., and C. Ruiz-Rivas. 1986. Outliers in circular data: A Bayesian approach. *Questillo* 10(1):1–6.

Upton, Graham J. G. 1973. Single-sample tests for the von Mises distribution. *Biometrika* 60:87–99.

———. 1976. More multisample tests for the von Mises distribution. *Journal of the American Statistical Association* 71:675–8.

———. 1986a. Approximate confidence intervals for the mean direction of a von Mises distribution. *Biometrika* 73:525–7.

———. 1986b. Distance and directional analyses of settlement patterns. *Economic Geography* 62:167–79.

Upton, Graham J. G., and B. Fingleton. 1989. *Spatial data analysis by example. Volume 2: Categorical and directional data*. New York: John Wiley and Sons.

von Mises, R. 1918. Uber Die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift* 19:490–500.

https://doi.org/10.1093/pan/mpq009 Published online by Cambridge University Press