Circular interpretation of projected normal regression coefficients:
Supplementary material

This is supplementary material for the paper: ‘Circular interpretation of projected normal regression coefficients’. Code for the simulations, empirical data analysis and plots can be found at the following GitHub archive: https://github.com/Circular-Data/Circular-Interpretation-BJMSP.

1 Signed Shortest Distance to the Origin

To understand the intuition behind the sign of the shortest distance to the origin, $SDO$, we created Figure 1. In this figure we show two bivariate regression lines:

\[
\begin{align*}
y^I &= \beta^I_0 + \beta^I_1 x \\
y^{II} &= \beta^{II}_0 + \beta^{II}_1 x.
\end{align*}
\] (1)

and

\[
\begin{align*}
y^I &= -\beta^I_0 + \beta^I_1 x \\
y^{II} &= -\beta^{II}_0 + \beta^{II}_1 x.
\end{align*}
\] (2)

The unit circle in Figure 1 is split into a positive (grey) and a negative side according to the line

\[
\begin{align*}
y^I &= \beta^I_1 x \\
y^{II} &= \beta^{II}_1 x.
\end{align*}
\] (3)

The intersection points of this line with the circle (squares) are equal to $\text{atan2}(\beta^{II}_1, \beta^I_1)$ and $\text{atan2}(\beta^{II}_1, \beta^I_1) - \pi$. The direction of the line in Equation 3 determines which side of the circle is the positive side. Imagine the direction of the line is North, then the West side is the positive side of the circle. To determine the sign of the SSDO of a regression line we compute $a_c = \text{atan2}(\beta^{II}_1, \beta^I_1)$. In this case we can see this as turning the regression line anti-clockwise by $\text{atan2}(\beta^{II}_1, \beta^I_1)$. See Figure 2 for the turning of regression line 1. By computing $\sin a_c - \text{atan2}(\beta^I_1, \beta^I_1)$ we can then determine whether the regression line is on the positive or negative side of the unit circle. We can thus compute the SSDO as:

$$SSDO = \text{sign}[[\sin (a_c - \text{atan2}(\beta^{II}_1, \beta^I_1))]SDO.$$
Figure 1: Plot illustrating the computation of the SSDO showing a unit circle and two regression lines in bivariate space. Squares are $\text{atan2}(\beta_{II}^{1}, \beta_{I}^{1})$ and $\text{atan2}(\beta_{II}^{1}, \beta_{I}^{1}) - \pi$, dots are the $a_c$ for the two regression lines. The dotted line is the regression line for which the linear intercepts, $\beta_{I0}^{I}$ and $\beta_{II0}^{II}$, equal 0. The shaded area represents the area where the SSDO of a regression line has a positive sign.
Figure 2: Plot illustrating the computation of the SSDO showing a unit circle and a regression line in bivariate space. The triangle represents $a_c - \text{atan2}(\beta^I_1, \beta^I_1)$. The dotted line is the regression line for which the linear intercepts, $\beta^I_0$ and $\beta^{II}_0$, equal 0. The shaded area represents the area where the SSDO of a regression line has a positive sign.
2 Plots for the Pointing North Data

Figure 3: Predicted circular regression curve for the relation between SBSOD and the pointing error together with the original datapoints. The square indicates the inflection point of the regression curve.
Figure 4: Figure showing the relation between SBSOD and the concentration of the predicted values on the circle.
Figure 5: Predicted circular regression curve for the relation between Experience and the pointing error together with the original datapoints. The inflection point of the regression curve is not shown as this point lies outside the range of the data.
Figure 6: Figure showing the relation between Experience and the concentration of the predicted values on the circle.

Figure 7: Predicted circular regression curve for the relation between Age and the pointing error together with the original datapoints. The square indicates the inflection point of the regression curve.
3 Simulation study

To assess the performance of the circular coefficients $b_c$, $SAM$ and $AS$ and the ability to distinguish between location and accuracy effects we conducted a simulation study with 1225 designs with one predictor. Of these designs 1056 were classified as location designs, 144 as accuracy and 25 as having no effect based on their population values. A description of the designs is given in Section 3.1. In Section 3.2 the simulation results are given.

3.1 Design

In each design different population values were chosen for the linear intercepts $\beta_0^I$ and $\beta_0^{II}$ and the regression coefficients $\beta_1^I$ and $\beta_1^{II}$. Chosen values were: -3, -1, 0, 1, 3 for the linear intercepts and -2, -1, -0.5, 0, 0.5, 1, 2 for the linear regression coefficients. From these values, the population values of the parameters $a_x$, $a_c$, $b_c$, $SAM$, $AS$, $SDO$ and $SSDO$ were computed.

We simulated 1000 datasets for each design. There are 500 datasets with $N = 50$ and 500 where $N = 200$ for each design. Each dataset contains one circular outcome $\theta$ and one linear predictor $x \sim N(0, 1)$. The relation between predictor and outcome was determined by the chosen values for the linear intercepts and coefficients. Before analysis of a dataset the linear predictor was centered at 0 such that the simulation study corresponds to the recommendations in Section 5 of the paper. For a random selection of datasets the convergence of the MCMC sampler used for estimation was checked. Typical posterior histograms are shown in Figure 9. For all of the selected datasets 5000 iterations, from which a burn-in of 1000 iterations was substracted, were deemed sufficient for convergence.
Figure 9: Typical posterior histograms for $\beta_0^I, \beta_1^I, \beta_0^{II}$ and $\beta_1^{II}$ of one dataset in the design with population values 0, 2, 3 and 1 respectively.

For each simulated dataset we computed several statistics from the posterior distributions of the parameters $\beta_0^I, \beta_1^I, \beta_0^{II}, \beta_1^{II}, a_x, a_c, b_c, SAM, AS$ and $SSDO$. We estimated the posterior mode, the deviation defined as population value minus estimated mode and the 95% HPD interval. The posterior mode and HPD interval were estimated using the methods proposed by (?) . We also recorded whether the population value of each parameter was located within the HPD interval. Furthermore, we checked whether the HPD intervals of the parameters $b_c, SSDO, \beta_1^I$ and $\beta_1^{II}$ contained 0.

For datasets of the same sample size and from the same design we then computed summary statistics. We averaged the deviation of the estimated mode from the population value over all datasets to compute the bias. The relative bias for the location designs is computed as the ratio of the absolute population value of the coefficient divided by the absolute bias. The relative bias for the accuracy designs is the absolute difference between the population value and the bias. We computed the frequentist coverage of the HPD intervals by computing the percentage of datasets of which the HPD interval contained the population value. We computed the Average Interval Width (AIW) by averaging all lower and upper bounds of the HPD intervals and then computing the distance between these averaged bounds. Lastly, we computed the amount of datasets in which the HPD interval of the parameters $b_c, SSDO, \beta_1^I$
and $\beta_1^{II}$ contained 0.

### 3.2 Results

For each design we created cross-tabulations with the percentage of datasets in which a location effect was detected and we counted the percentage of datasets in which any effect, location or accuracy, was detected. The cross-tabulations contain two variables: a location effect detected by $b_c$ (Yes/No) and a location effect detected by $SSDO$ (Yes/No). For the percentage of datasets in which any effect was detected we used $\beta_1^{I}$ and $\beta_1^{II}$ (Yes/No). Because we simulated datasets for a large amount of designs we decided to average the percentages over the designs of three different categories: the accuracy designs, the designs with a location effect and $SDO \leq 1$ and the designs with a location effect and $SDO > 1$. Because there were only 25 designs in which we did not simulate a circular effect we decided to not include results for this type. Tables 1 and 2 show the average percentages in combination with their standard deviations for designs with a sample size of 50. Tables 3 and 4 show the average percentages in combination with their standard deviations for designs with a sample size of 200.

**Table 1:** Average percentage of datasets with standard deviation for which the any effect indicator indicates that there is (Yes) or is no effect (No) per type of design with a sample size of 50.

| Accuracy | Location $SDO \leq 1$ | Location $SDO > 1$ |
|----------|-----------------------|---------------------|
| Yes      | No                    | Yes                 | No                 |
| 91.14 (17.46) | 8.86 (17.46)  | 99.53 (1.27) | 0.47 (1.27) |
| 95.27 (12.16) | 4.73 (12.16)          |

**Table 2:** Cross-tabulations showing the average percentage of datasets with standard deviation for which the $SSDO$ and $b_c$ do (Yes) or do not (No) indicate a location effect per type of design with a sample size of 50.

| Location $SSDO$ | Accuracy | Location $b_c$ | Yes | No |
|-----------------|----------|----------------|-----|----|
| Location $SDO \leq 1$ | Yes | 5.51 (1.22) | 1.40 (0.66) |
|                 | No      | 0.39 (0.79) | 92.71 (1.38) |
| Location $SDO > 1$ | Yes | 68.34 (29.95) | 2.47 (1.92) |
|                 | No      | 0.11 (0.33) | 29.08 (28.53) |

**Table 3:** Average percentage of datasets with standard deviation for which the any effect indicator indicates that there is (Yes) or is no effect (No) per type of design with a sample size of 200.

| Accuracy | Location $SDO \leq 1$ | Location $SDO > 1$ |
|----------|-----------------------|---------------------|
| Yes      | No                    | Yes                 | No                 |
| 99.17 (2.39) | 0.83 (2.39) | 100.00 (0.00) | 0.00 (0.00) |
| 99.67 (1.62) | 0.33 (1.62)          |
Table 4: Cross-tabulations showing the average percentage of datasets with standard deviation for which the SSDO and $b_c$ do (Yes) or do not (No) indicate a location effect per type of design with a sample size of 200.

| Location | Location $b_c$ | Yes | No |
|----------|----------------|-----|----|
| Accuracy | $b_c$ | 5.25 (1.06) | 1.80 (0.62) |
|          | No   | 0.04 (0.15) | 92.91 (1.14) |
| Location | SDO $\leq$ 1 | 91.07 (17.91) | 1.11 (1.94) |
|          | $b_c$ | 0.00 (0.10) | 0.90 (16.12) |
| Location | SDO $> 1$ | 98.89 (6.81) | 0.16 (0.87) |
|          | $b_c$ | 0.05 (0.31) | 7.82 (5.68) |

Figures 10 and 11 show several plots with simulation results for location (grey) and accuracy (black) designs, for the three circular coefficients $b_c$, $AS$ and $SAM$ from left to right. Note that each dot in these figures represents one design of our simulation. Again the no effect category was excluded.

Figure 10 shows the relative bias plotted against the population value of SSDO. The relative bias for the location designs is computed as the ratio of the absolute population value of the coefficient divided by the absolute bias. The relative bias for the accuracy designs is the absolute difference between the population value and the bias. We see that for location designs the relative bias for all three parameters increases for SSDOs closer to zero. The coefficients $AS$ and $SAM$ have lower relative biases than $b_c$. In accuracy designs the coefficients have smaller relative biases compared to location designs with an SSDO close to 0.

Figure 10 also shows the coverage plotted against the population value of the SSDO. For $b_c$ the location designs show coverages around 0.95 or higher. For the accuracy designs coverages for this parameter also lie around 0.95 but sometimes they are slightly lower. For $AS$ the location designs show some slight undercoverage, except those with an SSDO close to 0. Most accuracy designs show undercoverage for $AS$. For $SAM$ location designs show a slight undercoverage. Accuracy designs show some overcoverage for $SAM$.

Figure 11 shows the log of the AIWs plotted against the population value of the SSDO. We see that the largest log AIWs occur in the accuracy designs of the parameters $b_c$ and $SAM$. For the location designs $b_c$ shows the largest and $SAM$ shows the smallest log AIWs. The designs with smaller SSDO perform worse. The log of the AIW is larger for these designs.

The results for datasets with a sample size of 200 are shown in Figures 12 and 13. These results show that the coverages are better and the biases are lower for larger datasets, as is to be expected.
Figure 10: Relative bias and Coverage of the circular regression coefficients $b_c$, $AS$ and $SAM$, from left to right, for location (grey) and accuracy (black) designs plotted against the true $SSDO$. 
Figure 11: log(AIW) of the circular regression coefficients $b_c$, $AS$ and $SAM$, from left to right, for location (grey) and accuracy (black) designs plotted against the true $SSDO$. 
Figure 12: Relative bias and Coverage of the circular regression coefficients $b_c$, $AS$ and $SAM$, from left to right, for location (grey) and accuracy (black) designs plotted against the true $SSDO$. 
Figure 13: log(AIW) of the circular regression coefficients $b_c, AS$ and $SAM$, from left to right, for location (grey) and accuracy (black) designs plotted against the true $SSDO$. 