STATISTICAL SCIENCES
AND
DATA ANALYSIS

PROCEEDINGS OF THE
THIRD PACIFIC AREA STATISTICAL CONFERENCE

EDITORS
KAMEO MATUSITA
MADAN L. PURI
TAKESI HAYAKAWA

///VSP///

CIRCULAR REGRESSION

Y. R. SARMA and S.RAO JAMMALAMADAKA

Indian Statistical Institute, Calcutta, India
Department of Statistics and Applied Probability, University of California,
Santa Barbara, California, USA.

Abstract. The classical methods of regression analysis cannot be applied when observations are directions, measured as angles in a plane with reference to a fixed sense of rotation and a fixed zero direction. In this paper, a concept of circular regression is introduced and estimation of the regression parameters is discussed. Some empirical methods and asymptotic tests for the determination of the degree of the regression equation, are discussed and some recursive algorithms developed. Tests and estimation of regression parameters in some parametric models are also presented.

Keywords: Directions, circular data, regression, trigonometric polynomials, tests of hypotheses, circular models.

1. INTRODUCTION

Observations in the form of directions on two or more variables, are made in several experiments. Sometimes one may consider fixing the direction of one (independent) variable and observe the direction of the other (dependent) variable. In such situations, one is naturally interested in the problem of regression or predicting one variable using the values of the other. It is clear that the classical regression analysis developed for linear variables cannot be applied in this situation because of the natural restrictions one has for directional variables. In what follows, a method of predicting one circular variable based on another circular variable is presented. The concept of circular regression has not received much attention and only recently attention is drawn to this problem (see Jupp and Mardia [1], Batschelet [2], Rivest [3]). In Section 2, we present a concept of circular regression and discuss how the parameters of the model may be estimated. In Section 3, some criteria for the determination of the degree of the trigonometric polynomials introduced in Section 2 are given along with some empirical

approaches and large sample tests for the same. Some recursive algorithms for the computation of the estimates are also presented. In Section 4, regression parameters in some specific parametric models as well as some tests of regression parameters in such models are presented. For a discussion of a "correlation coefficient" which can be used as a measure of association between two circular variables, the reader is referred to Jammalamadaka and Sarma [4].

2. CIRCULAR REGRESSION

Let (α, β) be a pair of random variables which are directions, both measured with reference to the same zero direction and the same sense of rotation. The angles can be treated as unit vectors in the plane in terms of their sine and cosine components. Let $f(\alpha, \beta)$ be the joint density of (α, β) on the torus $0 \leq \alpha, \beta < 2\pi$.

To predict β for a given value of α, the vector corresponding to β is predicted by the conditional expectation (or regression) of $e^{i\beta}$ given α, namely

(i.e.) $\qquad E(e^{i\beta}|\alpha) = g(\alpha) = g_1(\alpha) + ig_2(\alpha) = \rho(\alpha)\, e^{i\mu(\alpha)}$, say. $\qquad$ (2.1)

Or equivalently,

$$E(\cos \beta|\alpha) = g_1(\alpha)$$

and

$$E(\sin \beta|\alpha) = g_2(\alpha)$$

$\qquad$ (2.2)

from which β is predicted as

$$\mu(\alpha) = \hat{\beta} = \begin{cases} \tan^{-1} \dfrac{g_2(\alpha)}{g_1(\alpha)} & \text{if} \quad g_1(\alpha) \geq 0 \\[2ex] \pi + \tan^{-1} \dfrac{g_2(\alpha)}{g_1(\alpha)} & \text{if} \quad g_1(\alpha) \leq 0 \\[2ex] \text{undefined} & \text{if } g_1(\alpha) = g_2(\alpha) = 0 \,. \end{cases} \qquad (2.3)$$

Here $\mu(\alpha)$ represents the conditional mean direction of β given α and $0 \leq \rho \leq 1$ the conditional concentration towards this direction. Predicting β as in (2.1) is optimal in the sense that it minimizes the distance $E\|e^{i\beta} - g(\alpha)\|^2$ and is very similar to the

"least squares" idea.

Note that in this approach, we do not need the knowledge of the joint distribution of (α, β), but rather the conditional distribution of β given α. This allows us to take the conditional distribution for instance, to be a wrapped stable law, as discussed in Section 4.

In the absence of further specifications on the structure of $(g_1(\alpha), g_2(\alpha))$ it is, in general, difficult to estimate them from the data although one might adopt nonparametric curve estimation methods. Even with further specifications, the determination can be quite difficult, a situation often encountered even in linear analysis. So, as in linear analysis, $g_1(\alpha)$ and $g_2(\alpha)$ are approximated by suitable functions. Since the functions $g_1(\alpha)$ and $g_2(\alpha)$ are both periodic functions with period 2π, they can be expressed in terms of their Fourier series expansions. This leads to the method of approximating $g_1(\alpha)$ and $g_2(\alpha)$ by trigonometric polynomials of "degree m", viz.

$$g_1(\alpha) \quad \approx \quad \sum_{k=0}^{m} (A_k \cos k\alpha + B_k \sin k\alpha)$$

and
$$(2.4)$$

$$g_2(\alpha) \quad \approx \quad \sum_{k=0}^{m} (C_k \cos k\alpha + D_k \sin k\alpha)$$

for a suitable choice of m. From this, one has the following model:

$$\cos \beta \quad = \quad \sum_{k=0}^{m} (A_k \cos k\alpha + B_k \sin k\alpha) + \epsilon_1$$

and
$$(2.5)$$

$$\sin \beta \quad = \quad \sum_{k=0}^{m} (C_k \cos k\alpha + D_k \sin k\alpha) + \epsilon_2$$

where $\epsilon' = (\epsilon_1 \ \epsilon_2)$ is the vector of errors which is assumed to be random with mean vector zero and covariance matrix Σ, which in general is unknown. The problem of estimating the various parameters $\{A_k, B_k, C_k, D_k, k = 0, 1, ..., m\}$, their standard errors as well as the matrix Σ, is considered here using the generalized least squares. Large sample tests are proposed for the regression coefficients and for determining the degree m. When specific parametric models are assumed for the conditional distribution of β given α, like the von Mises, wrapped Cauchy etc, $g_1(\alpha)$ and

$g_2(\alpha)$ have a known parametric form. Problems of estimation and tests of hypotheses about the regression parameters, in these cases are explored.

Let $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$ be a random sample of size n . The observational equations can be written as

$$Y_{1i} = \cos \beta_i = \sum_{k=0}^{m} (A_k \cos k\alpha_i + B_k \sin k\alpha_i) + \epsilon_{1i}$$

(2.6)

$$Y_{2i} = \sin \beta_i = \sum_{k=0}^{m} (C_k \cos k\alpha_i + D_k \sin k\alpha_i) + \epsilon_{2i}$$

for $i = 1, \dots, n$. We assume that $B_0 = D_0 = 0$, to ensure identifiability. Writing

$$\underline{Y}^{(1)} = (Y_{11}, \cdots, Y_{1n})'$$
$$\underline{Y}^{(2)} = (Y_{21}, \cdots, Y_{2n})'$$
$$\underline{\epsilon}^{(1)} = (\epsilon_{11}, \cdots, \epsilon_{1n})'$$
$$\underline{\epsilon}^{(2)} = (\epsilon_{21}, \cdots, \epsilon_{2n})',$$

(2.7)

$$\underset{n \times (2m+1)}{X} = \begin{bmatrix} 1 & \cos \alpha_1 & \cdots & \cos m\alpha_1 & \sin \alpha_1 & \cdots & \sin m\,\alpha_1 \\ 1 & \cos \alpha_2 & \cdots & \cos m\alpha_2 & \sin \alpha_2 & \cdots & \sin m\,\alpha_2 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \cos \alpha_n & \cdots & \cos m\alpha_n & \sin \alpha_n & \cdots & \sin m\,\alpha_n \end{bmatrix}$$

and

$$\underline{\lambda}^{(1)} = (A_0, A_1, \dots, A_m, B_1, \dots, B_m)'$$
$$\underline{\lambda}^{(2)} = (C_0, C_1, \dots, C_m, D_1, \dots, D_m)'$$

the observational equations (2.6) can be written in matrix notation as

$$\underline{Y}^{(1)} = X \underline{\lambda}^{(1)} + \underline{\epsilon}^{(1)}$$
$$\underline{Y}^{(2)} = X \underline{\lambda}^{(2)} + \underline{\epsilon}^{(2)}$$

(2.8)

or

$$\underset{\sim}{Y}^{*} = X^{*} \underset{\sim}{\lambda}^{*} + \underset{\sim}{\epsilon}^{*} \tag{2.9}$$

where $\underset{\sim}{Y}^{*\prime} = (\underset{\sim}{Y}^{(1)\prime} : \underset{\sim}{Y}^{(2)\prime})$, $\underset{\sim}{\epsilon}^{*\prime} = (\underset{\sim}{\epsilon}^{(1)\prime} : \underset{\sim}{\epsilon}^{(2)\prime})$

$$X^{*} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}, \lambda^{*\prime} = (\underset{\sim}{\lambda}^{(1)\prime} : \underset{\sim}{\lambda}^{(2)\prime}) .$$

The covariance matrix of $\underset{\sim}{\epsilon}^{*}$ can be written as $\Sigma \otimes I_n$ where $\otimes$ denotes the usual Kronecker product and I_n is the identity matrix of order n.

One can now obtain the generalized least squares estimates (possibly multistage) of $\underset{\sim}{\lambda}^{*}$ from the observational equations (2.9).

However, from the structure of equations (2.8), it can be seen (C.R. Rao [5], Section 8C.2, Schmidt [6], Section 2.6), that the generalized least squares estimates obtained from (2.9) coincide with the ordinary least squares estimates of $\underset{\sim}{\lambda}^{(1)}$ and $\underset{\sim}{\lambda}^{(2)}$ obtained separately from (2.8). The least squares estimates are given, on the assumption that X is of full rank $(2m + 1 < n)$, by

$$\underset{\sim}{\hat{\lambda}}^{(1)} = (X' X)^{-1} X' \underset{\sim}{Y}^{(1)}$$

$$\underset{\sim}{\hat{\lambda}}^{(2)} = (X' X)^{-1} X' \underset{\sim}{Y}^{(2)} \tag{2.10}$$

<u>Remark 2.1:</u>   If the values of α's are taken to be equally spaced over the circle, the computations become much simpler (see eg. Guest [7]). For instance, if $\alpha_i = \frac{2\pi(i-1)}{n}$, $1 \leq i \leq n$, we have the explicit estimates (for $1 \leq j \leq m$):

$$\hat{A}_0 = \frac{1}{n} \sum_{i=1}^{n} \cos\beta_i , \quad \hat{A}_j = \frac{1}{n} \sum_{i=1}^{n} \cos\beta_i \cos j\alpha_i , \quad \hat{B}_j = \frac{1}{n} \sum_{i=1}^{n} \cos\beta_i \sin j\alpha_i$$

$$\hat{C}_0 = \frac{1}{n} \sum_{i=1}^{n} \sin\beta_i , \quad \hat{C}_j = \frac{1}{n} \sum_{i=1}^{n} \sin\beta_i \cos j\alpha_i , \quad \hat{D}_j = \frac{1}{n} \sum_{i=1}^{n} \sin\beta_i \sin j\alpha_i .$$

<u>Remark 2.2:</u>   If $(\alpha, \beta_1), ..., (\alpha, \beta_n)$ is a sample for the same given value of α ,

then the regression coefficients (i.e. the coefficients in the Fourier series expansions of $g_1(\alpha)$ and $g_2(\alpha)$) will all be zero except A_0 and C_0 and these are given by $\hat{A}_0 = \frac{1}{n} \Sigma \cos\beta_i$ and $\hat{C}_0 = \frac{1}{n} \Sigma \sin \beta_i$ by taking the unique Moore–Penrose inverse of $X' X$ since X is not of full rank. Thus, when several β's are observed corresponding to the same α, then their resultant mean direction is obtained as the best predictor of β, as one would expect. Suppose for each α_i, $i = 1,..., n$ there are n_i observations $(\alpha_i, \beta_{i1}),...,(\alpha_i, \beta_{in_i})$ . . In view of Remark 2.2 we can first obtain the resultant mean direction β_i^* for each α_i and consider (α_i, β_i^*), $i = 1,..., n$ to obtain the regression equation.

<u>Remark 2.3</u>:   The parameters can also be estimated by minimizing the quantity

$$\sum_{i=1}^{n} \left[\cos \beta_i - \sum_{k=0}^{m} (A_k \cos k\alpha_i + B_k \sin k\alpha_i) \right]^2$$

$$+ \sum_{i=1}^{n} \left[\sin \beta_i - \sum_{k=0}^{m} (C_k \cos k\alpha_i + D_k \sin k\alpha_i) \right]^2 . \tag{2.11}$$

This method was called the method of "minimum distance" (J.S. Rao [8]) and it was shown there that the estimators obtained by this method are consistent and asymptotically normal under mild conditions. Since the estimators (2.10) coincide with the minimum distance estimators, one can conclude that these estimators are consistent and asymptotically normal. In fact, it is explicitly shown in the next section that $\hat{\underset{\sim}{\lambda}}^{(1)}$ and $\hat{\underset{\sim}{\lambda}}^{(2)}$ are jointly asymptotically normal and asymptotic tests regarding the regression coefficients are carried out for the determination of the degree m .

The covariance matrix Σ can be estimated from the least squares theory as follows writing:

$$\begin{aligned} R_0(i, j) &= \underset{\sim}{Y}^{(i)'} \underset{\sim}{Y}^{(j)} - \underset{\sim}{Y}^{(i)'} X(X'X)^{-1} X' \underset{\sim}{Y}^{(j)} \\ &= \underset{\sim}{Y}^{(i)'} (I - M) \underset{\sim}{Y}^{(j)} \end{aligned} \tag{2.12}$$

where $M = X(X'X)^{-1}X'$ and $R_0 = ((R_0(i, j))_{i,j=1,2}$, we have $\hat{\Sigma} =$

$(n - 2(2m + 1))^{-1} R_0$ is an unbiased and consistent estimator of Σ. From this, the standard errors of the estimators can also be obtained.

Remark 2.4:   Since the number of parameters to be estimated is quite large even for small values of m, the conventional method of using the residual sum of squares and cross–product matrix to estimate Σ and the standard errors of the estimates may not provide sufficiently accurate estimators. It is shown by several authors that the bootstrap method provides much better estimates of the standard errors of the estimates if resampling is done a large number of times (see eg. Wu [9]). For the bootstrap, we start with the estimates $\hat{\underset{\sim}{\lambda}}^{*}$ obtained above in (2.10) and using these estimates, the error vectors $\underset{\sim}{\epsilon}' = (\epsilon_{1i}, \epsilon_{2i})$ are computed from the model (2.6). From the emprical distribution of $\underset{\sim}{\epsilon}_1, ..., \underset{\sim}{\epsilon}_n$ a random sample of size n, say $\underset{\sim}{\epsilon}_1^{*}, ..., \underset{\sim}{\epsilon}_n^{*}$ is drawn. Using these, pseudo data $(\underset{\sim}{Y}^{(1)'}, \underset{\sim}{Y}^{(2)'})$ can be generated from

$$\underset{\sim}{Y}^{(1)} = X \hat{\underset{\sim}{\lambda}}^{(1)} + \underset{\sim}{\epsilon}^{*(1)}$$
$$\underset{\sim}{Y}^{(2)} = X \hat{\underset{\sim}{\lambda}}^{(2)} + \underset{\sim}{\epsilon}^{*(2)}$$

where $\underset{\sim}{\epsilon}^{*(i)} = (\epsilon_{i1}^{*}, ..., \epsilon_{in}^{*})'$ for $i = 1, 2$.

Assuming now that the data $(\underset{\sim}{Y}^{(1)'}, \underset{\sim}{Y}^{(2)'})$ are generated from the model (2.6), the regression coefficients $\underset{\sim}{\lambda}^{*}$ are estimated as before. This procedure is repeated a large number of times. This simulation preserves the assumptions that the errors are independent and identically distributed. In each repetition, a new set of sample of error vectors from the empirical distribution is generated and pseudo data are generated from model (2.6). Using these, a new set of estimators of $\underset{\sim}{\lambda}^{*}$ are obtained. From these sets thus obtained, the sample covariance matrix can be obtained, providing an estimate of the dispersion matrix Σ.

Remark 2.5:   It is shown in Bhimasankaram and Jammalamadaka [10] how recursive estimation and testing in linear models can be carried out when the sample size is increased. The methods developed there can be adapted to the present case also and the details are omitted as they can be directly worked out from those in the above cited paper.

<u>Remark 2.6</u>:   The hypothesis that β does not depend on α or $g_1(\alpha) = A_0$ and $g_2(\alpha) = C_0$ can be tested by an asymptotically χ^2 test, following the discussion in Section 3.2.

3.   DETERMINATION OF m IN THE REGRESSION EQUATION

A general problem in fitting any polynomial regression is the determination of the degree m of the polynomial. In what follows, some criteria for choosing m are suggested. Large sample tests for the validity of the criteria are also given. A further problem would be to update the estimates every time the degree is changed. In the usual polynomial regression for linear variables this is easily done with the use of orthogonal polynomials. Such orthogonality holds good in the present case also if α_i's are taken to be equally spaced (see Remark 2.1). In the more general case, the recursive method suggested below can be used to determine the degree m as well as to update the estimates without having to compute all the coefficients afresh when the degree is increased. This updating involves minimal additional computations. It is of interest to note that the different criteria discussed below leading to the determination of the degree, all lead to the same computations.

3.11.   One of the considerations in determining the degree m could be to see the reduction in error sum of squares by increasing m . The effect of augmenting the design matrix X on the residual sum of squares involves standard calculations (see Seber [11] and is done in our special case as follows (see also Bhimasankaram and Jammalamadaka [10] when X is not of full rank).

Deciding to take $(m + 1)^{th}$ degree trigonometric polynomial amounts to adding columns $(\cos(m + 1)\alpha_1 ,..., \cos(m + 1) \alpha_n)'$ and $(\sin(m + 1) \alpha_1 ,..., \sin(m + 1) \alpha_n)'$ as the $(2m + 2)^{th}$ and the last columns to the design matrix X given in (2.7). The augmented matrix can be written as

$$X_{(1)} = (X : W) P \quad \text{where} \quad W = \begin{bmatrix} \cos (m + 1) \alpha_1, \sin (m + 1) \alpha_1 \\ \vdots \qquad\qquad \vdots \\ \cos (m + 1) \alpha_n, \sin (m + 1) \alpha_n \end{bmatrix}$$

and P is a suitable permutation matrix which keeps the columns of W in correct

places. The corresponding model is now

$$\underline{Y}^{(1)} = X_{(1)} \underline{\lambda}_{(1)}^{(1)} + \underline{\epsilon}^{(1)}$$

$$\underline{Y}^{(2)} = X_{(1)} \underline{\lambda}_{(1)}^{(2)} + \underline{\epsilon}^{(2)}$$

(3.1)

where $\underline{\lambda}_{(1)}^{(1)}$ and $\underline{\lambda}_{(1)}^{(2)}$ are $(2m + 3) \times 1$ vectors of coefficients which can be estimated by the method of least squares. The least squares estimates are given for $i = 1, 2$, by

Now,

$$\underline{\hat{\lambda}}_{(1)}^{(i)} = [X'_{(1)} X_{(1)}]^{-1} X'_{(1)} \underline{Y}^{(i)} .$$

$$X'_{(1)} X_{(1)} = P' \begin{bmatrix} X'X & X'W \\ W'X & W'W \end{bmatrix} P .$$

In order to invert $X'_{(1)} X_{(1)}$ the following result is used.

<u>Theorem</u> (Rohde [12]). For any partitioned symmetric matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ we have

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BF^{-1}CA^{-1} & -A^{-1}BF^{-1} \\ -F^{-1}CA^{-1} & F^{-1} \end{bmatrix}$$

where $F = D - C A^{-1} B$. Using this result, we can write

$$\begin{bmatrix} X'X & X'W \\ W'X & W'W \end{bmatrix}^{-1} = \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} (X'X)^{-1} & X'W \\ & I \end{bmatrix} H^{-1} \begin{bmatrix} (X'X)^{-1} & X'W \\ & I \end{bmatrix}^{-1}$$

(3.2)

where

$$H = W'W - W'X(X'X)^{-1}X'W$$

$$= W'(I - M)W \text{ writing } M = X(X'X)^{-1}X' .$$

The least squares estimates can then be written as

$$\hat{\lambda}\binom{i}{(1)} = P'\{X'_{(1)}X_{(1)}\}^{-1} P P'\begin{bmatrix} X' \\ \cdots \\ W' \end{bmatrix}\underset{\sim}{Y}^{(i)}$$

$$= P'\begin{bmatrix} (X'X)^{-1} X' \underset{\sim}{Y}^{(i)} - (X'X)^{-1} X'N(I-M) Y^{(i)} \\ H^{-1} W'(1-M) \underset{\sim}{Y}^{(i)} \end{bmatrix} \qquad (3.3)$$

where $N = W H^{-1} W'$. The error sum of squares will now be

$$\underset{\sim}{Y}^{(i)'}\underset{\sim}{Y}^{(i)} - \underset{\sim}{Y}^{(i)'}(X \vdots W) \hat{\lambda}\binom{i}{(1)} = Y^{(i)'}(I-M)\underset{\sim}{Y}^{(i)} - Y^{(i)'}(I-M)N(I-M)Y^{(i)} \qquad (3.4)$$

on substitution from (3.3) and simplifying. From (3.4) it can be seen that the reduction in $R_0(i,i)$, the error sum of squares corresponding to $\hat{\lambda}^{(i)}$, by taking the additional terms is

$$\underset{\sim}{Y}^{(i)'}(I-M) N(I-M) \underset{\sim}{Y}^{(i)} \geq 0 . \qquad (3.5)$$

Thus before deciding to take the $(m+1)^{th}$ degree terms, we may first compute (3.5) and if this is sufficiently large from a practical point of view, we may decide to include the $(m+1)^{th}$ degree terms. When n is sufficiently large, a test for testing whether the proposed additional parameters are significantly different from zero, is available and is discussed in section 3.2. Notice that the computations (3.5) use $(I-M) \underset{\sim}{Y}^{(i)}$ which are already available and we only need to find N for which only a second order matrix H has to be inverted. Further, if the additional terms are deemed necessary, the computations we have already made are enough to estimate $\hat{\lambda}\binom{i}{(1)}$ using (3.3).

To decide whether to include yet another term, we start with $X_{(1)}$ and add two columns $(\cos(m+2)\alpha_1, \cdots, \cos(m+2)\alpha_n)'$ and $(\sin(m+2)\alpha_1, \cdots, \sin(m+2)\alpha_n)'$ as $(2m+4)^{th}$ and the last columns of $X_{(1)}$ and proceed as before. This procedure can be repeated successively.

Remark 3.1.1: In the special case when the column vector(s) W that are adjoined to the original design matrix X, are orthogonal to the columns of the design matrix X, the coefficients that we have already obtained are unaffected as the correction term

$-(X'X)^{-1}X'WH^{-1} W'(I-M)\underline{Y}^{(i)} = 0$ and the additional coefficient(s) is (are) given by $H^{-1}W'(I-M)\underline{Y}^{(i)} = (W'W)^{-1}W'\underline{Y}^{(i)}$ since $W'M = 0$ and $H = W'(I-M)W = W'W$.

3.1.2: An alternative approach which is algebraically equivalent to the above in the determination of m is to assess the increase in the estimate of the concentration parameter of the distribution of β given α.

Having computed the mth degree trigonometric polynomial approximations, say $g_{im}(\alpha)$ of $g_i(\alpha)$ for $i = 1,2$ the concentration parameter is given by

$$\rho_m(\alpha) = \{g_{1m}^2(\alpha) + g_{2m}^2(\alpha)\}^{\frac{1}{2}}. \tag{3.6}$$

Notice that $\rho_m(\alpha)$ is an increasing function of m and $\rho_m(\alpha) \uparrow \rho(\alpha)$ as $m \to \infty$ for any given α. Since $0 \leq \rho(\alpha) \leq 1$, we have $0 \leq \rho_m(\alpha) \leq 1$ for any m and any given α. Computing $\rho_m(\alpha_i)$ for each α_i, $i = 1, \cdots, n$ after an mth degree trigonometric polynomial is fitted, we can decide whether an additional term is to be included depending on whether there is a significant change in the estimate of the concentration parameter or not. Note

$$\hat{\rho}_m^2 = \frac{1}{n} \sum_{i=1}^{n} \rho_m^2(\alpha_i) = \frac{1}{n} \sum_{i=1}^{n} [g_{1m}^2(\alpha_i) + g_{2m}^2(\alpha_i)]$$

$$= \frac{1}{n} \sum_{j=1}^{2} \underline{Y}^{(j)'} M \underline{Y}^{(j)} = \frac{1}{n} \sum_{j=1}^{2} [\underline{Y}^{(j)'}\underline{Y}^{(j)} - R_0(j,j)]. \tag{3.7}$$

Working with the augmented matrix as above, it can be seen that the increase in the estimate of the square of the concentration parameter is given by

$$\frac{1}{n} \sum_{j=1}^{2} [\underline{Y}^{(j)'}(I-M) N (I-M) \underline{Y}^{(j)}],$$

which is equal to the reduction in error sum of squares given in (3.5). This shows that this criterion is equivalent to the one discussed in Section 3.1.1.

3.1.3.   The extent to which the accuracy of prediction of β would be improved by bringing in additional terms in the Fourier series approximation, can also be ascertained in a way similar to the linear analysis where partial correlation ratio is considered (C.R. Rao, [5] p. 268). It may be recalled that in linear analysis, if we are considering only linear predictors, the proportional reduction in the mean square error when variables $X_1, \cdots, X_p$ are used to predict Y rather than $X_1, \cdots, X_k$ where $p > k$, is given by

$$\rho^2_{0(k+1,\ldots,p)(1,\ldots,k)} = \frac{\rho^2_{0(1,\ldots,p)} - \rho^2_{0(1,\ldots,k)}}{1 - \rho^2_{0(1,\ldots,k)}} \tag{3.8}$$

where $\rho^2_{0(1,\ldots,m)}$ is the square of the multiple correlation coefficient between Y and $(X_1, \ldots, X_m)$. If σ_0 the vector of covariances of Y and $X_1, \ldots, X_m$ and C is the dispersion matrix of $(X_1, \ldots, X_m)$ assumed positive definite, then we have $\rho^2_{0(1,\ldots,m)}$ $= \sigma_0' C^{-1} \sigma_0 / \sigma_Y^2$. An estimate of this is obtained by taking the sample variances and covariances. In the present situation, we consider

$$\rho^{*2}_{iom} = \frac{Y^{(i)'} M Y^{(i)}}{Y^{(i)'} Y^{(i)}} = \frac{\hat{\lambda}^{(i)'}(X'X)\hat{\lambda}^{(i)}}{Y^{(i)'} Y^{(i)}} \tag{3.9}$$

for $i = 1, 2$ which is analogous to the sample multiple correlation coefficient in the linear case. Notice that $0 \leq \rho^{*2}_{iom} \leq 1$ since $Y^{(i)'} M Y^{(i)} \leq Y^{(i)'} Y^{(i)}$. This is easily seen since M is idempotent and for any positive definite matrix A, $\sup Z'AZ/Z'Z \leq \lambda_{max}$ where λ_{max} is the largest eigen value of A. The stated result follows since the eigen values of M are all $+1$. Consider

$$\rho^{*2}_{iom(m+1)} = \frac{\rho^{*2}_{io(m+1)} - \rho^{*2}_{iom}}{1 - \rho^{*2}_{iom}} \qquad \text{for } i = 1, 2 \tag{3.10}$$

which is the proportional reduction in the regression sum of squares to ascertain the accuracy of the prediction of β. If the quantities are significantly large, we decide to include the higher order term also. A large sample test to see whether this is

significantly large is give in Section 3.2. From (3.9) it can be deduced that

$$\rho^{*2}_{\text{iom}(m+1)} = \frac{\underset{\sim}{Y}^{(i)'}(I-M)\,N(I-M)\,\underset{\sim}{Y}^{(i)}}{\underset{\sim}{Y}^{(i)'}(I-M)\underset{\sim}{Y}(i)} \tag{3.11}$$

using the computations made earlier. It may be noticed that here again we are led to the same computations as in the above two criteria.

3.2 Asymptotic tests for the determination of the degree m:

From the discussion in Section 3.1 it is seen that the degree m can be determined on the basis of $S_i = \underset{\sim}{Y}^{(i)'}(I-M)N(I-M)\underset{\sim}{Y}^{(i)}$ or some suitable function thereof. If the sample size is sufficiently large the following asymptotic tests can be used for the determination of m.

<u>Proposition</u>: Let $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ and let $\underset{n\to\infty}{\lim}\left[\dfrac{X'X}{n}\right] = Q$ which is finite and nonsingular. Then for any fixed m, (1) $n^{-1/2}X'\underset{\sim}{\epsilon}^{(i)}$ converges in distribution to $N(0, \sigma_i^2\,Q)$ as $n\to\infty$ for $i = 1,2$ (2) $\sqrt{n}(\underset{\sim}{\hat{\lambda}}^{(i)} - \underset{\sim}{\lambda}^{(i)})$ converges in distribution to $N(0, \sigma_i^2\,Q^{-1})$ as $n\to\infty$ for $i = 1, 2$.

This proposition can be proved as in Schmidt [6] section 2.4 and the details are omitted.

Since $R_0(i, i)/[n - (2m + 1)]$ is a consistent estimator of σ_i^2, for sufficiently large n, we can approximate the distribution of

$$\frac{n - (2m+1)}{R_0(i,\,i)}(\underset{\sim}{\hat{\lambda}}_n^{(i)} - \underset{\sim}{\lambda}^{(i)})'(X'X)(\underset{\sim}{\hat{\lambda}}_n^{(i)} - \underset{\sim}{\lambda}^{(i)})$$

as $\chi^2(2m + 1)$.

<u>Remark 3.2.1</u>:   If one uses a bootstrap estimator σ_i^{*2} for σ_i^2, then the distribution

of $\frac{1}{\sigma_i^{*2}} (\hat{\lambda}_n^{(i)} - \underset{\sim}{\lambda}^{(i)})'(X'X)(\hat{\lambda}_n^{(i)} - \underset{\sim}{\lambda}^{(i)})$ can be approximated as $\chi^2(2m + 1)$ and

similar modification can be made in all the computations discussed below.

Further, the marginal distribution of any subset of $\hat{\underset{\sim}{\lambda}}^{(i)}$ is also asymptotically normal, and tests for determining the degree m can be based on this result.

Let $_m\underset{\sim}{\lambda}_{(1)}^{(i)} = (\lambda_{(1),2m+2}^{(i)}, \lambda_{(1),2m+3}^{(i)})$ the last two components of $\underset{\sim}{\lambda}_{(1)}^{(i)}$. Let their least squares estimates be denoted by $_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)}$.

From (3.3), we have $_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)} = H^{-1}W'(I - M)\underset{\sim}{Y}^{(i)}$ which is unbiased for $_m\underset{\sim}{\lambda}_{(1)}^{(i)}$. By direct computation we have the dispersion matrix of $_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)}$ as $V(_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)}) = \sigma_i^2 H^{-1}$. To examine whether an $(m + 1)^{th}$ degree polynomial improves the prediction significantly, we test,

$$H_0^{(i)} : {_m\underset{\sim}{\lambda}}_{(1)}^{(i)} = 0 \text{ against } H_1^{(i)} : {_m\underset{\sim}{\lambda}}_{(1)}^{(i)} \neq 0 .$$

From the above discussion, we have

$$T^{(i)} = \frac{n - (2m + 1)}{R_0(i, i)} \, {_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)}}' \, H \, {_m\hat{\underset{\sim}{\lambda}}_{(1)}^{(i)}} \qquad (3.12)$$

is asymptotically distributed as a $\chi^2(2)$ under $H_0^{(i)}$. Substituting the estimates, the test statistic has the form

$$T^{(i)} = (n - (2m + 1))\frac{\underset{\sim}{Y}^{(i)'}(I - M)N(I - M)\underset{\sim}{Y}^{(i)}}{\underset{\sim}{Y}^{(i)'}(I - M)\underset{\sim}{Y}^{(i)}} . \qquad (3.13)$$

Thus for all the three criteria proposed in Section (3.1) to determine the degree m, asymptotic tests can be obtained using $T^{(i)}$'s as defined in (3.13). We settle for an mth degree polynomial, only if both $H_0^{(i)}$ $i = 1, 2$ are accepted.

Remark 3.2.2: In the above discussion we have used a χ^2 test assuming the

consistency of of $R_0(i, i)$. For moderately large n, one can also justify the use of an F test for $H_0^{(i)}$ as follows:

<u>Lemma</u>: Under $H_0^{(i)}$, $\underset{\sim}{Y}^{(i)'}(I - M)\underset{\sim}{Y}^{(i)}$ and $\underset{\sim}{Y}^{(i)'}(I - M)N(I - M)\underset{\sim}{Y}^{(i)}$ are asymptotically independent.

Proof. Writing $R_{0k}(i, i)$ as the residual sum of squares when a kth degree polynomial is fitted, we have

$$\frac{R_{0m}(i,i)}{\sigma_i^2} = \frac{\underset{\sim}{Y}^{(i)'}(I - M)\,\underset{\sim}{Y}^{(i)}}{\sigma_i^2} \quad \text{asymptotically} \quad \chi^2(n - (2m + 1)) \quad \text{as} \quad n \to \infty$$

for $i = 1, 2$. Further, by direct computation using (3.3)

$$\frac{R_0(m + 1)^{(i,i)}}{\sigma_i^2} = \frac{\underset{\sim}{Y}^{(i)'}(I - M)\underset{\sim}{Y}^{(i)}}{\sigma_i^2} - \frac{\underset{\sim}{Y}^{(i)'}(I - M)N(I - M)\underset{\sim}{Y}^{(i)}}{\sigma_i^2}$$

which is asymptotically $\chi^2(n - (2m + 3))$. Moreover, under $H_0^{(i)}$, $\dfrac{\underset{\sim}{Y}^{(i)'}(I - M)N(I - M)\underset{\sim}{Y}^{(i)}}{\sigma_i^2}$ is asymptotically $\chi^2(2)$. From these we conclude the stated result.

Using the above lemma

$$F^{(i)} = \frac{\underset{\sim}{Y}^{(i)'}(I - M)N(I - M)\underset{\sim}{Y}^{(i)}}{\underset{\sim}{Y}^{(i)'}(I - M)\underset{\sim}{Y}^{(i)}}\, \frac{n - (2m + 1)}{2} \tag{3.14}$$

has an $F(2, n - (2m + 1))$ distribution under $H_0^{(i)}$. Thus $H_0^{(i)}$ can be tested using $F^{(i)}$ and the degree m can be determined.

<u>Remark 3.2.3</u>: To decide the degree m, it is suggested above to consider $H_0^{(i)}$ separately for $i = 1, 2$. It may be pointed out that we can derive the joint asymptotic distribution of $(\hat{\underset{\sim}{\lambda}}^{(1)}, \hat{\underset{\sim}{\lambda}}^{(2)})$, and a test can be obtained for $H_0 : {}_m\underset{\sim}{\lambda}^{(1)} = 0$ and

$$m\overset{\lambda}{\underset{(1)}{\sim}}{}^{(2)} = 0 \ .$$

4.   REGRESSION IN SOME PARAMETRIC MODELS

In this section exact forms of $g_1(\alpha)$ and $g_2(\alpha)$ in some special cases when the conditional distribution of β given α is of known form are obtained and the problems of estimation and testing hypotheses about regression parameters in these cases are considered.

4.1:   Suppose the conditional distribution of β given α is von Mises with density function

$$f(\beta|\alpha) = \frac{1}{2\pi I_0(\kappa)} \exp \kappa \cos\{\beta - \mu(\alpha)\} \tag{4.1}$$

where $\mu(\alpha)$ is the conditional mean and κ is the concentration parameter which is also possibly dependent on α. In this case

$$g_1(\alpha) = A(\kappa) \cos \mu(\alpha)$$
$$g_2(\alpha) = A(\kappa) \sin \mu(\alpha) \tag{4.2}$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. Here $\mu(\alpha)$ is of interest and without further structure, one may have to resort to the nonparametric regression methods or estimate $g_i(\alpha)$'s by the method described in Section 2.

(a)   The hypothesis that β does not depend on α is equivalent to

$$H_0 : g_1(\alpha) = A_0 \text{ and } g_2(\alpha) = C_0 \ . \tag{4.3}$$

Estimating A_0 and C_0 as described in Section 2,

$$\hat{A}_0 = \frac{1}{n} \sum_1^n \cos \beta_i \text{ and } \hat{C}_0 = \frac{1}{n} \sum_1^n \sin \beta_i \ .$$

Writing $\underset{\sim}{\xi}_i' = (\cos \beta_i - \hat{A}_0 , \sin \beta_i - \hat{C}_0)$ and $\bar{\underset{\sim}{\xi}} = \frac{1}{n} \sum \underset{\sim}{\xi}_i$, one has that $\sqrt{n}\, \hat{\Sigma}^{-1/2}\bar{\underset{\sim}{\xi}}$ is asymptotically distributed as $N_2(0, I)$ under H_0 and consequently

$$S = n \, \bar{\epsilon}' \, \hat{\Sigma}^{-1} \bar{\epsilon} \qquad (4.4)$$

is asymptotically distributed as a χ^2 variable on 2 degrees of freedom under H_0, where $\hat{\Sigma}$ is an unbiased, consistent estimate of Σ obtained from (2.12).

If data are available in the form that for each α_i there are $(\beta_{i1}, ..., \beta_{in_i})$, $i = 1, ..., p$ then $\hat{E}(\beta|\alpha_i) = \hat{\mu}(\alpha_i) = \bar{\beta}_i$ where $\bar{\beta}_i$ is the circular mean direction of $(\beta_{i1}, ..., \beta_{in_i})$. Here one can test the hypothesis that $\mu(\alpha)$ does not depend on α (i.e., the regression is constant) by the appropriate Anova procedure described in Mardia [13] p. 163).

(b) Suppose $\mu(\alpha) = \mu + a\alpha \pmod{2\pi}$ where μ and a are both unknown. A test for $a = 0$ corresponds to the regression being independent of α.

For the more general hypothesis

$$\begin{aligned} H_0 : a &= a_0 \text{ (not an integer)}, \\ \mu &= \mu_0, \text{ and } \kappa = \kappa_0 \end{aligned} \qquad (4.5)$$

one can proceed as follows. Under H_0,

$$\begin{aligned} g_1(\alpha)' &= A(\kappa_0) \cos(\mu_0 + a_0\alpha) \\ &= A(\kappa_0) \cos \mu_0 \cos a_0\alpha - A(\kappa_0) \sin \mu_0 \sin a_0\alpha \end{aligned} \qquad (4.6)$$

$$\begin{aligned} g_2(\alpha) &= A(\kappa_0) \sin(\mu_0 + a_0\alpha) \\ &= A(\kappa) \cos \mu_0\alpha + A(\kappa_0) \sin \mu_0 \cos a_0\alpha. \end{aligned}$$

But

$$\cos a_0\alpha = \frac{2a_0}{\pi} \sin a_0\pi \left[\frac{1}{2a_0^2} + \frac{\cos \alpha}{1^2 - a_0^2} - \frac{\cos 2\alpha}{2^2 + a_0^2} + ... + (-1)^{m-1} \frac{\cos m\alpha}{m^2 + a_0^2} + ... \right]$$

and

$$\sin a_0 \alpha = \frac{2}{\pi} \sin a_0 \, a_0 \pi \left[\frac{\sin \alpha}{1^2-a_0^2} - \frac{2 \sin 2\alpha}{2^2-a_0^2} + \dots + (-1)^{m-1} \frac{m \, \sin \, m\alpha}{m^2-a_0^2} + \quad (4.7)\right.$$

(cf. Carslaw [14] p. 265, problem 6). From (4.7) one can obtain the Fourier series expansions for $g_1(\alpha)$ and $g_2(\alpha)$. From this one can test the given hypothesis by comparing the estimates of the Fourier coefficients as obtained from the least squares theory with the values obtained from the above expansions (after fixing a suitable value of m) by using the proposition in Section 3.2.

(ii) If a_0 is an integer, say r, then we have

$$g_1(\alpha) = A(\kappa_0) \cos \mu_0 \cos r\alpha - A(\kappa_0) \sin \mu_0 \sin r\alpha$$
$$g_2(\alpha) = A(\kappa_0) \sin \mu_0 \cos r\alpha + A(\kappa_0) \cos \mu_0 \sin r\alpha .$$

In this case the hypothesis becomes

$$H_0: \quad A_r = A(\kappa_0) \cos \mu_0, \quad B_r = -A(\kappa_0) \sin \mu_0$$

$$C_r = A(\kappa_0) \sin \mu_0, \quad D_r = A(\kappa_0) \cos \mu_0 \tag{4.8}$$

and all the other coefficients are zero. This again can be tested using the results of the proposition in Section 3.2.

4.2: Suppose the conditional distribution of β given α is wrapped Cauchy with frequency function

$$f(\beta|\alpha) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho \cos(\beta-\mu(\alpha))} \tag{4.9}$$

where $0 \leq \rho \leq 1$, which may depend on α. From the characteristic function

$$\phi_p(\alpha) = e^{i\mu(\alpha)p-ap}$$

where $\rho = e^{-a}$, for all positive integer values of p, one has

$$g_1(\alpha) = \rho \cos \mu(\alpha)$$

$$g_2(\alpha) = \rho \sin \mu(\alpha) . \tag{4.10}$$

Here too one can consider the problems of testing as in the above case.

4.3: Suppose the conditional distribution of β given α is wrapped normal with the characteristic function

$$\phi_p(\alpha) = e^{i\mu(\alpha)p - (\sigma^2/2)p^2} \tag{4.11}$$

for all positive integer values of p. In this case

$$g_1(\alpha) = e^{-\sigma^2/2} \cos \mu(\alpha) , \quad g_2(\alpha) = e^{-\sigma^2/2} \sin \mu(\alpha) . \tag{4.12}$$

4.4: As a final example, suppose the conditional distribution of β given α is, more generally, a wrapped version of a stable law on the line with characteristic function given by

$$\phi_p(\alpha) = \exp\{i\mu(\alpha)p - c \, |p|^r \, (1 - \delta \, \text{sgn}(p) \, w(p,r)\} \tag{4.13}$$

for some $\mu(\alpha)$ real, $c > 0$, $|\delta| < 1$ and $0 < r \leq 2$ which may also depend on α, and

$$w(p,r) = \begin{cases} \tan \frac{r\pi}{2} & \text{if } r \neq 1 \\ -\frac{2}{\pi} \log |p| & \text{if } r = 1 \end{cases}$$

for any integral value of p. Special cases of this include the wrapped Cauchy distribution when $r = 1$, and the wrapped Normal distribution if $r = 2$. When $r = 0$ one has $c = 0$ and one gets the characteristic function of the degenerate distribution with mass concentrated at $\mu(\alpha)$. In the general case,

$$g_1(\alpha) = e^{-c} \cos\{\mu(\alpha) + \delta \tan \frac{r\pi}{2}]$$

$$\tag{4.14}$$

$$g_2(\alpha) = e^{-c} \sin[\mu(\alpha) + \delta \tan \frac{r\pi}{2}] .$$

Again several estimation and testing problems concerning various parameters can be handled by taking suitable trigonometric polynomial expansions of $g_1(\alpha)$ and $g_2(\alpha)$.

Remark: Throughout the discussion, we have considered the regression of one circular variable on another. The problem of regressing one circular variable on more than one circular variable, a case analogous to the multiple regression in the linear context, can also be dealt with along the lines of this paper. The results are more involved and will be considered elsewhere.

REFERENCES

[1] P.E. Jupp and K.V. Mardia, A General Correlation Coefficient for Directional Data and Related Regression Problems, *Biometrika*, **67**, 163–173, (1980).

[2] E. Batschelet, *Circular Statistics in Biology*, Acad. Press, New York. (1981).

[3] P. Rivest, Some statistical methods for bivariate circular data. *J. Roy. Statist. Soc.*, Ser. B, **37**, 81–90, (1982).

[4] S.R. Jammalamadaka and Y.R. Sarma, A Correlation Coefficient for Angular Variables, *Statistical Theory and Data Analysis II*. Ed. K. Matusita, North Holland, New York, 349–364, (1988).

[5] C.R. Rao, *Linear Statistical Inference and Its Applications*, 2nd edition, John Wiley, New York (1973).

[6] P. Schmidt, *Econometrics*. Marcel Dekker, New York, (1976).

[7] P.H. Guest, *Numerical Methods of Curve Fitting*, Cambridge University Press, Cambridge (1961).

[8] J.S. Rao, *Some contributions to the analysis of circular data*. Unpublished Ph.D. thesis, Indian Statistical Institute, Calcutta, (1969).

[9] C.F.J. Wu, Jacknife bootstrap and other resmpling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261–1350, (1986).

[10] P. Bhimasankaram and S.R. Jammalamadaka, *Recursive Estimation and Testing in General Linear Models with Applications to Regression Diagnostics*, Technical Report No. 29, Department of Statistics, University of California, Santa Barbara, (1987).

[11] G.A.F. Seber, *Applied regression analysis*. John Wiley, New York, (1977).

[12] C.A. Rohde, Generalized inverses of partitioned matrices. *SIAM J. Appl. Math*, **13**, 1033–1035, (1965).

[13] K.V. Mardia, *Statistics of Directional Data*, Acad. Press, New York (1972).

[14] H.S. Carslaw, *Introduction to the Theory of Fourier Series and Integrals*, 3rd revised edition, Dover, New York (1930).