

# Adaptive Estimation of Directional Trend

Rudolf Beran\*

Department of Statistics  
University of California, Berkeley  
Berkeley, CA 94720–3860, USA

July 2000

Consider a one-way layout with one directional observation per factor level. Each observed direction is a unit vector in  $R^p$  measured with random error. Information accompanying the measurements suggests that the mean directions, normalized to unit length, follow a trend: the factor levels are ordinal and mean directions at nearby factor levels may be close. Measured positions of the paleomagnetic north pole in time illustrate this design. The directional trend estimators studied in this paper stem from penalized least squares (PLS) fits in which the penalty function is the squared norm of first-order or second-order differences of mean vectors at adjacent factor levels. Expressed in spectral form, such PLS estimators suggest a much larger class of monotone shrinkage estimators that use the orthogonal basis implicit in PLS. Penalty weights and, more generally, monotone shrinkage factors are selected to minimize estimated risk. The possibly large risk reduction achieved by such adaptive monotone shrinkage estimators reflects the economy of the PLS orthogonal basis in representing the actual trend and the flexibility of unconstrained monotone shrinkage.

*AMS classification:* 62H11, 62J99

*Keywords and phrases:* directional data, penalized least squares, monotone shrinkage, economical basis, risk estimation, superefficiency, symmetric linear smoother.

## 1. INTRODUCTION

Consider  $n$  independent measurements taken successively in time on the varying position of the earth's north magnetic pole. Each measured position may be represented as a unit vector in  $R^3$  that gives direction from the center of the earth to the north magnetic pole. Because of measurement errors, it is plausible to model the data as a realization of independent random unit vectors  $\{y_i: 1 \leq i \leq n\}$  whose mean vectors are  $\eta_i = E(y_i)$ . The subscript  $i$  labels time. The *mean direction* of  $y_i$  is defined to be the unit vector  $\mu_i = \eta_i/|\eta_i|$ . In this example, the mean directions  $\{\mu_i\}$  follow a trend, by which we mean that the subscript order matters and that the distance between  $\mu_i$  and  $\mu_j$  may be relatively small when  $i$  is close to  $j$ .

The naive estimator of  $\mu_i$  is  $\hat{\mu}_{N,i} = y_i$ . It can be derived as the maximum likelihood

---

\* This research was supported in part by National Science Foundation Grant DMS99–70266.

estimator of  $\mu_i$  when the distribution of  $y_i$  is Fisher-Langevin with mean direction  $\mu_i$  and precision parameter  $\kappa$ . Unless measurement error is very small,  $\{\hat{\mu}_{N,i}\}$  is not a satisfactory estimator of the directional trend  $\{\mu_i\}$ . If we foresee that the trend in the means  $\{\eta_i\}$  may possess some degree of smoothness, not known to us, it is natural to look for more efficient estimators within classes of smoothers. In an instructive data analysis, Irving (1977) suggested forming local symmetric weighted averages of the  $\{y_i\}$ , normalizing these to unit length so as to obtain a more revealing estimator of directional trend.

A symmetric weighted average is a particular symmetric linear smoother in the sense of Buja, Hastie, and Tibshirani (1989). We may consider any large class of symmetric linear smoothers as candidate estimators for the mean vectors  $\{\eta_i\}$  and then proceed as follows: (a) estimate the quadratic risk of each such estimator without assuming any smoothness in the sequence of unknown mean vectors; (b) choose the candidate estimator that minimizes estimated risk; (c) normalize the estimated mean vectors to unit length so as to estimate the directional trend  $\{\mu_i\}$ . The candidate symmetric linear smoothers treated in this paper generalize certain penalized least squares (PLS) trend estimators. Details of the methodology are presented in Sections 2 and 3. Computational experiments reported in Sections 2 and 4 bring out how the proposed estimators reduce risk through constructive interplay between basis economy and unconstrained monotone shrinkage. Asymptotic theory developed in Section 5 supports key details of the methodology and quantifies how basis economy reduces risk.

Other directional trend estimators proposed by Watson (1985), Fisher and Lewis (1985), and Jupp and Kent (1987) rely on analogs of cubic-spline or kernel methods for curve smoothing in Euclidean spaces. Tacit in these treatments are assumptions on the smoothness of the unknown trend. The methods of this paper assume no smoothness in the unknown directional trend but take advantage of any smoothness present to reduce estimation risk.

## 2. CONSTRUCTION OF ESTIMATORS

As in the north magnetic pole example, choose subscripts so that  $y_i$  is the directional observation associated with the  $i$ -th smallest factor level. The directional trend estimators in this paper stem from penalized least squares (PLS) estimators for the mean vectors  $\{\eta_i: 1 \leq i \leq n\}$  in which the penalty function is the squared norm of first-order or second-order differences of the  $\{\eta_i\}$ . It will be convenient in the exposition to suppose that the measured directions are unit vectors in  $R^p$ . The practically important spherical and circular cases correspond to  $p = 3$  and  $p = 2$ , respectively.

### 2.1. Candidate Estimator Classes

The  $n \times p$  data matrix formed from the observed unit vectors  $\{y_i: 1 \leq i \leq n\}$  in  $R^p$  is

$$Y = \begin{pmatrix} y'_1 \\ \vdots \\ y'_n \end{pmatrix} = (y_{(1)}, \dots, y_{(p)}). \quad (1)$$

Here  $y_{(j)}$  denotes the  $j$ -th column of  $Y$ . The analogously organized matrix of mean vectors is then

$$H = \mathbb{E}(Y) = \begin{pmatrix} \eta'_1 \\ \vdots \\ \eta'_n \end{pmatrix} = (\eta_{(1)}, \dots, \eta_{(p)}). \quad (2)$$

*First-order Penalized Least Squares.* Let  $|\cdot|$  denote Euclidean matrix norm, so that  $|A|^2 = \text{tr}(AA') = \text{tr}(A'A)$ . For any  $r \times n$  matrix  $B$ , where  $r \leq n$ , define

$$d(H, B, \gamma) = |Y - H|^2 + \gamma|BH|^2. \quad (3)$$

Let  $D$  be the  $(n - 1) \times n$  first-difference matrix

$$D = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \quad (4)$$

The first-order PLS candidate estimators for  $H$  are the one-parameter family of symmetric linear smoothers  $\{\hat{H}_D(\gamma): \gamma \geq 0\}$ , where

$$\hat{H}_D(\gamma) = \underset{H}{\text{argmin}} d(H, D, \gamma) = (I + \gamma D'D)^{-1}Y. \quad (5)$$

For positive  $\gamma$ , the estimated means in  $\hat{H}_D(\gamma)$  are more nearly constant in  $i$  than the measured directions  $y_i$ .

A spectral decomposition of  $D'D$  is  $U_D \Lambda_D U_D'$ , where  $U_D$  is an orthogonal matrix,  $\Lambda_D = \text{diag}\{\lambda_{D,n-i+1}\}$ , and  $\lambda_{D,1} \geq \dots \geq \lambda_{D,n} = 0$ . The eigenvectors that form the columns of  $U_D$  are ordered so that the successive diagonal elements of  $\Lambda_D$  are nondecreasing. It follows from (5) that

$$\hat{H}_D(\gamma) = U_D F_D(\gamma) U_D' Y \quad \text{with} \quad F_D(\gamma) = (I + \gamma \Lambda_D)^{-1}. \quad (6)$$

Let  $f_D(\gamma) = \{f_{D,i}(\gamma): 1 \leq i \leq n\}$  denote the diagonal vector of the diagonal matrix  $F_D(\gamma)$ . Evidently  $1 \geq f_{D,1}(\gamma) \geq f_{D,2}(\gamma) \geq \dots \geq f_{D,n}(\gamma) \geq 0$ . Formula (6) plus modern algorithms for spectral decomposition provide a numerically stable method for computing the first-order PLS candidate estimators  $\{H_D(\gamma): \gamma > 0\}$ . Other computational methods for  $\hat{H}_D(\gamma)$  are discussed in Press, Teukolsky, Vetterling and Flannery (1992), section 18.5.

*Second-order Penalized Least Squares.* Let  $E$  be the  $(n - 2) \times n$  second-difference matrix

$$E = \begin{pmatrix} -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \end{pmatrix}. \quad (7)$$

The second-order PLS candidate estimators for  $H$  are the one-parameter family of symmetric linear smoothers  $\{\hat{H}_E(\gamma): \gamma \geq 0\}$ , where

$$\hat{H}_E(\gamma) = \underset{H}{\text{argmin}} d(H, E, \gamma) = (I + \gamma E'E)^{-1}Y. \quad (8)$$

For positive  $\gamma$ , the estimated means in  $\hat{H}_E(\gamma)$  are more nearly linear in  $i$  than the measured directions  $y_i$ . Replacing  $D$  in (6) with  $E$  yields a computationally useful alternative formula for  $\hat{H}_E(\gamma)$ . Here too, the diagonal elements of the matrix  $F_E(\gamma) = \text{diag}\{f_E(\gamma)\}$  satisfy  $1 \geq f_{E,1}(\gamma) \geq f_{E,2}(\gamma) \geq \dots \geq f_{E,n}(\gamma) \geq 0$ .

*Monotone Shrinkage Smoothers.* Abstracting the structure in formula (6) suggests larger families of candidate symmetric linear estimators for  $H$ . Let

$$\mathcal{F}_{Mon} = \{f \in [0, 1]^n: f_1 \geq f_2 \geq \dots \geq f_n\} \quad (9)$$

and let  $F = \text{diag}\{f\}$ . The class of monotone shrinkage candidate estimators for  $H$  associated with a specified orthogonal matrix  $U$  is

$$\mathcal{C}_{Mon}(U) = \{\hat{H}(f, U): f \in \mathcal{F}_{Mon}\} \quad \text{with} \quad \hat{H}(f, U) = UFU'Y. \quad (10)$$

Evidently, the first-order PLS candidate estimators are a proper subset of  $\mathcal{C}_{Mon}(U_D)$  in which the shrinkage vectors are restricted to  $\{f_D(\gamma): \gamma > 0\}$ . Similarly, the second-order PLS candidate estimators are a proper subset of  $\mathcal{C}_{Mon}(U_E)$  in which the shrinkage vectors are restricted to  $\{f_E(\gamma): \gamma > 0\}$ . The development in this paper emphasizes superefficient estimators of directional trend constructed from the candidate estimators  $\mathcal{C}_{Mon}(U_D)$  and  $\mathcal{C}_{Mon}(U_E)$  rather than from their PLS subsets. Enlarging the class of candidates can only decrease the risk of the best candidate and proves to be advantageous computationally.

## 2.2. Choice of Candidate Estimator and Normalization

If the risk function were known, it would be reasonable to choose the candidate monotone shrinkage estimator of  $H$  that minimizes risk, using a quadratic loss function for algebraic tractability. Because risk is not known, this oracle estimator is not realizable. Instead, we will select the candidate monotone shrinkage estimator that minimizes *estimated* risk and verify that the asymptotic performance of this estimator matches that of the oracle estimator.

For the risk calculations, we will assume that the measured directions  $\{y_i: 1 \leq i \leq n\}$  are independent column vectors in  $R^p$ , each having unit length. The distribution of  $y_i$  is Fisher-Langevin with mean direction  $\mu_i$ , a unit vector, and precision  $\kappa > 0$ . Properties of this probability model are developed in Watson (1983) or Mardia and Jupp (2000) and are summarized in Fisher, Lewis and Embleton (1987).

Let  $Z = U'Y$  and write, in analogy to (1),

$$Z = \begin{pmatrix} z'_1 \\ \vdots \\ z'_n \end{pmatrix} = (z_{(1)}, \dots, z_{(p)}). \quad (11)$$

For any vector  $h$ , let  $\text{ave}(h)$  denote the average of the components of  $h$ . Section 3 develops an estimator for the quadratic risk of  $\hat{H}(f, U)$  that is uniformly consistent over all  $f \in \mathcal{F}_M$  as  $\kappa$  and  $n$  tend to infinity:

$$\hat{\rho}(f, U) = \text{ave}[\hat{\kappa}^{-1}qf^2 + (z^2 - \hat{\kappa}^{-1}q)(1 - f)^2] \quad \text{with} \quad z^2 = \sum_{j=1}^p z_{(j)}^2. \quad (12)$$

Here  $q = p - 1$  while  $\hat{\kappa}^{-1}$  is a suitably consistent estimator for the dispersion  $\kappa^{-1}$ . Section 3.2 offers one possible construction of  $\hat{\kappa}^{-1}$ .

For specified orthogonal matrix  $U$ , define the adaptive monotone shrinkage estimator of  $H$  to be

$$\hat{H}_{Mon}(U) = \hat{H}(\hat{f}_{Mon}(U), U) \quad \text{with} \quad \hat{f}_{Mon}(U) = \underset{f \in \mathcal{F}_{Mon}}{\operatorname{argmin}} \hat{\rho}(f, U). \quad (13)$$

The first and second-order monotone shrinkage estimators  $\hat{H}_{Mon}(1)$  and  $\hat{H}_{Mon}(2)$  are specific instances of (13) with  $U = U_D$  and  $U_E$  respectively. In the notation that follows (6), the adaptive first-order PLS estimator of  $H$  is defined to be

$$\hat{H}_{PLS}(1) = \hat{H}(f_D(\hat{\gamma}_D), U_D) \quad \text{with} \quad \hat{\gamma}_D = \underset{\gamma > 0}{\operatorname{argmin}} \hat{\rho}(f_D(\gamma), U_D). \quad (14)$$

Replacing the first-difference matrix  $D$  in (14) with the second-difference matrix  $E$  defines the second-order PLS estimator  $\hat{H}_{PLS}(2)$  of  $H$ . Normalizing to unit length the rows of these respective estimators of  $H$  yields the monotone shrinkage estimators  $\hat{M}_{Mon}(k)$  and the PLS estimators  $\hat{M}_{PLS}(k)$  of the directional trend  $\{\mu_i: 1 \leq i \leq n\}$ .

[Figure 1A goes near here]

*A Paleomagnetic Example.* The directional data fitted in Figure 1A consists of  $n = 26$  measured positions of the paleomagnetic north pole taken from rock specimens at various sites in Antarctica of various ages. Kent and Jupp (1987, pp. 42–45) give the data and its provenance. Each subplot uses the Schmidt net, an area-preserving projection of the northern hemisphere onto the plane (cf. Section 4.2). The perimeter of each circle represents the equator while the center corresponds to the geographical north pole. Linear interpolation between successive mean directions or estimated mean directions is used to indicate the time sequence.

The subplot in the first row exhibits the measured directions, which coincide with the naive trend estimator. Even with linear interpolation between successive observations, it is difficult to see a pattern, especially in the most recent observations near the geographic north pole. Cells (2,1) and (2,2) display the first-order estimates  $\hat{M}_{PLS}(1)$  and  $\hat{M}_{Mon}(1)$  while cells (3,1) and (3,2) exhibit the second-order estimates  $\hat{M}_{PLS}(2)$  and  $\hat{M}_{Mon}(2)$ . Both monotone shrinkage fits and the second-order PLS fit are similar in appearance. Which should we use? On the basis of estimated risks and diagnostic plots, we will argue in Section 4.1 that the best of the competing estimates for this data is  $\hat{M}_{Mon}(1)$ . Next best, though with substantially larger estimated risks, are  $\hat{M}_{Mon}(2)$  and  $\hat{M}_{PLS}(2)$ , in that order.

In their analysis of the data, Kent and Jupp (1987, Figs. 1 and 2) unwrapped the sphere and data onto a plane, used a cubic spline fit to the planar data, then wrapped this fit back onto the sphere. Their spline fit on the sphere is similar to  $\hat{M}_{PLS}$  but has a surprising kink in the tail near the left edge of the Schmidt net plot. They noted that this kink lacks physical significance and is an artifact of the spline-fitting technique. That the three

estimates  $\hat{M}_{Mon}(1)$ ,  $\hat{M}_{PLS}(2)$  and  $\hat{M}_{Mon}(2)$  agree in broad visual features with the Kent-Jupp spline estimate but lack its suspect kink is a point in favor of the shrinkage estimates.

Section 3 treats estimation of dispersion  $\kappa^{-1}$ , risk estimation, and computational algorithms. Diagnostic plots for competing PLS or monotone shrinkage estimators and computational experiments are the subject of Section 4. Asymptotic theory in Section 5 brings out three important properties. First, adaptation works for the PLS and monotone shrinkage candidate estimators in the sense that minimizing estimated risk also minimizes risk asymptotically. Second, the asymptotic risk of the estimators  $\hat{M}_{Mon}(k)$  and  $\hat{M}_{PLS}(k)$  never exceeds that of the naive estimator and can be much smaller. Third, for greatest superefficiency of these estimators, the projection of the mean vectors  $\{\eta_i\}$  on the first few columns of  $U$  should yield an accurate approximation to the  $\{\eta_i\}$ . A diagnostic plot is available for identifying this favorable situation.

### 3. ESTIMATED RISKS AND ALGORITHMS

This section motivates the risk estimator  $\hat{\rho}(f, U)$  defined in (12) and discusses methods for computing the directional trend estimators  $\hat{M}_{Mon}(k)$  and  $\hat{M}_{PLS}(k)$  defined above.

#### 3.1. Estimating Risks

We suppose in our analysis that the directions  $\{y_i: 1 \leq i \leq n\}$  are independent unit random vectors in  $R^p$ . The distribution of  $y_i$  is Fisher-Langevin  $(\mu_i, \kappa)$ . As  $\kappa$  tends to infinity, it is known that, for  $q = p - 1$ ,

$$\begin{aligned}\eta_i &= E(y_i) = [1 - (2\kappa)^{-1}q]\mu_i + o(\kappa^{-1}) \\ \text{Cov}(y_i) &= \kappa^{-1}(I - \mu_i\mu_i') + o(\kappa^{-1})\end{aligned}\tag{15}$$

and that

$$\kappa^{-1/2}(y_i - \eta_i) \Rightarrow N_p(0, I - \mu_i\mu_i')\tag{16}$$

(see Watson (1983), chapter 4). The limiting normal distribution on the right side of (16) is singular, supported on the  $q$  dimensional subspace orthogonal to  $\mu_i$ . From (15) and independence of the rows in  $Y$ ,

$$E(y_{(j)}) = \eta_{(j)}, \quad \text{Cov}(y_{(j)}) = \text{diag}\{\sigma_{ij}^2: 1 \leq i \leq n\},\tag{17}$$

where  $\sigma_{ij}^2 = \kappa^{-1}(1 - \mu_{ij}^2) + o(\kappa^{-1})$  and  $\mu_{ij}$  denotes the  $j$ -th component of the unit vector  $\mu_i$ . Hence,

$$E \sum_{j=1}^p y_{(j)}y_{(j)}' = \sum_{j=1}^p \eta_{(j)}\eta_{(j)}' + \kappa^{-1}qI + o(\kappa^{-1})\tag{18}$$

as  $\kappa$  tends to infinity.

The performance of any directional trend estimator  $\{\hat{\mu}_i = \hat{\eta}_i/|\hat{\eta}_i|\}$  will be measured indirectly through the normalized quadratic loss

$$L_n(\hat{H}, H) = n^{-1}|\hat{H} - H|^2 = n^{-1} \sum_{i=1}^n |\hat{\eta}_i - \eta_i|^2 = n^{-1} \sum_{j=1}^p |\hat{\eta}_{(j)} - \eta_{(j)}|^2,\tag{19}$$

which compares the  $\{\hat{\eta}_i\}$  with the means  $\{\eta_i\}$ . This loss function leads to tractable formulae for risk. Tacit in our use of this loss is the supposition that a good estimator of the trend in means  $\{\eta_i\}$  will map, by normalizations to unit length, into a good estimator of the directional trend  $\{\mu_i\}$ . Experiments reported in Section 4 offer empirical support for this assumption.

For specified orthogonal matrix  $U$ , let  $Z = U'Y$  as in (11),  $\Xi = E(Z) = U'H$  and  $\hat{\Xi}(f, U) = U'\hat{H}(f, U) = FZ$ . By analogy with equation (11),

$$\Xi = \begin{pmatrix} \xi'_1 \\ \vdots \\ \xi'_n \end{pmatrix} = (\xi_{(1)}, \dots, \xi_{(p)}), \quad (20)$$

From (18),

$$E \sum_{j=1}^p z_{(j)} z'_{(j)} = \sum_{j=1}^p \xi_{(j)} \xi'_{(j)} + \kappa^{-1} q I + o(\kappa^{-1}) \quad (21)$$

as  $\kappa$  tends to infinity because  $z_{(j)} = U'y_{(j)}$  and  $\xi_{(j)} = U'\eta_{(j)}$ .

Under loss (19) and the Fisher-Langevin model, the risk of candidate estimator  $\hat{H}(f, U)$  is

$$R_n(\hat{H}(f, U), H, \kappa) = n^{-1} E |\hat{H}(f, U) - H|^2 = n^{-1} E |FZ - \Xi|^2 = R_n(\hat{\Xi}(f, U), \Xi, \kappa). \quad (22)$$

Let

$$\rho(f, \xi^2, \kappa) = \text{ave}[\kappa^{-1} q f^2 + \xi^2 (1 - f)^2] \quad \text{with} \quad \xi^2 = \sum_{j=1}^p \xi_{(j)}^2. \quad (23)$$

Here all operations on vectors are performed componentwise as in the S language (cf. Becker and Chambers (1984)). Applying (21) to (22) yields

$$\begin{aligned} R_n(\hat{\Xi}(f, U), \Xi, \kappa) &= n^{-1} \sum_{j=1}^p E |Fz_{(j)} - \xi_{(j)}|^2 \\ &= n^{-1} \sum_{j=1}^p \text{tr}[F^2 \text{Cov}(z_{(j)}) + (I - F)^2 \xi_{(j)} \xi'_{(j)}] \\ &= \rho(f, \xi^2, \kappa) + o(\kappa^{-1}) \end{aligned} \quad (24)$$

as  $\kappa$  tends to infinity.

We argue in Section 5 that, for any choice of the orthogonal matrix  $U$ ,

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\kappa n^{-1} |H|^2 \leq r} \kappa |R_n(\hat{H}_{Mon}(U), H, \kappa) - \tau_{Mon}(\xi^2, \kappa)| = 0, \quad (25)$$

where

$$\tau_{Mon}(\xi^2, \kappa) = \min_{f \in \mathcal{F}_{Mon}} \rho(f, \xi^2, \kappa). \quad (26)$$

This entails that the asymptotic risk of  $\hat{H}_{Mon}(U)$  matches that of the oracle estimator, which is the monotone shrinkage estimator that minimizes actual risk. In particular, this asymptotic risk cannot exceed  $\kappa^{-1}q$ , the asymptotic risk of the naive trend estimator, and is often much smaller.

To estimate the risk function in (22), it suffices for large  $\kappa$  to estimate the function  $\rho(f, \xi^2, \kappa)$ . It follows from (21) and the definition of  $z^2$  in (12) that

$$\begin{aligned} \text{E ave}[z^2(1-f)^2] &= n^{-1} \text{E tr}[(I-F)^2 \sum_{j=1}^p z_{(j)} z'_{(j)}] \\ &= n^{-1} \text{tr}[(I-F)^2 \{ \sum_{j=1}^p \xi_{(j)} \xi'_{(j)} + \kappa^{-1} q I \}] + o(\kappa^{-1}) \\ &= \text{ave}[(\xi^2 + \kappa^{-1} q)(1-f)^2] + o(\kappa^{-1}). \end{aligned} \tag{27}$$

This calculation, (23), and (24) motivate estimating the risk of  $\hat{H}(f, U)$  by the function  $\hat{\rho}(f, U)$  defined in (12). Section 5 gives this risk estimator much stronger theoretical support.

Scrutiny of formulae (23) and (24) throws light on ideal choice of the orthogonal basis matrix  $U$ . We say that the basis provided by the columns of  $U$  is *economical* in representing  $\Xi$  if all but the first few components of  $\xi^2$  are very nearly zero. In that case, setting the first few components of  $f$  equal to one and the remaining components to zero yields a monotone shrinkage candidate estimator of  $H$  whose risk, for large  $\kappa$ , is much smaller than that of the naive estimator  $Y$ .

### 3.2. Estimating Dispersion

A simple first-difference estimator of dispersion  $\kappa^{-1}$  may be constructed from the norm of first-differences among the observed directions  $\{y_i\}$ . The asymptotic approximations below indicate that the bias of this estimator is modest if the norm of first differences among the mean vectors  $\{\eta_i\}$  is relatively small. When this is not the case, analogous estimators of dispersion can be constructed from the norm of higher-order differences.

The first-difference dispersion estimator is

$$\hat{\kappa}^{-1} = (2q)^{-1} \sum_{i=2}^n |y_i - y_{i-1}|^2. \tag{28}$$

If

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \kappa n^{-1} \sum_{i=2}^n |\eta_i - \eta_{i-1}|^2 = 0, \tag{29}$$

then  $\hat{\kappa}^{-1}$  is a consistent estimator in the sense that

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \kappa \text{E} |\hat{\kappa}^{-1} - \kappa^{-1}| = 0. \tag{30}$$

To verify this, let  $T = \sum_{i=2}^n |y_i - y_{i-1}|^2$ ,  $e_i = \kappa^{1/2}(y_i - \eta_i)$  and  $d_i = \kappa^{1/2}(\eta_i - \eta_{i-1})$ . Evidently

$$\kappa T = n^{-1} \sum_{i=2}^n |e_i - e_{i-1}|^2 + n^{-1} \sum_{i=2}^n |d_i|^2 + n^{-1} \sum_{i=2}^n d'_i (e_i - e_{i-1}). \tag{31}$$

Because of (16), Skorokhod's theorem and Vitali's theorem, there exist versions of the  $\{e_i\}$  and independent random column vectors  $\{w_i\}$  such that the distribution of  $w_i$  is  $N(0, I - \mu_i \mu_i')$  and  $\lim_{\kappa \rightarrow \infty} \mathbb{E}|e_i - w_i|^2 = 0$ . These facts and (29) imply

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \mathbb{E}|\kappa T - n^{-1} \sum_{i=2}^n |w_i - w_{i-1}|^2| = 0. \quad (32)$$

On the other hand,

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \mathbb{E}|n^{-1} \sum_{i=2}^n |w_i - w_{i-1}|^2 - 2q| = 0. \quad (33)$$

Limits (32) and (33) imply the consistency property (30) for the original random variables.

### 3.3. Computational Aspects

The following remarks concern computation of the directional trend estimators  $\hat{M}_{Mon}(k)$  and  $\hat{M}_{PLS}(k)$ . Let  $\hat{g} = (z^2 - \hat{\kappa}^{-1}q)/z^2$ . Because the estimated risk function (12) satisfies

$$\hat{\rho}(f, U) = \text{ave}[(f - \hat{g})^2 z^2] + \text{ave}(\hat{\kappa}^{-1}q\hat{g}), \quad (34)$$

definition (13) of  $\hat{f}_{Mon}(U)$  is equivalent to the constrained isotonic weighted least squares evaluation

$$\hat{f}_{Mon}(U) = \underset{f \in \mathcal{F}_{Mon}}{\text{argmin}} \text{ave}[(f - \hat{g})^2 z^2]. \quad (35)$$

This expression reveals that  $\hat{f}_{Mon}(U)$  is a regularization of the raw shrinkage vector  $\hat{g} \in (-\infty, 1]^n$ .

Let  $\mathcal{H} = \{h \in R^n: h_1 \geq h_2 \geq \dots \geq h_n\}$ , a superset of  $\mathcal{F}_M$ . An argument in Beran and Dümbgen (1998) shows that

$$\hat{f}_{Mon}(U) = \check{f}_+ \quad \text{with} \quad \check{f} = \underset{f \in \mathcal{H}}{\text{argmin}} \text{ave}[(f - \hat{g})^2 z^2]. \quad (36)$$

The pool-adjacent-violators algorithm for isotonic regression (see Robertson, Wright and Dykstra (1988)) finds  $\check{f}$  expeditiously in a finite number of steps. The positive-part clipping in (36) arises because  $\hat{g}$  is restricted to  $(-\infty, 1]^n$  rather than to  $[0, 1]^n$ .

Similarly, definition (14) of  $\hat{\gamma}_D$  for first-order PLS is equivalent to the constrained non-linear weighted least squares evaluation

$$\hat{\gamma}_D = \underset{\gamma > 0}{\text{argmin}} \text{ave}[(f_D(\gamma) - \hat{g})^2 z^2], \quad (37)$$

where  $f_D(\gamma)$  is given in the discussion that follows (6). The S-Plus function `nls()` may be summoned to solve this problem iteratively, in the manner described on p. 244 of Venables and Ripley (1999). Simple grid search provides the necessary starting approximation to  $\hat{\gamma}_D$ . With minor changes in the code, the R function `nls()` also iterates to  $\hat{\gamma}_D$ . Computation of  $\hat{\gamma}_E$  for second-order PLS is entirely analogous.

In numerical experiments, computation of the monotone shrinkage estimator  $\hat{M}_{Mon}(k)$  was considerably faster than computation of the motivating PLS estimator  $\hat{M}_{PLS}(k)$ . This finding together with the theoretical superiority in risk of  $\hat{M}_{Mon}(k)$  over  $\hat{M}_{PLS}(k)$  provides strong grounds for considering only the former.

## 4. EXPERIMENTS AND DIAGNOSTICS

Section 4.1 discusses estimated risks and diagnostic plots for the competing fits to the paleomagnetic data presented in Section 2.2. Further experiments with artificial data, described in Sections 4.2 and 4.3, suggest that the orthogonal matrices  $U_D$  and  $U_E$  implicit in the PLS fits provide economical bases for a range of directional trends. Consequently, the PLS estimators described in Section 2 have much smaller estimated risk than the naive trend estimator; and the associated monotone shrinkage fits reduce risk further. The experimental results support theoretical conclusions developed in Section 5 about the benefits of basis economy while revealing aspects of estimator performance not covered by the asymptotics.

### 4.1 Paleomagnetic Data

For this data,  $\hat{\kappa} = 19.7$  and the *rescaled* risk estimates  $\hat{\kappa}\hat{\rho}(\hat{f}, U)$  for the competing fits displayed in Figure 1A are:

$\hat{M}_{PLS}(1)$	$\hat{M}_{Mon}(1)$	$\hat{M}_{PLS}(2)$	$\hat{M}_{Mon}(2)$	Naive
.477	.203	.351	.292	2.000

The estimate  $\hat{M}_{Mon}(1)$  is the clear winner in having smallest estimated risk, far smaller than that of the naive estimator. It is noteworthy that the relatively small estimated risk of  $\hat{M}_{Mon}(1)$  is coupled with a pleasing visual appearance. The estimate gives a clear picture of the time-trend in the position of paleomagnetic north pole as measured from Antarctica.

The diagnostic plots in Figure 1B provide further insight into the behavior of these directional trend estimates. Let  $v = \hat{\kappa}^{1/2}|z|$ . In cells (1,1) and (1,2), the plots of  $v_i^{1/2}$  versus  $i$  suggest that  $U_D$  provides a more economical basis for the unknown trend than  $U_E$ : the  $\{v_i\}$  for the first basis tend to zero faster than for the second basis. The square root transformation enhances visibility of the smaller components. Greater basis economy explains why  $\hat{M}_{Mon}(1)$  has smaller estimated risk than  $\hat{M}_{Mon}(2)$ .

Cell (2,1) displays, with linear interpolation, the successive components of the shrinkage vectors  $\hat{f}_{Mon}(U_D)$  (dashed line) and  $f_D(\hat{\gamma}_D)$  (solid line) that enter into the constructions of  $\hat{M}_{Mon}(1)$  and  $\hat{M}_{PLS}(1)$  respectively. We see that  $f_D(\hat{\gamma}_D)$  provides only a rough approximation to the better  $\hat{f}_{Mon}(U_D)$  and gives more weight to higher “frequencies.” This observation explains both the ragged visual appearance of  $\hat{M}_{PLS}(1)$  and the substantially smaller estimated risk of  $\hat{M}_{Mon}(1)$ . The free-floating points in cell (2,1) are the components of the raw shrinkage vector  $\hat{g}$  plotted against  $i$  without interpolation. It is these highly irregular values that monotone and PLS shrinkage vectors approximate in constrained fashion through (35) and (37) respectively. In cell (2,2), the analogous plots of the shrinkage vectors  $\hat{f}_{Mon}(U_E)$  (dashed line) and  $f_E(\hat{\gamma}_E)$  (solid line) reveal that the latter is a good approximation to the

former. This explains why the estimated risk of  $\hat{M}_{Mon}(2)$  is not much smaller than that of  $\hat{M}_{PLS}(2)$ .

Both the economy of the orthogonal basis and the quality of the shrinkage strategy affect the risk of the directional trend estimate. In this example, first and second-order PLS generate orthogonal bases that are plausibly economical. However, the strongly constrained one-parameter shrinkage strategy implicit in PLS can fail to exploit basis economy. Adaptive monotone shrinkage takes full advantage of basis economy and is computationally faster than adaptive PLS. There seems little reason to use PLS trend estimators except as a source of potentially economical orthogonal bases.

[Figure 1B goes near here]

## 4.2. Generating and Plotting Trend Data

This subsection summarizes ideas used to generate and plot pseudo-random directional trend data in three dimensions. All calculations and plots were done in Windows S-Plus 3.2 with `set.seed(2)`. As a software check, the computations were repeated in Unix S-Plus 3.4. Very similar results were obtained, after small changes in the code, in Unix R 1.00.

*Cartesian and Polar Representations.* A direction in  $R^3$  is a unit vector  $u = (a, b, c)'$  that has an equivalent polar coordinate representation  $(\theta, \phi)$ , where  $\theta \in [0, \pi]$  and  $\phi \in [0, 2\pi)$ . Direction  $(0, 0, 1)'$  is the north pole of the coordinate system. On the one hand,

$$a = \sin(\theta) \cos(\phi), \quad b = \sin(\theta) \sin(\phi), \quad c = \cos(\theta). \quad (38)$$

On the other hand,

$$\theta = \cos^{-1}(c) \in [0, \pi], \quad \phi = \tan^{-1}(b/a) \in [0, 2\pi). \quad (39)$$

These values may be computed by using S-Plus functions `acos()` and `atan(, )`.

*Generating a Fisher-Langevin  $(\mu, \kappa)$  random direction.* Let  $V_1, V_2$  be independent random variables, each uniformly distributed on  $[0, 1]$ . Define

$$\begin{aligned} \theta &= \cos^{-1}(\kappa^{-1}\delta - 1) \quad \text{with} \quad \delta = \log[1 + (\exp(2\kappa) - 1)V_1] \\ \phi &= 2\pi V_2. \end{aligned} \quad (40)$$

The random unit vector  $u$  with polar coordinates  $(\theta, \phi)$  has a Fisher-Langevin distribution with mean direction  $\nu_0 = (0, 0, 1)'$  and precision  $\kappa$  (e.g. Mardia and Jupp (2000)). For any unit vector  $\mu$ , the orthogonal matrix

$$O(\mu) = (1 + \nu_0' \mu)^{-1} (\nu_0 + \mu)(\nu_0 + \mu)' - I \quad (41)$$

rotates  $\nu_0$  into  $\mu$  (Watson (1983, p. 28)). Thus, the random unit vector  $O(\mu)u$  has a Fisher-Langevin distribution with mean vector  $\mu$  and precision  $\kappa$ .

*Generating and Plotting Trend Data.* Let  $f$  and  $g$  be functions that map  $[0, 1]$  into, respectively,  $[0, \pi]$  and  $[0, 2\pi)$ . The pairs

$$\theta_{\mu,i} = f[i/(n+1)], \quad \phi_{\mu,i} = g[i/(n+1)], \quad 1 \leq i \leq n \quad (42)$$

determine in polar coordinates a trend of  $n$  successive mean direction vectors  $\{\mu_i\}$ . Let  $\{u_i: 1 \leq i \leq n\}$  be independent unit random vectors, each constructed using (40) and (38) to have a Fisher-Langevin  $(\nu_0, \kappa)$  distribution. Then the

$$y_i = O(\mu_i)u_i, \quad 1 \leq i \leq n \quad (43)$$

are independent and  $y_i$  has a Fisher-Langevin  $(\mu_i, \kappa)$  distribution. This method, applied to pseudo-random Uniform  $(0, 1)$  variates, generated the data for the experiments in the next subsection.

The figures in this paper use the Schmidt net to plot directions in three dimensions. In this area-preserving projection of the sphere into the plane, the three-dimensional direction  $(\theta, \phi)$  is plotted as the planar point having polar coordinates  $(r, \phi)$ , where

$$\begin{aligned} r &= 2 \sin(\theta/2) & \text{if } 0 \leq \theta \leq \pi/2 \\ r &= 2 \sin[(\pi - \theta)/2] & \text{if } \pi/2 < \theta \leq \pi. \end{aligned} \quad (44)$$

The north pole  $\nu_0$  maps into the point  $(0, 0)$  and the equator of the sphere maps into a circle of radius  $\sqrt{2}$  about that point. Watson (1983), pp. 21–22, gives further details.

### 4.3. Artificial Data

To probe how the nature of the trend might affect what we discovered in analyzing the paleomagnetic data, we will consider three sets of pseudo-random directional trend data. Each is generated as described in Section 4.2 with  $n = 300$  observed directions and precision  $\kappa = 40$ . The functions  $f$  and  $g$  that determine the actual trend in mean direction are:

*Wobble:*  $g(t) = 4\pi t; f(t) = .3\pi[t + .2 + .15 \sin(36\pi)t]$ .

*Bat:*  $g(t) = .4\pi \sin(6\pi t); f(t) = .8\pi(t - .5)$ .

*Jumps:*  $g(t) = 2\pi t; f(t) = .2\pi$  if  $0 \leq t \leq .15$ ,  $= .1\pi$  if  $.15 < t \leq .3$ ,  $= .4\pi$  if  $.3 < t \leq .45$ ,  $= .2\pi$  if  $.45 < t \leq .65$ ,  $= .3\pi$  if  $.65 < t \leq .8$ , and  $= .4\pi$  if  $.8 < t \leq 1$ .

The following display reports, for each set of data, the rescaled risk estimate  $\hat{\kappa}\hat{\rho}(\hat{f}, U)$ , where  $U$  is the orthogonal matrix and  $\hat{f}$  is the shrinkage vector that define the superefficient estimator.

	$\hat{M}_{PLS}(1)$	$\hat{M}_{Mon}(1)$	$\hat{M}_{PLS}(2)$	$\hat{M}_{Mon}(2)$	Naive
Wobble	.209	.109	.208	.107	2.000
Bat	.143	.051	.051	.035	2.000
Jump	.196	.164	.194	.165	2.000

The artificial Wobble data was constructed to resemble observations on the Chandler-wobble of the geographic north pole, blown up to wander over a larger portion of the northern hemisphere and given greater measurement errors. Brillinger (1973) analyzed actual Chandler-wobble data using time series techniques in the tangent plane to the north pole.

In Figure 2, cells (1,1) and (1,2) present, with linear interpolation, the true Wobble mean directions and the observed directions. Cells (2,1) and (2,2) display the superefficient first-order estimates  $\hat{M}_{PLS}(1)$  and  $\hat{M}_{Mon}(1)$ . Cells (3,1) and (3,2) give the second-order estimates  $\hat{M}_{PLS}(2)$  and  $\hat{M}_{Mon}(2)$ . The interpolated true mean directions are superposed as a dotted curve on top of each estimate. Visually, each monotone fit improves on the respective PLS fit that provided the orthogonal basis used; the first-order and second-order monotone fits are similar; and each smoothed estimate improves greatly upon the naive estimate of directional trend.

Estimated risks for the fits to the Wobble data, reported above, support these assessments. For this data, using orthogonal matrix  $U_E$  in place of  $U_D$  scarcely affects estimated risk. However, using monotone shrinkage in place of PLS halves the estimated risks for both choices of basis. Diagnostic plots (not given) akin to Figure 1B suggest that  $U_D$  and  $U_E$  provide comparably economical bases here. For this data, the shrinkage vector  $f_D(\hat{\gamma}_D)$  only roughly approximates the better  $\hat{f}_{Mon}(U_D)$  and gives more weight to higher “frequencies”. This explains both the ragged visual appearance of  $\hat{M}_{PLS}(1)$  and the substantially smaller estimated risk of  $\hat{M}_{Mon}(1)$ . The shrinkage vector defining  $\hat{M}_{PLS}(2)$  is likewise a rough approximation to that for  $\hat{M}_{Mon}(2)$ , though it gives less weight to higher “frequencies.”

[Figures 2, 3, 4 go near here]

The graphics and estimated risks for the Bat data behave differently. In the diagnostic plots (not given), the basis  $U_E$  appears to be more economical than  $U_D$ . Consequently,  $\hat{M}_{Mon}(2)$  has smaller estimated risk than  $\hat{M}_{Mon}(1)$ . The clear winner, visually in Figure 3 as well as in estimated risk, is  $\hat{M}_{Mon}(2)$ . In this example, as for Wobble, the PLS shrinkage vector  $f_D(\hat{\gamma}_D)$  has difficulty approximating the better  $\hat{f}_{Mon}(U_D)$ . Second-order PLS is closer in performance to the second-order monotone fit, but is still inferior.

In diagnostic plots (not given) for the Jumps data, the  $\{v_i\}$  damp down to zero more slowly than in the preceding examples and at comparable rates for both orthogonal bases. Neither basis seems more economical than the other, a circumstance reflected in the estimated risks of the monotone shrinkage estimates. In this example,  $f_D(\hat{\gamma}_D)$  approximates  $\hat{f}_{Mon}(U_D)$  fairly well and  $f_E(\hat{\gamma}_E)$  approximates  $\hat{f}_{Mon}(U_E)$  fairly well. Consequently, the differences among the monotone shrinkage and PLS estimates are not visually impressive in Figure 4.

## 5. SOME ASYMPTOTICS

The following theorem shows that the loss and risk are asymptotically equal in this estimation problem and that the estimated risk converges to their common asymptotic value  $\rho(f, \xi^2, \kappa)$ . These findings reflect the ill-posed character of estimating directional trend and suggest an extended analysis that formally justifies selecting candidate estimator to minimize estimated risk and quantifies the effect of basis economy.

*Theorem.* Suppose that for  $r > 0$ ,

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\kappa n^{-1} |H|^2 \leq r} \kappa \mathbf{E} |\hat{\kappa}^{-1} - \kappa^{-1}| = 0. \quad (45)$$

For  $W$  equal to  $L_n(\hat{H}, H)$ ,  $R_n(\hat{H}, H, \kappa)$  or  $\hat{\rho}(f, U)$ ,

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\kappa n^{-1} |H|^2 \leq r} \kappa \mathbf{E} |W - \rho(f, \xi^2, \kappa)| = 0. \quad (46)$$

*Proof.* Fix  $n$ . Let  $S(r) = \{\Xi: \kappa n^{-1} |\Xi|^2 \leq r\}$ . Note that  $\Xi \in S(r)$  if and only if  $\kappa n^{-1} |H|^2 \leq r$ . Applying (45) to definitions (12) and (23) yields

$$\begin{aligned} \kappa \mathbf{E} |\hat{\rho}(f, U) - \rho(f, \xi^2, \kappa)| &= \kappa \mathbf{E} |\text{ave}[(z^2 - \kappa^{-1} q - \xi^2)(1 - f)^2]| + o(1) \\ &= V + o(1) \quad \text{say.} \end{aligned} \quad (47)$$

The remainder term tends to zero uniformly over  $\Xi \in S(r)$  as  $\kappa$  tends to infinity. Let  $A = UFU'$  and  $B = (I - A)^2$ . Using notation from Section 2 and Section 3.2 and the identity  $\text{ave}(h^2) = n^{-1} \text{tr}(hh') = n^{-1} \text{tr}(h'h)$ ,

$$\begin{aligned} V &= \kappa n^{-1} \left| \sum_{j=1}^p y'_{(j)} B y_{(j)} - \kappa^{-1} q \text{tr}(B) - \sum_{j=1}^p \eta'_{(j)} B \eta_{(j)} \right| \\ &= n^{-1} \left| \sum_{j=1}^p e'_{(j)} B e_{(j)} - q \text{tr}(B) + 2 \sum_{j=1}^p \kappa^{1/2} \eta'_{(j)} B e_{(j)} \right| \\ &= n^{-1} \left| \sum_{j=1}^p [e'_{(j)} B e_{(j)} - \mathbf{E}(e'_{(j)} B e_{(j)})] + 2 \sum_{j=1}^p \kappa^{1/2} \eta'_{(j)} B e_{(j)} \right| + o(1). \end{aligned} \quad (48)$$

The last step uses the calculation

$$\sum_{j=1}^p \mathbf{E}(e'_{(j)} B e_{(j)}) = \sum_{j=1}^p \text{tr}[B(I - \mu_{ij}^2)] + o(1) = q \text{tr}(B) + o(1). \quad (49)$$

On the one hand,

$$n^{-1} \mathbf{E} |e'_{(j)} B e_{(j)} - \mathbf{E}(e'_{(j)} B e_{(j)})| \leq n^{-1} \mathbf{E} \left| \sum_{i=1}^n [b_{ii} e_{ij}^2 - \mathbf{E}(b_{ii} e_{ij}^2)] + \sum_{k=1}^n \sum_{i \neq k} b_{ik} e_{ij} e_{kj} \right|. \quad (50)$$

Let  $2_+$  stand for  $2 + o(1)$  as  $\kappa$  tends to infinity. For increasing  $\kappa$ , the first term in (50) is bounded above by

$$n^{-1} \text{Var}^{1/2} \left[ \sum_{i=1}^n b_{ii} e_{ij}^2 \right] \leq 2_+^{1/2} n^{-1} \left[ \sum_{i=1}^n b_{ii}^2 \right]^{1/2} \leq 2_+^{1/2} n^{-1} [\text{tr}(B^2)]^{1/2} \leq 2_+^{1/2} n^{-1/2}. \quad (51)$$

For  $\kappa$  tending to infinity, the second term in (50) is bounded above by

$$\begin{aligned} n^{-1}\text{Var}^{1/2}\left[\sum_{k=1}^n \sum_{i \neq k} b_{ik} e_{ij} e_{kj}\right] &\leq n^{-1}\left[\sum_{k=1}^n \sum_{i \neq k} b_{ik}^2\right]^{1/2} \\ &\leq n^{-1}[\text{tr}(B^2)]^{1/2} \leq n^{-1/2}. \end{aligned} \quad (52)$$

Thus, for every  $n$  and every  $r > 0$ ,

$$\lim_{\kappa \rightarrow \infty} \sup_{\Xi \in \mathcal{S}(r)} n^{-1} \mathbb{E} \left| \sum_{j=1}^p [e'_{(j)} \widehat{B} e_{(j)} - \mathbb{E}(e'_{(j)} B e_{(j)})] \right| = O(n^{-1/2}). \quad (53)$$

On the other hand, because the largest eigenvalue of  $B^2$  lies between 0 and 1,

$$\begin{aligned} n^{-1} \mathbb{E} |\kappa^{1/2} \eta'_{(j)} B e_{(j)}| &\leq n^{-1} \text{Var}^{1/2}[\kappa^{1/2} \eta'_{(j)} B e_{(j)}] \\ &\leq n^{-1} [\kappa \eta'_{(j)} B^2 \eta_{(j)}]^{1/2} \leq n^{-1/2} [n^{-1} \kappa |\eta_{(j)}|^2]^{1/2}. \end{aligned} \quad (54)$$

Thus,

$$\lim_{\kappa \rightarrow \infty} \sup_{\Xi \in \mathcal{S}(r)} n^{-1} \mathbb{E} \left| \sum_{j=1}^p \kappa^{1/2} \eta'_{(j)} B e_{(j)} \right| = O(n^{-1/2}). \quad (55)$$

Combining (47), (48), (53) and (55) yields Theorem assertion (46) for  $W = \hat{\rho}(f, U)$ . Because of (24), the result also holds for  $W = R_n(\hat{H}, H, \kappa)$ .

From definitions (19) and (23),  $\kappa \mathbb{E} |L_n(\hat{H}, H) - \rho(f, \xi^2, \kappa)|$  can be expressed as

$$\begin{aligned} \kappa \mathbb{E} \left| \sum_{j=1}^p \text{ave}[f^2(z_{(j)} - \xi_{(j)})^2 - \kappa^{-1} q f^2 + 2f(1-f)\xi_{(j)}(z_{(j)} - \xi_{(j)})] \right| \\ = n^{-1} \mathbb{E} \left| \sum_{j=1}^p [e'_{(j)} C e_{(j)} - q \text{tr}(C) + 2\kappa^{1/2} \eta'_{(j)} G e_{(j)}] \right|, \end{aligned} \quad (56)$$

where  $C = A^2$  and  $G = A(I - A)$ . Analysis of the the right-side of (56) by the method used for (48) establishes assertion (46) when  $W = L_n(\hat{H}, H)$ .

*Extensions.* A more elaborate empirical process argument, akin to the analysis in Section 6 of Beran and Dümbgen (1998), shows that

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\kappa n^{-1} |H|^2 \leq r} \kappa \mathbb{E} \sup_{f \in \mathcal{F}_{Mon}} |W - \rho(f, \xi^2, \kappa)| = 0 \quad (57)$$

for the three values of  $W$  in the Theorem above. This strengthening of (46) clarifies the performance of the adaptive estimator  $\hat{H}_{Mon}(U)$ . It follows from (57) that  $\hat{\rho}(\hat{f}_{Mon}, U)$  is a consistent estimator for the risk of  $\hat{H}_{Mon}(U)$  and that

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\kappa n^{-1} |H|^2 \leq r} \kappa |R_n(\hat{H}_{Mon}(U), H, \kappa) - \tau_{Mon}(\xi^2, \kappa)| = 0, \quad (58)$$

where

$$\tau_{Mon}(\xi^2, \kappa) = \min_{f \in \mathcal{F}_{Mon}} \rho(f, \xi^2, \kappa) \leq \kappa^{-1}q. \quad (59)$$

To quantify the affect of basis economy on the risk of the adaptive estimator  $\hat{H}_{Mon}(U)$ , let

$$S(r, b) = \{\Xi: n^{-1}\kappa|\Xi|^2 \leq r, \xi_i = 0 \text{ if } i > bn\} \quad (60)$$

for  $b \in [0, 1]$  and  $r > 0$ . The basis  $U$  is highly economical for the mean matrix  $H$  if  $\Xi \in S(r, b)$  for a small value of  $b$ . By specialization of a theorem in Pinsker (1980),

$$\sup_{\Xi \in S(r, b)} \min_{f \in \mathcal{F}_{Mon}} \rho(f, \xi^2, \kappa) = \min_{f \in \mathcal{F}_{Mon}} \sup_{\Xi \in S(r, b)} \rho(f, \xi^2, \kappa) = qrb/(r + b). \quad (61)$$

For details, see Theorems 1 and 4 in Beran (2000), noting that the result depends only on the function  $\rho(f, \xi^2, \kappa)$ , not on the model in the background. Combining (58), (59) with (61) establishes

$$\lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow \infty} \sup_{\Xi \in S(r, b)} \kappa R_n(\hat{H}_{Mon}(U), H, \kappa) = qrb/(r + b) \leq qb. \quad (62)$$

Thus, the maximum asymptotic risk of the directional trend estimator  $\hat{M}_{Mon}(U)$  is small if the basis  $U$  is highly economical in the formal sense that  $b$  is small. Though this formulation of basis economy is overly simplified for the sake of mathematical analysis, result (62) supports the experimental finding in Section 4 that a quick decrease in the higher order components of  $\xi^2$  reduces the risk of  $\hat{M}_{Mon}(U)$ .

## REFERENCES

- Becker, R. A. and Chambers, J. M. (1984), *S: An Interactive Environment for Data Analysis and Graphics*, Belmont, CA: Wadsworth.
- Beran, R. (2000) "REACT Scatterplot Smoothers: Superefficiency Through Basis Economy," *Journal of the American Statistical Society*, 95, 155–171.
- Beran, R., and Dümbgen, L. (1998), "Modulation of Estimators and Confidence Sets," *Annals of Statistics*, 26, 1826–1856.
- Brillinger, D. R. (1973), "An Empirical Investigation of the Chandler Wobble and Two Proposed Excitation Processes," *Bulletin of the International Statistical Institute*, Book 3, 413–433.
- Buja, A., Hastie, T., and Tibshirani, R. (1989), "Linear Smoothers and Additive Models" (with discussion), *The Annals of Statistics*, 17, 453–555.
- Fisher, N. I. and Lewis, T. (1987), "A note on Spherical Splines," *Journal of the Royal Statistical Society, Series B*, 47, 482–488.
- Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1985), *Statistical Analysis of Spherical Data*, Cambridge: Cambridge University Press.

- Irving, E. (1977), “Drift of the Major Continental Blocks Since the Devonian,” *Nature*, 270, 304–309.
- Jupp, P. E. and Kent, J. (1987), “Fitting Smooth Paths to Spherical Data,” *Applied Statistics*, 36, 34–46.
- Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics* (second ed.), New York: Wiley
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (second ed.), Cambridge: Cambridge University Press
- Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, New York: Wiley.
- Venables, W. N., and Ripley, B. D. (1999), *Modern Applied Statistics with S-PLUS* (third ed.), New York: Springer.
- Watson, G. S. (1983), *Statistics on Spheres*, University of Arkansas Lecture Notes in the Mathematical Sciences, Volume 6, New York: Wiley-Interscience.
- Watson, G. S. (1985), “Interpolation of Directed and Undirected Line Data,” in *Multivariate Analysis VI*, ed. P. R. Krishnaiah, New York: Academic Press, pp. 613–625.

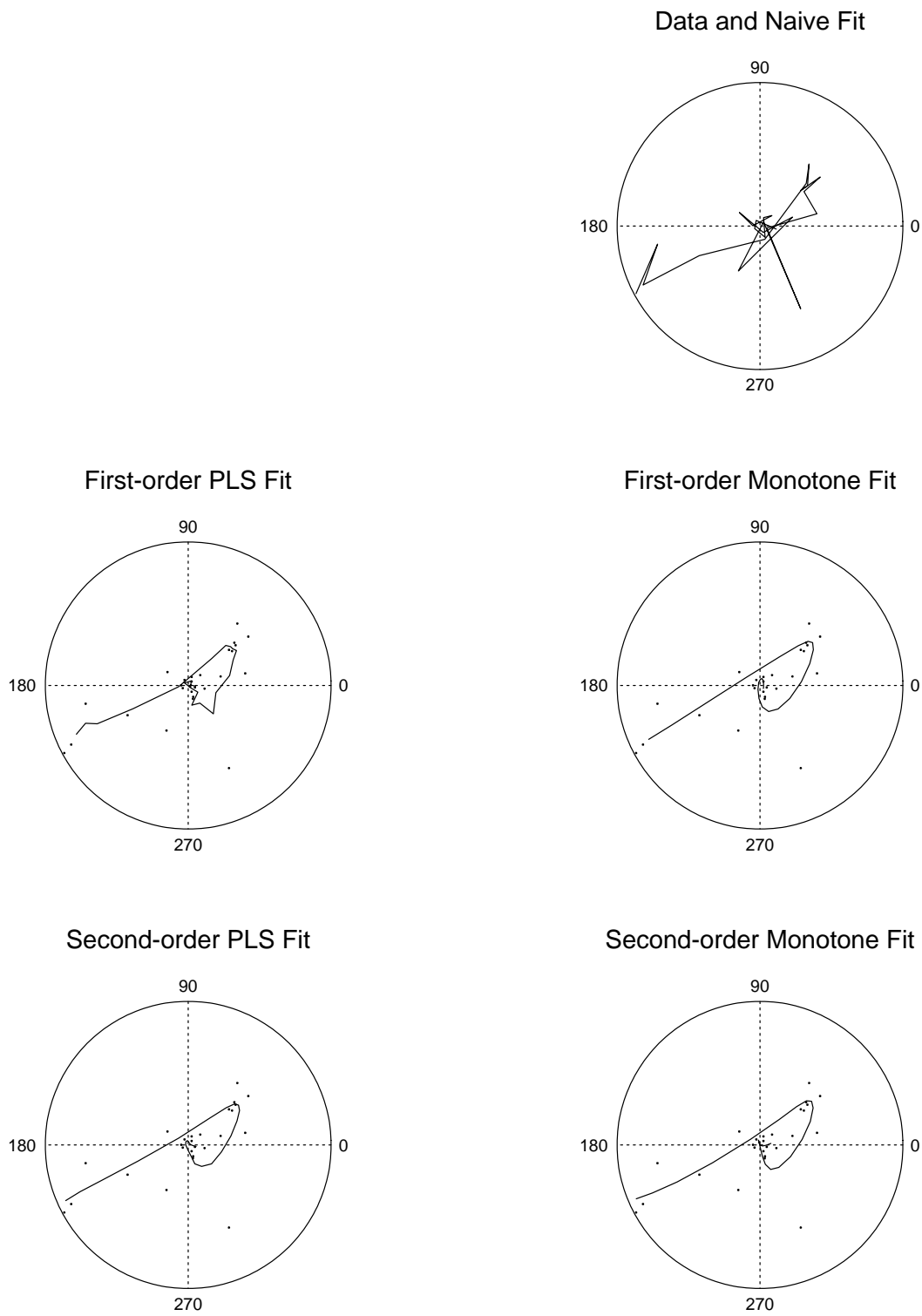


FIGURE 1A. Competing fits to time-ordered measured positions of the paleomagnetic north pole. Linear interpolation in the top subplot shows the time-sequence of the observed directions.

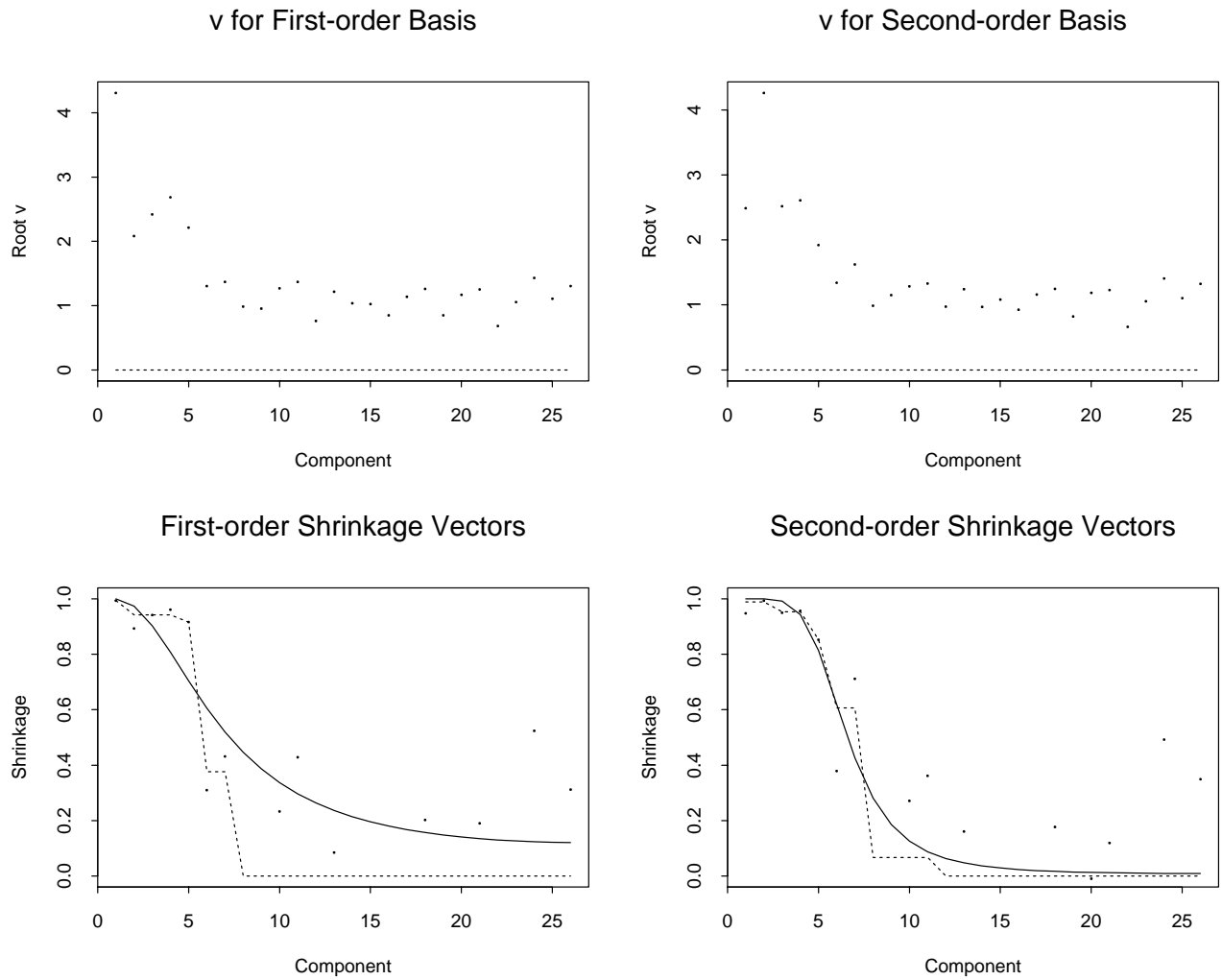


FIGURE 1B. Diagnostic plots for fits to the paleomagnetic north pole data. Top row displays the components of  $v = \hat{\kappa}^{1/2}|z|$  for each orthogonal basis. Bottom row displays the shrinkage vectors defining  $\hat{M}_{PLS}(k)$  (solid line) and  $\hat{M}_{Mon}(k)$  (dashed line).

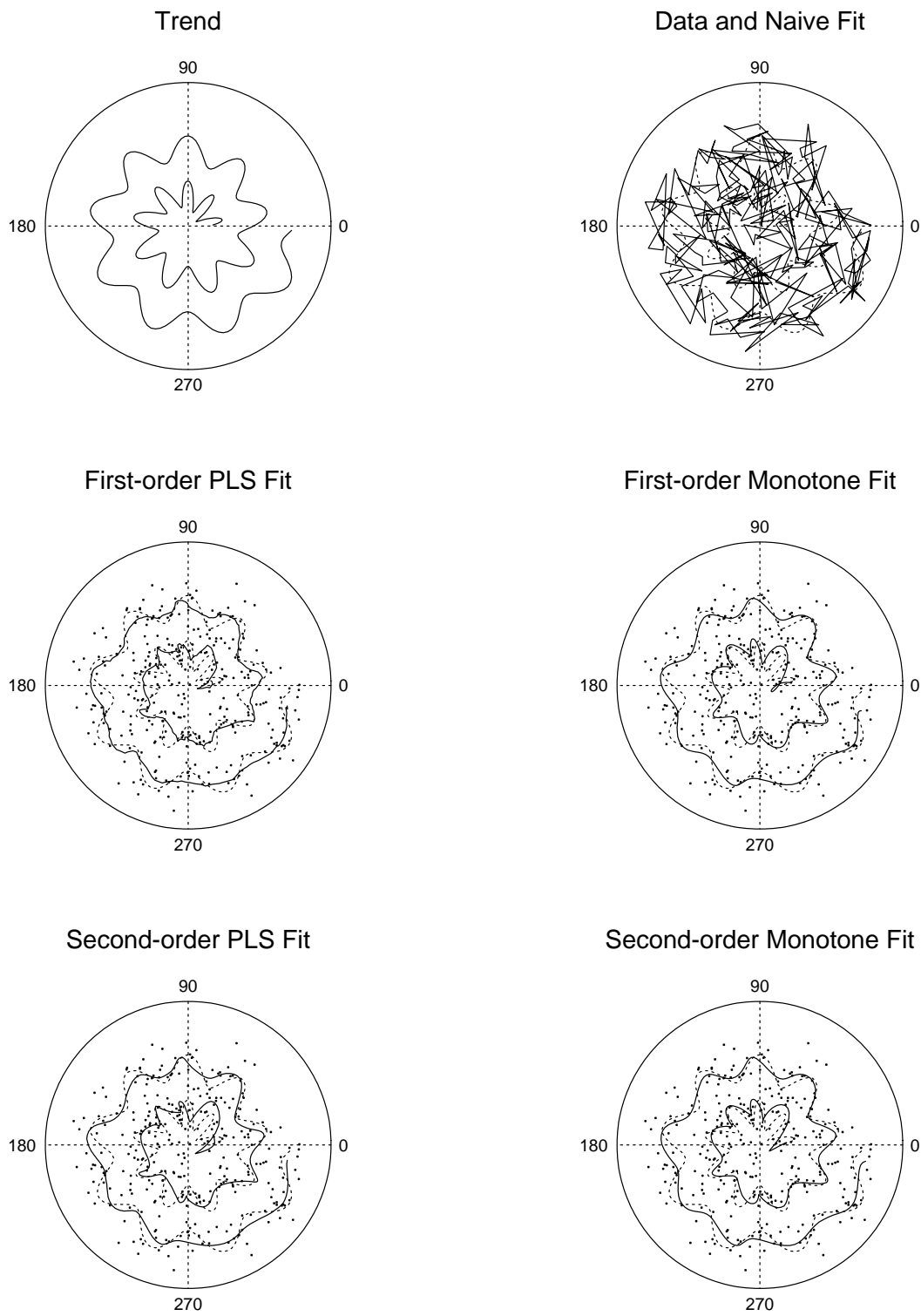


FIGURE 2. Competing fits to the Wobble directional trend. Linear interpolation in cell (1,2) shows the time-sequence of the observed directions.

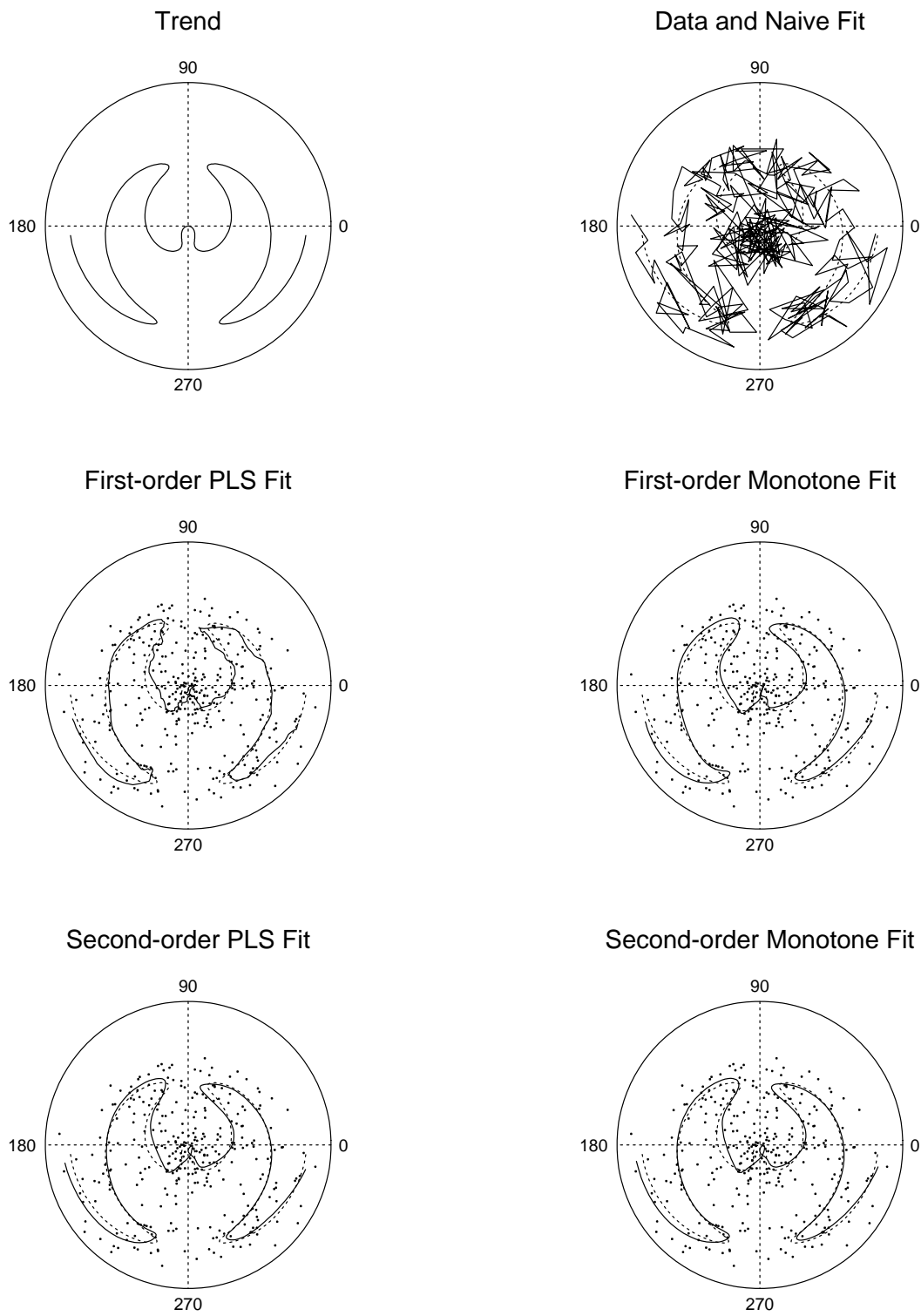


FIGURE 3. Competing fits to the Bat directional trend. Linear interpolation in cell (1,2) shows the time-sequence of the observed directions.

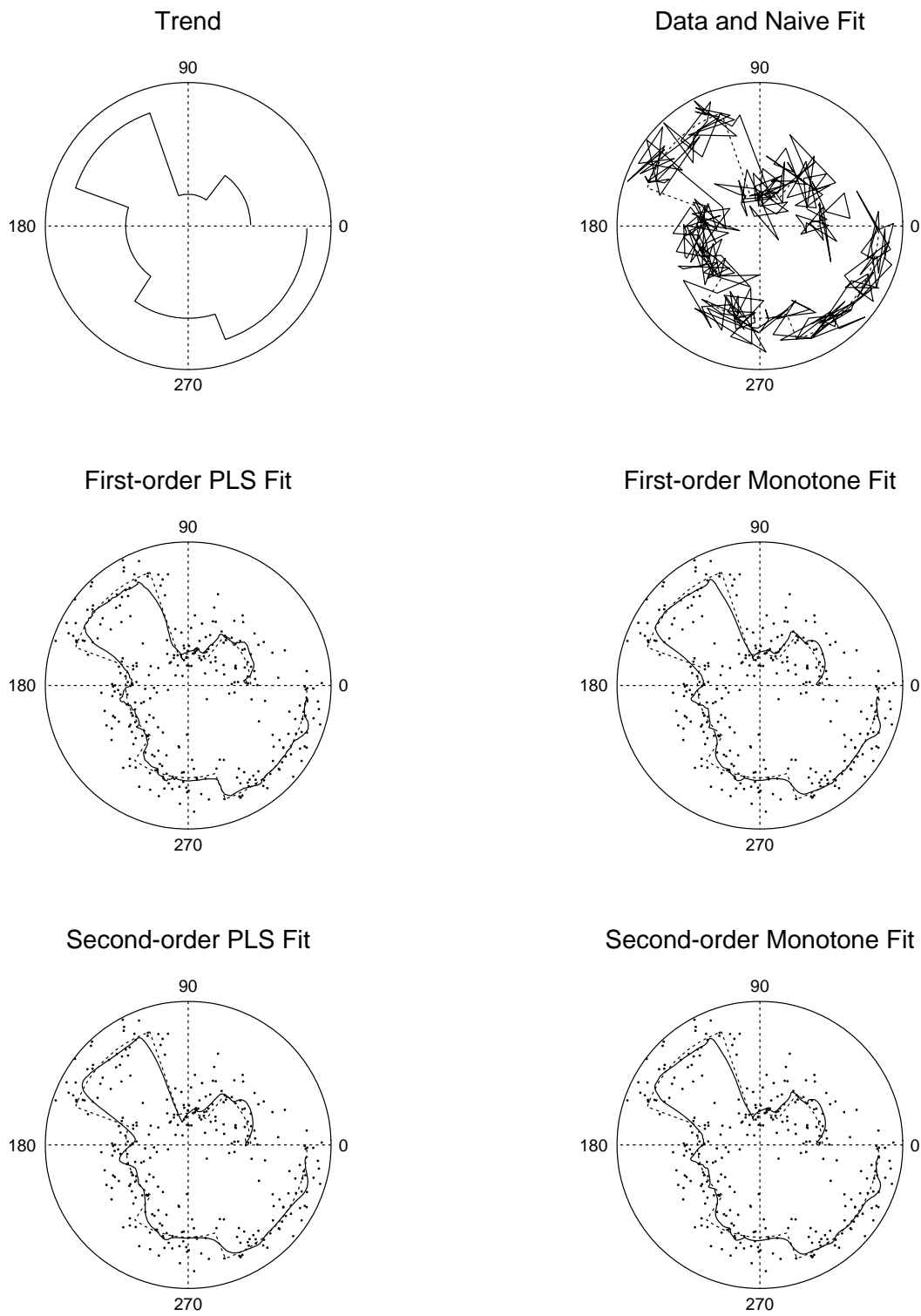


FIGURE 4. Competing fits to the Jumps directional trend. Linear interpolation in cell (1,2) shows the time-sequence of the observed directions.