Du, J. and Dryden, Ian L. and Huang, X. (2015) Size and shape analysis of error-prone shape data. Journal of the American Statistical Association, 110 (509). pp. 368-377. ISSN 1537-274X

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Size and Shape Analysis of Error-Prone Shape Data

Jiejun Du, Ian L. Dryden, and Xianzheng Huang

We consider the problem of comparing sizes and shapes of objects when landmark data are prone to measurement error. We show that naive implementation of ordinary Procrustes analysis that ignores measurement error can compromise inference. To account for measurement error, we propose the conditional score method for matching configurations, which guarantees consistent inference under mild model assumptions. The effects of measurement error on inference from naive Procrustes analysis and the performance of the proposed method are illustrated via simulation and application in three real data examples. Supplementary materials for this article are available online.

KEY WORDS: Complex normal; Configuration; Landmark; Ordinary Procrustes analysis; Quaternion.

1. INTRODUCTION

Data capturing the size and shape of an object are of great interest in many branches of science. For instance, facial recognition as a routine task in forensics relies on shape data of faces; chemists study shapes of molecules to understand and manipulate chemical properties; and study on shapes of complex organisms is an important part of research in biology and medicine. One approach that has a long history for characterizing sizes and shapes entails defining landmarks on an object, the collection of which is referred to as the configuration of the object (Kendall 1984; Bookstein 1991; Dryden and Mardia 1998). Then the shape data for this object consist of the geometric information of these landmarks in the configuration after removing location, rotation, and (possibly) scale. This way of measuring shape data has been adopted in biology, for example, among many other fields where applied scientists understand well the choice and interpretation of landmarks. Even though usually there are solid scientific grounds or mathematical motivations for the choice of landmarks, locating them on a subject and/or measuring their relative locations is often prone to error. This results in shape data as an error-contaminated measure of the true underlying shape of an object.

Assuming shape data free of measurement error, ordinary Procrustes analysis (Goodall 1991), referred to as OPA, is a conventional approach to match one configuration onto the other, which is an important step in size and shape comparisons among different objects. This method involves translation, rotation, and scaling of one configuration to match it onto the other configuration as closely as possible. In the presence of measurement error, naive implementation of OPA matches the noisy version of a configuration onto the other one. A natural question is whether or not this naive matching can reveal the same information regarding size and shape comparisons as when one matches the error-free configurations. For example, in the study of rat skulls (Bookstein 1991; Kenobi, Dryden, and Le 2010; Mardia et al. 2013), X-rays of rat skulls were recorded from age 7 days to 150 days, and it is of interest to estimate the change in size and shape of the skull as a rat grows. Assuming configurations of skulls are measured precisely, if one implements OPA to match the skull of a rat recorded at an earlier time onto the other skull of the same rat recorded later, then the amount of scaling entailed in OPA reflects the amount of growth of the skull during this time window. Suppose the reality is that configurations of skulls cannot be measured precisely, and thus the observed configurations are noisy surrogates of the unobserved true configurations. To study the growth pattern of rat skulls, it is important to understand the effects of measurement error on inference from naive OPA. We will show in Section 3 that matching two error-contaminated configurations via naive OPA can mask important distinctions between two configurations. This motivates new proposals that can account for measurement error when comparing sizes and shapes of objects. Note that for nontrivial matching the number of landmarks is more than the number of dimensions in which the data lie.

The remainder of the article is organized as follows. In Section 2 models for shape data accounting for measurement error are formulated for the purpose of matching two true configurations in two or three dimensions. Results regarding the effects of measurement error on inference from naive OPA are presented in Section 3. In Section 4, we propose the conditional score method. Simulation studies are reported in Section 5 to illustrate the performance of the proposed methods in comparison with naive OPA. These methods are applied to three real data examples in Section 6. We conclude the article with discussions of our findings and future research directions in Section 7.

2. MEASUREMENT ERROR MODELS

2.1 Models for Two-Dimensional Size-and-Shape Data

For a two-dimensional configuration, it is mathematically convenient to denote the location of a landmark by a complex

Jiejun Du is Statistician, K&L Consulting Inc., Fort Washington, PA 19034 (E-mail: *jiejun.du@gmail.com*). Ian L. Dryden is Professor, School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK (E-mail: *Ian.Dryden@nottingham.ac.uk*). Xianzheng Huang is Associate Professor, Department of Statistics, University of South Carolina, NC 29208 (E-mail: *huang@stat.sc.edu*). The authors are grateful to the associate editor and two anonymous referees for their helpful comments on the article. Ian L. Dryden acknowledges the support of a Royal Society Wolfson Research Merit Award WM110140 and the Engineering and Physical Sciences Research Council grant EP/K022547/1. Xianzheng Huang acknowledges the supported of NSF grant DMS-1006222.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

Published with license by Taylor & Francis
Journal of the American Statistical Association
March 2015, Vol. 110, No. 509, Theory and Methods
DOI: 10.1080/01621459.2014.908779

number, with the real and imaginary parts representing the x- and y-coordinates of the landmark, respectively. Let $\mathbf{X}$ and $\mathbf{Y}$ be two configurations of interest, each consisting of $K(\geq 3)$ landmarks. With the complex-value representation, both $\mathbf{X}$ and $\mathbf{Y}$ are elements in the K-dimensional complex space, $\mathbb{C}^K$. More specifically, $\mathbf{X} = (X_1, \ldots, X_K)^t$ and $\mathbf{Y} = (Y_1, \ldots, Y_K)^t$, where X_l, $Y_l \in \mathbb{C}^1$ correspond to the lth landmark in the configurations, for $l = 1, \ldots, K$. To match $\mathbf{X}$ onto $\mathbf{Y}$ via OPA, the following linear model that relates $\mathbf{Y}$ and $\mathbf{X}$ is assumed,

$$\mathbf{Y} = \beta_0 \mathbf{1}_K + \beta_1 \mathbf{X} + \boldsymbol{\epsilon}, \tag{1}$$

where $\beta_0 \in \mathbb{C}^1$ is the translation parameter, $\beta_1 (\neq 0) \in \mathbb{C}^1$ is the scale-and-rotation parameter, and multiplying β_1 by $\mathbf{X}$ in effect scales and rotates $\mathbf{X}$, $\mathbf{1}_K$ is the $K \times 1$ vector of ones, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_K)^t \in \mathbb{C}^K$ is the mean-zero random error. The interpretation of the complex multiplication in (1), $\beta_1 \mathbf{X}$, can be made more transparent by looking into the lth landmark ($l = 1, \ldots, K$), for which $\beta_1 X_l$ is equivalent to $\|\beta_1\| e^{i\theta} X_l$, and in real arithmetic,

$$\|\beta_1\| \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \mathrm{Re}(X_l) \\ \mathrm{Im}(X_l) \end{bmatrix},$$

where, for $t \in \mathbb{C}^1$, $\|t\|$ is the norm of t, Re (t) and Im (t) denote the real and imaginary part of t, respectively, and $\theta \in [0, 2\pi)$ is the rotation parameter. Under the model given by (1), matching $\mathbf{X}$ onto $\mathbf{Y}$ as closely as possible can be formulated as a least-square problem, where one minimizes the squared distance given by $\|\mathbf{Y} - \beta_0 \mathbf{1}_K - \beta_1 \mathbf{X}\|^2$ over $\boldsymbol{\beta} = (\beta_0, \beta_1)^t \in \mathbb{C}^2$. Here, for $\mathbf{A} \in \mathbb{C}^K$, denote by $\mathbf{A}^*$ the transpose of the conjugate of $\mathbf{A}$, $\|\mathbf{A}\|^2 = \mathbf{A}^*\mathbf{A}$ is the squared Euclidean norm of $\mathbf{A}$.

Solutions to the above least-square problem are the outcomes of OPA (Dryden and Mardia 1998, sec. 5.2). In fact, OPA for two-dimensional shape data can be viewed as the complex version of the least-square method for real-value linear regression. This intimate connection between OPA and real-value linear regression leads us to formulate the upcoming measurement error models for shape data, and further inspires our proposal of the conditional score method, which has successes in drawing inference based on error-contaminated data modeled by real-value regression models.

Instead of assuming $\mathbf{X}$ and $\mathbf{Y}$ are observed directly as in most existing literature on shape analysis, we consider the scenario where the actual observed configurations result from contaminating the true configurations with measurement error. Measurement error in $\mathbf{Y}$ does not cause complications from a modeling perspective, as one may view such measurement error part of $\boldsymbol{\epsilon}$ in (1). Caution needs to be taken regarding measurement error in $\mathbf{X}$ however, for which the reasons will become clear in Section 3. For notational convenience, henceforth, we only consider $\mathbf{X}$ as the error-prone unobserved configuration, and denote $\mathbf{W} = (W_1, \ldots, W_K)^t \in \mathbb{C}^K$ as an error-contaminated measure of $\mathbf{X}$. More specifically, we assume classical measurement error (Carroll et al. 2006, sec. 1.2) and $\mathbf{W}$ relates to $\mathbf{X}$ according to

$$\mathbf{W} = \mathbf{X} + \mathbf{U}, \tag{2}$$

where $\mathbf{U} = (U_1, \ldots, U_K)^t \in \mathbb{C}^K$ is the mean-zero nondifferential measurement error (Carroll et al. 2006, sec. 2.5), independent of $\mathbf{X}$ and $\boldsymbol{\epsilon}$. Together (1) and (2) give the two component models of the measurement error model for the observed configurations ($\mathbf{Y}$, $\mathbf{W}$).

2.2 Models for Three-Dimensional Size-and-Shape Data

Conventionally, a three-dimensional configuration of $K(\geq 4)$ landmarks is denoted by a $K \times 3$ real matrix. Take configuration $\mathbf{X}$ as an example, the lth row of $\mathbf{X}$, for $l = 1, \ldots, K$, is $\mathbf{X}_l^t$, where $\mathbf{X}_l$ is an element in the three-dimensional real space $\mathbb{R}^3$ that consists of the x-, y-, and z-coordinates of the lth landmark, respectively. A more compact representation of $\mathbf{X}$ is attained by using quaternions (Horn 1987; Zhang 1997), which can be viewed as a generalization of complex numbers. Let $q = a + bi + cj + dk$ be a quaternion number in the one-dimensional quaternion space $\mathbb{Q}^1$, where i, j, k are three imaginary units, a, b, c, and d are in $\mathbb{R}^1$, among which a and (b, c, d) are the real part and imaginary parts of q, respectively. Denote by $\mathcal{X} \in \mathbb{Q}^K$ the quaternion version of $\mathbf{X}$, then $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_K)^t$, where the real part of $\mathcal{X}_l \in \mathbb{Q}^1$ is zero and the imaginary parts of it correspond to the three elements in $\mathbf{X}_l$, for $l = 1, \ldots, K$. Similarly, let $\mathcal{Y}$, $\mathcal{E}$, $\mathcal{W}$, and $\mathcal{U}$ be the $K \times 1$ quaternion versions of $K \times 3$ real matrices $\mathbf{Y}$, $\boldsymbol{\epsilon}$, $\mathbf{W}$, and $\mathbf{U}$, respectively. Then the classical measurement error model for the observed three-dimensional configurations ($\mathcal{Y}$, $\mathcal{W}$) consists of the following two component models,

$$\mathcal{Y} = \gamma \mathbf{1}_K + q\mathcal{X}\bar{q} + \mathcal{E}, \tag{3}$$

$$\mathcal{W} = \mathcal{X} + \mathcal{U}, \tag{4}$$

where $\gamma \in \mathbb{Q}^1$ is the translation parameter with the real part equal to zero, $q \in \mathbb{Q}^1$ is the scale-and-rotation parameter, and $\bar{q}$ is the conjugate of q. For a quaternion $q = a + bi + cj + dk$, its conjugate is defined as $\bar{q} = a - bi - cj - dk$. Appendix A of the supplementary materials provides a brief tutorial of quaternion arithmetic and the geometric interpretation of quaternion multiplication.

Alternatively, one may convert the above measurement error model to its real-value version by using the fact that, for the lth landmark ($l = 1, \ldots, K$), $q\mathcal{X}_l\bar{q}$ is equivalent to $\mathbf{Q}\mathbf{X}_l$ (Kunze and Schaeben 2004), where

$$\mathbf{Q} = \begin{bmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(ac+bd) \\ 2(bc+ad) & a^2-b^2+c^2-d^2 & 2(dc-ab) \\ 2(bd-ac) & 2(ab+cd) & a^2-b^2-c^2+d^2 \end{bmatrix}.$$

This yields the following model equivalent to (3),

$$\mathbf{Y} = \mathbf{1}_K \boldsymbol{\beta}_0^t + \mathbf{X}\mathbf{Q}^t + \boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^3$ is the translation parameter, of which the three elements are the imaginary parts of γ in (3), and $\boldsymbol{\epsilon}$ is the $K \times 3$ random error. The real-value version of (4) is simply (2) with $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{U}$ now all being $K \times 3$ real-value matrices.

Finally, the conventional OPA for three-dimensional shape data is developed based on the following model,

$$\mathbf{Y} = \mathbf{1}_K \boldsymbol{\beta}_0^t + \beta_1 \mathbf{X}\boldsymbol{\Gamma} + \boldsymbol{\epsilon}, \tag{6}$$

where $\beta_1 (> 0) \in \mathbb{R}^1$ is the scale parameter, $\boldsymbol{\Gamma}$ is a 3×3 rotation matrix in SO(3), and SO(3) denotes the special orthogonal group (Dryden and Mardia 1998, sec. 4.1). To match $\mathbf{X}$ onto $\mathbf{Y}$ via OPA, one minimizes the squared distance given by $\|\mathbf{Y} - \mathbf{1}_K \boldsymbol{\beta}_0^t - \beta_1 \mathbf{X}\boldsymbol{\Gamma}\|^2$ over $\boldsymbol{\beta}_0$, β_1, and $\boldsymbol{\Gamma}$. Here, for a $K \times 3$ real matrix $\mathbf{A}$, $\|\mathbf{A}\|^2$ is the squared Euclidean Frobenius norm of $\mathbf{A}$ defined by

$\text{tr}(\mathbf{A}^t\mathbf{A})$, where "tr" refers to the trace of a matrix. Comparing (5) and (6) reveals that $\mathbf{XQ}^t$ in (5) corresponds to the same operations on $\mathbf{X}$ implemented by $\beta_1\mathbf{X\Gamma}$ in (6), and $\beta_1 = a^2 + b^2 + c^2 + d^2$.

3. NAIVE PROCRUSTES ANALYSIS

3.1 Bias Analysis in Two-Dimensional Size-and-Shape Matching

Matching two-dimensional $\mathbf{X}$ onto $\mathbf{Y}$ via OPA yields an estimator of $\boldsymbol{\beta}$ given by $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0,\ \hat{\beta}_1)^t = (\mathbf{X}_D^*\mathbf{X}_D)^{-1}\mathbf{X}_D^*\mathbf{Y}$, where $\mathbf{X}_D = [\mathbf{1}_K\ \mathbf{X}]$ is the $K \times 2$ complex-value design matrix associated with (1) (Dryden and Mardia 1998, sec. 3.2). Naive implementation of OPA using the observed configurations $(\mathbf{Y},\ \mathbf{W})$ results in a naive estimator of $\boldsymbol{\beta}$ given by $\hat{\boldsymbol{\beta}}_W = (\hat{\beta}_{0,W},\ \hat{\beta}_{1,W})^t = (\mathbf{W}_D^*\mathbf{W}_D)^{-1}\mathbf{W}_D^*\mathbf{Y}$, where $\mathbf{W}_D = [\mathbf{1}_K\ \mathbf{W}]$.

To study the effects of measurement error on naive OPA in a more concrete setting, we assume landmarks in $\mathbf{X}$, $\boldsymbol{\epsilon}$, and $\mathbf{U}$ in (1) and (2) follow complex normal ($\mathcal{CN}$) distributions (Goodman 1963; Kent 1994; Konno 2007). More specifically, it is assumed that, for $l = 1, \ldots, K$, $X_l \sim \mathcal{CN}(\mu_l,\ 2\sigma_x^2)$, $\epsilon_l \sim \mathcal{CN}(0,\ 2\sigma_\epsilon^2)$, and $U_l \sim \mathcal{CN}(0,\ 2\sigma_u^2)$, independently. For a complex normal random variable, its variance being $2\sigma^2$ explicitly implies that the variance of the real and imaginary parts of it are both σ^2, and these two parts are uncorrelated (Goodman 1963, Example 3.1). Although the assumption of isotropic variance for each of these complex random quantities may not hold in all applications, it is a sensible starting point from which we were able to discover the following results that provide some practically important insights on the impact of measurement error.

Denote by $\boldsymbol{\mu}_X = (\mu_1, \ldots, \mu_K)^t$ the mean of $\mathbf{X}$, where $K \geq 3$. Proposition 3.1 given next is derived under a special case where $\boldsymbol{\mu}_X = \mu\mathbf{1}_K$ for $\mu \in \mathbb{C}^1$. The follow-up Proposition 3.2 is established under a general setting where one allows μ_l differ across $l = 1, \ldots, K$. Expectations appear in both propositions are defined with respect to the distribution of $(\mathbf{Y},\ \mathbf{W})$ under the above normality assumptions on landmarks.

Proposition 3.1. Under the above complex normality assumptions on X_l, ϵ_l, and U_l, for $l = 1, \ldots, K$, if $\boldsymbol{\mu}_X = \mu\mathbf{1}_K$, where $\mu \in \mathbb{C}^1$, then

$$E(\hat{\beta}_{0,W}) = \beta_0 + \left(1 - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\right)\mu\beta_1, \tag{7}$$

$$E(\hat{\beta}_{1,W}) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta_1. \tag{8}$$

Results in Proposition 3.1 are the same in spirit as those derived for real-value simple linear regression with classical measurement error (Fuller 1987, sec. 1.1). The proof, omitted here, is parallel with that given in Fuller (1987, sec. 1.1) except for the change from real normal distributions to complex normal distributions for relevant random variables in our setting (see Du 2012). An important quantity arising from Fuller's derivations that also emerges in (7) and (8) is the so-called reliability ratio, $\sigma_x^2/(\sigma_x^2 + \sigma_u^2)$, denoted by λ, which quantifies the severity of error contamination on the unobserved scalar predictor in simple linear regression. The same interpretation of λ carries over to two-dimensional shape analyses if one assumes that error contamination on all landmarks of a configuration in both x- and y-coordinates are comparable, which is a realistic assumption in many applications. Because $\lambda \in [0, 1]$, (8) indicates an attenuation effect of measurement error on the naive estimator of β_1, which is a well-known consequence of ignoring measurement error when estimating the slope parameter in real-value simple linear regression (Carroll et al. 2006, sec. 3.2). In our context of matching two configurations, the attenuation effect translates to underestimating the amount of scaling when matching $\mathbf{X}$ onto $\mathbf{Y}$. Because the expectation in (8) is a scalar multiple of β_1, the naive rotation is not compromised by measurement error in terms of unbiasedness. Finally, according to (7), measurement error does not compromise (naive) inference on β_0 either when $\mu = 0$.

Although it is theoretically reassuring to have Proposition 3.1 in agreement with existing findings in the context of simple linear regression, the assumption of $\boldsymbol{\mu}_X = \mu\mathbf{1}_K$ is overly restrictive for shape data as it forces an object shrink to a point after random noise is removed. It is practically and theoretically more interesting to relax this assumption. However, with $\boldsymbol{\mu}_X \neq \mu\mathbf{1}_K$, $E(\hat{\boldsymbol{\beta}}_W)$ cannot be easily derived in closed form. Proposition 3.2 provides the dominating terms of this expectation when $\sigma_x^2 + \sigma_u^2$ is small.

Proposition 3.2. Under the above complex normality assumptions on X_l, ϵ_l, and U_l, for $l = 1, \ldots, K$, if $\boldsymbol{\mu}_X = (\mu_1, \ldots, \mu_K)^t \neq \mu 1_K$, where $\mu \in \mathbb{C}^1$, then

$$E(\hat{\beta}_{0,W}) = \beta_0 + \frac{2\sigma_u^2\breve{\mu}_X}{\|\boldsymbol{\mu}_{X,c}\|^2}\left(K - 1 - \frac{1}{K}\right)\beta_1 + o\left(\sigma_x^2 + \sigma_u^2\right), \tag{9}$$

$$E(\hat{\beta}_{1,W}) = \beta_1 - \frac{2\sigma_u^2}{\|\boldsymbol{\mu}_{X,c}\|^2}\left\{K - 5 + \left(2 + \frac{2}{K}\right)\frac{\|\boldsymbol{\mu}_X\|^2}{\|\boldsymbol{\mu}_{X,c}\|^2}\right\}\beta_1 + z\text{`}o\left(\sigma_x^2 + \sigma_u^2\right), \tag{10}$$

where $\breve{\mu}_X = \mathbf{1}_K^t\boldsymbol{\mu}_X/K$ and $\boldsymbol{\mu}_{X,c} = \boldsymbol{\mu}_X - \breve{\mu}_X\mathbf{1}_K$.

The proof for Proposition 3.2 is relegated to Appendix B of the supplementary materials. Note that, for a centered configuration $\mathbf{X}$, one has $\breve{\mu}_X = 0$ and thus $\boldsymbol{\mu}_X = \boldsymbol{\mu}_{X,c}$. In this case, the dominating bias in $\hat{\beta}_{0,W}$ in (9) vanishes. Moreover, according to (10), the dominating bias in $\hat{\beta}_{1,W}$ reduces to

$$-\frac{2\sigma_u^2}{\|\boldsymbol{\mu}_{X,c}\|^2}\left(K - 3 + \frac{2}{K}\right)\beta_1, \tag{11}$$

which is negative for $K \geq 3$, suggesting a negative (dominating) bias in $\hat{\beta}_{1,W}$. This is reminiscent of the attenuation effect of measurement error on naive scale estimation implied by Proposition 3.1. Furthermore, (11) indicates that, with K, β_1, and σ_u^2 fixed, the attenuation is less noticeable when $\|\boldsymbol{\mu}_{X,c}\|$ is larger, or, equivalently, when the landmarks in the mean configuration $\boldsymbol{\mu}_X$ spread out more around the center. In summary, Proposition 3.2 suggests that ignoring measurement error in the error-prone-centered configuration is less harmful if this unobserved configuration comes from a population whose mean configuration consists of more diffuse landmarks. Otherwise, naive OPA can substantially underestimate β_1. As centering configurations (in two or three dimensions) is routinely done in practice, we assume centered $\mathbf{X}$ and $\mathbf{Y}$ henceforth.

3.2 Bias Analysis in Three-Dimensional Size-and-Shape Matching

The traditional OPA that matches three-dimensional configurations, $\mathbf{X}$ and $\mathbf{Y}$, yields the following estimators of the matching parameters appearing in (6) (Dryden and Mardia 1998, sec. 5.2),

$$\hat{\boldsymbol{\beta}}_0 = \mathbf{0}, \qquad \hat{\boldsymbol{\Gamma}} = \mathbf{T}\mathbf{V}^t, \qquad \hat{\beta}_1 = \frac{\operatorname{tr}(\mathbf{Y}^t\mathbf{X}\hat{\boldsymbol{\Gamma}})}{\operatorname{tr}(\mathbf{X}^t\mathbf{X})}, \tag{12}$$

where $\mathbf{T}$, $\mathbf{V} \in \mathrm{SO}(3)$ result from the singular value decomposition, $\mathbf{Y}^t\mathbf{X} = \|\mathbf{Y}\|\|\mathbf{X}\|\mathbf{V}\boldsymbol{\Lambda}\mathbf{T}^t$, in which $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)$, and $\lambda_1 \geq \lambda_2 \geq |\lambda_3|$ are square roots of the eigenvalues of $\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t\mathbf{X}$, among which only λ_3 is not necessarily positive (Dryden and Mardia 1998, sec. 4.2). Note that λ_3 is negative if and only if $\det(\mathbf{X}^t\mathbf{Y}) < 0$, and the optimal rotation is unique if $\lambda_2 + \lambda_3 > 0$, where the eigenvalues are nondegenerate and optimally signed (Kent and Mardia 2001), which we assume throughout. It follows that naive OPA matches $\mathbf{W}$ onto $\mathbf{Y}$ and yields the following counterpart naive estimators,

$$\hat{\boldsymbol{\beta}}_{0,W} = \mathbf{0}, \qquad \hat{\boldsymbol{\Gamma}}_W = \mathbf{T}_W\mathbf{V}_W^t, \qquad \hat{\beta}_{1,W} = \frac{\operatorname{tr}(\mathbf{Y}_\mathbf{W}^t\hat{\boldsymbol{\Gamma}}_W)}{\operatorname{tr}(\mathbf{W}^t\mathbf{W})}, \tag{13}$$

where $\mathbf{T}_W$ and $\mathbf{V}_W$ are similarly defined as $\mathbf{T}$ and $\mathbf{V}$ in (12) with $\mathbf{X}$ replaced by $\mathbf{W}$.

In what follows, we view the estimators in (12) as the ideal estimators, which can be computed only when the error-free configurations, that is, the true configurations ($\mathbf{Y}$, $\mathbf{X}$), are observed. Assuming common measurement error variance, σ_u^2, for all $K(\geq 4)$ landmarks of $\mathbf{X}$ in all three (x-, y-, and z-) coordinates, we investigate how the naive estimators compare with the ideal estimators given the true configurations. Our findings are summarized in the following proposition, where the expectation is conditional on ($\mathbf{Y}$, $\mathbf{X}$), which makes the distributional assumption on $\mathbf{X}_l$ and $\boldsymbol{\epsilon}_l$ ($l = 1, \ldots, K$) imposed in Propositions 3.1 and 3.2 irrelevant here.

Proposition 3.3. Under the assumption of small σ_u,

$$\hat{\boldsymbol{\Gamma}}_W = \hat{\boldsymbol{\Gamma}} + \sigma_u\mathbf{T}\mathbf{D}\mathbf{V}^t + \sigma_u^2\mathbf{T}\mathbf{D}^2\mathbf{V}^t/2 + O(\sigma_u^3), \tag{14}$$

$$E(\hat{\beta}_{1,W}) = \hat{\beta}_1 - \frac{(3K-2)\sigma_u^2}{\|\mathbf{X}\|^2}\hat{\beta}_1 + O(\sigma_u^3), \tag{15}$$

where $\mathbf{D} = [d_{ij}]_{i,j=1,2,3}$ has elements given by $d_{ij} = (c_{ji} - c_{ij})/(\lambda_i + \lambda_j)$ if $i \neq j$, and $d_{ij} = 0$ otherwise, in which c_{ij}, for $i, j = 1, 2, 3$, are elements in $\mathbf{C} = \mathbf{V}^t\mathbf{Y}^t\mathbf{Z}\mathbf{T}/(\|\mathbf{X}\|\|\mathbf{Y}\|)$, and $\mathbf{Z} = \mathbf{U}/\sigma_u$.

The result in (14) is a direct extension of Proposition 3 in Kent and Mardia (2001), and (15) follows from (13) and (14), which are elaborated in the proof provided in Appendix C of the supplementary materials. Note that, given ($\mathbf{Y}$, $\mathbf{X}$), $\mathbf{Z}$ is the only random quantity in $\mathbf{C}$, thus the dominating bias terms in (14) are random merely due to the dependence of $\mathbf{D}$ on $\mathbf{Z}$ (via $\mathbf{C}$). Noticing that $\mathbf{Z}$ has mean zero and d_{ij}'s are linear in c_{ij}'s ($i, j = 1, 2, 3$), the second term on the right-hand side of (14) has mean zero given ($\mathbf{Y}$, $\mathbf{X}$). This implies that $\hat{\boldsymbol{\Gamma}}_W$ is expected to be close to the ideal estimator $\hat{\boldsymbol{\Gamma}}$ when error contamination is not substantial. In contrast, the dominating bias in $\hat{\beta}_{1,W}$ according to (15) can be more noticeable. More specifically, this dominating bias is always negative, and is smaller in absolute value when $\|\mathbf{X}\|$ is larger, with K, $\hat{\beta}_1$, and σ_u^2 fixed. These findings bear obvious resemblance with the conclusions drawn based on the dominating bias in (11) under Proposition 3.2.

The consent of Propositions 3.1–3.3 is that the estimator of the scale parameter resulting from naive OPA is most affected in terms of consistency by measurement error among all estimators involved in matching configurations. This raises concerns especially when the size of objects is the focal point of a study, such as in the study of rat skulls' growth described in Section 1, which was recently revisited by Mardia et al. (2013) who carried out a Bayesian analysis with size being a key concept in their investigation. In the upcoming section, we derive unbiased score functions for the measurement error models formulated in Section 2, followed by estimating equations based on these scores. The goal is to obtain consistent estimators of the parameters involved in matching $\mathbf{X}$ onto $\mathbf{Y}$ using error-contaminated data ($\mathbf{Y}$, $\mathbf{W}$). In the sequel, denote by $\boldsymbol{\Theta}$ the collection of unknown parameters in the first component model, (1) and (5) (or (3)), of the measurement error model.

4. CONDITIONAL SCORE METHOD

4.1 Conditional Score for Two-Dimensional Shape Data

Following the derivations of conditional score for real-value linear measurement error models in Carroll et al. (2006, sec. 7.2), but using complex normal whenever real normal is used in their derivations, we first establish that, if one views β_1, σ_ϵ^2, and σ_u^2 as known constants and X_l as unknown parameters, then, for $l = 1, \ldots, K$, $\Delta_l = W_l + \bar{\beta}_1 Y_l \sigma_u^2/\sigma_\epsilon^2$ is a sufficient statistic for X_l, where $\bar{t}$ denotes the conjugate of t for $t \in \mathbb{C}^1$. Then, under the normality assumption on ϵ_l and U_l, we derive the first two moments of Y_l conditioning on Δ_l given by $E(Y_l|\Delta_l, \boldsymbol{\Theta}) = (\beta_0 + \beta_1\Delta_l)/(1 + \|\beta_1\|^2\sigma_u^2/\sigma_\epsilon^2)$ and $\operatorname{var}(Y_l|\Delta_l, \boldsymbol{\Theta}) = 2\sigma_\epsilon^2/(1 + \|\beta_1\|^2\sigma_u^2/\sigma_\epsilon^2)$, for $l = 1, \ldots, K$. Finally, using the idea of generalized method of moments (Hansen 1982), we define the following complex-value score function, referred to as the conditional score, for $l = 1, \ldots, K$,

$$\boldsymbol{\psi}(Y_l, \Delta_l, \boldsymbol{\Theta}) = \begin{bmatrix} \dfrac{Y_l - E(Y_l|\Delta_l, \boldsymbol{\Theta})}{\{Y_l - E(Y_l|\Delta_l, \boldsymbol{\Theta})\}\Delta_l} \\ \dfrac{K-p}{K} - \dfrac{\|Y_l - E(Y_l|\Delta_l, \boldsymbol{\Theta})\|^2}{\operatorname{var}(Y_l|\Delta_l, \boldsymbol{\Theta})} \end{bmatrix},$$

where p is the number of parameters in $\boldsymbol{\Theta}$ excluding σ_ϵ^2. By construction, $\boldsymbol{\psi}(Y_l, \Delta_l, \boldsymbol{\Theta})$ is an unbiased complex vector-value score. Following M-estimation theory (Huber 1967), the solution to the system of score equations, $\sum_{l=1}^K \boldsymbol{\psi}(Y_l, \Delta_l, \boldsymbol{\Theta}) = \mathbf{0}$, is a consistent estimator of $\boldsymbol{\Theta}$, denoted by $\hat{\boldsymbol{\Theta}}$ and referred to as the conditional score estimator. A variance estimator for $\hat{\boldsymbol{\Theta}}$ can be straightforwardly derived following the sandwich variance construction for M-estimators (Stefanski and Boos 2002).

An appealing feature of the above line of derivation is that its validity does not depend on the distribution of the unobserved configuration $\mathbf{X}$, as $\{X_l\}_{l=1}^K$ are viewed as unknown parameters and conditioned out by introducing $\{\Delta_l\}_{l=1}^K$.

4.2 Conditional Score for Three-Dimensional Shape Data

To derive the conditional score associated with the three-dimensional measurement error model, we alternate between the quaternion version of the model, that is, (3) along with (4), and the real version given by (5) in conjunction with (2). Similar to the normality assumptions in Sections 3.1 and 4.1, using the real-value representation for a configuration, we assume $\boldsymbol{\epsilon}_l \sim N(\mathbf{0},\ \boldsymbol{\Sigma}_\epsilon)$ and $\mathbf{U}_l \sim N(\mathbf{0},\ \boldsymbol{\Sigma}_u)$, for $l = 1, \ldots, K$, independently, where $\boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma}_u$ are 3×3 variance-covariance matrices that can be anisotropic. As commented at the end of Section 4.1, which is also implied in the proof in Appendix D in the supplementary materials, no distributional assumption on $\mathbf{X}_l$ is needed for the validity of the conditional score method.

Under the above distributional assumptions, we first prove in Appendix D that

$$\boldsymbol{\Delta}_l = \mathbf{W}_l + \boldsymbol{\Sigma}_u \mathbf{Q}^t \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{Y}_l \tag{16}$$

is a sufficient statistic for $\mathbf{X}_l$ if one views $\mathbf{X}_l$ as a parameter whereas all parameters in (16) are known constants. Then we derive the first two moments of $\mathbf{Y}_l$ given $\boldsymbol{\Delta}_l$, which realizes conditioning out $\mathbf{X}_l$ due to the sufficiency of $\boldsymbol{\Delta}_l$. These conditional moments are, for $l = 1, \ldots, K$, $E(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta}) = \boldsymbol{\beta}_0 + \mathbf{Q}(\mathbf{I}_3 + \boldsymbol{\Sigma}_u \mathbf{Q}^t \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{Q})^{-1}(\boldsymbol{\Delta}_l - \boldsymbol{\Sigma}_u \mathbf{Q}^t \boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\beta}_0)$ and $\operatorname{cov}(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta}) = \boldsymbol{\Sigma}_\epsilon - \mathbf{Q}(\mathbf{I}_3 + \boldsymbol{\Sigma}_u \mathbf{Q}^t \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{Q})^{-1} \boldsymbol{\Sigma}_u \mathbf{Q}^t$, where $\mathbf{I}_3$ is the 3×3 identity matrix. Next, following the same strategy employed in Section 4.1 for score construction, we obtain the following conditional score for three-dimensional shape data, for $l = 1, \ldots, K$,

$$\boldsymbol{\psi}(\mathbf{Y}_l,\ \boldsymbol{\Delta}_l,\ \boldsymbol{\Theta}) = \begin{bmatrix} \mathbf{Y}_l - E(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta}) \\ \{\mathcal{Y}_l - E(\mathcal{Y}_l|\mathcal{D}_l, \boldsymbol{\Theta})\}\mathcal{D}_l \\ \left(\dfrac{K-p}{K}\right) \operatorname{cov}(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta}) - \{\mathbf{Y}_l - E(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta})\}^t \{\mathbf{Y}_l - E(\mathbf{Y}_l|\boldsymbol{\Delta}_l, \boldsymbol{\Theta})\} \end{bmatrix}, \tag{17}$$

where p is the number of parameters in $\boldsymbol{\Theta}$ excluding those in $\boldsymbol{\Sigma}_\epsilon$, and the second (block) component of the score vector uses quaternion multiplication to attain a concise presentation, in which $\mathcal{D}_l$ is the quaternion version of $\boldsymbol{\Delta}_l$. For two quaternions, $q_r = a_r + b_r i + c_r j + d_r k$, for $r = 1, 2$, the quaternion multiplication used in (17) is defined by $q_1 q_2 = a_1 a_2 - b_1 b_2 - c_1 c_2 - d_1 d_2 + (b_1 a_2 + a_1 b_2 - d_1 c_2 + c_1 d_2) i + (-b_1 d_2 + a_1 c_2 + d_1 b_2 + c_1 a_2) j + (-c_1 b_2 + d_1 a_2 + a_1 d_2 + b_1 c_2) k$. Finally, we have the system of estimating equations, $\sum_{l=1}^K \boldsymbol{\psi}(\mathbf{Y}_l,\ \boldsymbol{\Delta}_l,\ \boldsymbol{\Theta}) = \mathbf{0}$, the solution to which is a consistent estimator of $\boldsymbol{\Theta}$ referred to as the conditional score estimator, whose variance estimator is derived according to the sandwich variance construction for M-estimators.

4.3 Measurement Error Variance Estimation

In Sections 4.1 and 4.2, parameters in the first component model of a measurement error model, (1) and (5) (or (3)), are the only unknown parameters estimated by solving the conditional score equations. In practice, the measurement error variance(s) involved in the second component model, (2) or (4), is (are) typically unknown. Hence, it is necessary to estimate σ_u^2 or $\boldsymbol{\Sigma}_u$ first to implement the proposed methods. It is well understood in the measurement error community that, for linear models, when there is only one error-contaminated measure for each value of the true covariate, the measurement error variance is intrinsically unidentifiable using data $(\mathbf{Y}, \mathbf{W})$. When there are multiple measures for each value of the true covariate, Carroll et al. (2006) provided an estimator of the measurement error variance in the context of real-value linear measurement error models (Carroll et al. 2006, eq. (4.3)).

Tailored for shape data, we develop a new strategy for estimating the measurement error variance when replicate measures are available. For notational simplicity, two-dimensional shape data are used next to illustrate the estimation, where σ_u^2 is the only unknown variance parameter in (2). Moreover, it is assumed that there are two replicate measures for the true configuration $\mathbf{X}$, denoted by $\mathbf{W}_{1,w}$ and $\mathbf{W}_{2,w}$, where the second subscript "w" is added to distinguish them from the notation for the lth landmark of $\mathbf{W}$, W_l, used in Section 2.1. Viewing each replicate measure as a result of some transformation of $\mathbf{X}$, one may assume that

$$\mathbf{W}_{1,w} = \alpha_{0,1} \mathbf{1}_K + \alpha_{1,1} \mathbf{X} + \mathbf{U}_{1,w}, \tag{18}$$

$$\mathbf{W}_{2,w} = \alpha_{0,2} \mathbf{1}_K + \alpha_{1,2} \mathbf{X} + \mathbf{U}_{2,w}, \tag{19}$$

where $\alpha_{0,r} \in \mathbb{C}^1$ is the translation parameter, $\alpha_{1,r} (\neq 0) \in \mathbb{C}^1$ is the scale-and-rotation parameter, and $\mathbf{U}_{r,w} \in \mathbb{C}^K$ is the measurement error in the rth replicate, for $r = 1, 2$. Combining (18) and (19) to cancel $\mathbf{X}$ yields

$$\mathbf{W}_{2,w} = \gamma_0 \mathbf{1}_K + \gamma_1 \mathbf{W}_{1,w} + \mathbf{U}_w, [6pt] \tag{20}$$

where $\gamma_0 = \alpha_{0,2} - \alpha_{0,1}\alpha_{1,2}/\alpha_{1,1}$, $\gamma_1 = \alpha_{1,2}/\alpha_{1,1}$, and $\mathbf{U}_w = \mathbf{U}_{2,w} - \mathbf{U}_{1,w}\alpha_{1,2}/\alpha_{1,1}$. Under the assumption that all $2K$ landmarks in $\{\mathbf{U}_{r,w},\ r = 1, 2\}$ are independent and identically distributed according to $\mathcal{CN}(0,\ 2\sigma_u^2)$, the K landmarks in $\mathbf{U}_w$ are independent and identically distributed according to $\mathcal{CN}\{0,\ 2\sigma_u^2(1 + \|\gamma_1\|^2)\}$. Now one may implement OPA to match $\mathbf{W}_{1,w}$ onto $\mathbf{W}_{2,w}$ and obtain estimates of γ_0 and γ_1, denoted by $\hat{\gamma}_0$ and $\hat{\gamma}_1$, respectively. Then the error variance in (20), that is, $2\sigma_u^2(1 + \|\gamma_1\|^2)$ as a whole, can be estimated using the mean residual squared distance given by $\|\mathbf{W}_{2,w} - \hat{\gamma}_0 \mathbf{1}_K - \hat{\gamma}_1 \mathbf{W}_{1,w}\|^2/K$. It follows that an estimator of σ_u^2 is $\hat{\sigma}_u^2 = \|\mathbf{W}_{2,w} - \hat{\gamma}_0 \mathbf{1}_K - \hat{\gamma}_1 \mathbf{W}_{1,w}\|^2/\{2K(1 + \|\hat{\gamma}_1\|^2)\}$.

We conducted extensive simulation studies to experiment with this strategy (results reported in Du 2012). Empirical evidence from these experiments suggests that this strategy can yield very accurate estimate of σ_u^2 when K is moderate or large (say, $K \geq 30$) and the reliability ratio λ is above 0.5. When there are $R(> 2)$ replicate measures for $\mathbf{X}$, one may follow the above procedure to estimate σ_u^2 repeatedly using all different pairs of replicates, then take the average of these resulting estimates as the final $\hat{\sigma}_u^2$. After a final estimate of σ_u^2 is computed, one may treat one of the replicate measures of $\mathbf{X}$ as $\mathbf{W}$ in Section 4.1 and plug in $\hat{\sigma}_u^2$ as σ_u^2. Alternatively, one may use the average of all R replicate measures of $\mathbf{X}$ as $\mathbf{W}$ and plug in $\hat{\sigma}_u^2/R$ as σ_u^2 when implementing the method proposed in Section 4.1. Parallel arguments carry over to the case with three-dimensional shape data to obtain an estimate of the measurement error variance to be plugged in the conditional scores in Section 4.2. To focus on comparing the conditional score method with naive OPA, we

assume the measurement error variance known in the simulation studies presented in the upcoming section.

5. EMPIRICAL EVIDENCE

5.1 Simulation Studies for Two-Dimensional Shape Data

In this section, we present simulation studies under the general setting in Proposition 3.2. We also include simulation studies under the special setting with $\boldsymbol{\mu}_X = \mu \mathbf{1}_K$ as in Proposition 3.1 in Appendix E in the supplementary materials. The goal of these experiments is to empirically illustrate properties of the estimators from naive OPA and the performance of the conditional score method described in Section 4.1. All simulations reported in this article are conducted using R (R Development Core Team 2012) or in SAS 9.2, and the `shapes` library (Dryden 2012) is used to implement (naive) OPA.

To create true configurations, we first generate $\{\mathrm{Re}(\mu_l), \mathrm{Im}(\mu_l)\}$, for $l = 1, \ldots, 10$, independently from uniform$(-g, g)$, where $g = 5, 10$. Using one set of realizations, $\{\mu_l\}_{l=1}^{10}$, we create a configuration $\mathbf{X}$ with landmarks $X_l \sim \mathcal{CN}(\mu_l, 2)$, for $l = 1, \ldots, 10$. Given one simulated $\mathbf{X}$, another configuration $\mathbf{Y}$ is generated according to (1) with $\beta_0 = 1 + 2i$, $\beta_1 = 2 + i$, and $\{\epsilon_l\}_{l=1}^{10}$ generated independently from $\mathcal{CN}(0, 2)$. Finally, an error-contaminated version of $\mathbf{X}$, namely, $\mathbf{W}$, is created according to (2) with $\{U_l\}_{l=1}^{10}$ independently simulated from $\mathcal{CN}(0, 2\sigma_u^2)$, where σ_u^2 is set at different values to attain the reliability ratio $\lambda = 0.5$, 0.8, and 1. Note that $\mathbf{W}$ coincides with $\mathbf{X}$ when $\lambda = 1$. Naive estimates, $\hat{\boldsymbol{\Theta}}_W = (\hat{\beta}_{0,W}, \hat{\beta}_{1,W}, \hat{\sigma}^2_{\epsilon,W})^t$, and conditional score estimates, $\hat{\boldsymbol{\Theta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2_{\epsilon})^t$, of the parameters in (1) are obtained based on $(\mathbf{Y}, \mathbf{W})$ at each g–λ combination. Summary statistics of these estimates from 1000 Monte Carlo (MC) replicates are provided in Table 1.

When λ is as low as 0.5, results in Table 1 suggest significant bias in $\hat{\beta}_{0,W}$ and substantial attenuation in $\hat{\beta}_{1,W}$. As λ increases, that is, as error contamination lessens, the bias in $\hat{\boldsymbol{\beta}}_W$ becomes less significant. The comparison between the naive estimates when $g = 5$ (upper half of Table 1) with those when $g = 10$ (lower half of Table 1) indicates that the former $\hat{\boldsymbol{\beta}}_W$ are more biased than the latter. This comparison reinforces the implication of the dominating bias in (11) under Proposition 3.2, which is that $\hat{\boldsymbol{\beta}}_W$ is less compromised by measurement error when μ_l's are more variable across $l = 1, \ldots, K$. In contrast, the conditional score estimates $\hat{\boldsymbol{\beta}}$ exhibit performance one would expect for a consistent estimator at all levels of g–λ combination.

In Appendix F in the supplementary materials, we present simulation studies to illustrate the sandwich standard error estimates for conditional score estimators as stated in Section 4. Empirical evidence there indicates that the sandwich standard error estimators are reliable when K is not small (say, $K \geq 30$).

5.2 Simulation Studies for Three-Dimensional Shape Data

For three-dimensional shape data, each landmark $\mu_l (\in \mathbb{Q}^1)$ in $\boldsymbol{\mu}_X$ consists of three imaginary parts, for $l = 1, \ldots, K$, resulting in a total of $3K$ real numbers in the entire mean configuration $\boldsymbol{\mu}_X$. To control the spread of μ_l's across $l = 1, \ldots, K$, we generate these $3K$ real numbers independently from uniform$(-g, g)$, where g varies from 0 to 10, allowing one to observe the effect of the spread among the landmarks in $\boldsymbol{\mu}_X$ on different estimates. Given a set of simulated $\{\mu_l\}_{l=1}^K$, $\mathbf{X}_l$ is generated from $N(E(\mathbf{X}_l), \mathbf{I}_3)$, where the three imaginary parts of μ_l constitute $E(\mathbf{X}_l)$, for $l = 1, \ldots, K$ with $K = 30$. Then, based on a realization of $\mathbf{X}$, we generate $\mathbf{Y}$ according to (6) with $\beta_1 = 2.25$, $\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \mathbf{I}_3$ with $\sigma_\epsilon^2 = 1$, and $\boldsymbol{\Gamma}$ as the rotation matrix corresponding to the unit axis of rotation given by $(0.5, 0.33, 0.8)^t$ and the angle of rotation equal to 60°. Finally, $\mathbf{W}$ is obtained based on (2) with $\boldsymbol{\Sigma}_u$ to be specified in detail next. Slightly different from simulations in Section 5.1, here we acknowledge the fact that the true configurations $(\mathbf{Y}, \mathbf{X})$ have mean zero and set $\boldsymbol{\beta}_0$ at zero in (5) and (6), rather than estimating $\boldsymbol{\beta}_0$. This leaves one with unknown parameters including the scale parameter β_1, the rotation $\boldsymbol{\Gamma}$, and the model error variance σ_ϵ^2 to be estimated. Using each of 1000 pairs of configurations $(\mathbf{Y}, \mathbf{W})$ at a fixed g–$\boldsymbol{\Sigma}_u$ combination, we implement naive OPA and the conditional score method to obtain two sets of estimates for the unknown parameters.

With $g = 5$, Figure 1 depicts the MC average of $\hat{\beta}_1$ and that of $\hat{\beta}_{1,W}$ versus λ (left panel), the same comparison for estimates of σ_ϵ^2 (middle panel), and finally $\|\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|$ and its naive counterpart (right panel), when $\boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_3$ with σ_u^2 varying over a range to produce λ increasing from 0.8 to 1 (in Figure 1(a)) or when $\boldsymbol{\Sigma}_u$ is equal to (in Figure 1(b))

$$h^{-1} \begin{bmatrix} 0.25 & 0.2 & 0.25 \\ 0.2 & 0.4 & 0.3 \\ 0.25 & 0.3 & 0.35 \end{bmatrix}, \tag{21}$$

Table 1. Averages of naive estimates and averages of conditional score estimates across 1000 Monte Carlo replicates from the simulation study in Section 5.1. Numbers in parentheses are Monte Carlo standard errors of the averages. True parameter values are $\mathrm{Re}(\beta_0) = 1$, $\mathrm{Im}(\beta_0) = 2$, $\mathrm{Re}(\beta_1) = 2$, $\mathrm{Im}(\beta_1) = 1$, and $\sigma_\epsilon^2 = 1$. "Naive" refers to naive estimates; "CSE" refers to conditional score estimates

	$\mathrm{Re}(\beta_0)$	$\mathrm{Im}(\beta_0)$	$\mathrm{Re}(\beta_1)$	$\mathrm{Im}(\beta_1)$	σ_ϵ^2
Elements in $\boldsymbol{\mu}_X$ are generated from uniform$(-5, 5)$					
			$\lambda = 0.5$		
Naive	0.96 (0.01)	2.10 (0.01)	1.82 (0.00)	0.91 (0.00)	5.42 (0.03)
CSE	1.01 (0.01)	2.03 (0.01)	1.99 (0.00)	1.00 (0.00)	1.26 (0.03)
			$\lambda = 0.8$		
Naive	1.00 (0.01)	2.02 (0.01)	1.95 (0.01)	0.98(0.00)	2.17 (0.01)
CSE	1.01 (0.01)	2.00 (0.01)	2.00 (0.00)	1.00(0.00)	1.09 (0.01)
			$\lambda = 1$		
Naive	0.99 (0.01)	2.00 (0.00)	1.99 (0.00)	1.00 (0.00)	0.98 (0.00)
CSE	0.99 (0.01)	2.00 (0.00)	2.00 (0.00)	1.00 (0.00)	1.04 (0.01)
Elements in $\boldsymbol{\mu}_X$ are generated from uniform$(-10, 10)$					
			$\lambda = 0.5$		
Naive	0.92 (0.01)	2.02 (0.01)	1.95 (0.01)	0.97 (0.01)	5.77 (0.03)
CSE	1.02 (0.01)	2.02 (0.01)	2.00 (0.00)	1.00 (0.01)	1.30 (0.03)
			$\lambda = 0.8$		
Naive	0.97 (0.01)	1.99 (0.01)	1.99 (0.01)	0.99 (0.00)	2.20 (0.01)
CSE	0.99 (0.01)	1.99 (0.01)	2.00 (0.01)	1.00 (0.00)	1.10 (0.01)
			$\lambda = 1$		
Naive	1.00 (0.01)	1.99 (0.00)	2.00 (0.00)	1.00 (0.00)	0.98 (0.00)
CSE	1.00 (0.00)	2.00 (0.00)	2.00 (0.00)	1.00 (0.00)	1.04 (0.00)

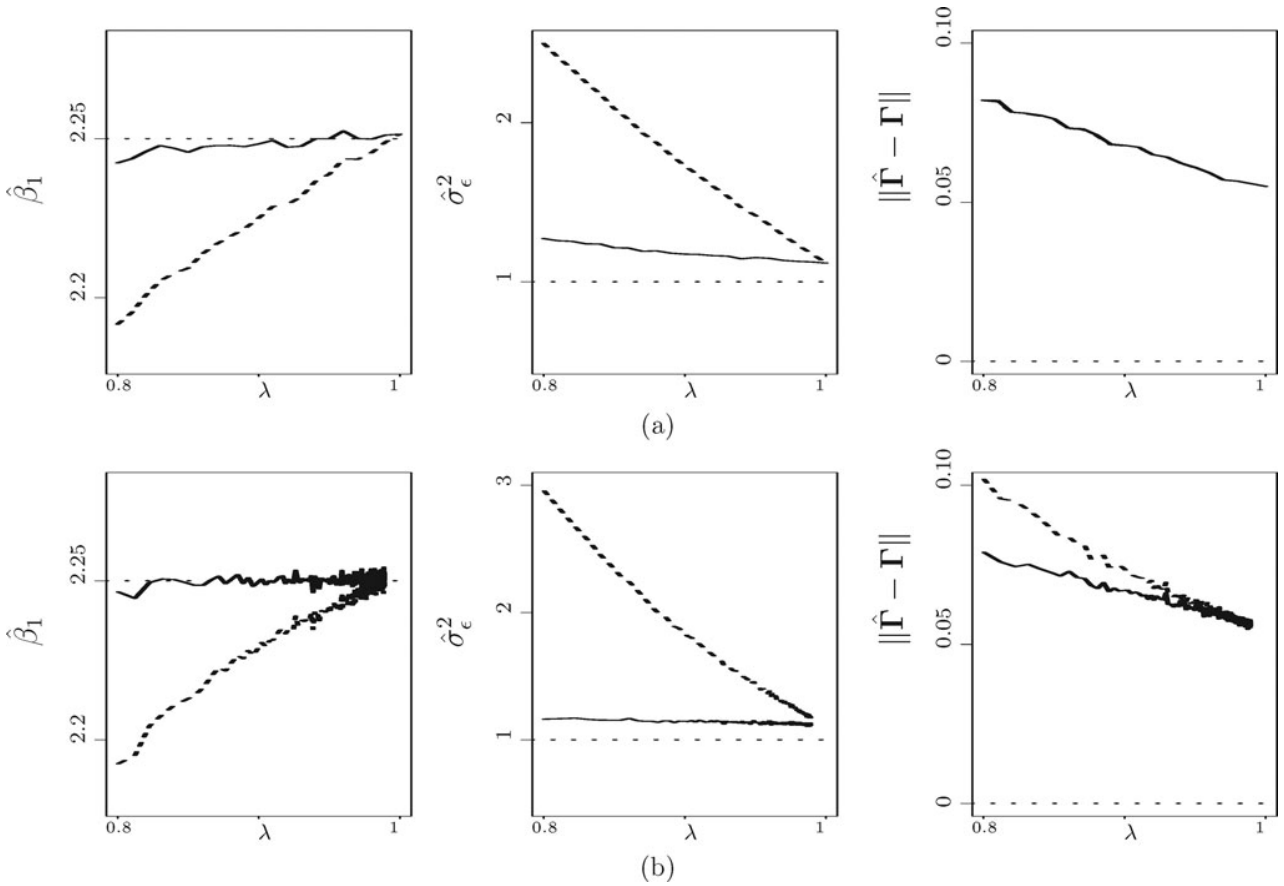


Figure 1. Averages of naive estimates (dashed lines) and averages of conditional score estimates (solid lines) across 1000 Monte Carlo replicates versus reliability ratio λ from simulations in Section 5.2, with $g = 5$, (a) $\mathbf{\Sigma}_u = \sigma_u^2 \mathbf{I}_3$, and (b) $\mathbf{\Sigma}_u$ given in (21). Dotted lines correspond to the true parameter values.

where h varies from 1 to 30. As a multivariate generalization of the reliability ratio, here we define $\lambda = \text{tr}\{(\mathbf{\Sigma}_x + \mathbf{\Sigma}_u)^{-1}\mathbf{\Sigma}_x\}/3$, where $\mathbf{\Sigma}_x$ is the variance-covariance of $\mathbf{X}_l$, which is equal to $\mathbf{I}_3$ in this experiment. As h increases from 1 to 30, $\mathbf{\Sigma}_u$ given in (21) results in λ varying from 0.800 to 0.989. Figure 2 presents the same comparisons between estimates versus g when $\lambda = 0.85$, which is obtained by first setting $\mathbf{\Sigma}_u = (3/17)\mathbf{I}_3$ (in Figure 2(a)), and second (in Figure 2(b)) setting $\mathbf{\Sigma}_u$ at

$$\begin{bmatrix} 0.19 & -0.1 & -0.2 \\ -0.1 & 0.2 & 0.15 \\ -0.2 & 0.15 & 0.25 \end{bmatrix}. \tag{22}$$

In both Figures 1(a) and 2(a), with an isotropic diagonal $\mathbf{\Sigma}_u$, estimates of $\mathbf{\Gamma}$ from two methods are very similar across the board, a phenomenon that reconciles with the comments made on (14) following Proposition 3.3. But when the assumption of isotropic and independent measurement error along all three coordinates is violated, as in both Figures 1(b) and 2(b), $\hat{\mathbf{\Gamma}}_W$ is more compromised by measurement error. As for the other parameters, Figure 1 suggests that naive estimates are more adversely affected by measurement error at lower levels of λ, and the conditional score estimates are much more robust and accurate across different λ. Moreover, Figure 2 reveals that a more diffuse $\boldsymbol{\mu}_X$ can substantially alleviate bias in the naive estimates of the matching parameters due to measurement error. This observation is consistent with the implication of (15) under Proposition 3.3.

6. APPLICATION TO REAL DATA

In what follows, we apply both naive OPA and the conditional score method to three real data examples, with the first two concerning two-dimensional configurations and the third considering three-dimensional configurations.

Example 1 (Rat skulls). A detailed description of the data of rat skulls brought up in Section 1 is given in Bookstein (1991). Besides Kenobi, Dryden, and Le (2010) and Mardia et al. (2013), many researchers have analyzed the data assuming the recorded skull configurations are precise (Goodall and Lange 1989; Monteiro 1999; Le and Kume 2000; Kent et al. 2001). For illustration purposes, we assume that the skull configurations, each consisting of eight landmarks, are contaminated with measurement error. The data include information on 18 rats, whose skulls were X-rayed at ages 7, 14, 21, 30, 40, 60, 90, and 150 days. For each rat, viewing the first (at age 7 days) recorded two-dimensional skull configuration as $\mathbf{W}$ and each of the latter seven recorded skull configurations as $\mathbf{Y}$ at that time point, we match $\mathbf{W}$ onto $\mathbf{Y}$ at each of the seven later time points, first by implementing OPA naively, then using the conditional score method. Because the skull of each rat was X-rayed only once at each time point,

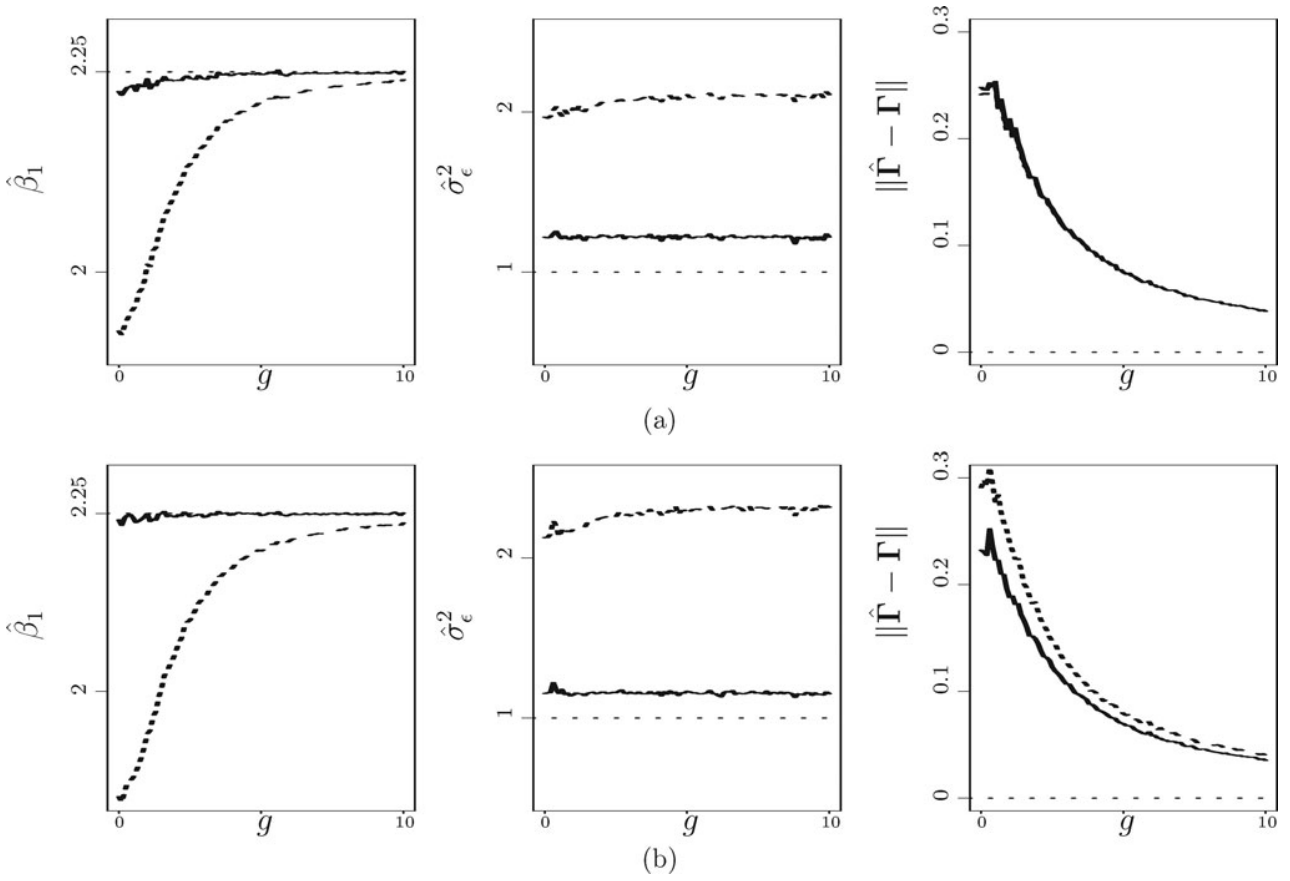


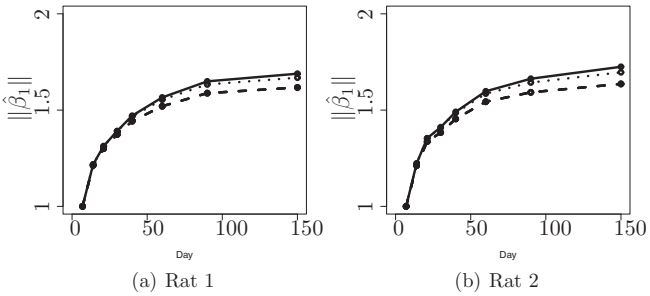
Figure 2. Averages of naive estimates (dashed lines) and averages of conditional score estimates (solid lines) across 1000 Monte Carlo replicates versus g from simulations in Section 5.2, with $\lambda = 0.85$ resulting from (a) $\mathbf{\Sigma}_u = (3/17)\mathbf{I}_3$, and (b) $\mathbf{\Sigma}_u$ given in (22). Dotted lines correspond to the true parameter values.

the measurement error variance is unidentifiable in this study. As typically done in the measurement error literature when σ_u^2 cannot be estimated, we conduct a sensitivity analysis by setting σ_u^2 at different values, and we monitor how the naive inference compares with inference from the conditional score method at different assumed levels of σ_u^2. More specifically, at each time point starting from age 14 days, we use data of all 18 rats to estimate $\sigma_x^2 + \sigma_u^2$. Then we set σ_u^2 at two levels to attain the estimated λ equal to 0.5 and 0.9. From a practical standpoint, $\lambda = 0.5$ suggests severe measurement error contamination, and $\lambda = 0.9$ implies moderate error contamination.

Figure 3 presents the estimated scale resulting from naive OPA and that from the conditional score method across seven time points at different levels of λ for two randomly selected rats. The pictorial comparison between $\|\hat{\beta}_1\|$ and $\|\hat{\beta}_{1,W}\|$ highlights one key finding in Section 3, which is that naive OPA tends to underestimate the scale parameter.

As pointed out by a referee, a natural extension of the above analysis is to embrace the longitudinal nature of the study and carry out a longitudinal size and shape analysis for the entire sample of 18 rats. For this extension, one may consider a longitudinal model in a very similar spirit as the longitudinal functional model proposed by Greven et al. (2010). This is beyond the scope of our current study, and we provide a brief discussion in Appendix G in the supplementary materials on a longitudinal model for configurations and its connection with the longitudinal functional model in Greven et al. (2010).

Example 2 (Brain templates). In this example, we consider template matching in medical imaging where candidate templates of brain images are generated by an automatic algorithm. For each axial MRI brain scan, eight landmarks are estimated on the corpus callosum according to a graphical template algorithm (Amit 1997; Dryden 2003), and the resulting candidate templates are subject to estimation error. Because, given one



(a) Rat 1 (b) Rat 2

Figure 3. Naive estimates (dashed lines) and conditional score estimates (solid lines for $\lambda = 0.5$, dotted lines for $\lambda = 0.9$) of the scale parameter $\|\beta_1\|$ at seven time points based on the rat data in Example 1.

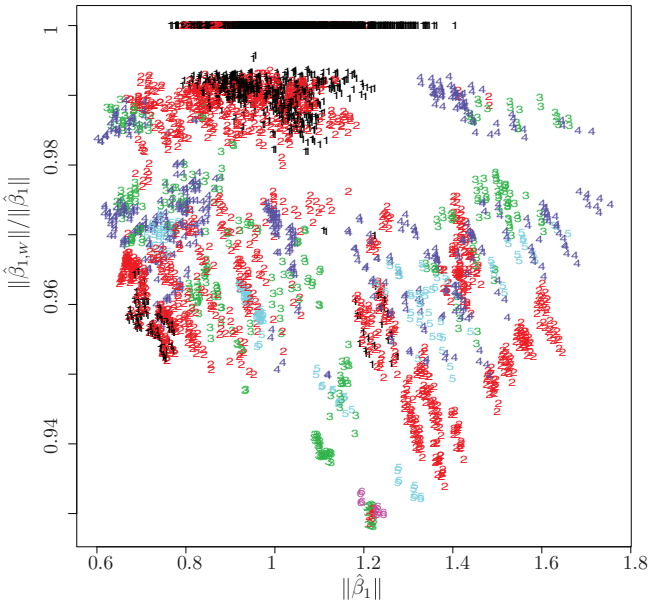


Figure 4. Ratio of $\|\hat{\beta}_{1,w}\|$ over $\|\hat{\beta}_1\|$ versus $\|\hat{\beta}_1\|$ based on 100 brain templates, resulting in 4950 pairs, in Example 2. Estimated standard errors of $\|\hat{\beta}_1\|$ (s.e.) associated with different numerical symbols are: s.e.$\in (0, 0.1)$ for "1," s.e.$\in [0.1, 0.5)$ for "2," s.e.$\in [0.5, 1)$ for "3," s.e.$\in [1, 2)$ for "4," s.e.$\in [2, 3)$ for "5," and s.e.$\in [3, 4.21)$ for "6."

slice of the brain scan, these candidate templates can be viewed as error-contaminated measures of the same underlying structure in the corpus callosum, the true scale parameter when matching one template onto the other, while appropriately accounting for measurement error, is equal to one. Using all available 604 candidate templates corresponding to the same slice of brain scan, we obtain $\hat{\sigma}_u^2 \approx 0.002$ and $\hat{\sigma}_x^2 + \hat{\sigma}_u^2 \approx 0.015$, yielding an estimated reliability ratio as 0.860. Then we choose a particular pair of templates as $\mathbf{Y}$ and $\mathbf{W}$. Matching $\mathbf{W}$ onto $\mathbf{Y}$ via naive OPA yields the estimated scale as 0.951 (0.061) and the estimated rotation as 2.116° (2.956), with sandwich standard error estimates in parentheses, and the counterpart conditional score estimates are 0.999 (0.062) and 2.116° (2.947). With such a small number of landmarks ($K = 8$), high variability in these estimates is inevitable. Although we cannot conclude statistically significant discrepancy between the naive scale estimate and conditional score estimate (with the high variability in estimates), we observe across all 182,106 pairs (from 604 templates with $\mathbf{Y}$ and $\mathbf{W}$ assigned at random for each pair) that the former is never larger than the latter. Figure 4 depicts $\|\hat{\beta}_{1,w}\|/\|\hat{\beta}_1\|$ versus $\|\hat{\beta}_1\|$ when matching 4950 pairs resulting from the first 100 templates.

Example 3 (Face data). Evison et al. (2010) analyzed three-dimensional facial images of healthy volunteers, which were collected using a Geometrix FaceVision® FV802 Biometric Camera (ALIVE Tech, Cumming, GA). To demonstrate the conditional score method for three-dimensional shape data in comparison with naive OPA, we implement these two methods using data from one volunteer, whose facial configuration consisting of 61 landmarks was measured twice by each of the two operators in the study.

More specifically, denote by $\mathbf{W}_{1,w}$ and $\mathbf{W}_{2,w}$ the two replicate measures for this volunteer's face collected by one operator, and by $\mathbf{Y}$ a facial configuration of the same volunteer measured by the other operator. Using the duplicate measures, $\mathbf{W}_{1,w}$ and $\mathbf{W}_{2,w}$, and applying the method related in Section 4.3, we obtain $\hat{\sigma}_u^2 \approx 0.007$. This estimate of the measurement error variance, in contrast to the estimate of $\sigma_x^2 + \sigma_u^2$ (≈ 5.216), results in the estimated reliability ratio equal to 0.999, suggesting minimal error contamination in the observed facial images. This is also implied by Evison et al. (2010), who stated that the three-dimensional coordinates of landmarks were collected with high precision using a software tool. The estimated scale resulting from matching $\mathbf{W}_{1,w}$ onto $\mathbf{Y}$ using naive OPA and the conditional score estimate are 0.988 (0.006) and 0.990 (0.006), respectively, with the corresponding estimated standard errors in parentheses. As expected when measurement error is scarce, these two estimates are very similar. To simulate a scenario with slightly more error contamination as expected from CCTV cameras in the field, we add isotropic measurement error to $\mathbf{W}_{1,w}$ to reduce the reliability ratio to 0.9, and repeat the two analyses using the further contaminated configuration. This round of analyses yields the naive estimate of the scale as 0.845 (0.020) and the conditional score estimate as 0.920 (0.024), which demonstrates that naive OPA leads to an attenuated scale estimate.

7. DISCUSSION

We have shown theoretically and empirically that comparisons of two error-contaminated configurations via naive OPA can be misleading, with the adverse effects of measurement error mostly manifested in the attenuated scale estimation. Besides the severity of error contamination, the interplay of several factors that affect the degree of deterioration in inference from naive OPA is revealed in our bias analyses in Section 3. One noteworthy finding is that ignoring measurement error does less harm when the unobserved true configuration consists of more diffuse landmarks. Comforting as this finding sounds, one may not know with absolute certainty if the true configuration has landmarks diffuse enough to counteract measurement error in one particular application. The proposed conditional score method that accounts for measurement error and yields consistent inference for all parameters of interest is a valuable addition to the methodology available for size and shape analyses.

Formulating the measurement error models for shape data using complex/quaternion random variables allows us to derive the conditional scores in ways that are similar to those employed in the context of real-value linear measurement error models. However, although it is assumed in most real-value regression analyses that $(Y_l,\ X_l,\ W_l)$ are independent across $l = 1, \ldots, K$, it is more natural to view $(Y_l,\ X_l,\ W_l)$ correlated across $l = 1, \ldots, K$ for shape data. This correlation creates some technical challenges in studying the bias of naive estimators but not in deriving the conditional scores, as knowledge about such correlation is neither needed nor used in constructing conditional scores. But now that correlation does exist, it is worth considering next how to construct new unbiased scores that account for such correlation. By taking into account the correlation among landmarks within a configuration, the new scores

can be more efficient than the conditional scores proposed in the current study.

When more than two shapes are of interest, generalized Procrustes analysis, or GPA, can be used to estimate the mean size and shape (Dryden and Mardia 1998, sec. 5.3). In the presence of measurement error, one may fit the conditional score method into the algorithm of GPA where each error-prone shape is matched onto the estimated mean shape obtained at each iteration. But this becomes unnecessary if the main goal is to estimate the mean size and shape of these error-prone configurations, as ignoring measurement error does not affect the outcome of (naive) GPA regarding mean shape estimation in terms of consistency. This argument is in line with that (naive) predictions are not affected by measurement error, except for efficiency, in real-value linear measurement error models. Regardless, conditional score estimates of the matching parameters can be useful in the case, say, when an accurately measured rat skull becomes available and one wishes to predict the size and shape of this skull at a future time point. Assuming this rat and the rats in Example 1 in Section 6 come from the same population, then using the conditional score estimates of the matching parameters obtained there for prediction of this new rat skull is clearly more appealing than using the estimates from naive OPA.

As suggested by a referee, instead of the additive measurement error in (2), one may view the true configuration $\mathbf{X}$ as a set of realizations of the function, $X(d)$, where $d(\in \mathcal{D})$ is an argument of the function representing the true location of a landmark; then one may view the observed configuration $\mathbf{W}$ as another set of realizations of the same function, $X(d_W)$, where $d_W(\in \mathcal{D})$ is a noisy version of the target location d. Adopting this viewpoint, one needs a model that relates d_W and d, which may lead to a model connecting $\mathbf{W}$ and $\mathbf{X}$. Either using (2) or this strategy, with appropriate measurement error models formulated for error-prone shape data, the large battery of methodologies in the existing literature on measurement error can be potentially useful hints for solving other problems in shape analyses in the presence of measurement error. These include the widely adopted regression calibration and SIMEX (Carroll et al. 2006, chaps. 4 and 5), and other methods under the Bayesian framework (Carroll et al. 2006, chap. 9). The study reported in this article opens up this new avenue of research in shape analysis.

SUPPLEMENTARY MATERIALS

The supplement to this article contains Appendices A–G referenced in Sections 2.2, 3, 4.2, 5.1, and 6.

[Received April 2013. Revised January 2014.]

REFERENCES

Amit, Y. (1997), "Graphical Shape Templates for Automatic Anatomy Detection With Applications to MRI Brain Scans," *IEEE Transactions on Medical Imaging*, 16, 28–40. [375]

Booktein, F. L. (1991), *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge: Cambridge University Press. [368,374]

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Non-Linear Models: A Modern Perspective* (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC. [369,370,371,372,377]

Dryden, I. L. (2003), "Statistical Shape Analysis in High-Level Vision," *Mathematical Methods in Computer Vision*, New York: Springer-Verlag. [375]

——— (2012), *shapes Package*, Vienna, Austria: R Foundation for Statistical Computing. Contributed package. Available at *http://www.R-project.org*. [373]

Dryden, I. L., and Mardia, K. V. (1998), *Statistical Shape Analysis*, Chichester: Wiley. [368,369,370,371,377]

Du, J. (2012), "Measurement Error Models in Shape Analysis," Ph.D. dissertation, Department of Statistics, University of South Carolina. [370,372]

Evison, M. P., Dryden, I. L., Fieller, N., Mallett, X. D., Morecroft, L., Schofield, D., and Bruegge, R. V. (2010), "Key Parameters of Face Shape Variation in 3D in a Large Sample," *Journal of Forensic Sciences*, 55, 159–162. [376]

Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley. [370]

Goodall, C. (1991), "Procrustes Methods in the Statistical Analysis of Shape," *Journal of the Royal Statistical Society,* Series B, 53, 285–339. [368]

Goodall, C. R., and Lange, N. (1989), "Growth Curve Models for Correlated Triangular Shapes" (with discussion), in *Proceedings of the 21st Symposium on the Interface Between Computing Science and Statistics*, pp. 445–454. [374]

Goodman, N. R. (1963), "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution" (an introduction), *The Annals of Mathematical Statistics*, 34, 152–177. [370]

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010), "Longitudinal Functional Principal Component Analysis," *Electronic Journal of Statistics*, 4, 1022–1054. [375]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [371]

Horn, B. P. (1987), "Closed-Form Solution of Absolute Orientation Using Unit Quaternions," *Journal of the Optical Society of America,* Series A, 4, 629–642. [369]

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the 5th Berkeley Symposium*, 1, 221–233. [371]

Kendall, D. G. (1984), "Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces," *Bulletin of the London Mathematical Society*, 16, 81–121. [368]

Kenobi, K., Dryden, I. L., and Le, H. (2010), "Shape Curves and Geodesic Modeling," *Biometrika*, 97, 567–584. [368,374]

Kent, J. T. (1994), "The Complex Bingham Distribution and Shape Analysis," *Journal of the Royal Statistical Society,* Series B, 56, 285–299. [370]

Kent, J. T., and Mardia, K. V. (2001), "Shape, Procrustes Tangent Projections and Bilateral Symmetry," *Biometrika*, 88, 469–485. [371]

Kent, J. T., Mardia, K. V., Morris, R. J., and Aykroyd, R. D. (2001), "Functional Models of Growth for Landmark Data," in *Proceedings in Functional and Spatial Data Analysis*, eds. K. V. Mardia, R. G. Aykroyd, Leeds: University of Leeds, pp. 109–115. [374]

Konno, Y. (2007), "Estimation of Normal Covariance Matrices Parameterized by Irreducible Symmetric Cones Under Stein's Loss," *Journal of Multivariate Analysis*, 98, 259–316. [370]

Kunze, K., and Schaeben, H. (2004), "The Bingham Distribution of Quaternions and Its Spherical Radon Transform in Texture Analysis," *Mathematical Geology*, 36, 917–943. [369]

Le, H., and Kume, A. (2000), "Detection of Shape Changes in Biological Features," *Journal of Microscopy*, 200, 140–147. [374]

Mardia, K. V., Fallaize, C. J., Barber, S. J., Jackson, R. M., and Theobald, D. L. (2013), "Bayesian Alignment of Similarity Shapes," *The Annals of Applied Statistics*, 7, 989–1009. [368,371,374]

Monteiro, L. R. (1999), "Multivariate Regression Models and Geometric Morphometrics: The Search for Causal Factors in the Analysis of Shape," *Systematic Biology*, 48, 192–199. [374]

R Development Core Team (2012), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [373]

Stefanski, L. A., and Boos, D. D. (2002), "The Calculus of M-Estimation," *The American Statistician*, 56, 29–38. [371]

Zhang, F. (1997), "Quaternions and Matrices of Quaternions," *Linear Algebra and Its Applications*, 251, 21–57. [369]