Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 119 (2004) 191–206

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Fitting a line segment to noisy data

O. Davidov*, A. Goldenshluger

Department of Statistics, University of Haifa, Mount Carmel, Haifa 31905, Israel

Received 27 January 2001; accepted 15 July 2002

Abstract

In this paper, we consider fitting a line segment to noisy data. We describe the structural segment model and establish conditions for its identifiability. The method of moments estimator (MME) is introduced and explored. The MME is easily computed and is invariant under translation rotation and reflection. We show that the MME follows, asymptotically, a normal distribution. The asymptotic efficiency of the MME relative to the maximum likelihood is investigated numerically and found to be high in most cases.

Keywords: Identifiability; Line-segment structural relationship; Measurement error; Method of moments

1. Introduction

In this paper, we investigate the problem of reconstructing a line segment from noisy observations. Consider the model

$$x_i = \lambda_i \xi_x + (1 - \lambda_i)\eta_x + \varepsilon_i, \tag{1}$$

$$y_i = \lambda_i \xi_y + (1 - \lambda_i)\eta_y + \upsilon_i, \tag{2}$$

where $(\varepsilon_1, \upsilon_1), \ldots, (\varepsilon_n, \upsilon_n)$ is a sequence of independent normally distributed random vectors with mean zero and diagonal variance matrix $\mathrm{Diag}(\sigma_\varepsilon^2, \sigma_\upsilon^2)$. The unobservable quantities $\lambda_1, \ldots, \lambda_n$ take values in the interval $[0, 1]$. The goal is to estimate $\eta = (\eta_x, \eta_y)$ and $\xi = (\xi_x, \xi_y)$, the endpoints of an interval in $\mathbb{R}^2$, using the observations (x_i, y_i), $i = 1, \ldots, n$.

* Corresponding author.
E-mail addresses: davidov@stat.haifa.ac.il (O. Davidov), goldensh@stat.haifa.ac.il (A. Goldenshluger).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00409-3

The model (1)–(2) is motivated by applications in image analysis and computer vision. Line segment fitting is frequently used in the transition from numerical to symbolic description of images. For example, many procedures for edge detection and image segmentation require accurate fitting of segments to form the boundary between distinct regions of the image. For applications, examples and a survey of current algorithms see Madsen and Christensen (1995), Vosselman and Haralick (1996), Stewart et al. (1999), Werman and Keren (2001), and the references therein. This literature focuses mostly on computational and algorithmic aspects of segment fitting and neglects important modeling and statistical considerations. Consequently the performance of seemingly natural and widely used algorithms may be poor in practice. In fact a widely used class of algorithms based on projections of the data on the estimated regression line results in inconsistent estimators (Madsen and Christensen, 1995).

In this paper, we study the statistical properties of (1) and (2) assuming that the quantities $\lambda_1, \ldots, \lambda_n$ are independent random variables with distribution G. This setup is commonly referred to as the structural model. If the quantities λ_i are treated as unknown incidental parameters (1) and (2) describes the so-called functional relationship (Pfanzagl, 1993). This model is closely related to regression with errors-in-variables (Kendall and Stuart, 1979; Fuller, 1987; Cheng and Van Ness, 1999). In particular, (1) and (2) can be rewritten as

$$x_i = \beta_0 + \beta_1 \lambda_i + \varepsilon_i,$$

$$y_i = \alpha_0 + \alpha_1 \lambda_i + \upsilon_i$$

which resembles the factor analysis model investigated by Fuller (1987, Section 1.4, Section 4.3). In this case λ_i plays the role of the common or latent factor. In factor analysis models the latent factor is assumed to be normally distributed and observed with an additive random error. In our setup the quantities $\{\lambda_i\}$ are not observable, and are randomly distributed on the interval $[0, 1]$. We focus on the endpoints of the line segment which can be expressed as linear combinations of β_0, β_1, and α_0, α_1. It will be shown that our modeling assumptions lead to results that differ from those for the factor analysis model described by Fuller (1987, Chapter 4). For completeness we note that the model is also closely related to the circular functional and structural models (Chan, 1965; Anderson, 1981; Berman and Culpin, 1986). In these papers methods for estimating the center and radius of an unknown circle were investigated.

The paper is organized in the following way. In Section 2 the identifiability of the model is established. In Section 3 the method of moments estimator is introduced. Its properties are investigated in Section 4. In Section 5 the efficiency of the moment estimator is compared numerically with the corresponding maximum likelihood estimator (MLE). Finally Section 6 serves as a brief summary and discussion.

2. Model identifiability

Assume that $\{\lambda_i\}$ are independent and identically distributed random variables with distribution function G. Let $\theta = (\eta, \xi)$. It is natural to ask whether the parameter

θ can be identified from the joint distribution of the observations $f(x,y)$. Define $\mathscr{S}=\{\theta,\sigma_\varepsilon^2,\sigma_\upsilon^2,G\}$ to be the structure of the model. Clearly the structure determines the distribution of the data. We say that a model is non-identifiable if there exist structures $\mathscr{S}\neq\mathscr{S}^*$ such that $f(x,y|\mathscr{S})=f(x,y|\mathscr{S}^*)$ for all x, y. Otherwise the model is said to be identifiable. The issue arises because the random variable (x,y) is itself a sum of unobserved random quantities that depend on the random variables λ, ε and υ.

The next theorem shows that the condition that the points 0 and 1 belong to the support of G is necessary and sufficient for identifiability. Recall that random variables λ_i model the sampling design along the line segment with endpoints η and ξ. Intuitively, η and ξ are identifiable if there are observations that come from, however small, neighborhoods of the endpoints. In other words, the support of G should include 0 and 1. The next theorem shows that if this is not the case, we can construct two different structures $\mathscr{S}$ and $\mathscr{S}^*$ such that the distribution of the points (x,y) derived under $\mathscr{S}$ and $\mathscr{S}^*$ is the same.

Theorem 1. *A necessary and sufficient condition for $\mathscr{S}$ to be identifiable is that* 0 *and* 1 *belong to the support of G.*

Proof. Let $\mathscr{S}=\{\theta,\sigma_\varepsilon^2,\sigma_\upsilon^2,G\}$, and $\mathscr{S}^*=\{\theta^*,\sigma_\varepsilon^{*2},\sigma_\upsilon^{*2},G^*\}$. Suppose $f(x,y|\mathscr{S})=f(x,y|\mathscr{S}^*)$ for all x,y, and 0 and 1 belong to the support of G and G^*. First we prove sufficiency by showing that $\mathscr{S}=\mathscr{S}^*$. The joint characteristic function of (x,y) under $\mathscr{S}$ is

$$\psi_{x,y}(s,t)=\exp\left\{-\frac{1}{2}\sigma_\varepsilon^2 s^2-\frac{1}{2}\sigma_\upsilon^2 t^2\right\}\bar{\psi}(s,t), \tag{3}$$

where

$$\bar{\psi}(s,t)=\int_0^1 \exp\{\{\mathrm{i}s[\lambda\xi_x+(1-\lambda)\eta_x]\}\exp\{\mathrm{i}t[\lambda\xi_y+(1-\lambda)\eta_y]\}\,\mathrm{d}G(\lambda).$$

Consider the one-dimensional characteristic function $\psi_x(s)$ obtained by putting $t=0$

$$\begin{aligned}\psi_x(s)&=\exp\left\{-\frac{1}{2}\,\sigma_\varepsilon^2 s^2\right\}\int_0^1 \exp\{\mathrm{i}s[\lambda\xi_x+(1-\lambda)\eta_x]\}\,\mathrm{d}G(\lambda)\\ &=\exp\left\{-\frac{1}{2}\,\sigma_\varepsilon^2 s^2\right\}\int_{\eta_x}^{\xi_x} \mathrm{e}^{\mathrm{i}sv}\,\mathrm{d}\tilde{G}(v).\end{aligned} \tag{4}$$

Here $\tilde{G}$ is the distribution function of $\lambda\xi_x+(1-\lambda)\eta_x$. Similarly, for $\mathscr{S}^*$ we have

$$\psi_x^*(s)=\exp\left\{-\frac{1}{2}\,\sigma_\varepsilon^{*2} s^2\right\}\int_{\eta_x^*}^{\xi_x^*} \mathrm{e}^{\mathrm{i}sv}\,\mathrm{d}\tilde{G}^*(v), \tag{5}$$

where $\tilde{G}^*$ is the distribution function of $\lambda\xi_x^*+(1-\lambda)\eta_x^*$. Since η_x,ξ_x and η_x^*,ξ_x^* are the endpoints of the supports of $\tilde{G}$ and $\tilde{G}^*$, respectively, the integrals on the RHS of (4) and (5) are analytic functions that do not have normal factors. Therefore, comparing the normal factors in (4) and (5) we conclude that $\sigma_\varepsilon^2 = \sigma_\varepsilon^{*2}$. Thus, if $\psi_x(s) = \psi_x^*(s)$ then,

$$\int_{\eta_x}^{\xi_x} e^{isv}\, d\tilde{G}(v) = \int_{\eta_x^*}^{\xi_x^*} e^{isv}\, d\tilde{G}^*(v), \qquad \forall s \in \mathbb{R}.$$

It follows from the uniqueness of the Fourier–Stieltjes transform that $\eta_x = \eta_x^*$, $\xi_x = \xi_x^*$, and $\tilde{G}(v) = \tilde{G}^*(v)$ for every v. Similarly, considering the one-dimensional characteristic functions $\psi_y(t)$ and $\psi_y^*(t)$ obtained by putting $s = 0$, we deduce that $\sigma_\upsilon^2 = \sigma_\upsilon^{*2}$, $\eta_y = \eta_y^*$, and $\xi_y = \xi_y^*$. Therefore if the structures $\mathscr{S}$ and $\mathscr{S}^*$ generate the same joint distribution for (x, y) then $\mathscr{S}$ and $\mathscr{S}^*$ are identical. Thus, if 0 and 1 belong to the support of G, then the corresponding structure is identifiable.

To show that the condition that 0 and 1 belong to the support of G is necessary for identifiability, we consider the following counter-example. Let $\mathscr{S} = \{\theta, \sigma_\varepsilon^2, \sigma_\upsilon^2, G\}$, where $\theta = (\eta, \xi)$, $\xi_x > \eta_x, \xi_y > \eta_y$, and G is the uniform distribution on the interval $[\frac{1}{2} - u, \frac{1}{2} + u]$ for some $0 < u < 1/2$. Consider the structure $\mathscr{S}^* = \{\theta^*, \sigma_\varepsilon^{*2}, \sigma_\upsilon^{*2}, G^*\}$, where $\theta^* = (\eta^*, \xi^*)$ satisfies the following conditions: $\xi_x^* > \eta_x^*, \xi_y^* > \eta_y^*$, and

$$\xi_x^* + \eta_x^* = \xi_x + \eta_x, \quad \xi_y^* + \eta_y^* = \xi_y + \eta_y, \quad \frac{\xi_x - \eta_x}{\xi_x^* - \eta_x^*} = \frac{\xi_y - \eta_y}{\xi_y^* - \eta_y^*} = \alpha < 1. \tag{6}$$

Let G^* be a uniform distribution on $[\frac{1}{2} - \alpha u, \frac{1}{2} + \alpha u]$. In words (6) means that (i) the segments under $\mathscr{S}$ and $\mathscr{S}^*$ have the same center and slope but the segment under $\mathscr{S}^*$ is shorter by a factor α; and (ii) the distribution functions G and G^* are both uniform but the support of G^* is shorter by a factor of α.

The joint characteristic function of (x, y) under $\mathscr{S}$ and $\mathscr{S}^*$ are given by (3) where

$$\bar{\psi}(s,t) = \frac{1}{2u} \int_{1/2-u}^{1/2+u} \exp\{is[\lambda\xi_x + (1-\lambda)\eta_x]\} \exp\{it[\lambda\xi_y + (1-\lambda)\eta_y]\}\, d\lambda,$$

$$\bar{\psi}^*(s,t) = \frac{1}{2\alpha u} \int_{1/2-\alpha u}^{1/2+\alpha u} \exp\{is[\lambda\xi_x^* + (1-\lambda)\eta_x^*]\} \exp\{it[\lambda\xi_y^* + (1-\lambda)\eta_y^*]\}\, d\lambda$$

respectively. Changing variables in the last integrals and using (6) we conclude that $\bar{\psi}(s,t) = \bar{\psi}^*(s,t)$. Thus the characteristic functions of (x, y) coincide under $\mathscr{S}$ and $\mathscr{S}^*$, which implies that $\mathscr{S}$ is not identifiable whenever 0 and 1 are outside the support of G. □

In general identifiability is a necessary but not sufficient condition to guarantee consistent estimation of the structural parameters. Typically (e.g. in the normal structural model) more conditions are imposed or required for any particular estimation method.

The identifiability conditions stated in Theorem 1 are assumed throughout the paper. Under these conditions estimating the endpoints (η, ξ) is equivalent to estimating the support of the nuisance distribution G. In Section 3 we develop the method of moments estimator (MME) which does not require full distributional assumptions on λ. It is important to note that the MME exists even if the conditions in Theorem 1 do not hold. However in that case the endpoints (η, ξ) cannot be estimated. Moreover, if G is known up to a finite number of parameters, the standard MLE can be implemented. The MLE will depend explicitly on the identifiability conditions. We discuss these issues in Section 6.

3. The method of moments estimator

In this section, we investigate the MME for the structural parameter θ indexing the models (1) and (2). Equating the sample and theoretical moments, up to the second order, results in the following system of equations:

$$\begin{aligned} \bar{x} &= \eta_x + (\xi_x - \eta_x)\mu_\lambda, \\ \bar{y} &= \eta_y + (\xi_y - \eta_y)\mu_\lambda, \\ s_{xx} &= (\xi_x - \eta_x)^2\sigma_\lambda^2 + \sigma_\varepsilon^2, \\ s_{yy} &= (\xi_y - \eta_y)^2\sigma_\lambda^2 + \sigma_\upsilon^2, \\ s_{xy} &= (\xi_x - \eta_x)(\xi_y - \eta_y)\sigma_\lambda^2, \end{aligned} \tag{7}$$

where μ_λ and σ_λ^2 are the mean and variance of the random variable λ, and the sample moments are $\bar{x} = n^{-1}\sum_{i=1}^n x_i$, $\bar{y} = n^{-1}\sum_{i=1}^n y_i$, and

$$\begin{aligned} s_{xx} &= n^{-1}\sum_{i=1}^n (x_i - \bar{x})^2, \\ s_{yy} &= n^{-1}\sum_{i=1}^n (y_i - \bar{y})^2, \\ s_{xy} &= n^{-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \end{aligned}$$

We consider estimating the points η and ξ assuming that the parameters $\mu_\lambda, \sigma_\lambda^2$, and $\sigma_\varepsilon^2, \sigma_\upsilon^2$ are known. The distribution of the points λ_i is related to sampling design along the line segment. In many applications it is reasonable to assume that the sampling design is symmetric, and typically chosen to be uniform on $[0, 1]$ (cf. Vosselman and Haralick, 1996; Werman and Keren, 2001). These assumptions can be weakened (see Section 6). The errors variances are related to the precision of the measuring device. In practice the errors variances can be often estimated independently from repeated observations. In particular, in computer vision applications the observations from the outer boundary of the image can be used for this purpose (Haralick and Shapiro, 1992).

In digitized images the errors standard deviations are usually small and may be measured in pixels. Madsen and Christensen (1995) argue that it is reasonable to assume that the edge pixels errors are independent in both directions and have the same variance, i.e. $\sigma_\varepsilon^2 = \sigma_\upsilon^2$. Given a real image, one can apply an estimation procedure for different values of the edge pixel variance and to choose the one yielding the best qualitative results. Although we develop our estimator assuming that both σ_ε^2 and σ_υ^2 are known, this assumption may be relaxed in several ways. In particular, one can assume that only the ratio $\sigma_\varepsilon^2/\sigma_\upsilon^2$ is known; for more details and other extensions see Section 6.

Let $\delta = (\delta_x, \delta_y) = (\xi_x - \eta_x, \xi_y - \eta_y)$ denote the signed lengths of the line segment along the x- and y-axis, respectively. The unsigned or positive lengths are $\delta_x^+ = |\delta_x|$ and $\delta_y^+ = |\delta_y|$. Without loss of generality we assume that $\xi_x \geqslant \eta_x$. Therefore we label the interval endpoint with the smaller x coordinate as η. If $\eta_x = \xi_x$, then the point with smaller y coordinate is denoted by η. Note that $\mathrm{Cov}[x, y] \neq 0$ if and only if $\xi_x \neq \eta_x$ and $\xi_y \neq \eta_y$. Hence nonzero covariance precludes intervals with horizontal or vertical orientations and intervals degenerating to a point. Alternatively, if $\mathrm{Cov}[x, y] = 0$ then either δ_x and/or δ_y equal 0.

The solution to (7) is the MME and is denoted $\hat\theta_n = (\hat\eta_x, \hat\eta_y, \hat\xi_x, \hat\xi_y)$. Note that if δ_x or δ_y are equal to zero only a subset of the system (7) need be considered. Hence the form of $\hat\theta_n$ depends on δ_x and δ_y and by extension on $\mathrm{Cov}[x, y]$. The estimators and their properties for the general case of $\mathrm{Cov}[x, y] \neq 0$ are investigated first. The estimating equations, and their properties, assuming $\mathrm{Cov}[x, y] = 0$ are discussed in Section 4.3. If $\mathrm{Cov}[x, y] \neq 0$, then (7) is a quadratic system of five equations in four unknowns which can be uniquely solved for θ. The MME is given by

$$
\begin{aligned}
\hat\eta_x &= \bar x - \mu_\lambda \hat\delta_x^+,\\
\hat\eta_y &= \bar y - \mu_\lambda \hat\delta_y^+ \operatorname{sgn}(s_{xy}),\\
\hat\xi_x &= \bar x + (1 - \mu_\lambda)\hat\delta_x^+,\\
\hat\xi_y &= \bar y + (1 - \mu_\lambda)\hat\delta_y^+ \operatorname{sgn}(s_{xy}),
\end{aligned}
\tag{8}
$$

where $\operatorname{sgn}(x)$ equals 1 if $x \geqslant 0$ and -1 otherwise, and

$$
\begin{aligned}
\hat\delta_x^+ &= \sigma_\lambda^{-1}\sqrt{(s_{xx} - \sigma_\varepsilon^2)_+},\\
\hat\delta_y^+ &= \sigma_\lambda^{-1}\sqrt{(s_{yy} - \sigma_\upsilon^2)_+},
\end{aligned}
$$

where $(x)_+ = \max(0, x)$. The signed lengths of the line segment along the x- and y-axis, respectively, are estimated by $\hat\delta_x = \hat\delta_x^+$ and $\hat\delta_y = \hat\delta_y^+ \operatorname{sgn}(s_{xy})$.

4. Algebraic and statistical properties of the MME

In this section, we investigate the statistical and algebraic properties of the MME $\hat{\theta}_n$. First note that the line segment connecting $\hat{\eta}$ and $\hat{\xi}$ is given by

$$\frac{y-\hat{\eta}_y}{x-\hat{\eta}_x} = \frac{\hat{\xi}_y-\hat{\eta}_y}{\hat{\xi}_x-\hat{\eta}_x}, \tag{9}$$

for $\hat{\eta}_x \leqslant x \leqslant \hat{\xi}_x$ and $\hat{\eta}_y \leqslant y \leqslant \hat{\xi}_y$. Substituting (8) into (9) we easily see that (9) passes through the point $(\bar{x}, \bar{y})$ consistent with simple regression as well as measurement error models.

4.1. Transformations

Next, we show that the MME $\hat{\theta}_n$ is location equivariant. Let $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$ be transformed to $x' = x + a1_n$ and $y' = y + b1_n$ where 1_n is the unit vector of length n. Note that

$$\bar{x}' = n^{-1}\sum_{i=1}^{n} x_i' = \bar{x} + a,$$

$$s_{x'x'} = n^{-1}\sum_{i=1}^{n}(x_i' - \bar{x}')^2 = s_{xx}$$

and therefore $\hat{\delta}_{x'} = \hat{\delta}_x$. Similarly $\bar{y}' = \bar{y} + b$, $s_{y'y'} = s_{yy}$ and $\mathrm{sgn}(s_{x'y'}) = \mathrm{sgn}(s_{xy})$. Thus,

$$\hat{\eta}_x(x', y'),= a + \hat{\eta}_x(x, y),$$

$$\hat{\eta}_y(x', y') = b + \hat{\eta}_y(x, y).$$

Analogous results pertain to $\hat{\xi}(x', y')$. Therefore the MME, $\hat{\theta}_n$, is location equivariant. Furthermore it can be easily verified that

$$f(x', y'|\theta') = f(x, y|\theta), \tag{10}$$

where $\theta' = (\eta_x + a, \eta_y + b, \xi_x + a, \xi_y + b)$. In addition

$$L(\hat{\theta}(x, y), \theta) = L(\hat{\theta}(x', y'), \theta'), \tag{11}$$

for any loss function $L(\cdot,\cdot)$ that is a function of $\hat{\theta} - \theta$. Hence the line segment estimation problem is location invariant. Consequently the variance and the bias of the MME depend only on the length of the line segment $\delta = (\delta_x, \delta_y)$ and do not depend on the "origin", η, of the line segment. The variance of the MME is most conveniently calculated at $\eta = 0$.

Similar calculations show that the line segment model satisfies (10) (with the proper induced parameter θ') for scale, rotation, and reflection transformations. Consider the transformation $x' = x\cos\alpha - y\sin\alpha$ and $y' = y\cos\alpha + x\sin\alpha$ which amounts to rotating the axes by the angle α. It is easily shown that $(\bar{x}', \bar{y}')$ the center of the transformed data is exactly the transformed center of the original data. Moreover

$$s^2_{x'x'} + s^2_{y'y'} = s^2_{xx} + s^2_{yy},$$

hence $\hat{\delta}^2_{x'} + \hat{\delta}^2_{y'} = \hat{\delta}^2_x + \hat{\delta}^2_y$. Thus the MME is also equivariant under rotation. Similarly, the MME is equivariant under reflection around the x $(x' = -x)$ and/or the y $(y' = -y)$ axes. Note that (11) does not hold for scale, rotation and reflection and the family of loss functions discussed above. Therefore, the variance matrix for the MME is not preserved under these transformations.

4.2. Asymptotic distribution

In the following it is shown that the MME is consistent and its asymptotic distribution is derived assuming $\mathrm{Cov}[x, y] \neq 0$. For $\mathrm{Cov}[x, y] = 0$ see Section 4.3.

Theorem 2. *Assume* $\mathrm{Cov}[x, y] \neq 0$. *Then the* MME, $\hat{\theta}_n$, *is consistent and asymptotically multivariate normal with mean* θ *and variance matrix* $V_\theta = M\Sigma_\theta M'$. *The matrix* Σ_θ *is given by*

$$\begin{pmatrix} \mu_{[1,1]}\delta_x^2 + \sigma_\varepsilon^2 & \mu_{[1,1]}\delta_x\delta_y & \mu_{[1,2]}\delta_x^3 + 2\mu_{[1]}\delta_x\sigma_\varepsilon^2 & \mu_{[1,2]}\delta_x\delta_y^2 \\ \mu_{[1,1]}\delta_x\delta_y & \mu_{[1,1]}\delta_y^2 + \sigma_\upsilon^2 & \mu_{[1,2]}\delta_x^2\delta_y & \mu_{[1,2]}\delta_y^3 + 2\mu_{[1]}\delta_y\sigma_\upsilon^2 \\ \mu_{[1,2]}\delta_x^3 + 2\mu_{[1]}\delta_x\sigma_\varepsilon^2 & \mu_{[1,2]}\delta_x^2\delta_y & \mu_{[2,2]}\delta_x^4 + 4\mu_{[2]}\delta_x^2\sigma_\varepsilon^2 + 2\sigma_\varepsilon^4 & \mu_{[2,2]}\delta_x^2\delta_y^2 \\ \mu_{[1,2]}\delta_x\delta_y^2 & \mu_{[1,2]}\delta_y^3 + 2\mu_{[1]}\delta_y\sigma_\upsilon^2 & \mu_{[2,2]}\delta_x^2\delta_y^2 & \mu_{[2,2]}\delta_y^4 + 4\mu_{[2]}\delta_y^2\sigma_\upsilon^2 + 2\sigma_\upsilon^4 \end{pmatrix},$$

where $\mu_{[s,t]} = E[\lambda^{s+t}] - E[\lambda^s]E[\lambda^t]$ *and* $\mu_{[s]} = E[\lambda^s]$. *The matrix* M *equals*

$$\begin{pmatrix} 1 + \frac{\mu_\lambda^2}{\sigma_\lambda^2} & 0 & -\frac{\mu_\lambda}{2\sigma_\lambda^2\delta_x} & 0 \\ 0 & 1 + \frac{\mu_\lambda^2}{\sigma_\lambda^2} & 0 & -\frac{\mu_\lambda}{2\sigma_\lambda^2\delta_y} \\ 1 - \frac{(1-\mu_\lambda)\mu_\lambda}{\sigma_\lambda^2} & 0 & \frac{(1-\mu_\lambda)}{2\sigma_\lambda^2\delta_x} & 0 \\ 0 & 1 - \frac{(1-\mu_\lambda)\mu_\lambda}{\sigma_\lambda^2} & 0 & \frac{(1-\mu_\lambda)}{2\sigma_\lambda^2\delta_y} \end{pmatrix}, \tag{12}$$

where $\mu_\lambda = \mu_{[1]}$ *and* $\sigma_\lambda^2 = \mu_{[1,1]}$.

Proof. The MME $\hat{\theta}_n$ is location invariant. Hence its variance matrix depends on η and ξ only through δ [cf. Theorem 1.1, Chapter 3, Lehmann and Casella (1998)]. Therefore, in the proof below we set $\eta = 0$. Define $S_n = (\bar{x}_1, \bar{y}_1, \bar{x}_2, \bar{y}_2)$, $u_n = I_{\{s_{xy} \geqslant 0\}}$ and $w_n = I_{\{s_{xx} > \sigma_\varepsilon^2, s_{yy} > \sigma_\upsilon^2\}}$ where $\bar{x}_k = \bar{x}_k(n) = n^{-1}\sum_i x_i^k$ and $\bar{y}_k = \bar{y}_k(n) = n^{-1}\sum_i y_i^k$ for

$k=1,2$. Let $h(x,y;\alpha,\beta)=x+\alpha\sqrt{y-x^2-\beta}$ and note that if $u_n=1$ and $w_n=1$ then

$$\hat{\theta}_n = H_1(S_n) = \begin{pmatrix} h(\bar{x}_1,\bar{x}_2;-\mu_\lambda/\sigma_\lambda,\sigma_\varepsilon^2) \\ h(\bar{y}_1,\bar{y}_2;-\mu_\lambda/\sigma_\lambda,\sigma_\upsilon^2) \\ h(\bar{x}_1,\bar{x}_2;(1-\mu_\lambda)/\sigma_\lambda,\sigma_\varepsilon^2) \\ h(\bar{y}_1,\bar{y}_2;(1-\mu_\lambda)/\sigma_\lambda,\sigma_\upsilon^2) \end{pmatrix}.$$

Similarly for $u_n=0$ and $w_n=1$ we have $\hat{\theta}_n=H_0(S_n)$. The function $H_0(S_n)$ is derived from $H_1(S_n)$ by replacing $-\mu_\lambda/\sigma_\lambda$ with $\mu_\lambda/\sigma_\lambda$ in row two and $(1-\mu_\lambda)/\sigma_\lambda$ with $-(1-\mu_\lambda)/\sigma_\lambda$ in row four. Clearly $w_n\xrightarrow{\text{p}}1$ because $\mathrm{Cov}[x,y]\neq 0$. Hence the MME may be written as

$$\hat{\theta}_n = u_n H_1(S_n) + (1-u_n)H_0(S_n) + \mathrm{o_p}(1). \tag{13}$$

Also $u_n\xrightarrow{\text{p}}u$ as $n\to\infty$ where

$$u = \begin{cases} 1 & \text{if } \mathrm{Cov}[x,y]>0, \\ 0 & \text{if } \mathrm{Cov}[x,y]<0. \end{cases}$$

Since S_n converges in probability to its mean $v_\theta = E[S_n]$ and h is continuous we see that $H_u(S_n)\xrightarrow{\text{p}} H_u(v_\theta)$ for $u=0,1$. Therefore from (13) $\hat{\theta}_n\xrightarrow{\text{p}}H_1(v_\theta)=\theta$. Alternatively, if $Cov[x,y]<0$ then $\hat{\theta}_n\xrightarrow{\text{p}}H_0(v_\theta)=\theta$. Hence $\hat{\theta}_n$ is consistent.

Applying the central limit theorem to S_n yields

$$\sqrt{n}(S_n - v_\theta)\xrightarrow{\mathscr{L}}\mathcal{N}(0,\Sigma_\theta),$$

where Σ_θ is the variance matrix of $S=(x,y,x^2,y^2)$. Let

$$M_u = \left.\frac{\partial H_u(S_n)}{\partial S_n}\right|_{S_n=v_\theta}, \quad u=0,1.$$

It is easily verified that $M_0=M_1=M$ where M is defined in (12). Applying a standard Taylor series expansion, i.e., the delta method, and Slutzky's theorem to (13) results in

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{\mathscr{L}}\mathcal{N}(0,V_\theta),$$

where $V_\theta = M\Sigma_\theta M'$. □

4.3. The MME for vertical and horizontal segments

If $\mathrm{Cov}[x,y]=0$ then either δ_x or δ_y equal 0 and the segment has a horizontal or vertical orientation. If both are equal to zero the segment degenerates to a point. In these cases only a subset of system of equations (7) need be considered. For example, if the segment is vertical, i.e., $\delta_x=0$ or equivalently $\eta_x=\xi_x$, then the MME of

$\theta = (\eta_x, \eta_y, \eta_x, \xi_y)$ is given by

$$\hat{\eta}_x = \bar{x},$$

$$\hat{\eta}_y = \bar{y} - \mu_\lambda \hat{\delta}_y^+,$$

$$\hat{\xi}_y = \bar{y} + (1 - \mu_\lambda)\hat{\delta}_y^+,$$

where $\hat{\xi}_x = \hat{\eta}_x = \bar{x}$ and $\hat{\delta}_y = \hat{\delta}_y^+$ by definition. Thus if $s_{yy} > \sigma_\upsilon^2$,

$$\hat{\theta}_n = H(S_n) = \begin{pmatrix} \bar{x}_1 \\ h(\bar{y}_1, \bar{y}_2; -\mu_\lambda/\sigma_\lambda, \sigma_\upsilon^2) \\ \bar{x}_1 \\ h(\bar{y}_1, \bar{y}_2; (1-\mu_\lambda)/\sigma_\lambda, \sigma_\upsilon^2) \end{pmatrix},$$

where S_n and h are defined in the proof of Theorem 2. Similarly the MME assuming $\delta_y = 0$ is given by

$$\hat{\eta}_x = \bar{x} - \mu_\lambda \hat{\delta}_x^+,$$

$$\hat{\eta}_y = \bar{y},$$

$$\hat{\xi}_x = \bar{x} + (1 - \mu_\lambda)\hat{\delta}_x^+$$

and again $\hat{\xi}_y = \hat{\eta}_y = \bar{y}$. If both δ_y and δ_x equal zero we obtain the trivial estimator $\hat{\theta}_n = (\bar{x}, \bar{y}, \bar{x}, \bar{y})$ with well-known properties. The asymptotic distributions of $\hat{\theta}_n$ when δ_x and/or δ_y equal zero is an easy consequence of Theorem 2. These asymptotic distributions are singular. Their exact forms are omitted for brevity.

Unfortunately, the form of the MME depends on the value of δ_x and δ_y which are generally unknown in advance. In the following we introduce a modified version of the MME which fixes this problem.

4.3.1. Modified MME

Suppose $\delta_x = 0, \delta_y \neq 0$ and the general form for $\hat{\theta}_n$, given by (8), is used to estimate θ. It is easily established that

$$\hat{\theta}_n - \theta = \begin{pmatrix} \bar{\varepsilon} - \mu_\lambda \hat{\delta}_x^+ \\ \bar{\upsilon} + \bar{\lambda}\delta_y - \mu_\lambda \hat{\delta}_y^+ \operatorname{sgn}(s_{xy}) \\ \bar{\varepsilon} + (1-\mu_\lambda)\hat{\delta}_x^+ \\ \bar{\upsilon} + (1-\mu_\lambda)\hat{\delta}_y^+ \operatorname{sgn}(s_{xy}) - (1-\bar{\lambda})\delta_y \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \mu_\lambda \delta_y (1 - v) \\ 0 \\ (1-\mu_\lambda)\delta_y (v - 1) \end{pmatrix} + \mathrm{o_p}(1), \tag{14}$$

where $\bar{\varepsilon} = n^{-1}\sum_{i=1}^{n} \varepsilon_i$ and $\bar{\upsilon} = n^{-1}\sum_{i=1}^{n} \upsilon_i$, and $\mathrm{sgn}(s_{xy}) \xrightarrow{\mathrm{p}} v$ as $n \to \infty$, v equals ± 1 with probability 1/2. Moreover, if $v = 1$ it follows from (14) that $\hat{\theta}_n = \theta + \mathrm{o_p}(1)$, otherwise $\hat{\theta}_n = \theta^* + \mathrm{o_p}(1)$, where θ^* is obtained from θ by switching its second and fourth elements, i.e., the points are simply relabeled. Similarly if $\delta_y = 0$. In these cases the general MME is not consistent. Fortunately the remedy to this problem is simple. Define

$$z_{xy} = I_{\{\hat{\delta}_x^+ > 0, \hat{\delta}_y^+ > 0\}} \,\mathrm{sgn}(s_{xy})$$

and let $\hat{\theta}_n^{\mathrm{mod}}$ be as in (8) with z_{xy} replacing $\mathrm{sgn}(s_{xy})$.

Theorem 3. *The modified* MME, $\hat{\theta}_n^{\mathrm{mod}}$, *is asymptotically equivalent to the* MME $\hat{\theta}_n$, *i.e.*, $\hat{\theta}_n^{\mathrm{mod}} = \hat{\theta}_n + \mathrm{o_p}(1)$ *as* $n \to \infty$.

The proof is straightforward in light of the above remarks and the behavior of z_{xy}. Clearly $z_{xy} \xrightarrow{\mathrm{p}} z$ as $n \to \infty$ where

$$z = \begin{cases} 1 & \text{if } \mathrm{Cov}[x, y] > 0, \\ z_0 & \text{if } \delta_x = 0 \text{ and/or } \delta_y = 0, \\ -1 & \text{if } \mathrm{Cov}[x, y] < 0, \end{cases}$$

where $P[|z_0| \leqslant 1] = 1$, which guarantees proper convergence. Hence $\hat{\theta}_n^{\mathrm{mod}}$ naturally accommodates the four possible values for (δ_x, δ_y). Consequently the modified MME $\hat{\theta}_n^{\mathrm{mod}}$ provides consistent asymptotically normal estimation regardless of the value of δ and $\mathrm{Cov}[x, y]$.

4.4. Optimality

Let $z_i' = (x_i, y_i, x_i^2, y_i^2, x_i y_i)$ and denote its mean by v_θ. It is easily shown (Kagan, 1976; Heyde, 1997) that the optimal linear estimating function for θ based on z is

$$\sum_{i=1}^{n} \left(\frac{\partial v_\theta}{\partial \theta}\right)' V_i^{-1}(z_i - v_\theta), \tag{15}$$

where $V_i = \mathrm{Var}[z_i]$. Because V_i are constant for all $i = 1, \ldots, n$, setting (15) to zero and solving for θ reduces to

$$\bar{z} = v_\theta$$

which is equivalent to system (7) used to solve for the MME $\hat{\theta}_n$. Hence $\hat{\theta}_n$ is optimal among all estimators which are based on estimating equations linear in z. Note that estimators based on non-linear functions of z will in general depend on higher order moments of λ which we leave unspecified.

5. Asymptotic relative efficiency

In this section, we compare the asymptotic efficiency of the MME with that of the MLE. The comparison is conducted for four models for the distribution of λ. The models are: (a) $\lambda \sim \mathrm{Bin}(1, 1/2)$; (b) $\lambda \sim U(0, 1)$; (c) $\lambda \sim \mathrm{Beta}(3, 1)$; (d) $\lambda \sim \mathrm{Beta}(3, 3)$. For λ distributed G, the density of (x, y) is given by

$$f(x, y; \theta) = \int_0^1 \phi_x(\lambda; \theta)\phi_y(\lambda; \theta)\, \mathrm{d}G(\lambda),$$

where

$$\phi_x(\lambda; \theta) = \frac{1}{\sigma_\varepsilon}\, \phi\left(\frac{x - \eta_x - \lambda(\xi_x - \eta_x)}{\sigma_\varepsilon}\right),$$

$$\phi_y(\lambda; \theta) = \frac{1}{\sigma_\upsilon}\, \phi\left(\frac{y - \eta_y - \lambda(\xi_y - \eta_y)}{\sigma_\upsilon}\right).$$

Here $\phi(\cdot)$ denotes the standard normal density. The elements of the Fisher information matrix I_θ are

$$I_{ij} = E\left[\frac{\partial f(x, y; \theta)}{\partial \theta_i}\, \frac{\partial f(x, y; \theta)}{\partial \theta_j}\, f^{-2}(x, y; \theta)\right] \qquad \text{for } i, j = 1, \ldots, 4. \tag{16}$$

A closed form for I_θ does not generally exist. The expectation on the RHS of (16) can be computed numerically using Monte-Carlo integration. In particular, $\lambda_1, \ldots, \lambda_N$ are sampled from the distribution G, and the corresponding sample (x_i, y_i), $i = 1, \ldots, N$ is generated according to the models (1)–(2) with $\sigma_\varepsilon = \sigma_\upsilon = 1$ and $\eta = 0$. The expectation (16) is then replaced by its sample average. Note that the partial derivatives $\partial f(x, y; \theta)/\partial \theta_i$ reduce to a simple sum when λ is distributed according to (a). Therefore (16) may be computed directly. Here we employ $N = 4 \times 10^5$ Monte-Carlo replications. Evaluating the partial derivatives $\partial f(x, y; \theta)/\partial \theta_i$ involves additional integration when λ is distributed according to (b), (c) or (d). This is done numerically. For each (x_i, y_i) where $i = 1, \ldots, 10^5$ we use a standard rectangular rule partitioning the interval $[0, 1]$ into 1000 subintervals. Our numerical experiments involving more Monte-Carlo replications and finer integration grids demonstrate that the reported results are accurate up to the third significant digit.

The asymptotic relative efficiency (ARE) of the MME compared with the MLE is defined as

$$\mathrm{ARE} = \left[\frac{\det(I_\theta^{-1})}{\det(V_\theta)}\right]^{1/4}.$$

Thus, the ARE compares the volumes of the concentration ellipsoids of the respective asymptotic distributions. In the following we study the effect of the length and orientation of the segment on the ARE.

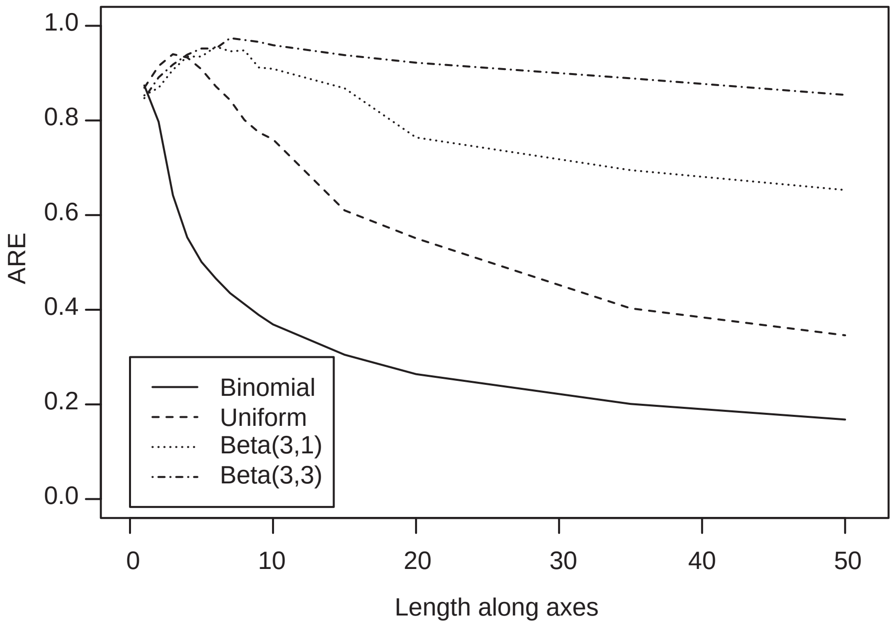

Fig. 1. ARE of the MME relative to the MLE assessed from the Monte-Carlo study as a function of the segment length.

First consider the effect of the length of the segment on the ARE. Let $\delta = \delta_x = \delta_y = 1/2, 1, 2, \ldots, 10, 15, 20, 35, 50$. In the following, we compute the ARE for each length and all four models. The results are summarized in Fig. 1.

The graph shows that for all models the asymptotic relative efficiency decreases as the length of the line segment increases. The decrease is particularly fast for model (a), and hardly noticible for model (d); models (b) and (c) fall in between. However, it is important to realize that the ARE does not provide the full picture. For example, in absolute terms the endpoints of the interval can be most precisely estimated when λ follows a binomial model and δ is large. The likelihood estimates take advantage of the fact that for large lengths, estimating the endpoints is similar to estimating the means of two bivariate normal populations. On the other hand, the MME is ignorant of the underlying distribution and uses only the information on the first two moments. This explains the sharp decrease in the ARE for model (a). Note that when the error variances are fixed, short line segments correspond to low signal-to-noise ratios. Thus, the graph demonstrates that the MME is almost as accurate as the MLE for small signal-to-noise ratios. Clearly these are more difficult cases to estimate.

Next, we study the effect of the segment orientation on the ARE. We consider segments of the length 1 with $\xi = (\cos\alpha, \sin\alpha)$ where the angle α takes values in the set $\pi j/20$, $j = 1, \ldots, 5$. Observe that it suffices to compute the ARE in the range of angles from 0 to $\pi/4$ due to symmetry considerations. The results are reported in Table 1.

Table 1
ARE of the MME relative to the MLE assessed from the Monte-Carlo study as a function of the segment orientation

α	Distribution of λ			
	Biomial	Uniform	Beta(3.3)	Beta(3.1)
$\pi/20$	0.487	0.475	0.470	0.471
$\pi/10$	0.669	0.657	0.649	0.643
$3\pi/20$	0.784	0.771	0.766	0.757
$\pi/5$	0.852	0.829	0.832	0.823
$\pi/4$	0.870	0.848	0.841	0.834

Table 1 indicates that the given the length of the segment the ARE depends primarily on the orientation and not on the distribution G. The explanation of this fact is quite intuitive. Recall that the MME is based on estimated projections of the line segment on the axes. When the segment is nearly parallel to one of the axes the corresponding projection is small compared with the standard deviation of the noise in that direction. Consequently the corresponding coordinates are estimated less accurately.

6. Discussion

6.1. Model assumptions

In this paper, we fit the structural segment model to noisy data. Our method does not require full distributional assumptions on λ. In contrast, the standard maximum likelihood approach can be implemented when G is known up to a finite number of parameters. However likelihood assumptions may be restrictive at times. Moreover, even under standard uniformly distributed $\{\lambda_i\}$ (Werman and Keren, 2001), the resulting estimating equations are difficult to solve. Alternatively, consistent estimators of the endpoints can be constructed under nonparametric assumptions on G. The densities of the random variables $\lambda\xi_x + (1-\lambda)\eta_x$ and $\lambda\xi_x + (1-\lambda)\eta_x$ can be estimated nonparametrically from the data. In fact, this is the standard density deconvolution problem, e.g., Carrol and Hall (1988) and Fan (1991). The estimate of θ is defined as the endpoints of the corresponding density estimates. Other nonparametric estimators based on the order statistics may be developed. However, it can be shown that these estimators converge at a very slow, logarithmic, rate. These results will be described elsewhere. Thus the MME is a useful compromise between fully specified likelihood and the completely nonparametric methods.

We study the MME under the assumption that the first two moments of G, and $\sigma_\varepsilon^2, \sigma_\upsilon^2$ are known. Using our assumptions we get parametric, root n, convergence and easily solvable estimating equations. It is worth noting also that the MME can be derived under other assumptions. For example it may be reasonable in applications to assume that: (1) $\mu_\lambda, \sigma_\lambda^2$ are known and the ratio of the variances $\rho = \sigma_\varepsilon^2/\sigma_\upsilon^2$ is known;

(2) $\mu_\lambda, \sigma_\lambda^2$ are known and one of the variances $\sigma_\varepsilon^2, \sigma_\upsilon^2$ is known; or (3) $\mu_\lambda, \sigma_\varepsilon^2, \sigma_\upsilon^2$ are known and the length of the line segment $\sqrt{\delta_x^2+\delta_y^2}$ is known. These assumptions define different structural parameters and lead to moments estimators which differ from those considered in this paper. Nevertheless similar analysis can be performed.

6.2. Inference

Some comments regarding inference for segments are in order. We start with the behavior of the estimator $\hat{\delta}_x$. Note that $\hat{\delta}_x = 0$ if $s_{xx} \leqslant \sigma_\varepsilon^2$. This event occurs with positive probability for all values of δ_x. Define $p_n = P[s_{xx} > \sigma_\varepsilon^2]$ to be the probability that $\hat{\delta}_x$ is positive. Note that

$$s_{xx} = \delta_x^2 s_{\lambda\lambda} + s_{\varepsilon\varepsilon} + 2\delta_x s_{\lambda\varepsilon},$$

where $s_{\lambda\lambda}$, $s_{\varepsilon\varepsilon}$, and $s_{\lambda\varepsilon}$ are defined in the obvious way. Suppose $\delta_x = 0$. Then s_{xx} is distributed as $\sigma_\varepsilon^2 \chi_{n-1}^2/n$. A straightforward calculation shows that p_n increases to $\frac{1}{2}$ as $n \to \infty$. Note that $\delta_x = 0$ implies that the distribution of s_{xx} is asymptotically symmetric around σ_ε^2. Hence $\hat{\delta}_x$ will be, mistakenly, positive roughly half of the time. Its value however converges in probability to zero. If $\delta_x \neq 0$ then

$$p_n = P[\delta_x^2 s_{\lambda\lambda} + \sigma_\varepsilon^2(\chi_{n-1}^2/n - 1) > 0] + \mathrm{O}(n^{-1})$$

since $s_{\lambda\varepsilon} \xrightarrow{\mathrm{p}} 0$. Moreover $\chi_{n-1}^2/n \xrightarrow{\mathrm{p}} 1$ and $s_{\lambda\lambda} \xrightarrow{\mathrm{p}} \sigma_\lambda^2$ imply that $p_n \to 1$. Consider testing the hypothesis $\mathrm{H}_0 : \delta_x = 0$ based on a sample of size n. If $\hat{\delta}_x = \delta_{xx}^* > 0$ the corresponding p-value is

$$P\left[\chi_{n-1}^2 \geqslant \frac{n\delta_{xx}^*}{\sigma_\varepsilon^2}\right]$$

which decreases quickly as $n \to \infty$. If however $\hat{\delta}_x = 0$ for sufficiently large n the conclusion that $\delta_x = 0$ has at most probability half. Similar conclusions hold for $\hat{\delta}_y$.

It is therefore sensible to provide inferences conditional on $(\hat{\delta}_x^+, \hat{\delta}_y^+)$. This amounts to using $\hat{\theta}_n^{\mathrm{mod}}$ to estimate θ. If both unsigned lengths are positive then inference can be based on the asymptotic distribution given in Theorem 2. If either or both are zero then the asymptotic distribution should be patterned along Section 4.3. The asymptotic distribution of the MME $\hat{\theta}_n$ depends, through Σ_θ, on the first four moments of λ. If the distribution G is completely specified, its moments are known and can be substituted directly to estimate Σ_θ. If only the first two moments are specified in advance one can consistently estimate Σ_θ by

$$\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_i - \bar{x}_1 \\ y_i - \bar{y}_1 \\ x_i^2 - \bar{x}_2 \\ y_i^2 - \bar{y}_2 \end{pmatrix}(\, x_i - \bar{x}_1, \quad y_i - \bar{y}_1, \quad x_i^2 - \bar{x}_2, \quad y_i^2 - \bar{y}_2 \,).$$

Given its large sample distribution inferences regarding θ are carried out in a standard way.

Acknowledgements

We thank Dr. Danny Keren for motivating our interest in this problem, and the referees for useful comments.

References

Anderson, D.A., 1981. The circular structural model. J. Roy. Statist. Soc. Ser. B 43, 131–141.

Berman, M., Culpin, D., 1986. The statistical behavior of some least-squares estimators for the center and radius of a circle. J. Roy. Statist. Soc. Ser. B 48, 183–196.

Carrol, R.J., Hall, P., 1988. Optimal rates of convergence for deconvolving a density. J. Amer. Statist. Assoc. 83, 1184–1186.

Chan, N.N., 1965. On circular functional model. J. Roy. Statist. Soc. Ser. B 27, 45–56.

Cheng, C.L., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Oxford University Press, New York.

Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19, 1257–1272.

Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.

Haralick, R.M., Shapiro, L.G., 1992. Computer and Robot Vision. Addison-Wesley, Reading, MA.

Heyde, C.C., 1997. Quasi-Likelihood and its Application. Springer, New York.

Kagan, A.M., 1976. The information contained in a finite dimensional linear space and a correctly posed version of the method of moments. Problems of Information Transmission 12, 20–42.

Kendall, M.G., Stuart, A., 1979. The Advanced Theory of Statistics, Vol. 2, 4th Edition. Griffin, London.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, 2nd Edition. Springer, New York.

Madsen, S.B., Christensen, H.I., 1995. Modeling and testing the stability of edge segments: length and orientation. In: Borgefors, G. (Ed.), Theory and Application in Image Processing II—Selected papers from the Ninth Scandinavian Conference on Image Analysis. World Scientific Press, Singapore, pp. 1–15.

Pfanzagl, J., 1993. Incidental versus random nuisance parameters. Ann. Statist. 21, 1663–1691.

Stewart, C.V., Bubna, K., Perera, A., 1999. Estimating model parameters and boundaries by minimizing adjoint robust objective function. In: Proceedings of the 1999 IEEE Conference on Computer Vision and Pattern Recognition, pp. 387–393. Fort Collins Colorado June 22–25 1999.

Vosselman, J., Haralick, R.M., 1996. Performance analysis of line and circle fitting in digital images. In: Proceedings of the Workshop on Performance Characteristics of Vision Algorithms, Cambridge, UK.

Werman, M., Keren, D., 2001. A bayesian method for fitting parametric and nonparametric models to noisy data. IEEE Trans. Pattern Anal. Mach. Intell. 23, 528–534.