ON NONPARAMETRIC ESTIMATION OF INTERCEPT AND SLOPE DISTRIBUTIONS IN RANDOM COEFFICIENT REGRESSION

By Rudolf Beran,[1] Andrey Feuerverger[2] and Peter Hall[2]

University of California, Berkeley, University of Toronto and Australian National University

An experiment records stimulus and response for a random sample of cases. The relationship between response and stimulus is thought to be linear, the values of the slope and intercept varying by case. From such data, we construct a consistent, asymptotically normal, nonparametric estimator for the joint density of the slope and intercept. Our methodology incorporates the radial projection-slice theorem for the Radon transform, a technique for locally linear nonparametric regression and a tapered Fourier inversion. Computationally, the new density estimator is more feasible than competing nonparametric estimators, one of which is based on moments and the other on minimum distance considerations.

1. Introduction. Consider the following experimental situation. For each case in a certain population, it is possible, in principle, to measure a response Y and a covariate X. These two measurements are thought to be linked through a linear relation $Y = A + BX$. The values of the intercept A and the slope B *depend on the case* and are unknown. Data are obtained by measuring (X, Y) for each of n cases that are drawn at random from the population. Thus, the data are in the form of independent pairs (X, Y) generated by the model $Y = A + BX$, where (A, B) is a random vector independent of X. We wish to estimate nonparametrically the joint density $f_{A, B}$ of (A, B) and possibly also the marginal densities.

Random coefficient linear regression models such as the one just described are discussed by Longford (1993). They occur naturally in econometric sampling, particularly in studies of panel data [cf. Raj and Ullah (1981) and Nicholls and Pagan (1985)]. In such contexts, Y and A may be vectors while B and X may then be matrices. A simple parametric assumption is that A and B are independent and Normally distributed, although of course the assertion of independence may be relaxed. Close statistical cousins are the random effects models of analysis of variance, in which, however, X is not random. Analogous nonlinear random coefficient regression models arise in analyzing patient response data from pharmaceutical trials. Here, for each

patient, the measured response satisfies $Y = g(X, C)$, where X is the drug dosage, g is a known function determined by pharmokinetic models and C is a vector of coefficients that varies with patient [cf. Liu (1994)]. If the patients form a random sample, then it is reasonable to regard C as a random coefficient vector independent of X.

Nonparametric estimators for the density or distribution of (A, B) are a recent development. To understand why, observe that the conditional density of Y given $X = x$ is

$$(1.1) \qquad f_{Y|X}(y|x) = \int_{-\infty}^{\infty} f_{A, B}(y - bx, b) \, db.$$

This says that the conditional density of Y given X is a Radon transform of the density $f_{A, B}$ [cf. Deans (1983)]. Thus, under tail assumptions on $f_{A, B}$, it is possible mathematically to recover $f_{A, B}$ as an inverse Radon transform of $f_{Y|X}$. While this observation helps to explain the nature of our problem, it does not directly yield a consistent estimator of $f_{A, B}$. We do not know $f_{Y|X}$ exactly, though it is estimable from our data. Unfortunately, the inverse Radon transform is highly sensitive to perturbations of its argument and lacks continuity in familiar metrics on conditional density estimators.

Problems of tomographic reconstruction have inspired effective algorithms for numerical inversion of Radon transforms in situations where the function that plays the role of $f_{Y|X}$ is measured accurately on a fine grid of (x, y) values. For example, see Devaney (1989), Shepp and Kruskal (1979) and Silverman, Jones, Wilson and Nychka (1990), and additional references cited there. Because the errors in estimating the conditional density $f_{Y|X}$ are not of this nature, the tomographic algorithms are not directly applicable to our problem.

Let $\phi_{Y|X}(\cdot|x)$ and $\phi_{A, B}$ be the characteristic functions of $f_{Y|X}(\cdot|x)$ and $f_{A, B}$, respectively. It follows from the model $Y = A + BX$ that

$$\phi_{Y|X}(t|x) = \phi_{A, B}(t, tx).$$

A variant of this identity, after changing variables to a particular polar-coordinate system, is the basis for our nonparametric estimator of $f_{A, B}$. The identity corresponds to the *projection-slice theorem* of Radon transform theory, which states: the one-dimensional Fourier transform (in the radial variable) of the Radon transform of a two-dimensional function is identical to the two-dimensional Fourier transform of that function.

Before describing the methodology of this paper, we briefly review two recent nonparametric methods for estimating the distribution of (A, B). The *conditional moment method* is suitable for the special case when A and B are independent. Then, letting $\alpha_k = E(A^k)$ and $\beta_k = E(B^k)$,

$$(1.2) \qquad E(Y^r|X) = \sum_{k=0}^{r} \binom{r}{k} \alpha_{r-k} \beta_k X^k.$$

This equation suggests a recursive procedure for estimating the moments of A and B. First, note that $\alpha_0 = \beta_0 = 1$. Use this fact, plus least-squares regression of Y on $(1, X)$ based on the data, plus (1.2) with $r = 1$, to deduce estimates of α_1 and β_1. Next, use the estimates just obtained, plus least-squares regression of Y^2 on $(1, X, X^2)$, plus (1.2) with $r = 2$, to deduce estimates of α_2 and β_2. Continue in this fashion to estimate the first m moments of A and B, where m is much smaller than the sample size n. From these moment estimators, we can devise consistent estimators for the marginal cumulative distribution functions or densities of A and B. For details of such an approach, see Beran and Hall (1992). In numerical practice this method suffers from round-off errors in estimation of the higher-order moments. Moreover, extending the conditional moment approach to the case where A and B are not independent seems impracticable.

The *minimum distance method* fits the empirical distribution of (X, Y) to the joint distribution of (X, Y) that is determined by the random coefficient regression model. More specifically, let $F_{A,B}$ and F_X be the distributions of (A, B) and X, respectively. Let $P(F_{A,B}, F_X)$ then denote the distribution of (X, Y) under the model $Y = A + BX$. Let $\hat{F}_X$ and $\hat{P}$ be the empirical distributions of X and (X, Y), based on the data. Finally, let d be a distance that metrizes weak convergence of distributions in R^2. A minimum distance estimator $\hat{F}_{A,B}$ for the distribution of (A, B) is then

$$(1.3) \qquad \hat{F}_{A,B} = \arg \inf_{F_{A,B}} d\left[\hat{P}, P\left(F_{A,B}, \hat{F}_X\right)\right].$$

These minimum distance estimators are consistent in the sense of weak convergence. They generalize readily to the case where A and B are not univariate and even further to nonlinear random coefficient regression models. In principle, they also provide a starting point for constructing estimators of the density $f_{A,B}$. For implementation of (1.3), it is convenient to let d be an L^2-metric acting on characteristic functions. See Beran and Millar (1994) for further details in the linear case and Liu (1994) for nonlinear extensions. Unfortunately, numerical approximation of a discretized version of the minimum distance estimator $\hat{F}_{A,B}$ is often plagued by numerous relative minima on the right-hand side of (1.3). These relative minima, occurring as they do in a space of high dimension, challenge simulated annealing, Nelder–Mead and other standard minimization techniques [cf. Liu (1994)].

Estimation techniques such as those just described implicitly require assumptions that ensure the identifiability of the distribution of (A, B). For example, we may require that the distribution of (A, B) have compact support and that the support of the distribution of X contain a cluster point. Or, as in this paper and in parametric treatments of random coefficient regression, we may require that f_X have full support, in which case no condition is needed on the support of $f_{A,B}$. That such assumptions cannot be avoided entirely is apparent from the conditions needed for existence of the inverse to the Radon transform (1.1).

The search for a good, numerically feasible way of estimating the density or distribution of (A, B) thus continues. Our methodology in this paper is based on the following observation: with $\Theta = \arctan X$ and $R = Y(1 + X^2)^{-1/2} = Y \cos \Theta$,

$$R = (A, B) \cdot (\cos \Theta, \sin \Theta).$$

Therefore, if $\phi_{R|\Theta}$ denotes the characteristic function of R conditional on Θ, then the joint characteristic function $\phi_{A,B}(u, v)$ of A and B is given by

$$(1.4) \qquad \phi_{A,B}(r \cos \theta, r \sin \theta) = \phi_{R|\Theta}(r|\theta)$$

for $-\infty < r < \infty$ and $-\pi/2 < \theta \leq \pi/2$. Note that $\phi_{R|\Theta}(\cdot | -\pi/2)$ equals the complex conjugate of $\phi_{R|\Theta}(\cdot | \pi/2)$, and that $\phi_{R|\Theta}(r|\theta) = E\{\exp(irR) \mid \Theta = \theta\}$, so that estimating $\phi_{R|\Theta}$ is a problem of nonparametric regression. Furthermore, the conditional density of R given Θ is precisely the Radon transform of $f_{A,B}$, and so (1.4) is then precisely the *projection-slice theorem*.

We may estimate $\phi_{R|\Theta}$ by any of several nonparametric methods—the approach discussed in Section 2 is an adaptation of local linear regression. Alternative approaches such as Nadaraya–Watson kernel estimation could be employed and would produce very similar results. However, depending on the method, the formula for the bias contribution to the estimator for the joint density could be considerably more complex than that given here. (In particular, the formula is more complex in the case of Nadaraya–Watson methods). Having obtained an estimator of $\phi_{R|\Theta}$, (1.4) suggests an approximation to $\phi_{A,B}$ by $\hat{\phi}_{A,B}$, say. The latter Fourier transform may be inverted to produce an approximation to the joint density $f_{A,B}$.

The convergence rate of our estimator of $f_{A,B}$ depends heavily on the behavior of the characteristic function $\phi_{R|\Theta}$ in the tails. This is to be expected, given similar results in more straightforward deconvolution problems [see, e.g., Carroll and Hall (1988) and Fan (1991)], and is not an artifact of our methodology or our mathematical analysis. One could expect, given the work of Carroll and Hall and of Fan, that our methods would produce optimal or near-optimal convergence rates. This topic will not be taken up in the present paper, however.

Finally, we note that identity (1.4) has extensions to random coefficient regression models with several covariates. Consider, for instance, the model $Y = A + BX_1 + CX_2$, where (A, B, C) is independent of (X_1, X_2). Let $R = Y(1 + X_1^2 + X_2)^{-1/2}$ and let (Θ, Φ) denote the standard polar coordinates of the unit vector

$$\left(X_1(1 + X_1^2 + X_2^2)^{-1/2}, X_2(1 + X_1^2 + X_2^2)^{-1/2}, (1 + X_1^2 + X_2^2)^{-1/2} \right).$$

Because the third component of this vector is positive, the range of Θ is $[0, \pi/2]$ while the range of Φ is $[0, 2\pi)$. The joint characteristic function $\psi_{A,B,C}(u, v, w)$ of (A, B, C) is given by

$$(1.5) \qquad \psi_{A,B,C}(r \cos \theta, r \sin \theta \cos \varphi, r \sin \theta \sin \varphi) = \psi_{R|\Theta,\Phi}(r|\theta, \phi)$$

for $-\infty < r < \infty$, $0 \le \theta \le \pi/2$ and $0 \le \phi < 2\pi$. The characteristic function of $\psi_{A,B,C}$ may thus be recovered by estimating the conditional characteristic function of R given (Θ, Φ) and then applying (1.5). This paper does not pursue the case of multiple covariates any further.

Sections 2 and 4 describe, in fuller detail, the new method and conditions under which it yields consistent, asymptotically normal estimates for $f_{A,B}$. Numerical algorithms are the subject of Section 3. We find that the new estimators are computationally feasible and more trustworthy numerically than their predecessors, reviewed above. Section 5 contains proofs. On the technical side, our use of local linear regression in a polar-coordinate context has novel aspects.

2. Methodology.

2.1. *Estimation of $\phi_{R|\Theta}$.* The observed data (X_i, Y_i) are assumed generated by the model $Y_i = A_i + B_i X_i$, $1 \le i \le n$, where the pairs (A_i, B_i) are independent and identically distributed as (A, B), and are independent also of the independent and identically distributed design sequence $\{X_1, \ldots, X_n\}$. We transform the observed data to pairs (Θ_j, R_j), where $\Theta_j = \arctan X_j$ and $R_j = Y_j(1 + X_j^2)^{-1/2}$.

Let K_κ, the kernel in our local linear regression, denote a function that is periodic with period π, is even on $(-\pi/2, \pi/2]$ and has its mode in that interval at $\theta = 0$. The kernel should be chosen so that increasing κ makes the mode more pronounced, and such that (after normalization) K_κ converges to the Dirac delta "comb" function (centered at the origin) as $\kappa \to \infty$. In our theoretical development we shall take

$$(2.1) \qquad K_\kappa(\theta) = K\{\kappa(1 - |\cos\theta|)\},$$

where the nonnegative function K is supported on the positive half-line. In this context, $\kappa > 0$ functions like the inverse of the square of a bandwidth. We could multiply K_κ by a normalizing constant $C(\kappa)$, chosen to ensure that K_κ integrates to 1 on $(-\pi/2, \pi/2]$, but then $C(\kappa)$ would cancel from the quantities that we consider or be incorporated into other constants. When $K(u) = \exp(-u)$ the function at (2.1) is closely related to the density of the von Mises distribution, restricted here to $(-\pi/2, \pi/2]$ rather than $(-\pi, \pi]$. The relationship is perhaps even closer if instead of (2.1) we define $K_\kappa(\theta) = K[\kappa\{1 - |\cos(2\theta)|\}]$, for which straightforward analogs of our main results also hold.

Write $\theta_1 \ominus \theta_2$ for $\theta_1 - \theta_2$ if this quantity lies in the interval $(-\pi/2, \pi/2]$, and for $\theta_1 - \theta_2 + \pi$ (respectively, $\theta_1 - \theta_2 - \pi$) if $\theta_1 - \theta_2 \le -\pi/2$ ($> \pi/2$). (This definition of difference is the natural one for arithmetic on the interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi]$, mod π.) Define

$$s_\kappa(\theta) = \sum_{j=1}^{n} K_\kappa(\theta - \Theta_j)(\theta \ominus \Theta_j)^k,$$

$$u_j(\theta) = K_\kappa(\theta - \Theta_j)\{s_2(\theta) - (\theta \ominus \Theta_j)s_1(\theta)\}$$

and

$$v_j(\theta) = u_j(\theta) \Big/ \left\{ \sum_{l=1}^{n} u_l(\theta) + \eta \right\},$$

where $\eta \geq 0$ denotes a ridge parameter. These will be the weights in our periodic version of local linear regression. Similar weights in more conventional settings have been discussed by Fan (1992, 1993), for example. We shall take $\eta = 0$, although identical results are obtainable in many other cases; their proofs follow lines established in Hall, Marron, Neumann and Titterington (1994).

Our estimator of $\phi_{R|\Theta}$ is

$$(2.2) \qquad \hat{\phi}_{R|\Theta}(r|\theta) = \sum_{j=1}^{n} v_j(\theta) \exp\{irR_j S_j(\theta)\},$$

where $S_j(\theta) = \pm 1$ according as $\theta \ominus \Theta_j$ equals $\theta - \Theta_j$ or not, respectively, and i denotes the square root of -1. [The factor $S_j(\theta)$ stems from the Möbius-like character of the conditional distribution of R, which tends in distribution to $\pm B$ according as $\theta \to \pm \pi/2$, respectively. This factor also ensures that all arithmetic is conducted strictly modulo π; without it, $\hat{\phi}_{R|\Theta}$ is not consistent for $\phi_{R|\Theta}$.] Our estimator of $\phi_{A,B}(u,v)$ is, for $u \neq 0$,

$$\hat{\phi}_{A,B}(u,v) = \hat{\phi}(u,v) \equiv \hat{\phi}_{R|\Theta}\big\{(u^2 + v^2)^{1/2}\mathrm{sgn}\, u \mid \arctan(v/u)\big\}.$$

2.2. *Density estimation.* With $\hat{\phi}_{R|\Theta}$ given by (2.2), our estimator of the joint density $f_{A,B}$ of A and B is

$$\hat{f}_{A,B}(a,b) = (2\pi)^{-2} \int\int \exp(-iau - ibv)\, \hat{\phi}_{A,B}(u,v) w\big\{(u^2 + v^2)^{1/2}\big\}\, du\, dv$$

$$(2.3) \qquad = (2\pi)^{-2} \int_{-\pi/2}^{\pi/2} d\theta \int \exp(-iar\cos\theta - ibr\sin\theta)$$

$$\times \hat{\phi}_{R|\Theta}(r|\theta) w(r)|r|\, dr,$$

where (here and below) unqualified integrals are over the entire real line. The taper, or weight function, w is introduced to ensure integrability. In general, one might allow w to depend on θ as well as r, and to enjoy elliptical-like contours in (r, θ) space. A simplified approach would be to ask that w be univariate and symmetric, and that for constants $0 < \tau_1 < \tau_2 < \infty$, both of them diverging to ∞ with n and typically $O(1)$ apart,

$$(2.4) \qquad w(u) \begin{cases} = 1, & \text{if } |u| < \tau_1, \\ \in [0,1], & \text{if } |u| \in (\tau_1, \tau_2], \\ = 0, & \text{if } |u| > \tau_2. \end{cases}$$

In practice, τ_1 and τ_2 would be functions of the data. For the purpose of the theory developed in Section 4 we shall, to simplify both our discussion and the technicalities, take $\tau_1 = \tau_2 = \tau$, say, to be a deterministic function.

Estimators of the marginal densities f_A and f_B are obtainable by inverting $\hat{\phi}_{A,B}(u, 0)$ and $\hat{\phi}_{A,B}(0, v)$, respectively:

$$(2.5a) \qquad \hat{f}_A(a) = (2\pi)^{-1} \int \exp(-iau)\, \hat{\phi}_{R|\Theta}(u \mid 0)\, w(u)\, du,$$

$$(2.5b) \qquad \hat{f}_B(b) = (2\pi)^{-1} \int \exp(-ibv)\, \hat{\phi}_{R|\Theta}(v|\pi/2)\, w(v)\, dv.$$

Alternatively, recognizing that f_A and f_B are the densities of R given $\Theta = 0$ and R given $|\Theta| = \pi/2$, respectively, one may estimate them using more traditional kernel methods, not involving Fourier inversion.

Note particularly that, since w is symmetric, both the joint density estimator [defined at (2.3)] and the marginal density estimators [at (2.5)] are assuredly real valued. However, they generally take negative values over a portion of their domains.

3. Numerical properties. The procedures discussed in Section 2 were implemented on an SGI Challenge computer using the S-PLUS statistical software package (version 3.3). See Becker, Chambers and Wilks (1988). In particular, we implemented the local linear regression algorithm at (2.2) to compute estimates of the characteristic function $\phi_{A,B}$, and then tapered, and finally inverted this estimated function using a two-dimensional fast Fourier transform (FFT) algorithm to obtain the bivariate density estimates for (A, B).

We carried out our computations over rectangular grid structures, and found, as expected, that the choice of tapering function was of considerable importance in numerical work. After some experimentation we decided upon the following class of tapering functions:

$$(3.1) \qquad \begin{aligned} w(r) &= \exp\left\{c\left[1 - \frac{1}{1 - (r/r_1)^2}\right]\right\} \\ &= \exp\left\{-c\sum_{j=1}^{\infty}\left(\frac{r}{r_1}\right)^{2j}\right\}, \qquad |r| < r_1, \end{aligned}$$

and $= 0$ otherwise, with obvious modifications in the event that elliptical-like contours (in θ) were desired. A typical plot of this function and its Fourier transform is shown in Figure 1(a) and (b).

Some of the data-analytic difficulties associated with deconvolution may be alleviated by ad hoc modeling of the characteristic function $\phi(r|\theta)$ in the tails, where it decreases to 0. We have not done that here, as it detracts
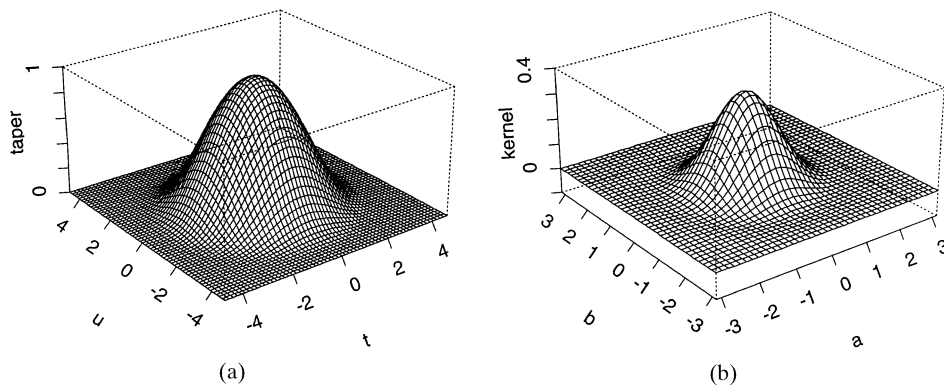
FIG. 1.   (a) *Plot of the taper in* (3.1) *with circular contours, and* $r_1 = 5$, $c = 4$. (b) *Fourier transform of the taper in* (a).

a little from our use of numerical examples to illustrate our theoretical points—such modeling ideas are difficult to quantify using the asymptotic development in this paper.

In our numerical work we used the same functional form for $K_\kappa(\cdot)$; specifically,

$$(3.2) \qquad K_\kappa(\theta') = \exp\left\{c\left[1 - \frac{1}{1 - \{\theta'/h(\theta)\}^2}\right]\right\}$$

for $|\theta'| < h(\theta)$, where $h(\theta) = \{2/\kappa(\theta)\}^{1/2}$. After some experimentation we took

$$(3.3) \qquad h(\theta) = c_1\left(1 + c_2\frac{1 - \cos(2\theta)}{2}\right)^\delta$$

for positive constants c_1, c_2 and δ.

Suppose now that we wish to use n_1 points for the t-dimension and n_2 points for the u-dimension in an FFT algorithm. If t and u are taken to range over $\pm T$ and $\pm U$, respectively, then the corresponding spacings will be $\Delta t = 2T/n_1$ and $\Delta u = 2U/n_1$. Specifically then, suppose that estimates of $\phi(-T + j\Delta t, -U + k\Delta u)$, for integers $1 \leq j \leq n_1$ and $1 \leq k \leq n_2$, will be available to use in an FFT algorithm.

Discrete approximation to

$$(3.4) \qquad f(a, b) = \left(\frac{1}{2\pi}\right)^2 \int \int \phi(t, u)\exp(-ita - iub)\, dt\, du$$

leads immediately to

$$f(-a' + j\,\Delta a, -b' + k\,\Delta b)$$

$$\approx \left(\frac{1}{2\pi}\right)^2 \frac{1}{\Delta t}\frac{1}{\Delta u}\sum_{t=1}^{n_1}\sum_{u=1}^{n_2}\left[\phi\left(\Delta t\left(-\frac{n_1}{2}+t\right), \Delta u\left(-\frac{n_2}{2}+u\right)\right)\right.$$

(3.5)
$$\times \exp\left\{-2\pi i\left[\frac{(-(n_1/2)+j)(-(n_1/2)+t)}{n_1}\right.\right.$$

$$\left.\left.\left.+\frac{(-(n_2/2)+k)(-(n_2/2)+u)}{n_2}\right]\right\}\right],$$

where $\Delta a = \pi/T$ and $\Delta b = \pi/U$, corresponding to ranges $\pm a'$ and $\pm b'$, where $a' = n_1\pi/2T = \pi/\Delta t$ and $b' = n_2\pi/2U = \pi/\Delta u$ for the a and b dimensions, respectively. Using standard methods in order to appropriately "center" the transform, the expression (3.5) may be evaluated by means of the two-dimensional FFT.

Figure 2 shows the results of our computations for a simulation based on 1000 data points (X_j, Y_j) and standard Cauchy X_j's. The distribution for the
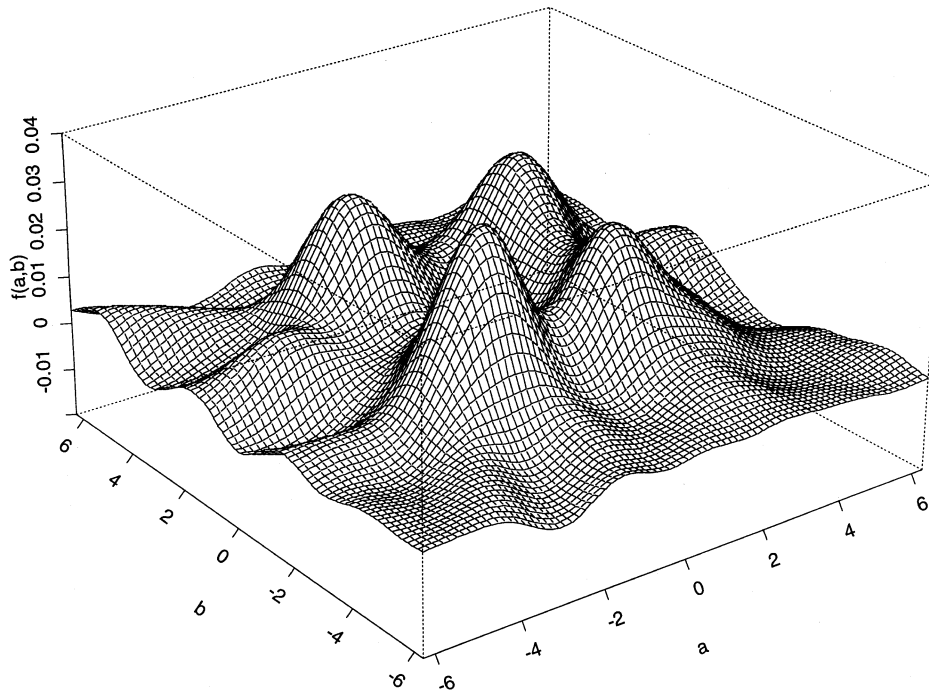

FIG. 2. *Estimated density $f_{A,B}$ using $n = 1000$ standard Cauchy-distributed X's and a mixture distribution for (A, B) consisting, in equal proportions, of four bivariate normal distributions having unit variances, zero covariances and means $(\pm 2, \pm 2)$.*

(A_j, B_j)'s here was chosen to be a mixture, in equal proportions, of four bivariate normal distributions having unit variances, zero covariances and means given by the four vectors $(\pm 2, \pm 2)$. We used $h(\theta) = 0.2$ in (3.2) and the taper function (3.1) with $c = 4$ and $r_1 = 5$. We also employed $n_1 = 256$, $n_2 = 256$, $T = 20$ and $U = 20$, so that $\Delta t = 0.156$, $\Delta u = 0.156$, $\Delta a = 0.157$, $\Delta b = 0.157$, $A = 20.1$ and $B = 20.1$. The total computing time required was 1.04 minutes. Figure 2 shows the resulting estimated bivariate density function, which has four clearly distinguishable modes. Of course, the combination of the substantial sample size here and the (optimal) uniform spacing in θ makes this estimation problem relatively easy.

Our second simulation example involves a sample of size 200 and a normal distribution with mean 0 and standard deviation 2 for the X_j's. The (A_j, B_j)'s were chosen to be standard bivariate normal. The very light tails of the X's and hence of the Θ's makes this estimation problem relatively difficult. The estimated density for (A, B) shown in Figure 3 was produced using the bandwidth function (3.3) with $c_1 = 0.25$, $c_2 = 2$ and $\delta = 1$ and the taper function (3.1) with $c = 4$ and $r_1 = 5$. We also used $n_1 = 256$, $n_2 = 256$, $T = 20$ and $U = 20$ as before. The total computing time required was 28 seconds.
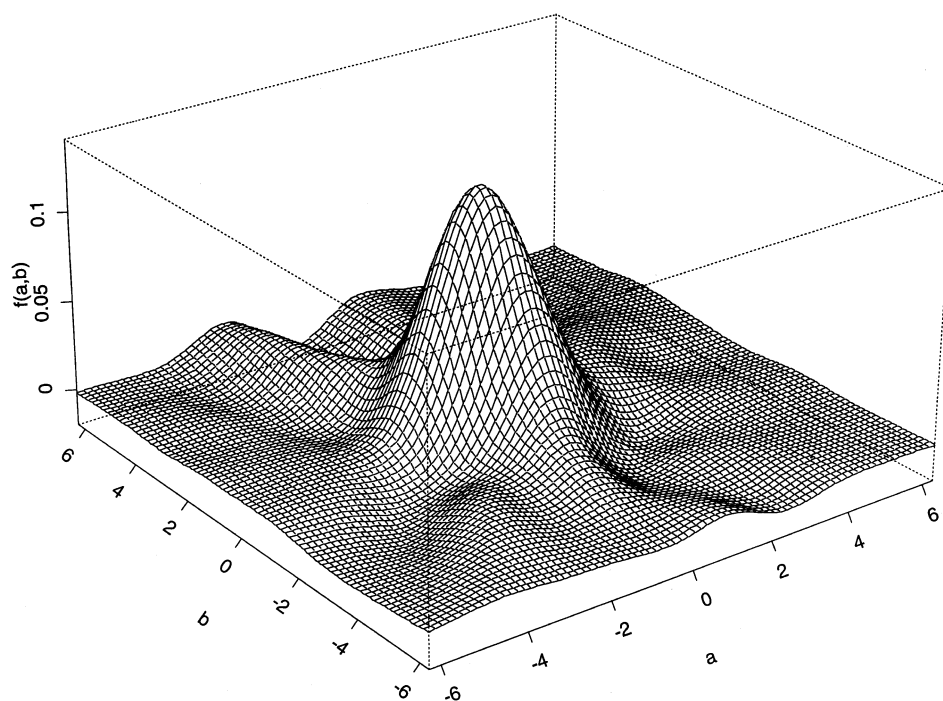

FIG. 3.   *Estimated density $f_{A, B}$ using $n = 200$ normal X's with mean 0 and standard deviation 2 and standard bivariate normal (A, B)'s.*

Finally, Figure 4(a)–(d) provides the results of an analysis of some real data. The data used here were Galton's tabulation of the heights of 928 adult children (the Y's) by averages of the heights of their two parents (the X's), as quoted in Stigler (1986), Table 8.1, page 286. It might be thought that such data could be modeled via a random coefficient regression. As these data are recorded on a 1-inch grid, we used the interval midpoints for our X and Y values as reasonable approximations for present purposes. Furthermore, we standardized both our X and Y values prior to analysis, and also regressed out Y on X so as to center the B distribution. (These operations could be recommended quite generally). We chose the bandwidth in (3.2) to be $h(\theta) = 0.7$, in order to accommodate the fairly wide discrete spacing of the Θ values that occur for these data, and used the taper function (3.1) with $c = 4$ and $r_1 = 10$. Figure 4(a) shows the resulting joint density estimate, while Figure 4(b) and (c) shows the marginal densities of A and B obtained by numerical summation. Finally, Figure 4(d) shows a cross section of the marginal of the Fourier transform of the tapering function used. Figure 4(a) [and also com-
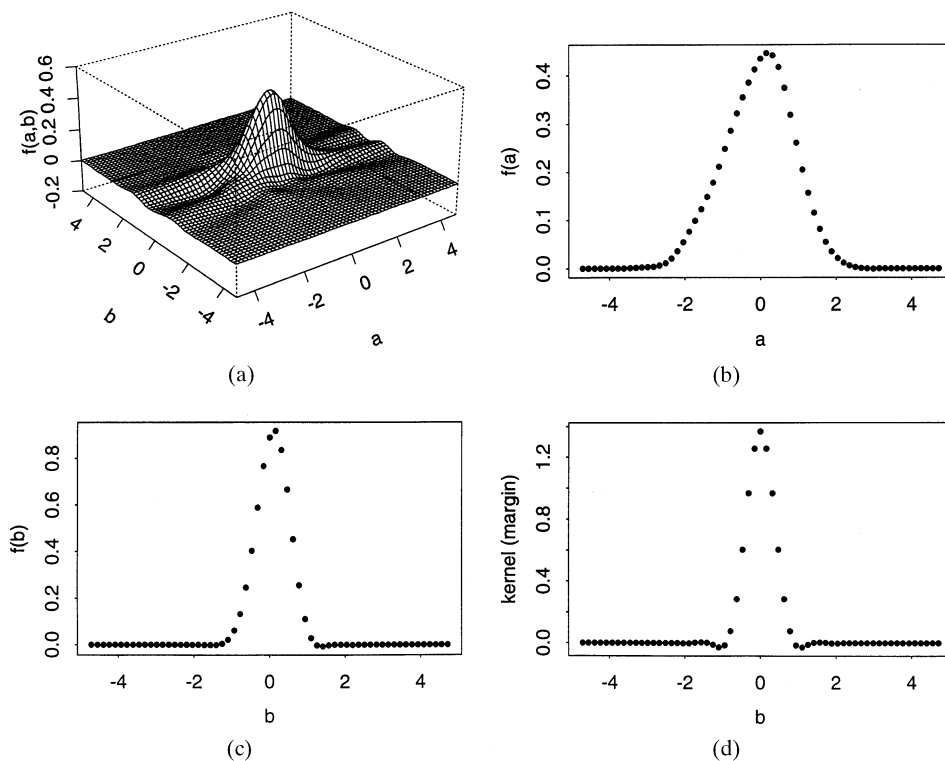

FIG. 4. (a) Density estimate for (A, B) in Galton's data. (b) Marginal density for A. (c) Marginal density for B. (d) Cross section of Fourier transform of taper used in analysis of Galton's data.

parison of Figure 4(c) and (d)] suggests that, for these data, B is degenerate or nearly so. We therefore conclude that, for these data, B is essentially constant, and that the random coefficient model fit to these data is in fact a classical regression with a (convolved) error distribution (as estimated by our procedure) given in Figure 4(b). It would seem that the population Galton sampled was genetically homogeneous.

4. Theory. Before developing technical details we outline the form of our main results. We shall show that the variance contribution to the mean-squared error of $\hat{f}_{A,B}(a, b)$ consists of two *multiplied* components, representing (a) the rate at which the design density, f_X, decreases to 0 in the upper and lower tails, and (b) the rate at which the cutoff, τ [$= \tau_1 = \tau_2$ in the notation of (2.4)], diverges to ∞. Indeed, we may express the asymptotic variance component as $D_1(nh_0^{1+\xi})^{-1}\tau^3$, where $D_1 > 0$ is a constant, depending on the functions f_X and $f_{A,B}$ but not on n; where $h_0 = h_0(n) \to 0$ is effectively a bandwidth, and is of the same size as $\kappa^{-1/2}$; where $\xi \geq 0$ depends on the tail properties of f_X, and is larger in cases where f_X decreases to 0 more rapidly in the tails (resulting in poorer overall convergence rates in such cases); and where $\tau = \tau(n)$ is the taper cutoff.

By way of contrast, the bias contribution to the mean-squared error derives essentially from two *added* terms, representing respectively the bandwidth h_0 and the cutoff τ. The smoothness of the distribution of (A, B), entering through properties of the tails of the characteristic function $\phi_{A,B}$, plays a significant role in determining the size of both contributions. Our regularity conditions will involve selecting the smoothing parameter, κ, and imposing conditions on $\phi_{A,B}$, such that the contribution of h_0 to the bias component is essentially h_0^2. However, alternative approaches are possible, and will be discussed in remarks following our main theorems. In the discussion below we shall represent the taper component by $\tau^{-\eta}$, where $\eta > 0$ depends on the rate of decrease of the tails of $\phi_{A,B}$. Thus, the bias contribution to the asymptotic mean-squared error is $(D_2 h_0^2 + D_3\tau^{-\eta})^2$, where D_2 and D_3 are constants; and so the overall asymptotic mean-squared error is

$$(4.1) \qquad D_1\big(nh_0^{1+\xi}\big)^{-1}\tau^3 + \big(D_2 h_0^2 + D_3\tau^{-\eta}\big)^2.$$

Next, we present a little notation, as a prelude to describing details behind the results discussed above. The smoothing parameter $\kappa = \kappa(\theta)$ will be taken to depend on θ. Owing to the decrease (typically) in design density (in θ space) toward the ends of the interval $\mathscr{I} = (-\frac{1}{2}\pi, \frac{1}{2}\pi)$, the amount of smoothing should be increased there if we are to achieve an effective balance between bias and variance contributions to mean-squared error. That is, κ should be decreased toward the ends of $\mathscr{I}$. This suggests defining $\kappa = 2\{h_0 \lambda(\theta, h_0)\}^{-2}$, say, where $h_0 = h_0(n) > 0$ tends to 0 as $n \to \infty$, does not depend on θ and represents the appropriate bandwidth away from the ends of $\mathscr{I}$ (i.e., for most of $\mathscr{I}$); and the function λ is virtually constant (with respect to both θ and h_0) when θ is not near the ends of $\mathscr{I}$, and increases

toward the ends of $\mathscr{I}$. [The smoothing parameter κ is naturally expressible in this form since $h = h(\theta) = \{2/\kappa(\theta)\}^{1/2}$ has many of the properties of bandwidth in a regular kernel estimator. In this notation, $h(\theta) = h_0 \lambda(\theta, h_0)$.]

There is a bewildering variety of possible choices for such a function λ, and we need to reduce them to a tractable and manageable array of useful candidates. It turns out to be appropriate to choose κ to be essentially constant (with respect to θ) if θ lies within a distance of order h_0 from the ends of $\mathscr{I}$, but to be decreasing before reaching that point, as θ approaches an end of $\mathscr{I}$. (For the sake of brevity we do not present the argument that leads us to this specification.) Even so, the possibilities for λ are still very numerous. To illustrate the effects of different rates of decrease of κ at the ends of $\mathscr{I}$, we shall consider only those that decrease in a polynomial way.

Together these considerations lead us to propose that

$$(4.2) \qquad \kappa = 2\{h_0 \lambda(\theta \bullet h_0)\}^{-2}.$$

Here, $n^{-1+\varepsilon} \leq h_0(n) \leq n^{-\varepsilon}$ for some $\varepsilon > 0$; the bounded, continuous function λ from $\mathscr{I}$ to $(0, \infty)$ satisfies $\lambda(\theta) \sim C_1(\frac{1}{2}\pi + \theta)^{-\gamma_1}$ as $\theta \downarrow -\frac{1}{2}\pi$, and $\lambda(\theta) \sim C_2(\frac{1}{2}\pi - \theta)^{-\gamma_2}$ as $\theta \uparrow \frac{1}{2}\pi$, for constants $C_1, C_2 > 0$ and $0 \leq \gamma_1, \gamma_2 < \frac{1}{2}$; and, for each $\theta \in \mathscr{I}$ and $\delta > 0$,

$$\theta \bullet \delta = \begin{cases} (\frac{1}{2}\pi + \theta) \vee \delta - \frac{1}{2}\pi, & \text{if } \theta \in \left(-\frac{1}{2}\pi, 0\right], \\ \frac{1}{2}\pi - \{(\frac{1}{2}\pi - \theta) \vee \delta\}, & \text{if } \theta \in \left(0, \frac{1}{2}\pi\right). \end{cases}$$

These conditions on κ will be referred to below as (C1). [Note: if $0 < h < \frac{1}{2}\pi$, then $\theta \bullet h$ equals θ if the latter is at least h away from the nearer of the two ends of the interval $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$, and otherwise equals $-\frac{1}{2}\pi + h$ or $\frac{1}{2}\pi - h$, whichever is the nearer to θ.]

Additionally we ask that K in (2.1) be bounded, nonnegative, Hölder continuous, supported on $[0, 1]$ and not identically 0 [these conditions will be referred to below as (C2)]; K_κ be given by (2.1); for $j, k \geq 0$ satisfying $j + k \leq 3$, and $D_4, \alpha_0 > 2$,

$$(4.3) \qquad \left|(\partial/\partial u)^j (\partial/\partial v)^k \phi_{A,B}(u, v)\right| \leq D_4(1 + u^2 + v^2)^{-(\alpha_0 + j + k)/2};$$

for a sequence of positive constants $\tau = \tau(n) \uparrow \infty$, $w(r) = 1$ for $|r| \leq \tau$ and $w(r) = 0$ for $|r| > \tau$ [these conditions on w and τ will be referred to below as (C3)]; and X have a nondegenerate, uniformly continuous density f_X that is bounded away from 0 and ∞ on compacts, and satisfies

$$f_X(x) \sim d_1 \alpha_1 x^{-\alpha_1 - 1}, \qquad f_X(-x) \sim d_2 \alpha_2 x^{-\alpha_2 - 1}$$

as $x \to \infty$, where $d_1, d_2, \alpha_1, \alpha_2 > 0$ [these conditions will be referred to below as (C4)].

These conditions on the distribution of X may be equivalently stated in terms of the distribution of Θ, in which case they ask that the density f_Θ of Θ be uniformly continuous, bounded away from 0 and ∞ on $(-\frac{1}{2}\pi + \varepsilon, \frac{1}{2}\pi - \varepsilon)$ for all $\varepsilon > 0$, and satisfy

$$f_\Theta\left(\frac{1}{2}\pi - \theta\right) \sim d_1 \alpha_1 \theta^{\alpha_1 - 1}, \qquad f_\Theta\left(-\frac{1}{2}\pi + \theta\right) \sim d_2 \alpha_2 \theta^{\alpha_2 - 1}$$

as $\theta \downarrow 0$.

Writing ϕ for $\phi_{A,B}$, define $\phi_{jk} = (\partial/\partial a)^j(\partial/\partial b)^k\phi(a,b)$ and

$$\beta(a,b) = -(8\pi^2)^{-1}\int \exp(-iar\cos\theta - ibr\sin\theta)|r|^3\,dr$$

$$\times\int_{-\pi/2}^{\pi/2}\lambda(\theta)^2\{\cos\theta\,\phi_{20}(r\cos\theta, r\sin\theta)$$

$$+\sin\theta\,\phi_{02}(r\cos\theta, r\sin\theta)$$

$$+2\sin\theta\cos\theta\,\phi_{11}(r\cos\theta, r\sin\theta)\}\,d\theta,$$

$$\nu(a,b) = -(2\pi)^{-2}\int_{-\pi/2}^{\pi/2}d\theta\int\Re\{\exp(-iar\cos\theta - ibr\sin\theta)$$

$$\times\phi(r\cos\theta, r\sin\theta)\}\{1 - w(r)\}|r|\,dr.$$

Put

$$c = \left\{\int K(u^2)u^2\,du\right\}\Big/\left\{\int K(u^2)\,du\right\},$$

$$\gamma - \alpha \equiv \min(\gamma_1 - \alpha_1, \gamma_2 - \alpha_2),$$

$$g_1(h_0) \equiv \begin{cases} h_0^{\gamma-\alpha+2}, & \text{if } \alpha - \gamma - 2 > 0, \\ \log h_0^{-1}, & \text{if } \alpha - \gamma - 2 = 0, \\ 1, & \text{if } \alpha - \gamma - 2 < 0, \end{cases}$$

$$g_2(h_0) \equiv \begin{cases} h_0^{3(\gamma-\alpha)+4}, & \text{if } 3(\alpha-\gamma) - 4 > 0, \\ \log h_0^{-1}, & \text{if } 3(\alpha-\gamma) - 4 = 0, \\ 1, & \text{if } 3(\alpha-\gamma) - 4 < 0. \end{cases}$$

We ask that the cutoff, τ, not increase too rapidly, the rate of increase satisfying the condition

$$(4.4) \qquad\qquad (nh_0)^{-1}g_2(h_0)g_1(h_0)^{-2}\tau(n) \to 0.$$

THEOREM 4.1. *Assume conditions* (C1) *on* κ, (C2) *on* K, (C3) *on* w, (C4) *on the design distribution,* (4.3) *on the distribution of* (A, B) *and* (4.4) *on* τ *and* h_0. *Furthermore, assume* $(nh_0)^{-1}\tau(h_0) = O(n^{-\varepsilon})$ *for some* $\varepsilon > 0$. *Then, for each* $(a, b) \in \mathbb{R}^2$, $\hat{f}_{A,B}(a, b)$ *is asymptotically Normally distributed with mean*

$$(4.5) \qquad\qquad f_{A,B}(a,b) + h_0^2\beta(a,b) + \nu(a,b) + o(h_0^2)$$

and variance

$$(4.6) \quad \left\{\int r^2 w(r)\,dr\right\}\left[D_5(a,b)(nh_0)^{-1}g_1(h_0) + o\{(nh_0)^{-1}g_1(h_0)\}\right],$$

where $D_5(a,b) = D_5(a,b; f_{A,B}, f_X, \lambda) > 0$ *does not depend on* n.

THEOREM 4.2. *Assume the conditions of Theorem 4.1, except for (4.4). Let $\mathscr{S} \subseteq \mathbb{R}^2$ be a bounded, open set, and let $\mathscr{X} = \{X_1, \ldots, X_n\}$ denote the collection of design points. Then, as $n \to \infty$,*

$$
\iint_{\mathscr{S}} E\left[\left\{\hat{f}_{A,B}(a,b) - f_{A,B}(a,b)\right\}^2 \Big| \mathscr{X}\right] da\, db
$$

$$
= \iint_{\mathscr{S}} \left\{h_0^2 \beta(a,b) + \nu(a,b)\right\}^2 da\, db
$$

$$
+ \left\{\int r^2 w(r)\, dr\right\}\left\{\iint_{\mathscr{S}} D_5(a,b)\, da\, db\right\}(nh_0)^{-1} g_1(h_0)
$$

$$
+ o_p\left[h_0^4 + \iint_{\mathscr{S}} \nu(a,b)^2\, da\, db + \left\{\int r^2 w(r)\, dr\right\}(nh_0)^{-1} g_1(h_0)\right].
$$

REMARK 4.1 (Generalization of the function λ). Our conditions on λ are tailored to produce relatively simple results, so as to make the main issues more transparent. They may be relaxed at the expense of more complex conclusions. In particular, if in the definition of λ it is asked only that $\gamma_1, \gamma_2 \geq 0$, and not in addition that $\gamma_1, \gamma_2 < \frac{1}{2}$, then the formula for the asymptotic variance remains valid but that for the asymptotic bias generally is not, since the function λ^2 is no longer integrable. More generally still, λ could be regularly varying at $\pm \frac{1}{2}\pi$ in a more general sense, for example, with $\lambda(\frac{1}{2}\pi - \delta) \sim |\frac{1}{2}\pi - \delta|^{-\gamma_2} L(\frac{1}{2}\pi - \delta)$ as $\delta \downarrow 0$, where the function L is a slowly varying function at the origin and is not necessarily asymptotically constant as $\delta \to 0$.

REMARK 4.2 [Conditions (C4)]. The assumption of simple regular variation of the tails of f_X may similarly be relaxed. Obviously, a slowly varying function L may be incorporated into the prescription, so that the regular variation is of a more general type. However, very different tail behavior—for example, design distributions such as the Normal with exponentially decreasing tails—may be considered. There the convergence rate of $\hat{f}_{A,B}$ to $f_{A,B}$ may be shown to be only logarithmic. (Nevertheless, our numerical procedures will not necessarily perform badly in such cases due to the typically good behavior of $\phi_{A,B}$ functions that arise in practice; compare, for example, our second simulation in Section 3.)

REMARK 4.3 (The constant D_5). In general, D_5 in (4.6) depends on f_X and $f_{A,B}$ in a highly complex way, and we shall not give a formula here. However, sufficient details will be provided in Section 5 to enable the interested reader to construct a general formula.

REMARK 4.4 [Example of calculation of $\nu(a,b)$]. Since $\nu(a,b)$ contributes significantly to the bias of $\hat{f}_{A,B}$, it is instructive to study an example which describes the possible size of this quantity. Let $A = \mu_A + A'$ and $B = \mu_B + B'$, where μ_A, μ_B are arbitrary constants and A', B' are independent and

identically distributed random variables with a symmetric "double gamma" distribution and characteristic functions $\phi_{A'}(t) = \phi_{B'}(t) \equiv (1 + t^2)^{-\alpha_3/2}$, where $\alpha_3 > 1$. Assume that $ab \neq 0$, put $\omega = -\arctan(a/b)$ and take $w(r) = 1$ for $|r| \leq \tau = \tau(n)$ and $w(r) = 0$ otherwise. We claim that

$$\nu(a + \mu_A, b + \mu_B) \sim D_6 \tau^{1 - 2\alpha_3}$$

as $\tau \to \infty$, where $D_6 = (2\pi b \cos \omega)^{-1} |\sin \omega \cos \omega|^{-\alpha_3}$. To appreciate why, observe that

$$
\begin{aligned}
2\pi^2 &\nu(a + \mu_B, b + \mu_B) \\
&= -\int_{-\pi/2}^{\pi/2} (a \cos \theta + b \sin \theta)^{-1} d\theta \int_{\tau}^{\infty} (1 + r^2 + r^4 \sin^2 \theta \cos^2 \theta)^{-\alpha_3/2} r \\
&\hspace{5cm} \times (\partial/\partial r) \sin\{r(a \cos \theta + b \sin \theta)\}\, dr \\
&\sim \tau \int_{-\pi/2}^{\pi/2} (a \cos \theta + b \sin \theta)^{-1} \sin\{\tau(a \cos \theta + b \sin \theta)\} \\
&\hspace{3cm} \times (1 + \tau^2 + \tau^4 \sin^2 \theta \cos^2 \theta)^{-\alpha_3/2}\, d\theta \\
&\sim \tau^{1 - 2\alpha_3} |\sin \omega \cos \omega|^{-\alpha_3} \int \{(b \cos \omega)x\}^{-1} \sin\{(b \cos \omega)x\}\, dx \\
&= \tau^{1 - 2\alpha_3} |\sin \omega \cos \omega|^{-\alpha_3} (b \cos \omega)^{-1} \pi.
\end{aligned}
$$

REMARK 4.5 (Optimal choice of h_0 and τ). Theorems 4.1 and 4.2 and Remark 4.4 verify (4.1) for mean-squared error, with $\xi = 0$ or $\alpha - \gamma - 2$ according as $\alpha - \gamma - 2 < 0$ or > 0, respectively, and with $\eta = 2\alpha_3 - 1 > 1$. Assuming for the sake of simplicity that D_2 and D_3 in (4.1) are of the same sign (the contrary case may be treated similarly), we see that the optimal orders of h_0 and τ are those that minimize $(nh_0^{1+\xi})^{-1} + h_0^4 + \tau^{-2\eta}$, and so are $n^{-\eta/\{6 + \eta(5+\xi)\}}$ (for h_0) and $n^{2/\{6 + \eta(5+\xi)\}}$ (for τ). Empirical selection of the optimal h_0 and τ is beyond the scope of this paper, but is feasible in very large data sets.

5. Proofs.

PROOF OF THEOREM 4.1.

Step 1: *Preliminaries.* Recalling the definitions of $\hat{\phi}_{R|\Theta}$ and $\hat{f}_{A,B}$ at (2.2) and (2.3), we see that

$$(5.1) \qquad \hat{f}_{A,B}(a, b) = (2\pi)^{-2} \sum_{j=1}^{n} \Psi_j(a, b),$$

where

$$(5.2) \quad \Psi_j(a, b) = \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \int \exp(-iar\cos\theta - ibr\sin\theta)$$

$$\times \exp\{irR_jS_j(\theta)\}w(r)|r|\, dr$$

$$(5.3) \quad = \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \int \cos\left[r\{R_jS_j(\theta) - (a\cos\theta + b\sin\theta)\}\right]$$

$$\times w(r)|r|\, dr.$$

Write $D_1, D_2, \ldots$ for generic positive constants, differing from those appearing in Section 4.

Step 2: *Conditional bias.* Recall that $\mathscr{X} = \{X_1, \ldots, X_n\}$, and observe that, by (1.1) and (5.2),

$$E\{\Psi_j(a, b)|\mathscr{X}\} = \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \int \exp(-iar\cos\theta - ibr\sin\theta)$$

$$\times \phi_{A,B}\{rS_j(\theta)\cos\Theta_j, rS_j(\theta)\sin\Theta_j\}w(r)|r|\, dr.$$

Furthermore,

$$(2\pi)^2 f_{A,B}(a, b) = \int_{-\pi/2}^{\pi/2} d\theta \int \exp(-iar\cos\theta - ibr\sin\theta)$$

$$\times \phi_{A,B}(r\cos\theta, r\sin\theta)|r|\, dr,$$

and so, by (5.1),

$$(2\pi)^2\left[E\{\hat{f}_{A,B}(a, b)|\mathscr{X}\} - f_{A,B}(a, b)\right]$$

$$= \sum_{j=1}^{n} E\{\Psi_j(a, b)|\mathscr{X}\} - (2\pi)^2 f_{A,B}(a, b)$$

$$(5.4) \quad = \int \exp(-iar\cos\theta - ibr\sin\theta)w(r)|r|\, dr$$

$$\times \int_{-\pi/2}^{\pi/2}\left[\sum_{j=1}^{n} v_j(\theta)\phi_{A,B}\{rS_j(\theta)\cos\Theta_j, rS_j(\theta)\sin\Theta_j\}\right.$$

$$\left. - \phi_{A,B}(r\cos\theta, r\sin\theta)\right] d\theta + \rho_1(a, b),$$

where $\rho_1(a, b) \equiv (2\pi)^2\nu(a, b)$.

Let cs denote either cosine or sine, and observe that

$$(5.5) \quad S_j(\theta)\text{cs }\Theta_j = \text{cs}\{\theta - (\theta \ominus \Theta_j)\}$$

$$= \text{cs }\theta - (\theta \ominus \Theta_j)\text{cs}'\theta + \tfrac{1}{2}(\theta \ominus \Theta_j)^2\text{cs}''\theta + o\left(|\theta \ominus \Theta_j|^2\right)$$

as $|\theta \ominus \Theta_j| \to 0$. Therefore,

$$\phi\{rS_j(\theta)\cos\Theta_j, rS_j(\theta)\sin\Theta_j\}$$

$$= \phi(r\cos\theta, r\sin\theta)$$

$$+ r(\theta \ominus \Theta_j)\{\sin\theta\,\phi_{10}(r\cos\theta, r\sin\theta)$$

(5.6)
$$- \cos\theta\,\phi_{01}(r\cos\theta, r\sin\theta)\}$$

$$- \tfrac{1}{2}r^2(\theta \ominus \Theta_j)^2\{\cos\theta\,\phi_{20}(r\cos\theta, r\sin\theta)$$

$$+ \sin\theta\,\phi_{02}(r\cos\theta, r\sin\theta)$$

$$+ 2\sin\theta\cos\theta\,\phi_{11}(r\cos\theta, r\sin\theta)\} + \rho_{2j}(\theta),$$

where, by (4.3),

(5.7)
$$\rho_{2j}(\theta) \le D_1|r(\theta \ominus \Theta_j)|^3(1+r^2)^{-(\alpha_0+3)/2}.$$

Since we have assumed $\eta = 0$ in the definition of the weights $v_j(\theta)$, then $\sum v_j(\theta) \equiv 1$. Furthermore, $\sum v_j(\theta)(\theta \ominus \Theta_j) \equiv 0$. It follows from these results, (5.4), (5.6) and (5.7) that

$$(2\pi)^2\big[E\{\hat{f}_{A,B}(a,b)\,|\,\mathscr{X}\} - f_{A,B}(a,b)\big]$$

$$= \tfrac{1}{2}\int \exp(-iar\cos\theta - ibr\sin\theta)w(r)|r|^3\,dr$$

(5.8)
$$\times \int_{-\pi/2}^{\pi/2}\left\{\sum_{j=1}^{n} v_j(\theta)(\theta \ominus \Theta_j)^2\right\}\{\cos\theta\,\phi_{20}(r\cos\theta, r\sin\theta)$$

$$+ \sin\theta\,\phi_{02}(r\cos\theta, r\sin\theta)$$

$$+ 2\sin\theta\cos\theta\,\phi_{11}(r\cos\theta, r\sin\theta)\}\,d\theta$$

$$+ \rho_1(a,b) + \rho_2(a,b),$$

where

(5.9)
$$\rho_2(a,b) \le D_2\int_{-\pi/2}^{\pi/2}\left(\sum_{j=1}^{n} v_j(\theta)|\theta \ominus \Theta_j|^3\right)d\theta.$$

By elementary algebra from the definitions of $u_j(\theta)$ and $v_j(\theta)$,

$$\sum_{j=1}^{n} v_j(\theta)(\theta \ominus \Theta_j)^2 = \{s_2(\theta)^2 - s_1(\theta)s_3(\theta)\}\big/\{s_0(\theta)s_2(\theta) - s_1(\theta)^2\}.$$

Since $s_k(\theta)$ equals a sum of independent, bounded random variables, standard methods, using Bernstein's inequality to bound large deviations of $s_k(\theta)$ on a fine grid and employing the assumed Hölder continuity of K to fill in the gaps between grid points, may be used to prove that, uniformly on $\mathscr{I} \equiv (-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi)$, $s_k(\theta) = E\{s_k(\theta)\} + o_p(nh^{k+1})$. Furthermore, with $h = h(\theta) =$

$(2/\kappa)^{1/2}$ and f_Θ denoting the density of Θ, we have, for $k = 0$ and 2,

$$n^{-1}E\{s_k(\theta)\} \sim E\Big(K\big[\kappa\{1 - |\cos(\theta - \Theta_j)|\}\big](\theta - \Theta_j)^k\Big)$$

$$\sim \int K\Big[\{h^{-1}(\theta - \omega)\}^2\Big](\theta - \omega)^k f_\Theta(\omega)\, d\omega$$

$$\sim h^{k+1}f_\Theta(\theta)\int K(u^2)u^k\, du = e_k h^{k+1}f_\Theta(\theta),$$

where $e_k \equiv \int K(u^2)u^k\, du$; while, for $k = 1$, $E\{s_k(\theta)\} = o(nh^2)$.

Recall from the paragraph preceding Theorem 4.1 that we defined $c = e_2/e_0$. Therefore, uniformly on $\mathscr{I}_\varepsilon \equiv (-\tfrac{1}{2}\pi + \varepsilon, \tfrac{1}{2}\pi - \varepsilon)$ for any $\varepsilon > 0$,

$$T(\theta) \equiv h(\theta)^{-2}\sum_{j=1}^n v_j(\theta)\big(\theta \ominus \Theta_j\big)^2 = c + o_p(1).$$

Furthermore, arguments of Hall, Marron, Neumann and Titterington (1994) may be used to show that $T(\theta) = O_p(1)$ uniformly in $\theta \in \mathscr{I}$ and, similarly,

$$h(\theta)^{-3}\sum_{j=1}^n v_j(\theta)|\theta \ominus \Theta_j|^3 = O_p(1)$$

uniformly in $\theta \in \mathscr{I}$. It now follows from (5.8) and (5.9) that

(5.10) $$E\{\hat{f}_{A,B}(a,b)\,|\,\mathscr{X}\} - f_{A,B}(a,b) = h_0^2\beta(a,b) + \nu(a,b) + o_p(h_0^2).$$

This establishes an analog of the formula at (4.5), with $o(h_0^2)$ there replaced by $o_p(h_0^2)$ here and the expression giving the *conditional* asymptotic mean of $\hat{f}_{A,B}(a,b)$.

Step 3: *Conditional variance.* Observe from (5.1) that

$$\mathrm{var}\{\hat{f}(a,b)\,|\,\mathscr{X}\} = (2\pi)^{-4}\sum_{j=1}^n \mathrm{var}\{\Psi_j(a,b)\,|\,\mathscr{X}\},$$

where, in view of (5.3) and in the case $k = 1$,

$$\mathrm{var}\{\Psi_j(a,b)\,|\,\mathscr{X}\}$$

(5.11)
$$= \int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2} v_j(\theta_1)v_j(\theta_2)\, d\theta_1\, d\theta_2$$

$$\times \int\int\{V_k(\Theta_j, \theta_1, \theta_2, r_1, r_2) - U(\Theta_j, \theta_1, \theta_2, r_1, r_2)\}$$

$$\times w(r_1)w(r_2)|r_1 r_2|\, dr_1\, dr_2,$$

with

$$V_1(\Theta, \theta_1, \theta_2, r_1, r_2) \equiv E\Big(\cos\big[r_1\{RS(\theta_1) - (a\cos\theta_1 + b\sin\theta_1)\}\big]$$

$$\times \cos\big[r_2\{RS(\theta_2) - (a\cos\theta_2 + b\sin\theta_2)\}\big]\Big|\Theta\Big),$$

$$U(\Theta, \theta_1, \theta_2, r_1, r_2) \equiv \prod_{k=1}^2 E\Big(\cos\big[r_k\{RS(\theta_k) - (a\cos\theta_k + b\sin\theta_k)\}\big]\Big|\Theta\Big),$$

and $(R, \Theta, S(\theta))$ denoting a generic version of $(R_j, \Theta_j, S_j(\theta))$. After a little manipulation it may be shown that (5.11) continues to hold with $k = 2$ and

$$V_2(\Theta, \theta_1, \theta_2, r_1, r_2) \equiv \Re\{\phi_{R|\Theta}(r_1 + r_2|\Theta)\exp(i\xi)\},$$

where $\xi = -(\eta_1 + \eta_2)$ and $\eta_k = r_k S(\theta_k)(a \cos \theta_k + b \sin \theta_k)$. Therefore,

$$\mathrm{var}\{\Psi_j(a, b) \,|\,\mathscr{X}\}$$

$$(5.12) \qquad = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} v_j(\theta_1) v_j(\theta_2)\, d\theta_1\, d\theta_2 \int\int V_3(\Theta_j, \theta_1, \theta_2, r_1, r_2)$$

$$\times w(r_1 - r_2) w(r_2)|(r_1 - r_2)r_2|\, dr_1\, dr_2 - J_j,$$

where

$$V_3(\Theta, \theta_1, \theta_2, r_1, r_2) \equiv \Re\big[\phi_{R|\Theta}(r_1|\Theta)\exp\{-ir_1 \zeta_1 + ir_2(\zeta_1 - \zeta_2)\}\big],$$

$$\zeta_k = S(\theta_k)(a \cos \theta_k + b \sin \theta_k)$$

and

$$J_j = \left(\int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta \int \Re\big[\phi_{R|\Theta}(r|\Theta_j)\exp\{-irS_j(\theta)(a \cos \theta + b \sin \theta)\}\big]\right.$$

$$\left.\times w(r)|r|\, dr\right)^2.$$

In view of the Taylor expansion at (5.5),

$$|S_j(\theta_k)\mathrm{cs}\,\theta_k - \mathrm{cs}\,\Theta_j| \le |\theta_k \ominus \Theta_j|.$$

Therefore, writing $\zeta_{kj} = S_j(\theta_k)(a \cos \theta_k + b \sin \theta_k)$ (for $k = 1, 2$) and $\zeta_{3j} = a \cos \Theta_j + b \sin \Theta_j$, we have

$$|\exp(-ir\zeta_{kj}) - \exp(-ir\zeta_{3j})| \le r|\zeta_{kj} - \zeta_{3j}| \le r(|a| + |b|)|\theta_k \ominus \Theta_j|.$$

Hence, by (5.12), defining

$$V_4(\Theta_j, r_1) \equiv \Re\{\phi_{R|\Theta}(r_1|\Theta_j)\exp(-ir_1 \zeta_{3j})\}$$

and

$$I_j = \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} v_j(\theta_1) v_j(\theta_2)\, d\theta_1\, d\theta_2$$

$$\times \int\int V_4(\Theta_j, r_1) w(r_1 - r_2) w(r_2)|(r_1 - r_2)r_2|\, dr_1\, dr_2$$

$$= \left\{\int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta\right\}^2 \int\int \Re\{\phi_{R|\Theta}(r_1|\Theta_j)\exp(-ir_1 \zeta_{3j})\}$$

$$\times w(r_1 - r_2) w(r_2)|(r_1 - r_2)r_2|\, dr_1\, dr_2,$$

we have

$$\left| \text{var}\{\Psi_j(a,b)|\mathscr{X}\} - I_j + J_j \right| \le L_j$$

$$\equiv 3(|a| + |b|)\left\{ \int_{-\pi/2}^{\pi/2} |\theta \ominus \Theta_j| v_j(\theta)\, d\theta \right\}$$

(5.13)

$$\times \left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta \right\}\left\{ \int |\phi_{R|\Theta}(r|\Theta_j)|(1 + |r|)\, dr \right\}$$

$$\times \left\{ \int w(r)(|r| + r^2)\, dr \right\}.$$

Now, $\sum_{j=1}^{n} I_j = \{1 + o_p(1)\}V(n)$, where

$$V(n) \equiv 2\pi\left\{ \int w(r) r^2\, dr \right\}\sum_{j=1}^{n}\left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta \right\}^2$$

$$\times f_{R|\Theta}(a \cos \Theta_j + b \sin \Theta_j | \Theta_j)$$

and

$$\sum_{j=1}^{n} J_j \le \left\{ \int D_3 |r|(1 + r^2)^{-\alpha_0/2}\, dr \right\}^2 \sum_{j=1}^{n}\left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta \right\}^2$$

$$= o_p\{V(n)\},$$

$$\sum_{j=1}^{n} L_j \le 3(|a| + |b|)\left\{ \int D_3(1 + |r|)(1 + r^2)^{-\alpha_0/2}\, dr \right\}$$

$$\times \left\{ \int w(r)(|r| + r^2)\, dr \right\}$$

$$\times \sum_{j=1}^{n}\left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta)\, d\theta \right\}\left\{ \int_{-\pi/2}^{\pi/2} |\theta \ominus \Theta_j| v_j(\theta)\, d\theta \right\}$$

$$= o_p\{V(n)\}.$$

Therefore, by (5.13),

(5.14)
$$\sum_{j=1}^{n} \text{var}\{\Psi_j(a,b)|\mathscr{X}\} = \{1 + o_p(1)\}V(n).$$

Next we introduce further notation. Recall the definition of $\theta \bullet \delta$ in Section 4, and for $\alpha_1 = \alpha_2$ put $\theta \circ \delta = \theta \diamond \delta = \theta \bullet \delta$. For $\alpha_1 < \alpha_2$ put

$$\theta \circ \delta = \begin{cases} \theta \bullet \delta, & \text{if } -\tfrac{1}{2}\pi < \theta \le \tfrac{1}{2}\pi - \delta, \\ -\tfrac{1}{2}\pi + \delta, & \text{if } \tfrac{1}{2}\pi - \delta < \theta < \tfrac{1}{2}\pi, \end{cases}$$

the same definition giving $\theta \diamond \delta$ if $\alpha_1 > \alpha_2$; for $\alpha_1 > \alpha_2$ define

$$\theta \circ \delta = \begin{cases} \theta \bullet \delta, & \text{if } -\tfrac{1}{2}\pi + \delta < \theta \le \tfrac{1}{2}\pi, \\ \tfrac{1}{2}\pi - \delta, & \text{if } -\tfrac{1}{2}\pi < \theta < -\tfrac{1}{2}\pi + \delta, \end{cases}$$

the same definition giving $\theta \diamond \delta$ if $\alpha_1 < \alpha_2$. (This notation is a periodic version of the $\bullet$ notation, designed to make appropriate allowance for the periodicity of the $\ominus$ notation.) For $0 \leq v \leq 1$ and $j = 1, 2$, define

$$q_{kj}(v) = \int K(u^2)u^k\{1 \wedge v - (1 \wedge v^{-1})u\}^{\alpha_j} \, du;$$

for $-1 \leq v \leq 1$ define

$$q_{k3}(v) = \int_{-1}^{v} K(u^2)u^k(u - v)^{\alpha_1 - 1} \, du,$$

$$q_{k4}(v) = \int_{v}^{1} K(u^2)u^k(u - v)^{\alpha_2 - 1} \, du;$$

for all α_1, α_2 define

$$q_k(\theta, \delta) = \begin{cases} q_{k1}\{(\tfrac{1}{2}\pi + \theta)\delta^{-1}\}, & \text{if } \theta \in (-\tfrac{1}{2}\pi + \delta, 0], \\ q_{k2}\{(\tfrac{1}{2}\pi - \theta)\delta^{-1}\}, & \text{if } \theta \in (0, \tfrac{1}{2}\pi - \delta]; \end{cases}$$

for $\alpha_1 = \alpha_2$ define

$$q_k(\theta, \delta) = \begin{cases} q_{k3}(v) + (c_2/c_1)q_{k4}(v), & \text{if } -\tfrac{1}{2}\pi < \theta \equiv -\tfrac{1}{2}\pi + \delta v \\ & \leq -\tfrac{1}{2}\pi + \delta, \\ (c_1/c_2)q_{k3}(v) + q_{k4}(v), & \text{if } \tfrac{1}{2}\pi - \delta < \theta \equiv \tfrac{1}{2}\pi + \delta v \leq \tfrac{1}{2}\pi; \end{cases}$$

for $\alpha_1 < \alpha_2$ define $q_k(\theta, \delta) = q_{k3}(v)$ if

$$\theta \equiv \pm\tfrac{1}{2}\pi + \delta v \in \left(-\tfrac{1}{2}\pi, -\tfrac{1}{2}\pi + \delta\right] \cup \left(\tfrac{1}{2}\pi - \delta, \tfrac{1}{2}\pi\right);$$

for $\alpha_1 > \alpha_2$ define $q_k(\theta, \delta) = q_{k4}(v)$ for the same range of values of θ; and generally define $V(\theta, \delta) \equiv q_2(\delta, \theta)q_0(\theta, \delta) - q_1(\theta, \delta)^2$.

Observe that $\sum u_j(\theta) = s_0(\theta)s_2(\theta) - s_1(\theta)^2$. With $h = (2/\kappa)^{1/2}$ and arguing as in Hall, Marron, Neumann, and Titterington (1994), we may show that

$$s_k(\theta) = \sum_{j=1}^{n} K\left[\{(\theta \ominus \Theta_j)/h\}^2\right](\theta \ominus \Theta_j)^k + o_p\{nh^{k+1}f_\Theta(\theta \circ h)\}$$

$$= nE\left(K\left[\{(\theta \ominus \Theta)/h\}^2\right](\theta \ominus \Theta)^k\right) + o_p\{nh^{k+1}f_\Theta(\theta \circ h)\}$$

$$= nh^{k+1}\int K(u^2)u^k f_\Theta(\theta \ominus hu) \, du + o_p\{nh^{k+1}f_\Theta(\theta \circ h)\}$$

$$= nh^{k+1}f_\Theta(\theta \circ h)q_k(\theta, h) + o_p\{nh^{k+1}f_\Theta(\theta \circ h)\}.$$

Therefore,

$$\sum_{j=1}^{n} u_j(\theta) = nh^3 f_\Theta(\theta \circ h)V(\theta, h) + o_p\{nh^{k+1}f_\Theta(\theta \circ h)\}.$$

Similarly it may be shown that, for a constant $D_4 > 0$,

$$\sum_{j=1}^{n} u_j(\theta) \geq D_4\{1 + o_p(1)\}nh^3 f_\Theta(\theta \diamond h),$$

the remainder terms in both relations being of the stated orders uniformly in $\theta \in \mathscr{I}$. These results lead to

(5.15)
$$\sum_{j=1}^{n} \left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \right\}^2 f_{R|\Theta}(a \cos \Theta_j + b \sin \Theta_j \mid \Theta_j)$$
$$= \{1 + o_p(1)\} D_5(a, b) n^{-1} \int_{-\pi/2}^{\pi/2} \{hf_\Theta(\theta \diamond h)\}^{-1} \, d\theta,$$

where $h = h(\theta)$ and $D_5(a, b) = D_5(a, b; K, f) > 0$ does not depend on n.

In view of conditions (C1) and (C4),

(5.16)
$$\int_{-\pi/2}^{\pi/2} \{hf_\Theta(\theta \diamond h)\}^{-1} \, d\theta \sim D_7 h_0^{-1} g(h_0)$$

as $n \to \infty$. Results (5.14)–(5.16) establish an analog of (4.6), in which the left-hand side is the *conditional* asymptotic variance of $\hat{f}_{A, B}(a, b)$, and on the right-hand side, $o(1)$ is replaced by $o_p(1)$.

Step 4: *Asymptotic normality.* Put $\Phi_j(a, b) \equiv \Psi(a, b) - E\{\Psi(a, b) \mid \mathscr{X}\}$. In view of (5.1) the fourth-order Liapounov condition for deriving conditional asymptotic Normality of the estimator $\hat{f}_{A, B}$ is

(5.17)
$$\left[\sum_{j=1}^{n} E\{\Phi_j(a, b)^4 \mid \mathscr{X}\} \right] \Big/ \left\{ (nh_0)^{-1} g_1(h_0) \tau^3 \right\}^2 \to 0$$

in probability. Using (5.3) for $\Phi_j(a, b)$ and arguing as in Step 3, we may show that

$$E\{\Phi_j(a, b)^4 \mid \mathscr{X}\} \le D_8 \left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \right\}^4 \tau^7$$

and

$$\sum_{j=1}^{n} \left\{ \int_{-\pi/2}^{\pi/2} v_j(\theta) \, d\theta \right\}^4 \le D_9 \{1 + o_p(1)\} n^{-3} \int_{-\pi/2}^{\pi/2} \{\hat{f}_\Theta(\theta \diamond h)\}^{-3} \, d\theta$$

$$\le D_{10} \{1 + o_p(1)\} n^{-3} h_0^{-3} g_2(h_0).$$

It follows that (5.17) is implied by (4.4).

Step 5: *Conclusion.* Combining the results of Steps 1–3, we see that, conditional on $\mathscr{X}$, $\hat{f}_{A, B}$ is asymptotically Normally distributed with bias and variance given by (4.5) and (4.6), respectively, provided that the $o(\dots)$ remainder terms there are altered to $o_p(\dots)$. Since the dominant terms in (4.5) and (4.6) are nonrandom and, in particular, do not depend on $X_1, \dots, X_n$, the unconditional distribution of $\hat{f}_{A, B}$ is asymptotically Normal with mean and variance given by (4.5) and (4.6). This completes the proof of Theorem 4.1. □

The proof of Theorem 4.2 is based on a relatively minor modification of Steps 2 and 3 in the proof of Theorem 4.1, and so is not given here.

REFERENCES

BECKER, R. A., CHAMBERS, J. M. and WILKS, A. R. (1988). *The New S Language*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

BERAN, R. J. and HALL, P. (1992). Estimating coefficient distributions in random coefficient regression. *Ann. Statist.* **20** 1110–1119.

BERAN, R. J. and MILLAR, P. W. (1994). Minimum distance estimation in random coefficient regression models. *Ann. Statist.* **22** 1976–1992.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

DEANS, S. R. (1983). *The Radon Transform and Some of Its Applications*. Wiley, New York.

DEVANEY, A. J. (1989). The limited-view problem in diffraction tomography. *Inverse Problems* **5** 501–521.

FAN, J. (1991). On the optimal rate of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.

FAN, J. (1993). Local linear smoothers and their minimax efficiency. *Ann. Statist.* **21** 196–216.

HALL, P., MARRON, S. J., NEUMANN, M. H. and TITTERINGTON, D. M. (1994). Curve estimation when the design density is low. *Ann. Statist.* To appear.

LIU, J. (1994). Minimum distance procedures in nonlinear random coefficient models. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

LONGFORD, N. T. (1993). *Random Coefficient Models*. Oxford Univ. Press.

NICHOLLS, D. F. and PAGAN, A. R. (1985). Varying coefficient regression. In *Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5**. North-Holland, Amsterdam.

RAJ, B. and ULLAH, A. (1981). *Econometrics, A Varying Coefficients Approach*. Croom-Helm, London.

SHEPP, L. A. and KRUSKAL, J. B. (1979). Computerized tomography: the new medical X-ray technology. *Amer. Math. Monthly* **85** 420–439.

SILVERMAN, B. W., JONES, M. C., WILSON, J. D. and NYCHKA, D. W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Statist. Soc. Ser. B* **52** 271–324.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Belknap Press, Cambridge, MA.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720
E-MAIL: beran@stat.berkeley.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
TORONTO, ONTARIO
CANADA M5S 3G3
E-MAIL: audrey@utstat.toronto.edu

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA, ACT 0200
AUSTRALIA
E-MAIL: hall@anu.edu.au