The Annals of Statistics
2002, Vol. 30, No. 5, 1480–1495

ON LOCAL LIKELIHOOD DENSITY ESTIMATION

BY B. U. PARK,[1] W. C. KIM[1] AND M. C. JONES

Seoul National University, Seoul National University and The Open University

This paper considers a class of local likelihood methods introduced by Eguchi and Copas. Unified asymptotic results are presented in the usual smoothing context of the bandwidth, h, tending to zero as the sample size tends to infinity. We present our results pointwise in the univariate case, but then go on to extend them to global properties and to indicate how to cope with the multivariate case. Specific members of the class due to Copas, and Hjort and Jones are seen to be members of a subset of the whole class with the same, and best, small h behavior. Further comparisons between members of the class are alluded to based on the complementary large h asymptotic results of Eguchi and Copas.

1. Introduction. Semiparametric density estimation attempts to combine parametric and nonparametric approaches to density estimation in such a way that the resulting method inherits the best properties of each, namely, efficient estimation if the proposed parametric family includes a good model for the data and the usual good behaviour of nonparametric density estimation if it does not.

In this paper, we are concerned only with one particular class of kernel-based semiparametric methods, those which are local likelihood methods. Let $X_1, \dots, X_n$ be a univariate random sample from the distribution with (unknown) density $f(\cdot)$ and let $f(\cdot, \theta)$, with θ a p-dimensional parameter vector, be a parametric model proposed for the data. Also, let K be a fixed symmetric kernel function with $\int K(y)\,dy = 1$ and, for convenience, $K(0) = 1$ and let h be the bandwidth. Then the formulation of kernel-based local likelihood estimation due to Copas (1995) is to take $\hat{f}(x) = f(x, \hat{\theta})$ where $\hat{\theta} = \hat{\theta}(x)$ is chosen to maximize

$$\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \log f(X_i, \theta) + \sum_{i=1}^{n} \left\{1 - K\left(\frac{X_i - x}{h}\right)\right\} \log\left\{1 - \int K\left(\frac{z - x}{h}\right) f(z, \theta)\, dz\right\}.$$

Received April 2000; revised February 2002.
[1]Supported in part by KOSEF Grant 1999-1-104-001-5 and the Brain Korea 21 Project.
AMS 2000 subject classifications. Primary 62G07; secondary 62G20.
Key words and phrases. Kernel smoothing, locally parametric density estimation, semiparametric inference, small bandwidth asymptotics.

An alternative formulation of kernel-based local likelihood estimation due to Hjort and Jones (1996) and Loader (1996) is to take $\hat{\theta}$ to maximize

$$\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \log f(X_i, \theta) - n \int K\left(\frac{z - x}{h}\right) f(z, \theta)\, dz.$$

Each formulation has as its first term the log-likelihood localized to x by means of the kernel function which gives high weight to terms corresponding to X_i close to x and less weight to terms with X_i far from x. Maximization of the first term alone is inadequate as it leads to inconsistent estimation. A second, correction, term is necessary but its precise details allow scope for variation. See Copas (1995) and Hjort and Jones [(1996), Section 2] for arguments leading to the correction terms above. A particularly useful choice is to take $\log f(x, \theta)$ to be a polynomial [Loader (1996)], although other simple standard forms for $f(x, \theta)$ have been suggested too.

In a very interesting paper, Eguchi and Copas (1998) gave a unified formulation of these two local likelihood approaches which also allows further variations on the theme. Eguchi and Copas go on to determine the *large h* properties of their class of estimators. This is very useful in describing how the semiparametric estimator behaves when the parametric estimator actually is, or is almost, a good model for the data and hence a large bandwidth, which results in at most small modifications of the parametric model when fitted to data, is appropriate. Note that as $h \to \infty$ the local log likelihoods tend to the usual, global, log likelihood (plus perhaps an irrelevant constant).

In this paper, we present the *small h* asymptotics of the same class of methods. This is useful for describing how the semiparametric estimator behaves when the parametric estimator is unsuitable for the data and the nonparametric aspect of the estimator takes over. This work unifies and extends that of Hjort and Jones (1996)—which gives small h results for their estimator and of Kim, Park and Kim (2001)—which gives small h results for the Copas estimator.

In Section 2, we describe the Eguchi and Copas (1998) unified formulation and make some initial remarks about choice of their general function ξ. The main development of the small h pointwise theoretical properties of the Eguchi and Copas class is contained in Section 3 with some proofs deferred to the Appendix. In Section 4, we discuss various aspects of our results including extensions to global properties and to the multivariate case, and make some discussion of the bandwidth selection issue. Finally, and in combination with the large h results of Eguchi and Copas (1998), we make some general comparison of estimators within the class (including some not mentioned thus far).

2. A class of estimating equations. Define $u(t, \theta) = (\partial/\partial\theta) \log\{f(t, \theta)\}$. Let $\xi(\cdot, \cdot)$ be an arbitrary function. Then the general form of local likelihood estimating equation introduced by Eguchi and Copas (1998) is to choose $\widehat{\theta}_n \equiv \widehat{\theta}_n(x)$ to satisfy $\Psi_n(\theta, h) = 0$ where $n\Psi_n(\theta, h) \equiv n\Psi_n(x, \theta, h)$ equals

$$\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) u(X_i, \theta) \\ - \sum_{i=1}^{n} \xi\left\{K\left(\frac{X_i - x}{h}\right), E_\theta K\left(\frac{X_1 - x}{h}\right)\right\} E_\theta\left\{K\left(\frac{X_1 - x}{h}\right) u(X_1, \theta)\right\}.$$

Here and below E_θ means expectation with respect to the parametric density $f(\cdot, \theta)$ while E with no subscript means expectation with respect to the true density f. The density estimator is, of course, $f(x, \widehat{\theta}_n(x))$.

The function ξ is required to satisfy

$$E_\theta \xi\left\{K\left(\frac{X_1 - x}{h}\right), E_\theta K\left(\frac{X_1 - x}{h}\right)\right\} = 1 \tag{2.1}$$

for the system of estimating equations to be unbiased when $f(\cdot) = f(\cdot, \theta)$ for some θ. For small h asymptotics we need some additional desirable properties of $\xi(y, z)$. We assume that for each z, $\xi(y, z)$ is linear in y. This is for simplicity of presentation, yet the class of such functions is rich enough to include all three versions discussed in Eguchi and Copas (1998). The theory in this paper indeed goes through with $\xi(y, z)$ being, for each z, a polynomial in y, but that includes more involved expansions and formulas.

We may write

$$\xi(y, z) = \xi_0(z) + \xi_1(z)y \tag{2.2}$$

for some functions $\xi_0(z)$ and $\xi_1(z)$. The condition (2.1) then implies that, writing $a_h = (1/h)E_\theta K\{(X_1 - x)/h\}$ which is $O(1)$,

$$\xi_0(ha_h) + ha_h\xi_1(ha_h) = 1 \tag{2.3}$$

from which we deduce that $\xi_0(z) = O(1) = z\xi_1(z)$ as $z \to 0$. It is convenient to introduce the two functions $\beta(z) = \xi_0(z)$ and $\gamma(z) = z\xi_1(z)$ so that we may rewrite (2.2) as

$$\xi(y, z) = \beta(z) + \gamma(z)(y/z). \tag{2.4}$$

For small h asymptotic analysis, we require that $\beta(z)$ and $\gamma(z)$ should be differentiable sufficiently many times in a neighborhood of $z = 0$. This assumption will be made throughout the work below.

Let $\beta_k = \beta^{(k)}(0)/k!$ and $\gamma_k = \gamma^{(k)}(0)/k!$ where, for a function g, $g^{(k)}$ denotes its kth derivative. In terms of β and γ, (2.3) may be rewritten as

$$\beta(ha_h) + \gamma(ha_h) = 1. \tag{2.5}$$

In particular, by letting $h \to 0$ we get $\beta_0 + \gamma_0 = 1$. In the work in Section 3 we exclude the case $\gamma_0 = 1$, however. The reason is that a function γ with $\gamma_0 = 1$

may produce an estimator which does not converge to a proper limit. To be specific, define $g(\theta, h) \equiv g(x, \theta, h) = h^{-1} E\Psi_n(x, \theta, h)$ for $h > 0$, and $g(\theta, 0) \equiv g(x, \theta, 0) = \lim_{h\to 0} g(x, \theta, h)$. The solution of the equation $\Psi_n(\theta, h) = 0$ may converge to a limit only when there exists a unique solution of the equation $g(\theta, h) = 0$ in a neighborhood of $h = 0$. By the implicit function theorem [see, e.g., Apostol (1975)], this happens when $g(\theta, 0)$ has a continuous first derivative at $\theta = \theta_0$ with nonzero determinant where θ_0 satisfies $g(\theta_0, 0) = 0$. However, we note that $g(\theta, 0) = (1 - \gamma_0)u(x, \theta)\{f(x) - f(x, \theta)\}$ which implies that $(\partial/\partial\theta)g(\theta, 0)$ is identically zero when $\gamma_0 = 1$.

Eguchi and Copas considered three special cases of their formulation, which they called the U-version [Hjort and Jones (1996), Loader (1996)], the C-version [Copas (1995)] and the T-version (a truncation-based local likelihood method). Note that for the U-version and C-version the condition (2.4) holds with $\beta(z) \equiv 1$, $\gamma(z) \equiv 0$ and $\beta(z) = 1/(1 - z)$, $\gamma(z) = -z/(1 - z)$, respectively. However, for the T-version, $\beta(z) \equiv 0$ and $\gamma(z) \equiv 1$. Thus we can expect that the T-version has undesirable small h asymptotic properties.

3. Theoretical properties.

3.1. *Stochastic expansion.* Suppose that, for each x, the solutions, denoted by θ_h and θ_0, respectively, of the equations $g(\theta, h) = 0$ and $g(\theta, 0) = 0$ are unique. Then, we may expect that under some additional regularity conditions $\widehat{\theta}_n \equiv \widehat{\theta}_n(x)$, a solution of the estimating equation $\Psi_n(\theta, h) = 0$, gets closer to θ_h as n grows. Specifically, assume that $g(\theta, h)$ converges to $g(\theta, 0)$ uniformly on a compact neighborhood of θ_0, say $\mathcal{D}$, that $g(\theta, 0)$ is continuous on $\mathcal{D}$, and that $f(x, \cdot)$ is bounded away from zero on $\mathcal{D}$ and $u(z, \theta)$ is bounded by a constant for z in a neighborhood of x and for $\theta \in \mathcal{D}$. Furthermore, suppose that $u(z, \theta)$ is two times differentiable with respect to θ for all z in a neighborhood of x and that there exists a function G which is continuous at x and satisfies for all z in a neighborhood of x,

$$\sup_{\theta\in\mathcal{D}} \left| \frac{\partial^2}{\partial\theta^2} u(z, \theta) \right| \le G(z). \tag{3.1}$$

Suppose also that $\xi(y, z)$ is two times continuously differentiable with respect to z in a neighborhood containing zero for all y in the range of the kernel function K. Finally, assume that the kernel function K is bounded and compactly supported. Call these conditions (S).

We describe stochastic expansions for $\widehat{\theta}_n$ and $\widehat{f}_n(x) = f(x, \widehat{\theta}_n)$ in the following theorem which may be proved under the conditions (S). To state the theorem, we write $\Psi_n(\theta)$ for $\Psi_n(\theta, h)$, and define $R_n \equiv R_n(x) = -(nh)^{1/2} \dot{f}(x, \theta_h)^t \times \{E\dot{\Psi}_n(\theta_h)\}^{-1} \Psi_n(\theta_h)$ where $\dot{f}(x, \theta)$ and $\dot{\Psi}_n(\theta)$ denote the first derivatives with respect to θ.

THEOREM 1. *Under the conditions* (S), *the estimator* $\widehat{\theta}_n$ *admits the stochastic expansion*

$$\widehat{\theta}_n - \theta_h = -\{E\dot{\Psi}_n(\theta_h)\}^{-1}\Psi_n(\theta_h) + O_p\big((nh)^{-1}\big)$$

as $n \to \infty$, $h \to 0$ *and* $nh \to \infty$. *Thus, the local likelihood density estimator* $\widehat{f}_n(x)$ *satisfies*

$$\widehat{f}_n(x) = f(x, \theta_h) + (nh)^{-1/2} R_n + O_p\big((nh)^{-1}\big). \tag{3.2}$$

The asymptotic bias of $\widehat{f}_n(x)$ comes from the deterministic part $f(x, \theta_h)$ since $ER_n = 0$ by the definition of θ_h, and the asymptotic variance may be obtained from the stochastic part R_n.

3.2. *Asymptotic variance.* In our second theorem we give a formula for $\mathrm{var}(R_n)$. To state the theorem, let $\mu_r = \int y^r K(y)\,dy$ and $\kappa_r = \int y^r K^2(y)\,dy$. Denote by M the $p \times p$ matrix which has, as its (r, s)th entry, $m_{r,s} = \kappa_{r+s} - \gamma_0\kappa_r\mu_s - \gamma_0\mu_r\kappa_s + \gamma_0^2\kappa_0\mu_r\mu_s$. Here and below the indices for vector or matrix entries count from 0. For example, the first diagonal entry of M equals m_{00}. Define another $p \times p$ matrix $N = (n_{r,s})$ with $n_{r,s} = \mu_{r+s} - \gamma_0\mu_r\mu_s$. For R_n to have a proper limit law the matrix N should be invertible. In fact, if N is singular then $\mathrm{var}(R_n)$ diverges and consequently the stochastic part in (3.2) has a rate slower than $O_p((nh)^{-1/2})$.

Now write $N = N_0 - \gamma_0\mu\mu^t$ where N_0 is a $p \times p$ matrix which has μ_{r+s} as its (r, s)th entry and $\mu = (\mu_0, \mu_1, \ldots, \mu_{p-1})^t$. Also write e_{r-1} for the unit vector with 1 appearing at the rth place and zeroes elsewhere. Recall that adding to one column of a matrix any multiple of another column does not affect the value of the determinant. Therefore,

$$\begin{aligned}\det(N) &= \det\left(N + \sum_{r=1}^{p-1}\gamma_0\mu_r\mu e_r^t\right) = \det(N + \gamma_0\mu\mu^t - \gamma_0\mu e_0^t)\\ &= \det(N_0 - \gamma_0\mu e_0^t) = (1-\gamma_0)\det(N_0)\end{aligned}$$

since $N_0 - \gamma_0\mu e_0^t$ is the same as N_0 except its first column which equals $(1-\gamma_0)$ times the first column of N_0. Note that the matrix N_0 is positive definite if K is nonnegative and its support contains an interval with nonempty interior. Under these conditions on the kernel, the matrix N is invertible if $\gamma_0 \neq 1$.

Suppose then that $\gamma_0 \neq 1$ and that we choose a kernel K so that the matrix N is invertible. Assume that the point of interest x is in the interior of the support of f, that is, $f(x) > 0$, that $f(\cdot)$ and $f(\cdot,\cdot)$ have two bounded and continuous (partial) derivatives, and that each component of $u_0(\cdot) = \lim_{h\to 0} u(\cdot, \theta_h)$ has p continuous derivatives in a neighborhood of x. Assume further that $u_0(x), u_0^{(1)}(x), \ldots, u_0^{(p-1)}(x)$ are linearly independent. Call these conditions (V).

THEOREM 2. *Under the conditions* (V), *the variance of* R_n *equals* $f(x)e_0^t \times N^{-1}MN^{-1}e_0 + o(1)$ *where* $e_0^t = (1, 0, \ldots, 0)$.

A proof is given in the Appendix. We will now show that this asymptotic variance expression does not depend on γ_0. Write M_0 for the $p \times p$ matrix with κ_{r+s} as its (r, s)th entry and $\kappa = (\kappa_0, \kappa_1, \ldots, \kappa_{p-1})^t$. Now,

$$N^{-1} = N_0^{-1} + \frac{\gamma_0}{1 - \gamma_0 \mu^t N_0^{-1} \mu} N_0^{-1} \mu \mu^t N_0^{-1} = N_0^{-1} + \frac{\gamma_0}{1 - \gamma_0} e_0 e_0^t.$$

The second equality follows from the fact that $N_0^{-1}\mu$ equals the first column of $N_0^{-1}N_0 = I$, namely e_0. Plugging this expression into $e_0^t N^{-1} M N^{-1} e_0$ and using the additional facts that the first column of M equals $(1 - \gamma_0)\kappa - \gamma_0\kappa_0(1 - \gamma_0)\mu$ and that the first diagonal element of M equals $(1 - \gamma_0)^2\kappa_0$, we obtain

$$e_0^t N^{-1} M N^{-1} e_0 = e_0^t N_0^{-1} M N_0^{-1} e_0 + 2\gamma_0 \kappa^t N_0^{-1} e_0 - \gamma_0^2 \kappa_0.$$

Plugging the expression $M = M_0 - \gamma_0 \kappa \mu^t - \gamma_0 \mu \kappa^t + \gamma_0^2 \kappa_0 \mu \mu^t$ into the above formula yields the following corollary.

COROLLARY 1. *Under the conditions* (V), *the variance of* R_n *equals* $f(x)e_0^t N_0^{-1} M_0 N_0^{-1} e_0 + o(1)$.

This corollary implies that the (first-order) asymptotic variance of a local likelihood density estimator does not depend on the functions β and γ at all except for the requirement that $\gamma_0 \neq 1$. The asymptotic variance of the various versions of local likelihood density estimation with $\gamma_0 \neq 1$ are all identical, and what is more, they are the same as the asymptotic variances of local polynomial regression where the order of polynomial to be fitted takes the role of $p - 1$ here. See Ruppert and Wand (1994) or Fan and Gijbels (1996), for example. Specifically, for $p = 1$ or 2, the asymptotic variance is κ_0, which also coincides with the usual asymptotic variance of the ordinary kernel density estimator. For $p = 3$ or 4, the asymptotic variance is $(\mu_4^2\kappa_0 - 2\mu_2\mu_4\kappa_2 + \mu_2^2\kappa_4)/(\mu_4 - \mu_2^2)^2$. The corollary reveals another interesting fact about how the asymptotic variance of $\widehat{f}_n(x)$ changes as p increases. Note that the rate which is $(nh)^{-1}$ does not depend on p. Moreover, if p is odd, then there is no inflation even in the constant factor when one increases the number of parameters from p to $p + 1$. This property of variance is due to the special structure of the matrices M and N: the entries $m_{r,s}$ and $n_{r,s}$ equal zero for (r, s) with $r + s$ odd, and when one increases the number of parameters the new matrices M and N are formed by adding a new column and a new row to the old ones. Formally, we give the following proposition which is not difficult to show.

PROPOSITION 1. *Let A and B be $k \times k$ matrices. Denote by $\tilde{A}$ and $\tilde{B}$ the $(k+1) \times (k+1)$ matrices such that $\tilde{A}_{r,s} = A_{r,s}$ and $\tilde{B}_{r,s} = B_{r,s}$ for $0 \le r, s \le k-1$. Suppose that $\tilde{A}_{r,s} = \tilde{B}_{r,s} = 0$ for (r, s) with $r + s$ odd (and so also for A and B). Suppose in addition that the matrices obtained by deleting all the odd-numbered columns and rows from $A, \tilde{A}$, and those obtained by deleting all the even-numbered columns and rows are invertible. Then $(\tilde{A}^{-1}\tilde{B}\tilde{A}^{-1})_{r,r} = (A^{-1}BA^{-1})_{r,r}$ when $k - r$ is odd.*

3.3. *Asymptotic bias.* Next, we investigate the asymptotic bias $f(x, \theta_h) - f(x)$. Recall that θ_h satisfies $E\{\Psi_n(\theta_h)\} = 0$; this is the equation governing the asymptotic bias. Write $\nu(y, \theta) = f(y, \theta) - f(y)$. Define $\zeta_1(x, \theta, h) = \int K(y)\nu(x + hy, \theta)\,dy$, and likewise define $\zeta_2(x, \theta, h), \zeta_3(x, \theta, h)$ by replacing $\nu(\cdot, \theta)$ with $u(\cdot, \theta)f, u(\cdot, \theta)\nu(\cdot, \theta)$, respectively. From (2.5) it follows that, writing $a_h = a(x, \theta, h)$ for a_h defined in Section 2,

$$E\xi\left\{K\left(\frac{X_1 - x}{h}\right), E_\theta K\left(\frac{X_1 - x}{h}\right)\right\} = 1 - \gamma\big(ha(x, \theta, h)\big)\left\{\frac{\zeta_1(x, \theta, h)}{a(x, \theta, h)}\right\}.$$

Plugging this expression into the governing equation and writing $\zeta_4(x, \theta, h) = \gamma(ha(x, \theta, h))$, we get $F(x, \theta_h, h) = 0$ where

$$\begin{aligned} F(x, \theta, h) &= \zeta_3(x, \theta, h)a(x, \theta, h) \\ &\quad - \zeta_1(x, \theta, h)\zeta_4(x, \theta, h)\big\{\zeta_2(x, \theta, h) + \zeta_3(x, \theta, h)\big\}. \end{aligned} \tag{3.3}$$

As a function of h, $F(x, \theta, h)$ admits a Taylor expansion if $f(y)$ and $f(y, \theta)$ are sufficiently smooth. Thus θ_h, being the unique solution to $F(x, \theta, h) = 0$, admits a Taylor expansion by the implicit function theorem. We can expand then $u(\cdot, \theta_h)$, $\nu(\cdot, \theta_h)$ and $F(x, \theta_h, h)$ as functions of h. Let $u_r(\cdot)$, $\nu_r(\cdot)$ and $c_r(x)$ be the coefficients of h^r in the Taylor expansions of $u(\cdot, \theta_h)$, $\nu(\cdot, \theta_h)$ and $F(x, \theta_h, h)$, respectively. We are interested in the bias $\nu(x, \theta_h) = f(x, \theta_h) - f(x)$, and thus in $\nu_r(x)$. These coefficients can be obtained by equating $c_r(x)$ to zero. Direct calculation of the coefficients $c_r(x)$ from (3.3) is extremely lengthy and complex. Our next theorem presents a simple easy-to-use formula for computing $c_r(x)$ in terms of u_r, ν_r and f.

To state the theorem, assume that $f(y)$, $f(y, \theta)$ and $u(y, \theta)$ are functions with r bounded and continuous (partial) derivatives. Let s and l be nonnegative integers. Define

$$\zeta_{1,s}(x) = \sum_{k=0}^{s}\{(s-k)!\}^{-1}\nu_k^{(s-k)}(x)\mu_{s-k},$$

$$\zeta_{3,s}(x) = \sum_{k=0}^{s}\{(s-k)!\}^{-1}(u \times \nu)_k^{(s-k)}(x)\mu_{s-k},$$

$$A_{s,l}(x) = \sum_{k=0}^{s-l}\{(s-k)!\}^{-1}\binom{s-k}{l}\nu_k^{(s-l-k)}(x)\mu_{s-k}.$$

We use the convention $(u \times v)_r$ to denote $\sum_{k=0}^{r} u_k v_{r-k}$. One should observe that $\zeta_{1,r}(x) = A_{r,0}(x)$ for all r. It is convenient to introduce a notation for the condition of our theorem. Let $r \geq 1$ be an integer. We call the following condition $\mathcal{C}_r$: $\zeta_{1,s}(x) = 0$ for all $0 \leq s \leq r-1$ and $A_{s,l}(x) = 0$ for all $0 \leq l \leq s$, $0 \leq s \leq r-1$. Note that the conditions $\mathcal{C}_r$'s are nested, that is, $\mathcal{C}_r$ is implied by $\mathcal{C}_{r+1}$.

THEOREM 3. *Let $r \geq 1$ be an integer. Suppose that the condition $\mathcal{C}_r$ holds. Then,*

$$c_r(x) = \left\{(1-\gamma_0)\zeta_{1,r}(x)u_0(x) + \sum_{l=1}^{r} A_{r,l}(x)u_0^{(l)}(x)\right\} f(x). \tag{3.4}$$

A proof is given in the Appendix. Define $f_0(y) = \lim_{h\to 0} f(y, \theta_h)$. Assume that $f_0(x)$, which equals $f(x) + \nu_0(x)$, is strictly positive. The condition $\mathcal{C}_r$ is then satisfied for all $1 \leq r \leq p+1$ if $u_0(x), \ldots, u_0^{(p-1)}(x)$ are linearly independent where p is the number of parameters in the local likelihood estimating equation. To see this, first we note that $c_0(x) = (1-\gamma_0)u_0(x)\nu_0(x)\{f(x) + \nu_0(x)\}$. From the equation $c_0(x) = 0$ we get $\nu_0(x) = 0$ since $f(x) > 0$. (One should note that this does not mean the function ν is identically zero, but just that the function is zero at the point x.) This implies that $A_{r,r}(x) = (r!)^{-1}\nu_0(x)\mu_r = 0$ for all $r \geq 0$. Thus, $c_r(x)$ in (3.4) is a linear combination of $u_0(x), \ldots, u_0^{(r-1)}(x)$ which are linearly independent when $p \geq r$. Hence, whenever the equation (3.4) is true for $r \leq p$, we have $\zeta_{1,r}(x) = 0$ and $A_{r,l}(x) = 0$ for all $1 \leq l \leq r$. Now we consider the condition $\mathcal{C}_1$. We observe that $\zeta_{1,0}(x) = A_{0,0}(x) = \nu_0(x) = 0$. Thus, the condition $\mathcal{C}_1$ is satisfied. This implies that the identity (3.4) holds for $r = 1$ so that $\zeta_{1,1}(x) = 0$. The condition $\mathcal{C}_2$ is then satisfied, and we repeat this argument until we conclude that the identity (3.4) holds for $r = p$. We obtain the following corollary.

COROLLARY 2. *Let p be the number of parameters in the local likelihood estimating equation $\Psi_n(\theta, h) = 0$. Then, $\zeta_{1,r}(x) = 0$ for all $0 \leq r \leq p$, and $A_{r,l}(x) = 0$ for all $0 \leq l \leq r$, $0 \leq r \leq p$ so that for all $1 \leq r \leq p+1$ the equation* (3.4) *holds.*

Theorem 3 and Corollary 2 can be used to obtain each term in an expansion of the bias. We illustrate this in the cases $p = 1, 2, 3, 4$. Below, we often omit "x," the point of interest, in the notation. For example, we write c_r for $c_r(x)$ and $\nu_k^{(j)}$ for $\nu_k^{(j)}(x)$. First, we consider the case $p = 1$. By Corollary 2, we get $\zeta_{1,1} = \nu_1 = 0$. Since $\nu_0 = 0$, this means that the leading bias term is $O(h^2)$. The coefficient of h^2 can be obtained by the equation $c_2 = 0$. We find that $\zeta_{1,2} = \nu_0^{(2)}\mu_2/2 + \nu_2$ and $A_{2,1} = \nu_0^{(1)}\mu_2$. By Corollary 2, we get

$$c_2/f = (1-\gamma_0)\{\nu_0^{(2)}\mu_2/2 + \nu_2\}u_0 + \nu_0^{(1)}u_0^{(1)}\mu_2.$$

Since $\nu_0(y) = f_0(y) - f(y)$, this implies that when $p = 1$,

$$f(x, \theta_h) - f(x) \\ = \left\{\frac{1}{2}(f - f_0)^{(2)}(x) + \frac{1}{1-\gamma_0}\frac{u_0^{(1)}(x)}{u_0(x)}(f - f_0)^{(1)}(x)\right\}\mu_2 h^2 + o(h^2).$$

For $p = 2$ the constant factor of the leading bias is simplified. By Corollary 2, $A_{2,1} = 0$ so that $\nu_0^{(1)} = 0$. Hence, $\nu_2 = -\nu_0^{(2)}\mu_2/2$ and we obtain for $p = 2$,

$$f(x, \theta_h) - f(x) = \tfrac{1}{2}(f - f_0)^{(2)}(x)\mu_2 h^2 + o(h^2).$$

The rate in terms of h is the same as that of the ordinary kernel density estimator in both cases. It is unclear, however, whether in any particular case (and for any particular x) $f^{(2)}(x)$ or $(f - f_0)^{(2)}(x)$ results in smaller squared bias.

Next, we consider the cases $p = 3, 4$. We first show that in these cases $\nu_2 = \nu_3 = 0$ so that the leading bias is $O(h^4)$. Applying Corollary 2, in particular $\zeta_{1,3} = 0$ and $A_{3,1} = 0$, we obtain

$$\nu_1^{(2)}\mu_2/2 + \nu_3 = 0, \qquad \nu_1^{(1)} = 0. \tag{3.5}$$

Consider now c_4. For this we note that $A_{4,3} = \nu_0^{(1)}\mu_4/6 = 0$ since $\nu_0^{(1)} = 0$. Thus from Corollary 2, c_4 is a linear combination of $u_0, u_0^{(1)}$ and $u_0^{(2)}$ only, so that all the coefficients of $u_0, u_0^{(1)}$ and $u_0^{(2)}$ are zero. We obtain

$$\zeta_{1,4} = \tfrac{1}{24}\nu_0^{(4)}\mu_4 + \tfrac{1}{2}\nu_2^{(2)}\mu_2 + \nu_4 = 0, \tag{3.6}$$

$$A_{4,1} = \tfrac{1}{6}\nu_0^{(3)}\mu_4 + \nu_2^{(1)}\mu_2 = 0, \tag{3.7}$$

$$A_{4,2} = \tfrac{1}{4}\nu_0^{(2)}\mu_4 + \tfrac{1}{2}\nu_2\mu_2 = 0. \tag{3.8}$$

Combining (3.8) and the fact that $\nu_2 = -\nu_0^{(2)}\mu_2/2$, we conclude that $\nu_2 = \nu_0^{(2)} = 0$. Next, we consider c_5. Here we apply Theorem 3. Condition $\mathfrak{C}_5$ is satisfied. Note that $A_{5,3} = \nu_1^{(1)}\mu_4/6 = 0$ by (3.5) and $A_{5,4} = \nu_1\mu_4/24 = 0$ so that c_5 includes only $u_0, u_0^{(1)}$ and $u_0^{(2)}$. Therefore all their coefficients are zero. In particular,

$$A_{5,1} = \tfrac{1}{6}\nu_1^{(3)}\mu_4 + \nu_3^{(1)}\mu_2 = 0, \tag{3.9}$$

$$A_{5,2} = \tfrac{1}{4}\nu_1^{(2)}\mu_4 + \tfrac{1}{2}\nu_3\mu_2 = 0. \tag{3.10}$$

Now combining (3.5) and (3.10), we obtain $\nu_3 = \nu_1^{(2)} = 0$.

An explicit expression for the constant factor ν_4 of the h^4 term may now be obtained. When $p = 3$, we obtain from (3.6) that

$$f(x, \theta_h) - f(x) = \left\{\tfrac{1}{24}(f - f_0)^{(4)}(x)\mu_4 - \tfrac{1}{2}\nu_2^{(2)}(x)\mu_2\right\}h^4 + o(h^4).$$

The constant factor is further simplified when $p = 4$. For this, consider c_6 and apply Theorem 3 with $r = 6$ since condition $\mathfrak{C}_6$ is satisfied. We can find in this case $A_{6,4} = A_{6,5} = 0$ so that c_6 is a linear combination of $u_0, u_0^{(1)}, u_0^{(2)}$ and $u_0^{(3)}$. Equating $A_{6,2}$ to zero in particular we obtain

$$A_{6,2} = \tfrac{1}{48}\nu_0^{(4)}\mu_6 + \tfrac{1}{4}\nu_2^{(2)}\mu_4 + \tfrac{1}{2}\nu_4\mu_2 = 0. \tag{3.11}$$

Solving the system of equations (3.6) and (3.11) yields that for $p = 4$,

$$f(x,\theta_h) - f(x) = -\frac{1}{24}\frac{\mu_2\mu_6 - \mu_4^2}{\mu_4 - \mu_2^2}(f - f_0)^{(4)}(x)h^4 + o(h^4).$$

4. Extensions and discussion.

4.1. *Global properties.* The results in Section 3 are given in pointwise form. They may be extended to some global results. First, the stochastic expansion (3.2) can be strengthened as

$$\sup_{x\in\mathfrak{I}}\left|\widehat{f}_n(x) - f(x,\theta_h) - (nh)^{-1/2}R_n(x)\right| = O_p\{(nh)^{-1}\log n\} \tag{4.1}$$

for a compact interval $\mathfrak{I}$ where we write $R_n(x)$ instead of R_n to stress its dependence on x. Property (4.1) may be shown to be valid with slight modifications of the conditions (S) stated in Section 3.1, as follows: add uniformity and continuity over $x \in \mathfrak{I}$ to the conditions on g, require $f(x,\theta)$ to be bounded away from zero for $\theta \in \mathcal{D}$ and $x \in \mathfrak{I}$, $u(z,\theta)$ to be bounded for $\theta \in \mathcal{D}$ and z in an interval containing $\mathfrak{I}$, say, $\mathfrak{I}_\varepsilon$, the inequality (3.1) to hold for all z in $\mathfrak{I}_\varepsilon$ for a function G which is continuous on $\mathfrak{I}$, $\xi(y,z)$ to be twice partially continuously differentiable with respect to y and z, K to be twice differentiable as well, and h to be asymptotic to $n^{-(1/3)+\delta}$ for some $0 < \delta < 1/3$.

Furthermore, the variance expansion given at Theorem 2 may be shown to hold uniformly on $x \in \mathfrak{I}$ if we require, in addition to the conditions (V) in Section 3.2, that $\inf_{x\in\mathfrak{I}} f(x) > 0$, that $u_0(\cdot)$ has p bounded and continuous (partial) derivatives on $\mathfrak{I}_\varepsilon$, and that its derivatives up to $(p-1)$th order are linearly independent for all $x \in \mathfrak{I}$. Also, all the expansions in the bias approximation are valid uniformly on $\mathfrak{I}$ if the corresponding functions involved in the expansions are continuous on $\mathfrak{I}$.

4.2. *Bandwidth selection.* Asymptotic results such as those in this paper can be utilized in a variety of ways to produce useful bandwidth selectors [e.g., Jones, Marron and Sheather (1996)]. For example, in the case of fitting two parameters locally, the asymptotic mean squared error is minimised by

$$h_0 = \left[\frac{\kappa_0}{\int\{(f - f_0)^{(2)}(x)\}^2\,dx\,\mu_2^2 n}\right]^{1/5}.$$

It is not immediately clear how to specify f as well as f_0 to provide a rule-of-thumb based on this expression, but plug-in methods are feasible. In particular, the bias term requires estimates of f'', which might be provided by ordinary kernel density estimation or by the appropriate coefficient(s) in a three or more parameter local likelihood fit, in each case using an appropriate larger bandwidth, and f_0'', which is easy.

Local likelihood density estimation throws up even more opportunities for bandwidth selection methods based on asymptotics than does ordinary kernel estimation because of the existence of both small h and large h asymptotics for their performance. Eguchi and Copas (1998) provide a simple rule-of-thumb based on their large h asymptotics. It is an interesting and as yet unexplored question as to which type of asymptotics yields the most useful formulas for practice. In addition, there remain methods such as least squares cross-validation which are not so tied to asymptotics and which are readily applicable to local likelihood density estimation. The bandwidth selection issue is a topic ripe for extensive methodological development and practical testing. Also, it should be noted that the asymptotics presented in this paper are for fixed h. Exploring the sampling properties for stochastic h is very important and is a challenging problem for future research.

4.3. *Multidimensional data.* Extension of the local likelihood method to the case of d-dimensional data is straightforward. Let K now be a d-variate kernel and H a symmetric positive definite bandwidth matrix. Then, the definition of $\Psi_n(\theta)$ and $\widehat{\theta}_n$ given in the first paragraph of Section 2 is generalized immediately by substituting $K(H^{-1}(\boldsymbol{X}_i - \boldsymbol{x}))$ for $K(h^{-1}(X_i - x))$ there. Development of relevant theory for the density estimator $\widehat{f}_n(\boldsymbol{x}) = f(\boldsymbol{x}, \widehat{\theta}_n(\boldsymbol{x}))$ is possible, too. First, one may obtain the following analogue of (3.2):

$$\widehat{f}_n(\boldsymbol{x}) = f(\boldsymbol{x}, \theta_H) + (n|H|)^{-1/2} R_n + O_p\{(n|H|)^{-1}\},$$

where $R_n \equiv R_n(\boldsymbol{x}) = -(n|H|)^{1/2} \dot{f}(\boldsymbol{x}, \theta_H)^t \{E\dot{\Psi}_n(\theta_H)\}^{-1} \Psi_n(\theta_H)$ with θ_H being the solution of the equation $E\{\Psi_n(\theta, H)\} = 0$, and $|H|$ denoting the determinant of H. Treatment of a general number of parameters in the multidimensional case to have simple expressions for the variance and bias requires very careful notation. Below, we focus on the special case of a $(d+1)$-dimensional parameter, analogue of local linear fitting in d dimensions.

Define $\mu_{r,s} = \int y_r y_s K(\boldsymbol{y})\, d\boldsymbol{y}$ for $r, s = 1, \dots, d$, $\mu_{0,s} = \int y_s K(\boldsymbol{y})\, d\boldsymbol{y}$ for $s = 1, \dots, d$ and $\mu_{0,0} = \int K(\boldsymbol{y})\, d\boldsymbol{y} = 1$. Likewise, define $\kappa_{r,s}$ for $r, s = 1, \dots, d$, $\kappa_{0,s}$ for $s = 1, \dots, d$ and $\kappa_{0,0}$ with K^2 in place of K. Let $\tilde{M}$ be the $(d+1) \times (d+1)$ matrix which has, as its (r, s)th entry, $\tilde{m}_{r,s} = \kappa_{r,s} - \gamma_0 \kappa_{0,r} \mu_{0,s} - \gamma_0 \mu_{0,r} \kappa_{0,s} + \gamma_0 \kappa_{0,0} \mu_{0,r} \mu_{0,s}$. Write $\tilde{N} = (\tilde{n}_{r,s})$ with $\tilde{n}_{r,s} = \mu_{r,s} - \gamma_0 \mu_{0,r} \mu_{0,s}$. The variance of R_n has exactly the same expression as in Theorem 2 except that M and N are replaced by $\tilde{M}$ and $\tilde{N}$, respectively. In particular, if all the odd-order moments of K vanish, then the expression for the variance of R_n reduces to $\kappa_{0,0} f(\boldsymbol{x}) + o(1)$ so that the asymptotic variance of $\widehat{f}_n(\boldsymbol{x})$ equals $n^{-1}|H|^{-1} f(\boldsymbol{x}) \int K^2(\boldsymbol{y})\, d\boldsymbol{y}$. If we add the additional condition on K that $\int \boldsymbol{y}\boldsymbol{y}^t K(\boldsymbol{y})\, d\boldsymbol{y} = \mu_{1,1} I$ with $\mu_{1,1} \neq 0$, then it may be shown that the asymptotic bias of $\widehat{f}_n(\boldsymbol{x})$ equals $\frac{1}{2}\mu_{1,1}\,\mathrm{tr}\{H^2 \mathcal{H}_{f-f_0}(\boldsymbol{x})\}$ where $\mathcal{H}_{f-f_0}(\boldsymbol{x})$ is the $d \times d$ Hessian matrix of the function $f - f_0$ at $\boldsymbol{x}$ and f_0 is the limit of $f(\cdot, \theta_H)$ when all the entries of H tend to zero. These results are valid under the obvious multivariate version of the conditions imposed on f and $f(\cdot, \theta)$. We only point out that linear independence is now required for $u_0(\boldsymbol{x}), (\partial/\partial x_1)u_0(\boldsymbol{x}), \ldots, (\partial/\partial x_d)u_0(\boldsymbol{x})$ where u_0 is the limit of $u(\cdot, \theta_H)$ when all the entries of H tend to zero.

4.4. *Effect of number of parameters.* Because asymptotic bias and variance depend on ξ when of the linear form (2.2) only through the value of γ_0, the small h asymptotic performance of the Copas (1995) and Hjort and Jones (1996) local likelihood density estimators is identical [and, setting $\gamma_0 = 0$, reduce to results given in Kim, Park and Kim (2001) and Hjort and Jones (1996), respectively]. For any $\gamma_0 \neq 1$, the behavior of the bias mimics that of local polynomial fitting [Ruppert and Wand (1994), Fan and Gijbels (1996)] in being of $O(h^2)$ for $p = 1, 2$ and $O(h^4)$ for $p = 3, 4$, with the even number of parameters resulting in simpler bias relative to its odd number companion. In addition, provided $\gamma_0 \neq 1$, the asymptotic variance reduces to that of the appropriate "equivalent kernel" of local polynomial fitting. In the current density estimation context, it might be just as useful to note the mimicking of the asymptotic bias and variance behavior of Loader's (1996) local polynomial fitting to $\log f$, except that that approach yields a particular choice of f_0 in the bias.

4.5. *Choice of ξ.* The C- and U-versions, by virtue of having $\gamma_0 \neq 1$, are, therefore, particularly attractive in small h terms, and as mentioned in Section 2, the T-version is not. [The latter concurs with Eguchi and Copas (1998), page 721.] It also seems aesthetically attractive that $\gamma_0 = 0$ for each of C- and U-versions, but only to decrease the effect of "additional" bias terms in the case of an odd number of parameters (and the additional bias may, at least for some x, be beneficial!).

A more complete comparison of choices of ξ comes from combining the small h results here with the large h results of Eguchi and Copas (1998). We believe that good choices should perform well both when the situation is "near-parametric" and when it is not. We therefore discount the T-version even though Eguchi and Copas show it has good large h properties. U- and C- (and T-) versions have equivalent large h performance to $O(h^{-2})$ as $h \to \infty$ but not, it seems, to $O(h^{-4})$.

We finish on a speculative note. We suspect, from page 720 of Eguchi and Copas (1998), that improved large h performance relative to the respectable large h performance of the U-version of Hjort and Jones (1996) may require $\gamma(z) > 0$ for $z > 0$ (unlike the C-version). Combined with the possible small h preference for $\gamma(0) = 0$, we tentatively suggest the L-version (L for linear) in which $\beta(z) = 1 - z$, $\gamma(z) = z$ and $\xi(y, z)$ therefore has the attractively simple form $1 - z + y$. It follows that the L-version behaves as well as the U- and C-versions when h is small. And since the comparison of Eguchi and Copas [(1998), page 720] which suggests some superiority of the T-version over the U-version when h is large depends on a function $\eta(t)$ which is, for large h, the same for a comparison of the L-version with the U-version, the L-version seems as good as the T-version when h is large. It is beyond the scope of this paper to investigate the practical behavior of this theoretically promising method.

APPENDIX

PROOF OF THEOREM 2. First, define two vector-valued functions $\psi_{in}(t, x, \theta)$ for $i = 1, 2$ by

$$\psi_{1n}(t, x, \theta) = h^{-1/2} K\left(\frac{t - x}{h}\right) u(t, \theta),$$

$$\psi_{2n}(t, x, \theta) = h^{-1/2} \xi\left\{K\left(\frac{t - x}{h}\right), E_\theta K\left(\frac{X_1 - x}{h}\right)\right\} E_\theta\left\{K\left(\frac{X_1 - x}{h}\right) u(X_1, \theta)\right\}.$$

Write $\psi_i(t) \equiv \psi_{in}(t, x, \theta_h)$. Then we may write $(n/h)^{1/2}\Psi_n(\theta_h) = n^{-1/2} \times \sum_{i=1}^n \{\psi_1(X_i) - \psi_2(X_i)\}$. Since $E\{\psi_1(X_1) - \psi_2(X_1)\} = 0$ by the definition of θ_h we have

$$(n/h)\operatorname{var}\{\Psi_n(\theta_h)\} = E\{\psi_1(X_1) - \psi_2(X_1)\}\{\psi_1(X_1) - \psi_2(X_1)\}^t. \tag{A.1}$$

We use the facts that $f(x, \theta_h) = f(x) + O(h^2)$ and that $f(t, \theta_h) = f(x, \theta_h) + O(h)$ and $u(t, \theta_h) = u_0(t) + O(h)$ where $O(h)$ is uniform over $t : |t - x| \le Ch$ for any constant $C > 0$. Define a $p \times p$ matrix U and a vector-valued function $z_h(\cdot)$ by

$$U = \bigl(u_0, u_0^{(1)}, \dots, u_0^{(p-1)}/(p-1)!\bigr)$$
$$z_h(y)^t = \bigl(1, hy, \dots, (hy)^{p-1}\bigr),$$

where $u_0^{(k)} \equiv u_0^{(k)}(x)$ is the kth derivative of u_0 evaluated at x, the point of interest, which is fixed. Below we often omit "x" if a function is evaluated at x. With these notations and conventions we may write $u_0(x + hy) = Uz_h(y) + O(h^p)$. It is easy to see then that

$$E\psi_1(X_1)\psi_1(X_1)^t = fU\left\{\int z_h(y)z_h(y)^t K^2(y)\,dy\right\}U^t\{1 + O(h)\}. \tag{A.2}$$

Furthermore from the condition (2.4) we obtain

$$\begin{aligned} E\psi_1(X_1)\psi_2(X_1)^t &= \gamma_0 fU\left\{\int z_h(y)K^2(y)\,dy \int z_h(y)^t K(y)\,dy\right\}U^t\{1 + O(h)\} \\ &\quad + h\beta_0 f^2 U\left\{\int z_h(y)K(y)\,dy \right. \\ &\qquad \left. \times \int z_h(y)^t K(y)\,dy\right\}U^t\{1 + O(h)\}, \end{aligned} \tag{A.3}$$

$$E\psi_2(X_1)\psi_2(X_1)^t$$
$$= \{\gamma_0^2\kappa_0 f + h(\beta_0^2 + 2\beta_0\gamma_0)f^2\} \tag{A.4}$$
$$\times U\left\{\int z_h(y)K(y)\,dy \int z_h(y)^t K(y)\,dy\right\}U^t\{1 + O(h)\}.$$

Let D_h be a $p \times p$ diagonal matrix having h^r as its rth diagonal entry. Then, combining (A.2), (A.3) and (A.4) we get

$$(n/h)\,\mathrm{var}\{\Psi_n(\theta_h)\} = fUD_h\{M + O(h)\}D_hU^t\{1 + O(h)\}. \tag{A.5}$$

It may be proved in a similar fashion that

$$-Eh^{-1}\dot{\Psi}_n(\theta_h) = fUD_h\{N + O(h)\}D_hU^t\{1 + O(h)\}. \tag{A.6}$$

For (A.6) one needs to use the fact that $\dot{\xi}(y, z) = \beta'(z) + \gamma'(z)(y/z) - \gamma(z)(y/z^2)$ where $\dot{\xi}(y, z) = (\partial/\partial z)\xi(y, z)$. The matrix U is invertible since $u_0, u_0^{(1)}, \dots, u_0^{(p-1)}$ are linearly independent. Theorem 2 follows from (A.5) and (A.6) and the fact that $\dot{f}(x, \theta_h)^t = f(x, \theta_h)u(x, \theta_h)^t = f(x)e_0^tU^t + O(h)$. $\square$

PROOF OF THEOREM 3. For a nonnegative integer s, define

$$a_s = (s!)^{-1}f^{(s)}\mu_s + \sum_{k=0}^{s}\{(s-k)!\}^{-1}\nu_k^{(s-k)}\mu_{s-k}.$$

As in the text, we use $(v \times w)_s$ to denote $\sum_{k=0}^{s} v_k w_{s-k}$ for any sequences $\{v_k\}$ and $\{w_k\}$. Likewise, we write $(v \times w \times z)_s$ for the sum of $v_k\, w_l\, z_m$ over the triples (k, l, m) with $k + l + m = s$, and so on. Furthermore we write $(v^2)_s$ for $(v \times v)_s$, and $(v^i)_s$ for its obvious extension. Define

$$\zeta_{2,s} = \sum_{k=0}^{s}\{(s-k)!\}^{-1}(u_k f)^{(s-k)}\mu_{s-k}$$

and $\zeta_{4,s} = \sum_{k=0}^{s}\gamma_k(k!)^{-1}(a^k)_{s-k}$. Then we can write

$$c_r = (\zeta_3 \times a)_r - \big(\zeta_1 \times \zeta_4 \times (\zeta_2 + \zeta_3)\big)_r.$$

The theorem follows if we show that

$$\zeta_{3,k} = \zeta_{1,k}u_0 + \sum_{l=1}^{k} A_{k,l}u_0^{(l)} \tag{A.7}$$

under condition $\mathcal{C}_k$. For, if we assume $\mathcal{C}_r$, then (A.7) and $\mathcal{C}_k$ hold for all $k \le r$ and consequently $\zeta_{3,s} = 0$ for all $s \le r - 1$, from which we can write

$$c_r = \zeta_{3,r}f - \gamma_0\zeta_{1,r}u_0 f.$$

The theorem then follows if we plug the expression for $\zeta_{3,r}$, as given by (A.7) with $k = r$, into the above identity.

We now prove (A.7). Define

$$\zeta_{3,k,0} = \sum_{j=0}^{k} \{(k-j)!\}^{-1} (u_0 \nu_j)^{(k-j)} \mu_{k-j},$$

$$\zeta_{3,k,1} = \sum_{j=1}^{k} \{(k-j)!\}^{-1} \left(\sum_{s=1}^{j} u_s \nu_{j-s} \right)^{(k-j)} \mu_{k-j}.$$

We can write $\zeta_{3,k} = \zeta_{3,k,0} + \zeta_{3,k,1}$. Using the relation $(u_0 \nu_j)^{(k-j)} = \sum_{l=0}^{k-j} \binom{k-j}{l} \times u_0^{(l)} \nu_j^{(k-j-l)}$ and interchanging the order of summations, we find that

$$\zeta_{3,k,0} = \zeta_{1,k} u_0 + \sum_{l=1}^{k} A_{k,l} u_0^{(l)}. \tag{A.8}$$

Thus (A.7) follows if we show $\zeta_{3,k,1} = 0$. By an argument similar to that leading to (A.8) we can show

$$\zeta_{3,k,1} = \sum_{s=1}^{k} \zeta_{1,k-s} u_s + \sum_{s=1}^{k-1} \sum_{l=1}^{k-s} A_{k-s,l} u_s^{(l)}. \tag{A.9}$$

The first term on the right-hand side of (A.9) is zero by condition $\mathcal{C}_k$, and the second term can be written as $\sum_{s=1}^{k-1} \sum_{l=1}^{s} A_{s,l} u_{k-s}^{(l)}$ which is zero, too, by $\mathcal{C}_k$. $\square$

Acknowledgments. The authors thank two reviewers and an Associate Editor for helpful comments and suggestions on the article.

REFERENCES

APOSTOL, T. M. (1975). *Mathematical Analysis*, 2nd ed. Addison-Wesley, London.

COPAS, J. B. (1995). Local likelihood based on kernel censoring. *J. Roy. Statist. Soc. Ser. B* **57** 221–235.

EGUCHI, S. and COPAS, J. B. (1998). A class of local likelihood methods and near-parametric asymptotics. *J. Roy. Statist. Soc. Ser. B* **60** 709–724.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

HJORT, N. L. and JONES, M. C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24** 1619–1647.

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407.

KIM, W. C., PARK, B. U. and KIM, Y. G. (2001). On Copas' local likelihood density estimator. *J. Korean Statist. Soc.* **30** 77–87.

LOADER, C. R. (1996). Local likelihood density estimation. *Ann. Statist.* **24** 1602–1618.
RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

B. U. PARK
W. C. KIM
DEPARTMENT OF STATISTICS
SEOUL NATIONAL UNIVERSITY
SEOUL 151-747
KOREA
E-MAIL: bupark@stats.snu.ac.kr

M. C. JONES
DEPARTMENT OF STATISTICS
THE OPEN UNIVERSITY
MILTON KEYNES MK7 6AA
UNITED KINGDOM