On Simultaneous Estimation of a Mean Response and its Derivatives

Richard Charnigo* and Cidambi Srinivasan†

University of Kentucky

Abstract

Standard nonparametric regression methods have the rather peculiar feature that the estimator of the mean response depends on whether we wish to recover derivatives of the mean response and, if so, how many. For instance, we may use a smoothing spline of order 5 if only one derivative is to be estimated, whereas a smoothing spline of order 9 may be employed if we are interested in three derivatives. This paper develops an "extended compound estimator" that can be differentiated any number of times and whose derivatives — all of them! — are consistent for the derivatives of an analytic mean response. The extended compound estimator is "invariant" to the number of derivatives to be recovered: its definition is the same whether or not we are interested in the first derivative, the third derivative, or any other derivative. The extended compound estimator is well suited to applications in which the number of derivatives to be estimated is determined only after data analysis begins.

Key words: Compound estimator, extended compound estimator, inductive estimators, invariance, nonparametric regression, self-consistency.

---

*Corresponding author. Postal address: 851 Patterson Office Tower, Department of Statistics, University of Kentucky, Lexington KY 40506-0027. E-mail address: RJCharn2@aol.com.

†Postal address: 869 Patterson Office Tower, Department of Statistics, University of Kentucky, Lexington KY 40506-0027. E-mail address: srini@ms.uky.edu.

1 Introduction

Consider the nonparametric regression model

$$Y_i = \mu(x_i) + \epsilon_i \quad \text{for } i \in \{1, \ldots, n\}, \tag{1}$$

where the x_i belong to a compact interval $\mathcal{X} \subset \mathbb{R}$, the ϵ_i are independent zero-mean random errors with variances bounded above by a positive constant M, and the mean response $\mu(x)$ is analytic. (A formal definition of analyticity is given in Section 2.) We assume without loss of generality that $\mathcal{X} = [-1, 1]$, and we write $\mu^{(j)}(x)$ for $\frac{d^j}{dx^j}\mu(x)$.

A problem of practical importance is the simultaneous estimation of $\mu(x)$ and its derivatives. Statistical modeling based on derivatives is natural in applications involving velocities, accelerations, and differential equations (Ramsay and Silverman, 2002). Derivatives are also useful in pattern recognition. For instance, although scattering profiles for evanescent waves and surface plasmons are useful in characterizing nanoparticles of unknown configuration (Francoeur et al, 2007), the derivatives of scattering profiles lead to more accurate characterizations (Charnigo et al, 2007).

The goal of this article is to develop a consistent estimator of $\mu(x)$ that can be differentiated any number of times to yield a consistent estimator of the corresponding derivative of $\mu(x)$. Thus, the estimator of $\mu(x)$ and its companion estimators of the derivatives are "self-consistent" in the sense of Property P1 below. Moreover, we wish for these estimators to be "invariant" as in Property P2.

Property P1: Self-Consistency. An estimator $\widehat{\mu}(x)$ and companion estimators $\widehat{\mu^{(j)}(x)}$, $j \in \mathcal{J}$, are self-consistent if $\frac{d^j}{dx^j}\widehat{\mu}(x)$ exists and equals $\widehat{\mu^{(j)}(x)}$ for every $j \in \mathcal{J}$, where $\mathcal{J} \subset \mathbb{N}$.

Property P2: Invariance. An estimator of $\mu^{(j)}(x)$, $j \geq 0$, is invariant if, for every positive integer k, the estimator of $\mu^{(j)}(x)$ does not depend on whether we calculate an estimator of $\mu^{(j+k)}(x)$.

Property P1 and Property P2 are not satisfied by standard nonparametric regression methods. Consider spline smooths (Schoenberg, 1964; Silverman, 1984; Wahba, 1990, and references therein), and suppose that initially we are interested in estimating $\mu^{(1)}(x)$. We may accomplish this using the first derivative of a spline smooth of order 5. However, suppose that estimation of $\mu^{(3)}(x)$ is requested subsequently. The third derivative of a spline smooth of order 5 may be too noisy, so we may employ the third derivative of a spline smooth of order 9 instead. Now we must make a decision: do we retain the original estimate of $\mu^{(1)}(x)$ from the spline smooth of order 5, or do we adopt a new estimate based on the spline smooth of order 9? The former option violates Property P1, while the latter choice — perhaps the preference of most data analysts — violates Property P2. We cannot really circumvent this predicament by starting with a spline smooth of order 9: someone may still ask us to estimate $\mu^{(9)}(x)$ or $\mu^{(27)}(x)$. Likewise, kernel smooths based on piecewise polynomial kernel functions (Nadaraya, 1964; Watson, 1964; Gasser and Muller, 1984; Hardle, 1990, and references therein) and local regression smooths (Cleveland, 1979; Stone, 1980; Loader, 1999, and references therein) do not satisfy Property P1 and Property P2.

Besides possessing conceptual appeal, self-consistency and invariance are important properties for data analysis. Having $\frac{d^j}{dx^j}\widehat{\mu}(x) \neq \widehat{\mu^{(j)}(x)}$ may yield irreconcilably conflicting conclusions if, say, we are interested in pinpointing a local maximum of $\mu^{(1)}(x)$: without self-consistency, there is no guaranteee that a local maximum of $\widehat{\mu^{(1)}(x)}$ will correspond to a stationary point of $\widehat{\mu^{(2)}(x)}$. Allowing $\widehat{\mu^{(j)}(x)}$ to depend on whether we estimate $\mu^{(j+k)}(x)$ is problematic because of the increased computational burden: without invariance, we may have to start from scratch in estimating $\mu^{(j)}(x)$ each time an estimate of a higher order derivative is requested.

We pursue our goal of consistent, self-consistent, and invariant estimation of $\mu(x)$ and its derivatives in two steps. First, we obtain invariant "local estimators" of $\mu(x)$ by solving calculus of variations problems. Each local estimator approximates $\mu(x)$ well over a small subinterval of $[-1, 1]$. Second, we knit the invariant local estimators into a "global estimator" of $\mu(x)$ (later referred to as the "extended compound estimator") that approximates $\mu(x)$ well over the full interval $[-1, 1]$. The global estimator inherits the invariance of the local estimators, and self-consistency is automatic because we define the corresponding global estimator of $\mu^{(j)}(x)$ to be the j^{th} derivative of the global estimator of $\mu(x)$.

The global estimator of $\mu^{(j)}(x)$ is, in fact, consistent for each $j \geq 0$. No similar statement applies to spline or local regression smooths. Local regression does not even define an estimator of $\mu^{(j)}(x)$ when j is greater than the degree of the local polynomial. If j is specified a priori (and if j is not too large relative to the sample size), then we can choose the degree of the local polynomial to guarantee that an estimator of $\mu^{(j)}(x)$ is defined. Yet, whatever degree is chosen for the local polynomial, infinitely many derivatives of $\mu(x)$ remain inestimable.

The rest of this paper is organized as follows. Section 2 derives the invariant local estimators of $\mu(x)$. Section 3 discusses the penalty terms in the calculus of variations problems and addresses computational considerations for the local estimators. Section 4 employs the local estimators to construct the global estimator of $\mu(x)$, whose derivatives are then shown to be consistent for the derivatives of $\mu(x)$. The convergence rates for the global estimator and its derivatives are slower than the optimal rates discovered by Stone (1980), when only finitely many derivatives are to be estimated and when this finite number is specified a priori. Whether better convergence rates than ours are possible, when $\mu(x)$ and all of its derivatives are to be consistently estimated in a self-consistent and invariant manner, remains an open question. Section 5 presents results from simulation studies assessing the global estimator and its derivatives. Section 6 provides closing comments about the simultaneous estimation of a mean response and its derivatives, including an analogy to follow-up comparisons in a one-way analysis of variance.

2 Invariant Local Estimation

Let a be an arbitrary but fixed point in $[-(1-\zeta),(1-\zeta)]$ for a small $\zeta > 0$. We seek a local estimator of the form

$$\sum_{j=0}^{\infty} \widetilde{c}_{j;a}(x-a)^j. \tag{2}$$

For (2) to be a reasonable estimator of $\mu(x)$ in a neighborhood of a, $\mu(x)$ should be analytic in the sense of Property P3 below and $\widetilde{c}_{j;a}$ should consistently estimate $\mu^{(j)}(a)/j!$ for each $j \geq 0$.

Property P3: Analyticity. A function $\nu(x)$ defined on $\mathcal{D} \subset \mathbb{R}$ is analytic if there exists an open interval $\mathcal{O} \subset \mathbb{R}$ such that $\mathcal{D} \subset \mathcal{O}$ and there exists a function $\bar{\nu}(x)$ defined on $\mathcal{O}$ whose restriction to $\mathcal{D}$ is $\nu(x)$ and for which

$$\bar{\nu}(x) = \sum_{j=0}^{\infty} \nu^{(j)}(b)(x-b)^j/j! \quad \text{for all } x \in \mathcal{O} \text{ and } b \in \mathcal{D}.$$

We may obtain $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$ from any existing nonparametric regression technique that consistently estimates $\mu(a)$ and $\mu^{(1)}(a)$. For instance, we may acquire $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$ from local regression with a local polynomial of degree $P > 1$. In this context, our reliance on local regression does not violate the invariance property because the definition of $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$ does not limit (to P or to any specific positive integer) the number of derivatives of $\mu(x)$ for which global estimators will be available in Section 4.

Let $\{h_j\}_{j=2}^{\infty}$ be a sequence of positive numbers. This sequence is in j, not in n, though each member of the sequence may depend on n. For example, we may put $h_2 := \kappa n^{-2/9}$ for a positive constant κ. Let

$$\widetilde{c}_{j;a} := \frac{n^{-1}\sum_{i=1}^{n}[Y_i - \widetilde{c}_{0;a} - (x_i - a)\widetilde{c}_{1;a}]\{h_j j(j-1) d_{j,i,h_j;a} - d_{j-2,i,h_j;a}\} + (\widetilde{c}_{j-2;a}/2)\ 1_{j\geq 4}}{(h_j/4)j(j-1)} \tag{3}$$

for $j \geq 2$, where

$$d_{j,i,h;a} := (2\pi h)^{-1/2} \exp\left[-\frac{(x_i-a)^2}{2h}\right] \frac{1}{j!}\ H_j\left(\frac{(x_i-a)}{\sqrt{2h}}\right)\ (2h)^{-j/2} \tag{4}$$

for $j \geq 0$. Above,

$$H_j(s) := (-1)^j \exp\left[s^2\right] \frac{d^j}{ds^j} \exp\left[-s^2\right] \tag{5}$$

is the j^{th} Hermite polynomial, defined for $s \in \mathbb{R}$.

We refer to the $\widetilde{c}_{j;a}$, $j \geq 2$, as "inductive estimators". The inductive estimators are invariant: formula (3) does not involve any parameter that limits the number of derivatives of $\mu(x)$ for which global estimators will be available in Section 4. Note that $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$ must be available before the inductive estimators can be computed. This is why an existing nonparametric regression technique must first be employed to obtain $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$.

Two questions arise. First, where do the inductive estimators in (3) come from? In other words, to what optimization problem do they constitute a (modified) solution? Second, do the inductive estimators actually work? That is, can we specify $\{h_j\}_{j=2}^{\infty}$ so that $\mu^{(j)}(a)/j!$ is consistently estimated by $\widetilde{c}_{j;a}$ for each $j \geq 2$? The developments in Sections 2.1 and 2.2 will answer both of these questions.

2.1 Formulating and solving an optimization problem

Until stated otherwise, we fix $a := 0$ without loss of generality. Let $C^2[-1,1]$ denote the normed vector space of twice continuously differentiable functions defined on $[-1,1]$ with $||u|| := \sup_{x\in[-1,1]} |u(x)| +$

$\sup_{x\in[-1,1]} |u'(x)| + \sup_{x\in[-1,1]} |u''(x)|$. We consider minimizing the functional

$$J[u] := \int_{-1}^{1} \sum_{i=1}^{n} w_{i;h}(x) \left[Y_i - u(x) - (x_i - x)u'(x)\right]^2 \, dx + \lambda \int_{-1}^{1} u'(x)^2 \, dx \tag{6}$$

with respect to $u \in C^2[-1,1]$, subject to the constraints that

$$K_0[u] := u(0) = \widetilde{c}_{0;0} \quad \text{and} \quad K_1[u] := u'(0) = \widetilde{c}_{1;0}. \tag{7}$$

Above, λ is a real-valued (not necessarily nonnegative) "penalty parameter", h is a positive "bandwidth-type parameter", and the $w_{i;h}(x)$ are "weight functions".

Our first theorem provides conditions under which the Euler-Lagrange equation (Pierre, 1986; Smith, 1998) for the minimization problem (6)-(7) admits a unique solution u^* that minimizes $J[u]$ among all $u \in C^2[-1,1]$ satisfying the constraints in (7) and $u(\pm 1) = u^*(\pm 1)$. The proofs of this result and all other technical results are in the Appendix.

Theorem 2.1 *Consider model (1) and the minimization problem (6)-(7). Suppose that the weight functions are analytic in* x *and satisfy* $\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 + \lambda > 0$, $\sum_{i=1}^{n} \{2w_{i;h}(x) - w'_{i;h}(x)\,(x_i - x)\} > 0$*. Then the Euler-Lagrange equation for the minimization problem (6)-(7) has a unique solution* u^* *with*

$$J[u] - \sum_{i=1}^{n} w_{i;h}(1)\;(x_i - 1)\left\{u(1) - u^*(1)\right\}^2 + \sum_{i=1}^{n} w_{i;h}(-1)\;(x_i + 1)\left\{u(-1) - u^*(-1)\right\}^2 \ge J[u^*] \tag{8}$$

for any $u \in C^2[-1,1]$ *satisfying (7). In particular, if* $u(\pm 1) = u^*(\pm 1)$*, then* $J[u] \ge J[u^*]$.

So far, no explicit restrictions have been placed on the x_i. However, to proceed further, we now specify that the x_i are either: (i) fixed and equispaced on $[-1,1]$; or, (ii) realizations of a random process for which Appendix formulas (33) and (36) are satisfied. In addition, we specify that

$$w_{i;h}(x) := (2\pi h)^{-1/2} \exp\left[-\frac{(x - x_i)^2}{2h}\right] \quad \text{and} \quad \lambda := -\frac{nh}{4}. \tag{9}$$

The specifications in (9), particularly the negative penalty parameter, are discussed in Section 3.1.

With the above specifications, the Euler-Lagrange equation for (6)-(7) determines the following inductive formulas for estimators of $\mu^{(j)}(0)/j!$, $j \ge 2$.

Theorem 2.2 *Consider model (1) and x_i either: (i) fixed and equispaced on $[-1,1]$; or, (ii) realizations of a random process for which Appendix formulas (33) and (36) are satisfied. Let the $w_{i;h}(x)$ and λ be as in (9). Put $\widehat{c}_{0;0} := \widetilde{c}_{0;0}$ and $\widehat{c}_{1;0} := \widetilde{c}_{1;0}$. Then the solution to the Euler-Lagrange equation for the minimization problem (6)-(7) has the form $u^*(x) = \sum_{j=0}^{\infty} \widehat{c}_{j;0}\ x^j$, where*

$$\widehat{c}_{2;0} := \frac{b_0^3 - [\widehat{c}_{0;0} b_0^2 + \widehat{c}_{1;0} b_0^1]}{2(b_0^0 - nh/4)} \quad \textit{and, for } j \geq 3, \tag{10}$$

$$\widehat{c}_{j;0} := \frac{b_{j-2}^3 - [\widehat{c}_{0;0} b_{j-2}^2 + \widehat{c}_{1;0}\{b_{j-2}^1 + b_{j-3}^2\} + \sum_{k=2}^{j-2} \widehat{c}_{k;0} t_k + \widehat{c}_{j-1;0}\{(j-2)(j-1) b_1^0 + (j-1) b_0^1\}]}{(j-1)j(b_0^0 - nh/4)}, \tag{11}$$

with the t_k, b_j^3, b_j^2, b_j^1, and b_j^0 as defined in Appendix formulas (30) to (32).

Remark 1. We use $\widehat{}$ notation for the estimators in Theorem 2.2 instead of $\widetilde{}$ notation because we do not adhere strictly to u^* when defining the inductive estimators in (3). The modifications to u^* and the rationale for these modifications are described in Section 2.2.

Remark 2. Inequality (8) is neither required for nor implied by Theorem 2.2. Under the assumptions of Theorem 2.2, we have

$$\sum_{i=1}^{n} w_{i;h}(x)\ (x_i - x)^2 + \lambda = \frac{nh}{4}\{1 + o(1)\} \quad \text{and} \quad \sum_{i=1}^{n}\{2w_{i;h}(x) - w'_{i;h}(x)\ (x_i - x)\} = \frac{n}{2}\{1 + o(1)\}$$

for each $x \in (-1,1)$, provided that $h \to 0$ and $nh^{5/2} \to \infty$ as $n \to \infty$. The $o(1)$ terms are uniform over $x \in [-(1-\zeta), (1-\zeta)]$, so conditions $\sum_{i=1}^{n} w_{i;h}(x)\ (x_i - x)^2 + \lambda > 0$ and $\sum_{i=1}^{n}\{2w_{i;h}(x) - w'_{i;h}(x)\ (x_i - x)\} > 0$ from Theorem 2.1 will hold on $[-(1-\zeta), (1-\zeta)]$ when n is sufficiently large. Even if these conditions fail near the boundaries ±1, application of the series method to solve the Euler-Lagrange equation in Appendix formula (26) — and, hence, to derive formulas (10) and (11) — is still permitted because the coefficient of $u^{*\prime\prime}(x)$ in the Euler-Lagrange equation only needs to be positive in a neighborhood of 0. The upper bound for $J[u^*]$ stated in (8) may not hold if the above conditions fail near the boundaries ±1. However, if the functional (6) is altered so that the integrals range from $-(1-\zeta)$ to $+(1-\zeta)$, then for sufficiently large n the upper bound for $J[u^*]$ stated in (8) will hold with ±1 replaced by $\pm(1-\zeta)$.

2.2 Consistency through a modified solution to the optimization problem

We make two modifications to the solution (10)-(11) of the Euler-Lagrange equation for (6)-(7). The first modification enables us to obtain consistent estimators of $\mu^{(j)}(0)/j!$, $j \geq 2$.

Modification #1. Allow h to change with each inductive step. More specifically:

(a) Put $\check{c}_{0;0} := \widehat{c}_{0;0}$ and $\check{c}_{1;0} := \widehat{c}_{1;0}$.

(b) Set $h := h_2$ when applying (10) to estimate $\mu^{(2)}(0)/2!$, and let the resulting estimator be denoted $\check{c}_{2;0}$.

(c) Set $h := h_3$ when applying (11) to estimate $\mu^{(3)}(0)/3!$, but use $\check{c}_{2;0}$ rather than $\widehat{c}_{2;0}$ in the calculation, and let the resulting estimator be denoted $\check{c}_{3;0}$.

(d) Continue in similar fashion, setting $h := h_j$ when applying (11) to estimate $\mu^{(j)}(0)/j!$, $j \geq 4$, but use $\check{c}_{2;0}, \ldots \check{c}_{j-1;0}$ rather than $\widehat{c}_{2;0}, \ldots, \widehat{c}_{j-1;0}$ in the calculation, and let the resulting estimator be denoted $\check{c}_{j;0}$.

An intuitive explanation for the first modification is that estimation of $\mu^{(j)}(0)/j!$ should require a larger value of the bandwidth-type parameter when j is large than when j is small: we should be able to rely heavily on data with $x_i \approx 0$ to estimate a lower order derivative at 0, but we should use more information to estimate a higher order derivative at 0.

The next theorem shows how to choose $\{h_j\}_{j=2}^{\infty}$ to obtain consistent estimators of $\mu^{(j)}(0)/j!$, $j \geq 2$. In fact, formulas (46) and (47) in the proof shed more light on the first modification: the bias of order h_{j-2} in estimating $\mu^{(j-2)}(0)/(j-2)!$, $j \geq 4$, propagates to the estimation of $\mu^{(j)}(0)/j!$ after multiplication by a factor of order h_j^{-1}. Thus, consistent estimation of $\mu^{(j)}(0)/j!$, $j \geq 4$, requires that $h_{j-2}h_j^{-1} \to 0$ as $n \to \infty$. This rules out using a single value of the bandwidth-type parameter in all inductive steps.

Theorem 2.3 *Consider model (1) and x_i either: (i) fixed and equispaced on $[-1, 1]$; or, (ii) realizations of a random process for which Appendix formulas (33) and (36) are satisfied. Let the $\check{c}_{j;0}$, $j \geq 0$, be as* *defined in Modification #1. Suppose that*

$$h_j \propto n^{-\left(\frac{2}{9}\right)\left(\frac{2}{3}\right)^{j/2}}$$

for $j \geq 2$*. If* $MSE[\check{c}_{0;0}] = O\left(n^{-1+2\epsilon}\right)$ *and* $MSE[\check{c}_{1;0}] = O\left(n^{-1+2\epsilon}\right)$ *for some* $\epsilon \in (0, 1/22)$*, then*

$$\check{c}_{j;0} - \mu^{(j)}(0)/j! = O_p\left(n^{-\left(\frac{2}{9}\right)\left(\frac{2}{3}\right)^{(j-2)}}\right).$$

The second modification to the solution (10)-(11) allows us to obtain slightly better convergence rates than those in Theorem 2.3.

Modification #2. Simplify the inductive formulas by removing terms that increase variability without a corresponding reduction in bias. More specifically:

For $j \geq 2$, let

$$\widetilde{c}_{j;0} := \frac{n^{-1}\sum_{i=1}^{n}[Y_i - \widetilde{c}_{0;0} - x_i\widetilde{c}_{1;0}]\{h_j j(j-1)d_{j,i,h_j;0} - d_{j-2,i,h_j;0}\} + (\widetilde{c}_{j-2;0}/2)\ 1_{j\geq 4}}{(h_j/4)j(j-1)}, \qquad (12)$$

where $d_{j,i,h_j;0}$ and $d_{j-2,i,h_j;0}$ are as defined in (4).

The following theorem shows how to choose $\{h_j\}_{j=2}^{\infty}$ so that, with the second modification, the exponent in the convergence rate is multiplied by (1/2) whenever j increases by 2, compared to multiplication by (4/9) in Theorem 2.3.

Theorem 2.4 *Consider model (1) and* x_i *either: (i) fixed and equispaced on* $[-1, 1]$*; or, (ii) realizations of a random process for which Appendix formulas (33) and (36) are satisfied. Let the* $\widetilde{c}_{j;0}$*,* $j \geq 2$*, be as defined in Modification #2. Suppose that*

$$h_2 \propto n^{-2/9}, \quad h_3 \propto n^{-2/11}, \quad \text{and} \ \ h_j \propto h_{j-2}^{1/2} \ \text{for } j \geq 4.$$

If $MSE[\widetilde{c}_{0;0}] = O\left(n^{-1+2\epsilon}\right)$ *and* $MSE[\widetilde{c}_{1;0}] = O\left(n^{-1+2\epsilon}\right)$ *for some* $\epsilon \in (0, 1/22)$*, then*

$$\widetilde{c}_{j;0} - \mu^{(j)}(0)/j! = O_p\left(n^{-\left(\frac{2}{9+2(j \mod 2)}\right)\left(\frac{1}{2}\right)^{\lfloor j/2\rfloor - 1}}\right)$$

for $j \geq 2$*, where* $\lfloor\cdot\rfloor$ *returns the integer part of its argument.*

Remark 3. As the $\widetilde{\ }$ notation in (12) suggests, and as one confirms by observing the equivalence of (12) to (3) in the special case that $a := 0$, we have now arrived at our desired recipe for the inductive estimators. Although Theorem 2.4 as written pertains to $a := 0$, the conclusion of Theorem 2.4 holds for a generic $a \in [-(1-\zeta), (1-\zeta)]$ with $\widetilde{c}_{j;a}$ and $\mu^{(j)}(a)/j!$ substituted for $\widetilde{c}_{j;0}$ and $\mu^{(j)}(0)/j!$, assuming that analogues to Appendix formulas (33) and (36) are satisfied for such a if the x_i are not fixed and equispaced on $[-1, 1]$.

Remark 4. Besides improving upon the convergence rates in Theorem 2.3, the second modification reduces computational effort. Calculating $\widetilde{c}_{j;a}$ for a given $j \geq 2$ requires only $\widetilde{c}_{0;a}$, $\widetilde{c}_{1;a}$, and $\widetilde{c}_{j-2;a}$ rather than all of $\widetilde{c}_{0;a}, \widetilde{c}_{1;a}, \ldots, \widetilde{c}_{j-1;a}$.

Remark 5. The rates at which h_2, h_3, etc., should approach 0 do not dictate how they should be chosen for a given data set, except perhaps to suggest $h_j \leq h_{j+1}$ for any $j \geq 2$. However, a potentially useful heuristic based on (9) is to view $\sqrt{h_j}$ as a standard deviation. For instance, choosing $h_2 := 1/100$ suggests that, when we are estimating $\mu^{(2)}(0)/2!$, we plan to rely mainly on those observations for which $|x_i| \leq 2\sqrt{1/100} = 0.2$. A formal method for choosing h_2, h_3, etc., will be pursued in another paper.

3 Issues in Invariant Local Estimation

Before proceeding to construct a global estimator of $\mu(x)$ in Section 4, we address some issues with the local estimators. Section 3.1 explains why the penalty parameter in the minimization problem (6)-(7) must be negative if we wish to obtain consistent estimators of $\mu^{(j)}(a)/j!$, $j \geq 2$ and $a \in [-(1-\zeta), (1-\zeta)]$. In particular, the fact that the penalty parameter in (9) is negative does not imply that the weight functions in (9) are somehow pathological. Section 3.2 describes how a data analyst can handle the "∞" in (2) and provides convenient updating formulas for the inductive estimators.

3.1 Negativity of the penalty parameter

The specifications of $w_{i;h}(x)$ and λ in (9) ultimately yield consistent estimators of $\mu^{(j)}(a)/j!$, $j \geq 2$ and $a \in [-(1-\zeta), (1-\zeta)]$, as indicated by Theorem 2.4 and Remark 3. However, one may wonder what would have happened with different choices for $w_{i;h}(x)$ and λ. In particular, one may ask whether consistent estimation would have been possible with $\lambda \geq 0$, if different weight functions had been used.

We now show that, under some assumptions about the weight functions, there is essentially only one way to choose λ so that consistent estimation of $\mu^{(j)}(a)/j!$, $j \geq 2$ and $a \in [-(1-\zeta), (1-\zeta)]$, is possible. That choice of λ depends on the weight functions but is always negative and proportional to nh.

Assume that

$$w_{i;h}(x) = Q\left(\frac{x - x_i}{\sqrt{h}}\right)/\sqrt{h}, \tag{13}$$

where $Q(s)$ is a nonnegative analytic function of $s \in \mathbb{R}$ satisfying

$$Q(s) = Q(-s) \quad \text{and} \quad \int_{-\infty}^{\infty} \left| \frac{d^{k_1}}{ds^{k_1}} Q(s)\ s^{k_2} \right| ds < \infty \text{ for any integers } k_1, k_2 \geq 0. \tag{14}$$

The specification of $w_{i;h}(x)$ in (9) has the form (13)-(14) with $Q(s) := (2\pi)^{-1/2} \exp[-s^2/2]$.

Condition 3.1 *Consider model (1) and* x_i *fixed, equispaced on* $[-1, 1]$. *Let the weight functions be as in (13)-(14). Suppose that we solve the Euler-Lagrange equation in Appendix formula (26) using the series method and allow* h *to change with each inductive step when calculating estimators of* $\mu^{(j)}(a)/j!$, $j \geq 2$ *and* $a \in [-(1-\zeta), (1-\zeta)]$. *Then a necessary condition for such estimators to be consistent is*

$$\lambda = -\frac{nh}{4} \int_{-\infty}^{\infty} Q(s)\ s^2\ ds\ \{1 + o(1)\}.$$

Having $\lambda < 0$ may seem counterintuitive, as penalty parameters in the statistical literature are generally positive. However, positive penalty parameters ordinarily accompany a non-integrated weighted sum of squares. These positive penalty parameters constrain the solution of the optimization problem to be smooth. In contrast, the negative penalty parameter in (6) accompanies an integrated weighted sum of squares. The integration of the weighted sum of squares already constrains the solution of the optimization problem to be smooth. The negative penalty parameter <u>relaxes</u> the solution of the optimization problem, allowing it to be less smooth.

Further insight can be gained by inspecting the Euler-Lagrange equation in Appendix formula (26). If $\lambda = 0$, then as $n \to \infty$ the Euler-Lagrange equation becomes $u^{*\prime\prime}(x) = o_p(1)$, so that $u^*(x) \approx a + bx$ for large n. Yet, having $u^*(x) \approx a + bx$ is unsatisfactory because this suggests near-zero estimates of

$\mu^{(j)}(0)/j!$, $j \geq 2$. With $\lambda < 0$, we look for $u^*(x)$ to be <u>less</u> smooth than a straight line, so that estimates of $\mu^{(j)}(0)/j!$, $j \geq 2$, are not forced to be near zero.

3.2 Computational considerations

A practical and theoretically defensible approach to handling the "∞" in (2) is to replace $\widetilde{c}_{j;a}$ with

$$\widetilde{c}_{j;a} 1_{n \geq N_j}, \tag{15}$$

where $\{N_k\}_{k=0}^{\infty}$ is a strictly increasing sequence of positive integers. In essence, this changes the "∞" in (2) to $K_{tr}(n) := \max\{k : n \geq N_k\}$. Importantly, $K_{tr}(n) \to \infty$ as $n \to \infty$, so that the theoretical developments in Section 4 are undisturbed.

Another practical item is that the inductive estimators have convenient updating formulas for new data. Suppose that we have calculated $\widetilde{c}_{j;a}$, $2 \leq j \leq K_{tr}(n_1)$ and $a \in [-(1-\zeta), (1-\zeta)]$, from a data set of size n_1 and that a new data set of size n_2 originating from the same mean response has arrived. If n_2 is not large relative to n_1, then we may be willing to retain the existing $\widetilde{c}_{0;a}$, $\widetilde{c}_{1;a}$, and $\{h_j\}_{j=2}^{K_{tr}(n_1)}$. In this case, we may update $\widetilde{c}_{j;a}$, $2 \leq j \leq K_{tr}(n_1)$ and $a \in [-(1-\zeta), (1-\zeta)]$, via

$$\begin{aligned}
\widetilde{c}_{j;a;NEW} \quad := \quad & \left(\frac{n_1}{n_1+n_2}\right)\left(\widetilde{c}_{j;a;OLD} - \frac{\widetilde{c}_{j-2;a;OLD}/2 \;\; 1_{j\geq 4}}{(h_j/4)j(j-1)}\right) \\
& + \left(\frac{1}{n_1+n_2}\right)\left(\frac{\sum_{i=n_1+1}^{n_1+n_2}[Y_i - \widetilde{c}_{0;a} - (x_i - a)\widetilde{c}_{1;a}]\{h_j j(j-1) d_{j,i,h_j;a} - d_{j-2,i,h_j;a}\}}{(h_j/4)j(j-1)}\right) \\
& + \left(\frac{\widetilde{c}_{j-2;a;NEW}/2 \;\; 1_{j\geq 4}}{(h_j/4)j(j-1)}\right).
\end{aligned} \tag{16}$$

The result of (16) is identical to what would be obtained from (3) if the full data set of size $n_1 + n_2$ had been processed at one time, apart from the fact that $\widetilde{c}_{0;a}$, $\widetilde{c}_{1;a}$, and $\{h_j\}_{j=2}^{K_{tr}(n_1)}$ might have been different if the full data set had arrived as a single batch.

4 Self-Consistent and Invariant Global Estimation

Having exhibited $\widetilde{c}_{j;a}$ that are invariant and consistent for $\mu^{(j)}(a)/j!$, $j \geq 0$ and $a \in [-(1-\zeta), (1-\zeta)]$, we now define a global estimator of $\mu(x)$ by

$$\mu^*(x) := \frac{\sum_{a \in I_n} \exp\left[-\beta_n (x-a)^2\right] \sum_{j=0}^{\infty} \widetilde{c}_{j;a}(x-a)^j}{\sum_{a \in I_n} \exp\left[-\beta_n (x-a)^2\right]}, \tag{17}$$

where I_n is a finite subset of $[-(1-\zeta),(1-\zeta)]$ and β_n is a positive tuning parameter. More specifically, let $\{M_n\}_{n=1}^{\infty}$ be a nondecreasing sequence of nonnegative integers. For a given n, partition $[-1,1]$ into $L_n := 3^{M_n}$ subintervals of equal length. Then let I_n be the set of all subinterval midpoints that fall inside $[-(1-\zeta),(1-\zeta)]$. The reason for defining I_n this way is so that $I_{n_1} \subset I_{n_1+n_2}$ for any positive integers n_1 and n_2, allowing us to exploit the updating formula (16).

The global estimator (17), constructed from the inductive estimators (3), is called the "extended compound estimator". The original "compound estimator" (Charnigo and Srinivasan, 2009) is

$$\mu_J^*(x) := \frac{\sum_{a\in I_n} \exp\left[-\beta_n(x-a)^2\right] \sum_{j=0}^{J} \widetilde{c}_{j;a}(x-a)^j}{\sum_{a\in I_n} \exp\left[-\beta_n(x-a)^2\right]}, \tag{18}$$

with J a positive integer and $\widetilde{c}_{j;a}$ obtained from a local regression smooth of degree J. The original compound estimator and its derivatives are consistent for $\mu(x)$ and its first $\lfloor J/2 \rfloor$ derivatives. The original compound estimator is not invariant because of its dependence on J.

The extended compound estimator, on the other hand, is invariant. Moreover, as the next theorem shows, every derivative of the extended compound estimator is consistent for the corresponding derivative of $\mu(x)$. Actually, the next theorem establishes more than this because it allows the global estimator (17) to be constructed from invariant $\widetilde{c}_{j;a}$ other than those in (3), which may arise if one solves the Euler-Lagrange equation (26) with $w_{i;h}(x)$ and λ chosen in a manner other than that indicated by (9) but still in accord with Condition 3.1.

Theorem 4.1 *Consider model (1) and the global estimator (17). Suppose that there exist a positive integer* $n_0 > 1$*, a nonincreasing sequence of positive numbers* $\{\alpha_j\}_{j=0}^{\infty}$*, and a sequence of positive numbers* $\{K_j\}_{j=0}^{\infty}$ *such that, for each* $j \geq 0$ *and* $n \geq n_0$*,*

$$\sup_{x\in I} \left|\mu^{(j)}(x)\right| / j! \leq K_0 2^{-j}, \tag{19}$$

$$\sup_{a\in I_n} MSE[\widetilde{c}_{j;a}] \leq K_j n^{-2\alpha_j}, \text{ and} \tag{20}$$

$$\sup_{a\in I_n} |\widetilde{c}_{j;a}| \leq K_0 2^{-j}. \tag{21}$$

Then there exist sequences $\{\beta_n\}_{n=n_0}^{\infty}$ *and* $\{M_n\}_{n=n_0}^{\infty}$ *such that, for each* $j \geq 0$*,*

$$\sup_{x\in I} \left|\frac{d^j}{dx^j}\mu^*(x) - \mu^{(j)}(x)\right| = O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right),$$

where $\xi \in (0,1)$ is arbitrary but fixed.

Remark 6. Implicit in condition (20) are assumptions about the x_i. If the inductive estimators in (3) are used, then the x_i should be: (i) fixed and equispaced on $[-1,1]$; or, (ii) realizations of a random process that satisfy Appendix formulas (33) and (36) as well as their analogues for a generic $a \in I_n$.

Remark 7. Condition (21) ensures that the infinite sums $\sum_{j=0}^{\infty} \widetilde{c}_{j;a}(x-a)^j$ converge (in the sense of calculus) because $|x-a| \leq 2(1-\zeta)$. Thus, estimator (17) is well-defined. Likewise, convergence of $\frac{d^k}{dx^k} \sum_{j=0}^{\infty} \widetilde{c}_{j;a}(x-a)^j$ is ensured for any $k > 0$, so that the derivatives of estimator (17) are well-defined. If not otherwise true, condition (21) can be enforced by truncating $\widetilde{c}_{j;a}$ at $\text{sign}\,[\widetilde{c}_{j;a}]\, K_0 2^{-j}$. Assuming that condition (19) holds, $\text{sign}\,[\widetilde{c}_{j;a}]\, K_0 2^{-j}$ cannot have larger mean square error than $\widetilde{c}_{j;a}$. Hence, condition (20) is preserved for $\text{sign}\,[\widetilde{c}_{j;a}]\, K_0 2^{-j}$ if satisfied by $\widetilde{c}_{j;a}$.

5 Empirical Assessment

Two simulation studies were conducted to assess the performance of the extended compound estimator. In Study #1, we used the oscillatory test function

$$\mu_O(x) := \cos(2\pi x) + \sin(2\pi x) + \cos(3\pi x) + \sin(3\pi x)$$

to generate 100 data sets of size $n = 125$ from model (1) with equispaced x_i and $\epsilon_i \overset{iid}{\sim} N(0, 0.2^2)$.

To each data set, we applied both the original compound estimator and the extended compound estimator. The original compound estimator was adopted as a benchmark since it outperformed local regression and spline smooths in the simulation studies of Charnigo and Srinivasan (2009).

The specifications for the original compound estimator were as follows: $J = 7$; $|I_n| = 27$; $\beta_n = 15$; $\widetilde{c}_{j;a}$ from local regression based on local polynomials of degree 7, rectangular weights, and nearest neighbor smoothing parameter 0.3, 0.4, or 0.5. Also, we carried out filtration and extrapolation as described in Charnigo and Srinivasan (2009).

The specifications for the extended compound estimator were as follows: $|I_n| = 27$; $\beta_n = 15$; $\widetilde{c}_{j;a}$ from (3) using adjustment (15) with $N_7 < 125 < 500 < N_8$; $h_2 = h_3 = 1/120$ and $h_4 = h_5 = h_6 = h_7 = 1/40$; $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$ from local regression with the specifications in the preceding paragraph. Again, we implemented filtration and extrapolation. Adjustment (15) with $N_7 < 125 < 500 < N_8$ ensured as fair as possible a comparison between the original compound estimator and the extended compound estimator.

We repeated all of the above with 100 data sets of size $n = 500$.

Study #2 was identical to Study #1 except that we used the monotone test function

$$\mu_M(x) := 2(2\pi)^{-1/2} \int_{-\infty}^{8x} \exp[-z^2/2]\ dz + 2\exp(x/2) - 3.$$

For evaluative purposes, we tabulated the average value over the 100 simulated data sets of

$$\frac{\sum_{i=1}^n \left[\mu^{(j)}(x_i)\right]^2 - \sum_{i=1}^n \left[\widehat{\mu^{(j)}(x_i)} - \mu^{(j)}(x_i)\right]^2}{\sum_{i=1}^n \left[\mu^{(j)}(x_i)\right]^2} \tag{22}$$

for each method, where $0 \le j \le 6$ and $\widehat{\mu^{(j)}(x)}$ is generic notation for an estimate of $\mu^{(j)}(x)$. We also recorded the average value over the 100 simulated data sets of

$$\sum_{j=0}^{6} \frac{\sum_{i=1}^n \left[\widehat{\mu^{(j)}(x_i)} - \mu^{(j)}(x_i)\right]^2}{\sum_{i=1}^n \left[\mu^{(j)}(x_i)\right]^2}, \tag{23}$$

the "rectangular Sobolev distance" (RSD) between $\widehat{\mu}(x)$ and $\mu(x)$ (Charnigo and Srinivasan, 2009).

Table 1 presents the results of Study #1. The extended compound estimator substantially outperforms the original compound estimator. When $n = 125$, the best RSD obtained with the extended compound estimator is 45% smaller than that acquired with the original compound estimator. When $n = 500$, the extended compound estimator provides a 47% reduction.

Table 2 displays the results of Study #2. Recovering derivatives of $\mu_M(x)$ is clearly more difficult than estimating derivatives of $\mu_O(x)$. Still, the extended compound estimator is competitive with the original compound estimator. When $n = 125$, the best RSD acquired with the extended compound estimator is 4% larger than that obtained with the original compound estimator. When $n = 500$, the extended compound estimator yields a 10% reduction.

6 Discussion

We have addressed the problem of simultaneously estimating a mean response $\mu(x)$ and its derivatives by presenting an extended compound estimator that can be differentiated any number of times and whose derivatives — all of them! — are consistent for the corresponding derivatives of $\mu(x)$. The extended compound estimator is invariant and hence well suited to applications in which we do not know a priori how many derivatives should be recovered.

An analogy can be drawn to the scenario of follow-up comparisons in a one-way analysis of variance. The Scheffe procedure allows a single critical value to be used for any number of follow-up comparisons, much like the extended compound estimator is a single entity that can (through repeated differentiations) be used to estimate any number of derivatives. The Bonferroni procedure may be used instead of and is more powerful than the Scheffe procedure if the number of follow-up comparisons is specified a priori and modest, much like $\mu_J^*(x)$ may be used instead of and is more efficient asymptotically than the extended compound estimator if the number of derivatives to be recovered is specified a priori and less than $\lfloor J/2 \rfloor$.

Since Stone (1980) established optimal convergence rates without constraints of self-consistency and invariance, at the present time there is no yardstick by which to fairly evaluate the convergence rates of the extended compound estimator and its derivatives. Whether there exist self-consistent and invariant estimators that achieve better convergence rates is an open question. If such estimators do exist, their form is not obvious.

Perhaps the first place one may think to look for them is among the family of kernel smooths. While a kernel smooth based on a piecewise polynomial kernel function is not invariant (because the order of the piecewise polynomial is chosen according to how many derivatives one plans to estimate), Proposition 3.1.2 of Hardle (1990) suggests that a kernel smooth of the form $n^{-1} \sum_{i=1}^{n} K\left((x - x_i)/h\right) Y_i$ may work if the kernel function $K(s)$ satisfies the following four conditions: (i) $K(s)$ is infinitely differentiable and supported on $[-1, 1]$; (ii) $\int K(s)\ ds = 1$; (iii) $\int K(s) s^j\ ds = 0$ for any positive integer j; and, (iv) $K^{(j)}(\pm 1) = 0$ for any nonnegative integer j. However, the first three conditions are incompatible. Condition (iii) implies that $sK(s)$ is orthogonal to all monomials on $[-1, 1]$. Condition (i) then implies that $sK(s)$ and hence $K(s)$ are identically 0 on $[-1, 1]$, rendering condition (ii) impossible.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. DMS-0706857. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science

Foundation. We thank Professor Brian Borchers for bringing reference [13] to our attention.

References

[1] Charnigo, R., Francoeur, M., Mengüç, M. P., Brock, A., Leichter, M., and Srinivasan, C. (2007). Derivatives of scattering profiles: tools for nanoparticle characterization. *Journal of the Optical Society of America A*, **24**: 2578–2589.

[2] Charnigo, R. and Srinivasan, C. (2009). Self-consistent estimation of mean response functions and their derivatives. Submitted for publication and available at {www.richardcharnigo.net/TechReports}.

[3] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**: 829–836.

[4] Francoeur, M., Venkata, P. G., and Mengüç, M. P. (2007). Sensitivity analysis for characterization of gold nanoparticles and 2D agglomerates via surface plasmon scattering patterns. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **106**: 44-55.

[5] Gasser, T. and Muller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandanavian Journal of Statistics*, **11**: 171–185.

[6] Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.

[7] Loader, C. (1999). *Local Regression and Likelihood*. Springer, New York.

[8] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**: 141–142.

[9] Pierre, D. A. (1986). *Optimization Theory with Applications*. Dover, New York.

[10] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.

[11] Schoenberg, I. J. (1964). Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences U.S.A.*, **52**: 947–950.

[12] Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, **12**: 898–916.

[13] Smith, D. R. (1998). *Variational Methods in Optimization.* Dover, New York.

[14] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**: 1348–1360.

[15] Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

[16] Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, **26**: 359–372.

Appendix

Proof of Theorem 2.1. Applying the Euler-Lagrange multiplier theorem (p. 78 of Smith, 1998), we find that any local extremum u^* of the functional $J[u]$ in (6) satisfying the constraints in (7) must abide

$$\delta J(u^*;\Delta u) = \lambda_0 \delta K_0(u^*;\Delta u) + \lambda_1 \delta K_1(u^*;\Delta u), \tag{24}$$

where $\delta J(u^*;\Delta u)$ is the variation of $J[u]$ from u^* in the direction of $\Delta u \in C^2[-1,1]$ and similarly for $\delta K_0(u^*;\Delta u)$, $\delta K_1(u^*;\Delta u)$. Note that λ_0 and λ_1 in relation (24) are distinct from λ in (6).

After evaluating the variations and integrating by parts, we can express relation (24) as

$$\begin{aligned}
- \quad & 4\int_{-1}^{1} \sum_{i=1}^{n} w_{i;h}(x)[Y_i - u^*(x) - (x_i - x){u^*}'(x)]\Delta u(x)\, dx \\
- \quad & 2\left\{ \sum_{i=1}^{n} w_{i;h}(1)[Y_i - u^*(1) - (x_i - 1){u^*}'(1)](x_i - 1) - \lambda {u^*}'(1) \right\} \Delta u(1) \\
+ \quad & 2\left\{ \sum_{i=1}^{n} w_{i;h}(-1)[Y_i - u^*(-1) - (x_i + 1){u^*}'(-1)](x_i + 1) - \lambda {u^*}'(-1) \right\} \Delta u(-1) \\
+ \quad & 2\int_{-1}^{1} \left\{ \sum_{i=1}^{n} w'_{i;h}(x)[Y_i - u^*(x) - (x_i - x){u^*}'(x)](x_i - x) \right. \\
& \qquad \left. - \sum_{i=1}^{n} w_{i;h}(x){u^*}''(x)(x_i - x)^2 - \lambda {u^*}''(x) \right\} \Delta u(x)\, dx \\
= \quad & \lambda_0 \Delta u(0) + \lambda_1 \Delta u'(0),
\end{aligned} \tag{25}$$

where $\Delta u'(x) := \frac{d}{dx}\Delta u(x)$.

For any $\Delta u \in C^2[-1,1]$ such that $\Delta u(0) = \Delta u'(0) = \Delta u(-1) = \Delta u(1) = 0$, relation (25) yields

$\int_{-1}^{1} g(x)\Delta u(x)\ dx = 0$ with

$$\begin{aligned} g(x) &:= 2\sum_{i=1}^{n} w_{i;h}(x)[Y_i - u^*(x) - (x_i - x)u^{*\prime}(x)] \\ &- \sum_{i=1}^{n} w'_{i;h}(x)[Y_i - u^*(x) - (x_i - x)u^{*\prime}(x)](x_i - x) \\ &+ \sum_{i=1}^{n} w_{i;h}(x)u^{*\prime\prime}(x)(x_i - x)^2 + \lambda u^{*\prime\prime}(x). \end{aligned}$$

A straightforward extension of the DuBois-Reymond lemma (p. 356 of Smith, 1998) then shows that $g(x)$ must be identically 0, yielding the Euler-Lagrange equation

$$\begin{aligned} & u^{*\prime\prime}(x)\left[\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 + \lambda\right] \\ + \;& u^{*\prime}(x)\left[\sum_{i=1}^{n}\left\{(x_i - x)^2 w'_{i;h}(x) - 2(x_i - x)w_{i;h}(x)\right\}\right] \\ + \;& u^*(x)\left[\sum_{i=1}^{n}\left\{(x_i - x)w'_{i;h}(x) - 2w_{i;h}(x)\right\}\right] \\ = \;& \sum_{i=1}^{n} Y_i\left\{(x_i - x)w'_{i;h}(x) - 2w_{i;h}(x)\right\}. \end{aligned} \tag{26}$$

Since the $w_{i;h}(x)$ are analytic and $\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 + \lambda > 0$, we can solve the ordinary differential equation (26) by the series method. The two requirements $u^*(0) = \widetilde{c}_{0;0}$ and $u^{*\prime}(0) = \widetilde{c}_{1;0}$ determine u^* uniquely because the ordinary differential equation (26) is of second order.

Now consider any $u \in C^2[-1,1]$ satisfying the constraints in (7). Put $\Delta u := u - u^*$, and note that $\Delta u(0) = \Delta u'(0) = 0$. Relation (24) implies that

$$J[u] - J[u^*] = \int_{-1}^{1}\sum_{i=1}^{n} w_{i;h}(x)\{\Delta u(x) + (x_i - x)\Delta u'(x)\}^2\ dx + \lambda\int_{-1}^{1}\{\Delta u'(x)\}^2\ dx. \tag{27}$$

Expanding the square within the first integral on the right side of (27) and integrating by parts, we obtain

$$\begin{aligned} J[u] - J[u^*] &= \sum_{i=1}^{n} w_{i;h}(1)\ (x_i - 1)\left\{u(1) - u^*(1)\right\}^2 - \sum_{i=1}^{n} w_{i;h}(-1)\ (x_i + 1)\left\{u(-1) - u^*(-1)\right\}^2 \\ &+ \int_{-1}^{1}\sum_{i=1}^{n}\{2w_{i;h}(x) - w'_{i;h}(x)\ (x_i - x)\}\{\Delta u(x)\}^2\ dx \qquad (28) \\ &+ \int_{-1}^{1}\left[\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 + \lambda\right]\{\Delta u'(x)\}^2\ dx. \qquad (29) \end{aligned}$$

Theorem 2.1 follows immediately since the assumptions $\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 + \lambda > 0$ and $\sum_{i=1}^{n}\{2w_{i;h}(x) - w'_{i;h}(x)\ (x_i - x)\} > 0$ imply the nonnegativity of quantities (28) and (29). ■

Proof of Theorem 2.2. Let

$$t_k := (k-1)kb^0_{j-k} + kb^1_{j-k-1} + b^2_{j-k-2}, \tag{30}$$

$$b^0_j := h^2 a^0_{j+2}(j+2)(j+1) + ha^0_j,\ b^1_j := h^2 a^0_{j+3}(j+3)(j+2)(j+1) + ha^0_{j+1}(j+1), \tag{31}$$

$$b^2_j := ha^0_{j+2}(j+2)(j+1) - a^0_j,\ b^3_j := ha^1_{j+2}(j+2)(j+1) - a^1_j, \tag{32}$$

$$a^l_j := \sum_{i=1}^{n} Y^l_i\ d_{j,i,h;0}, \quad \text{and} \quad d_{j,i,h;0} := (2\pi h)^{-1/2} \exp\left[-\frac{x_i^2}{2h}\right] \frac{1}{j!}\ H_j\left(\frac{x_i}{\sqrt{2h}}\right)\ (2h)^{-j/2}.$$

Above, the allowable values of j, k, and l are those for which the indexing subscripts and superscripts are nonnegative. Also, H_j is the j^{th} Hermite polynomial as defined in formula (5).

A Taylor expansion yields

$$w_{i;h}(x) = (2\pi h)^{-1/2} \exp\left[-\frac{x_i^2}{2h}\right] \sum_{j=0}^{\infty} x^j\ \frac{1}{j!}\ H_j\left(\frac{x_i}{\sqrt{2h}}\right)\ (2h)^{-j/2}.$$

Then, noting that

$$\sum_{i=1}^{n} w_{i;h}(x) = \sum_{j=0}^{\infty} a^0_j x^j, \quad \sum_{i=1}^{n} w_{i;h}(x) Y_i = \sum_{j=0}^{\infty} a^1_j x^j, \quad \sum_{i=1}^{n} w_{i;h}(x)(x_i - x) = \sum_{j=0}^{\infty} ha^0_{j+1}(j+1)x^j,$$

$$\sum_{i=1}^{n} w_{i;h}(x)(x_i - x)^2 = \sum_{j=0}^{\infty}\{h^2 a^0_{j+2}(j+2)(j+1) + ha^0_j\}x^j, \text{ and } \sum_{i=1}^{n} w_{i;h}(x)(x_i - x)Y_i = \sum_{j=0}^{\infty} ha^1_{j+1}(j+1)x^j,$$

we find that the coefficients and right-hand side of the Euler-Lagrange equation (26) are $\sum_{j=0}^{\infty}\{b^0_j - 1_{j=0}nh/4\}x^j$, $\sum_{j=0}^{\infty} b^1_j x^j$, $\sum_{j=0}^{\infty} b^2_j x^j$, and $\sum_{j=0}^{\infty} b^3_j x^j$ respectively. Applying the series method of solution to the Euler-Lagrange equation (26), we obtain

$$\sum_{j=0}^{\infty} x^j \left(\sum_{k=0}^{j} \widehat{c}_{k+2;0}(k+2)(k+1)\{b^0_{j-k} - 1_{j=k}nh/4\} + \widehat{c}_{k+1;0}(k+1)b^1_{j-k} + \widehat{c}_{k;0} b^2_{j-k}\right) = \sum_{j=0}^{\infty} x^j b^3_j.$$

Setting equal the coefficients for the powers of x yields the inductive formulas in Theorem 2.2. ■

Proof of Theorem 2.3. For ease of exposition, we divide the proof into three parts: Preliminary Computations, Convergence of $\check{c}_{2;0}$ and $\check{c}_{3;0}$, and Convergence of $\check{c}_{j;0}$ for $j \geq 4$.

Preliminary Computations. For any integers $p, j \geq 0$ we have

$$\begin{aligned}
& n^{-1} \sum_{i=1}^{n} x_i^p \; d_{j,i,h;0} \\
= \; & (2\pi h)^{-1/2} \, \frac{1}{j!} \, (2h)^{-j/2} \, (1/2) \int_{-1}^{1} x^p \; \exp\left[-\frac{x^2}{2h}\right] \; H_j\left(\frac{x}{\sqrt{2h}}\right) \, dx + R_1 \\
= \; & (2\pi h)^{-1/2} \, \frac{1}{j!} \, (2h)^{-(j-p-1)/2} \, (1/2) \int_{-\infty}^{\infty} s^p \; \exp[-s^2] \; H_j(s) \, ds + R_1 + R_2,
\end{aligned}$$

where the "Riemann approximation error" R_1 is bounded absolutely by

$$\begin{aligned}
& (2\pi h)^{-1/2} \, \frac{1}{j!} \, (2h)^{-j/2} \, (2/n) \sup_{x \in [-1,1]} \left| \frac{d}{dx} \left\{ x^p \; \exp\left[-\frac{x^2}{2h}\right] \; H_j\left(\frac{x}{\sqrt{2h}}\right) \right\} \right| \\
= \; & O\left(h^{-(j+2)/2} n^{-1}\right) \qquad (33)
\end{aligned}$$

and, for some positive constants C_1 and C_2, the "Laplace approximation error" R_2 is bounded absolutely by

$$\begin{aligned}
& (2\pi h)^{-1/2} \, \frac{1}{j!} \, (2h)^{-(j-p-1)/2} \int_{(2h)^{-1/2}}^{\infty} s^p \; \exp[-s^2] \; |H_j(s)| \, ds \\
\leq \; & C_1 h^{-C_2} \exp[-(2h)^{-1}] \\
= \; & o\left(h^{-(j+2)/2} n^{-1}\right). \qquad (34)
\end{aligned}$$

By orthogonality of the Hermite polynomial system,

$$(2\pi h)^{-1/2} \, \frac{1}{j!} \, (2h)^{-(j-p-1)/2} \, (1/2) \int_{-\infty}^{\infty} s^p \; \exp[-s^2] \, H_j(s) \, ds = \begin{cases} 0 & : \; p < j \\ 0 & : \; (p-j) \bmod 2 = 1 \\ (1/2) & : \; p = j \\ (h/4)(j+2)(j+1) & : \; p = (j+2) \\ O\left(h^k\right) & : \; p = (j+2k). \end{cases} \qquad (35)$$

Formulas (33), (34), and (35) characterize $n^{-1} \sum_{i=1}^{n} x_i^p \; d_{j,i,h;0}$ for any integers $p, j \geq 0$. For any integers $j_1, j_2 \geq 0$, similar calculations yield

$$n^{-1} \sum_{i=1}^{n} \; d_{j_1,i,h;0} \; d_{j_2,i,h;0} = O\left(h^{-(j_1+j_2+1)/2}\right). \qquad (36)$$

Convergence of $\check{c}_{2;0}$ and $\check{c}_{3;0}$. For any $j \geq 2$, n^{-1} times the denominator of $\check{c}_{j;0}$ is

$$j(j-1) n^{-1} \sum_{i=1}^{n} \{2h_j^2 d_{2,i,h_j;0} + h_j d_{0,i,h_j;0} - h_j/4\} = j(j-1)h_j/4 + O\left(h_j^2\right) + O\left(h_j^{-1} n^{-1}\right).$$

On the other hand, n^{-1} times the numerator of $\check{c}_{2;0}$ has expected value

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{\infty}(\mu^{(j)}(0)/j!)\ x_i^j\{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\} \\
& -E[\check{c}_{0;0}]\{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\} - E[\check{c}_{1;0}]\{6h_2^2 d_{3,i,h_2;0} + h_2 d_{1,i,h_2;0}\} \\
= \quad & -\mu(0)\ 1/2 + (\mu^{(2)}(0)/2!)\ h_2/2 + (\mu^{(4)}(0)/4!)\ O\left(h_2^2\right) + O\left(h_2^3\right) + O\left(h_2^{-1}n^{-1}\right) \\
& + \left[\mu(0) + O\left(n^{-1/2+\epsilon}\right)\right]\left\{1/2 + O\left(h_2^{-1}n^{-1}\right)\right\} + \left[\mu'(0) + O\left(n^{-1/2+\epsilon}\right)\right]\left\{O\left(h_2^{-1/2}n^{-1}\right)\right\} \\
= \quad & (\mu^{(2)}(0)/2!)\ h_2/2 + (\mu^{(4)}(0)/4!)\ O\left(h_2^2\right) + O\left(h_2^3\right) + O\left(n^{-1/2+\epsilon}\right) \\
= \quad & (\mu^{(2)}(0)/2!)\ h_2/2 + O\left(h_2^2\right).
\end{aligned}$$

Hence,

$$E[\check{c}_{2;0}] = \frac{(\mu^{(2)}(0)/2!)\ h_2/2 + O\left(h_2^2\right)}{h_2/2 + O\left(h_2^2\right) + O\left(h_2^{-1}n^{-1}\right)} = (\mu^{(2)}(0)/2!) + O\left(h_2\right). \tag{37}$$

Now consider

$$Var\left[\check{c}_{0;0}\ \frac{\sum_{i=1}^{n}\{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\}}{\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}}\right], Var\left[\check{c}_{1;0}\ \frac{\sum_{i=1}^{n}\{6h_2^2 d_{3,i,h_2;0} + h_2 d_{1,i,h_2;0}\}}{\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}}\right], \tag{38}$$

$$\text{and}\ \ Var\left[\frac{\sum_{i=1}^{n} Y_i\ \{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\}}{\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}}\right]. \tag{39}$$

The variances in (38) are

$$O\left(n^{-1+2\epsilon}\right)\frac{\left(n^{-1}\sum_{i=1}^{n}\{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\}\right)^2}{\left(n^{-1}\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}\right)^2} = O\left(h_2^{-2}n^{-1+2\epsilon}\right) \tag{40}$$

$$\text{and}\ O\left(n^{-1+2\epsilon}\right)\frac{\left(n^{-1}\sum_{i=1}^{n}\{6h_2^2 d_{3,i,h_2;0} + h_2 d_{1,i,h_2;0}\}\right)^2}{\left(n^{-1}\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}\right)^2} = O\left(h_2^{-1}n^{-1+2\epsilon}\right). \tag{41}$$

Since $Var[\epsilon_i] \le M < \infty$ in model (1), an upper bound for (39) is

$$Mn^{-1}\frac{n^{-1}\sum_{i=1}^{n}\{2h_2 d_{2,i,h_2;0} - d_{0,i,h_2;0}\}^2}{\left(n^{-1}\sum_{i=1}^{n} 2\{2h_2^2 d_{2,i,h_2;0} + h_2 d_{0,i,h_2;0} - h_2/4\}\right)^2} = O\left(h_2^{-5/2}n^{-1}\right). \tag{42}$$

The Cauchy-Schwarz inequality then yields

$$Var\left[\check{c}_{2;0}\right] \le O\left(h_2^{-5/2}n^{-1}\right) + O\left(h_2^{-2}n^{-1+2\epsilon}\right) + O\left(h_2^{-1}n^{-1+2\epsilon}\right) = O\left(h_2^{-5/2}n^{-1}\right). \tag{43}$$

Recalling that $h_2 \propto n^{-2/9}$ and applying the Markov inequality to (37) and (43), we see that $\check{c}_{2;0}$ is $n^{2/9}$-consistent for $\mu^{(2)}(0)/2!$.

Continuing, n^{-1} times the numerator of $\check{c}_{3;0}$ has expected value

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\sum_{j=0}^{\infty}(\mu^{(j)}(0)/j!)\ x_i^j\{6h_3 d_{3,i,h_3;0}-d_{1,i,h_3;0}\} \\
& -E[\check{c}_{0;0}]\{6h_3 d_{3,i,h_3;0}-d_{1,i,h_3;0}\}-E[\check{c}_{1;0}]\{24h_3^2 d_{4,i,h_3;0}+4h_3 d_{2,i,h_3;0}-d_{0,i,h_3;0}\} \\
& -E[\check{c}_{2;0}]\{24h_3^2 d_{3,i,h_3;0}+4h_3 d_{1,i,h_3;0}\} \\
= \; & -\mu'(0)\ 1/2+(\mu^{(3)}(0)/3!)\ 3h_3/2+(\mu^{(5)}(0)/5!)\ O\left(h_3^2\right)+O\left(h_3^3\right)+O\left(h_3^{-3/2}n^{-1}\right) \\
& +\left[\mu(0)+O\left(n^{-1/2+\epsilon}\right)\right]\left\{O\left(h_3^{-3/2}n^{-1}\right)\right\}+\left[\mu'(0)+O\left(n^{-1/2+\epsilon}\right)\right]\left\{1/2+O\left(h_3^{-1}n^{-1}\right)\right\} \\
& +\left[(\mu^{(2)}(0)/2!)+O\left(n^{-2/9}\right)\right]\left\{O\left(h_3^{-1/2}n^{-1}\right)\right\} \\
= \; & (\mu^{(3)}(0)/3!)\ 3h_3/2+O\left(h_3^2\right).
\end{aligned}$$

Hence,

$$E[\check{c}_{3;0}]=\frac{(\mu^{(3)}(0)/3!)\ 3h_3/2+O\left(h_3^2\right)}{3h_3/2+O\left(h_3^2\right)+O\left(h_3^{-1}n^{-1}\right)}=(\mu^{(3)}(0)/3!)+O\left(h_3\right). \tag{44}$$

Computations similar to those yielding (40), (41), and (42) reveal that

$$Var[\check{c}_{3;0}]\leq O\left(h_3^{-7/2}n^{-1}\right)+O\left(h_3^{-3}n^{-1+2\epsilon}\right)+O\left(h_3^{-2}n^{-1+2\epsilon}\right)+O\left(h_3^{-1}n^{-4/9}\right)=O\left(h_3^{-1}n^{-4/9}\right). \tag{45}$$

Recalling that $h_3 \propto n^{-4/27}$ and applying the Markov inequality to (44) and (45), we see that $\check{c}_{3;0}$ is $n^{4/27}$-consistent for $\mu^{(3)}(0)/3!$.

Convergence of $\check{c}_{j;0}$ for $j \geq 4$. Proceeding inductively for $j \geq 4$, we have

$$\begin{aligned}
E[\check{c}_{j;0}] &= \frac{-(\mu^{(j-2)}(0)/(j-2)!)\ 1/2+(\mu^{(j)}(0)/j!)j(j-1)h_j/4+O\left(h_j^2\right)+E[\check{c}_{j-2;0}]\ 1/2+O\left(h_j^{-j/2}n^{-1}\right)}{j(j-1)h_j/4+O\left(h_j^2\right)+O\left(h_j^{-1}n^{-1}\right)} \\
&= (\mu^{(j)}(0)/j!)+O\left(h_j\right). \qquad (46)
\end{aligned}$$

Above, we have used the facts that

$$E[\check{c}_{j-2;0}]=(\mu^{(j-2)}(0)/(j-2)!)+O\left(h_{j-2}\right)=(\mu^{(j-2)}(0)/(j-2)!)+O\left(h_j^{9/4}\right), \tag{47}$$

$$h_j^{-(j+2)/2}n^{-1}=O\left(h_j\right),$$

and the $\check{c}_{k;0}$ with $k<j, k\neq(j-2)$ contribute to the expectation only through Riemann and Laplace approximation errors whose absolute values are bounded above by an $O\left(h_j^{-(j+2)/2}n^{-1}\right)$ quantity.

The variance, on the other hand, is driven by $\check{c}_{j-1;0}$. From the Cauchy-Schwarz inequality we have

$$Var[\check{c}_{j;0}] \leq O\left(h_j^{-(2j+1)/2} n^{-1}\right) + \sum_{k=0}^{j-1} O\left(h_j^{k-j} Var[\check{c}_{k;0}]\right) = O\left(h_j^{-1} Var[\check{c}_{j-1;0}]\right) = O\left(h_j^2\right). \tag{48}$$

Above, we have used the facts that

$$h_j^{-(2j+1)/2} n^{-1} = O\left(h_j^2\right), \quad h_j^{-j} n^{-1+2\epsilon} = O\left(h_j^2\right),$$

$$h_j h_{j-1} \cdots h_{k+1} = O\left(h_j^{j-k}\right) \text{ for } 2 \leq k \leq j-2, \quad \text{and } Var[\check{c}_{j-1;0}] = O\left(h_{j-1}^2\right) = O\left(h_j^3\right).$$

Applying the Markov inequality to (46) and (48) completes the proof of Theorem 2.3. ■

Proof of Theorem 2.4. We have

$$E[\widetilde{c}_{2;0}] = (\mu^{(2)}(0)/2!) + O\left(h_2\right) \quad \text{and} \quad Var[\widetilde{c}_{2;0}] = O\left(h_2^{-5/2} n^{-1}\right)$$

by arguments similar to those given for $\check{c}_{2;0}$ in the proof of Theorem 2.3. Hence, $\widetilde{c}_{2;0}$ is $n^{2/9}$-consistent for $\mu^{(2)}(0)/2!$.

Also, we have

$$E[\widetilde{c}_{3;0}] = (\mu^{(3)}(0)/3!) + O\left(h_3\right). \tag{49}$$

However, (45) is replaced by

$$Var[\widetilde{c}_{3;0}] \leq O\left(h_3^{-7/2} n^{-1}\right) + O\left(h_3^{-3} n^{-1+2\epsilon}\right) + O\left(h_3^{-2} n^{-1+2\epsilon}\right) = O\left(h_3^{-7/2} n^{-1}\right) \tag{50}$$

since $\widetilde{c}_{2;0}$ is not involved in the computation of $\widetilde{c}_{3;0}$. Recalling that $h_3 \propto n^{-2/11}$ and applying the Markov inequality to (49) and (50), we see that $\widetilde{c}_{3;0}$ is $n^{2/11}$-consistent for $\mu^{(3)}(0)/3!$.

Proceeding inductively for $j \geq 4$, we have

$$E[\widetilde{c}_{j;0}] = (\mu^{(j)}(0)/j!) + O\left(h_j\right). \tag{51}$$

However, (48) is replaced by

$$\begin{aligned} Var[\widetilde{c}_{j;0}] &\leq O\left(h_j^{-(2j+1)/2} n^{-1}\right) + O\left(h_j^{-j} Var[\widetilde{c}_{0;0}]\right) + O\left(h_j^{-j+1} Var[\widetilde{c}_{1;0}]\right) + O\left(h_j^{-2} Var[\widetilde{c}_{j-2;0}]\right) \\ &= O\left(h_j^{-2} Var[\widetilde{c}_{j-2;0}]\right) \\ &= O\left(h_j^2\right). \end{aligned} \tag{52}$$

Above, we have used the facts that

$$h_j^{-(2j+1)/2} n^{-1} = O\left(h_j^2\right), \quad h_j^{-j} n^{-1+2\epsilon} = O\left(h_j^2\right), \quad \text{and} \quad Var[\widetilde{c}_{j-2;0}] = O\left(h_{j-2}^2\right) = O\left(h_j^4\right).$$

Applying the Markov inequality to (51) and (52) completes the proof of Theorem 2.4. ■

Derivation of Condition 3.1. Let $Q^{(j)}(s) := \frac{d^j}{ds^j} Q(s)$ for $j \geq 0$. Write

$$w_{i;h}(x) = \sum_{j=0}^{\infty} q_{i,j} x^j / \sqrt{h}, \quad \text{where} \quad q_{i,j} := (-1)^j Q^{(j)}(x_i/\sqrt{h})\ h^{-j/2}/j!.$$

Identifying $u^*(x)$ with $\sum_{j=0}^{\infty} \widehat{c}_{j;0} x^j$, we simplify the Euler-Lagrange equation (26) to

$$\begin{aligned}
& \left(\sum_{j=0}^{\infty} (j+2)(j+1)\widehat{c}_{j+2;0} x^j\right) \left(\sum_{j=0}^{\infty} \left[\sqrt{h}\lambda 1_{j=0} + \sum_{i=1}^{n} \left\{q_{i,j} x_i^2 - 2q_{i,j-1} x_i 1_{j\geq 1} + q_{i,j-2} 1_{j\geq 2}\right\}\right] x^j\right) \\
+ & \left(\sum_{j=0}^{\infty} (j+1)\widehat{c}_{j+1;0} x^j\right) \left(\sum_{j=0}^{\infty} (j+1) \sum_{i=1}^{n} \left\{q_{i,j+1} x_i^2 - 2q_{i,j} x_i + q_{i,j-1} 1_{j\geq 1}\right\} x^j\right) \\
+ & \left(\sum_{j=0}^{\infty} \widehat{c}_{j;0} x^j\right) \left(\sum_{j=0}^{\infty} \left\{(j+1)\sum_{i=1}^{n} q_{i,j+1} x_i - (j+2)\sum_{i=1}^{n} q_{i,j}\right\} x^j\right) \\
= & \sum_{j=0}^{\infty} \left\{(j+1)\sum_{i=1}^{n} q_{i,j+1} Y_i x_i - (j+2)\sum_{i=1}^{n} q_{i,j} Y_i\right\} x^j.
\end{aligned}$$

Equating powers of x yields, for each $j \geq 0$,

$$\begin{aligned}
& \sum_{k=0}^{j} \left[(j-k+2)(j-k+1)\widehat{c}_{j-k+2;0} \left(\sqrt{h}\lambda 1_{k=0} + \sum_{i=1}^{n} \left\{q_{i,k} x_i^2 - 2q_{i,k-1} x_i 1_{k\geq 1} + q_{i,k-2} 1_{k\geq 2}\right\}\right)\right. \\
+ & \ (j-k+1)\widehat{c}_{j-k+1;0} \left((k+1)\sum_{i=1}^{n} \left\{q_{i,k+1} x_i^2 - 2q_{i,k} x_i + q_{i,k-1} 1_{k\geq 1}\right\}\right) \\
+ & \left.\ \widehat{c}_{j-k;0} \left((k+1)\sum_{i=1}^{n} q_{i,k+1} x_i - (k+2)\sum_{i=1}^{n} q_{i,k}\right)\right] \\
= & \ (j+1)\sum_{i=1}^{n} q_{i,j+1} Y_i x_i - (j+2)\sum_{i=1}^{n} q_{i,j} Y_i.
\end{aligned} \tag{53}$$

The expected value of the right member of equation (53) is

$$\sum_{i=1}^{n} \mu(x_i)\left\{(j+1)q_{i,j+1} x_i - (j+2)q_{i,j}\right\} = \sum_{p=0}^{\infty} (\mu^{(p)}(0)/p!) \sum_{i=1}^{n} \left\{(j+1)q_{i,j+1} x_i^{p+1} - (j+2)q_{i,j} x_i^p\right\}. \tag{54}$$

Matching the $\widehat{c}_{j+2;0}$ term from the left member of equation (53) to the $\mu^{(j+2)}(0)/(j+2)!$ term from the right member of equation (54) yields

$$(j+2)(j+1)\left[\sqrt{h}\lambda\{1+o(1)\} + \sum_{i=1}^{n} q_{i,0} x_i^2\right] = \sum_{i=1}^{n} \left\{(j+1)q_{i,j+1} x_i^{j+3} - (j+2)q_{i,j} x_i^{j+2}\right\}. \tag{55}$$

A Riemann approximation converts (55) to

$$\begin{aligned}
& (j+2)(j+1)\left[\frac{2\sqrt{h}\lambda\{1+o(1)\}}{n}+\int_{-1}^{1} Q(t/\sqrt{h})\ t^2\ dt\right] \\
= \quad & (-1)^{j+1}(j+1)\int_{-1}^{1} Q^{(j+1)}(t/\sqrt{h})\ t^{j+3}h^{-(j+1)/2}/(j+1)!\ dt \\
- \quad & (-1)^{j}(j+2)\int_{-1}^{1} Q^{(j)}(t/\sqrt{h})\ t^{j+2}h^{-j/2}/j!\ dt,
\end{aligned}$$

from which algebraic manipulations and a Laplace approximation provide

$$\begin{aligned}
& (j+2)!\left[\frac{2\lambda\{1+o(1)\}}{nh}+\int_{-\infty}^{\infty} Q(s)\ s^2\ ds\ \right] \\
= \quad & (-1)^{j+1}\int_{-\infty}^{\infty} Q^{(j+1)}(s)\ s^{j+3}\ ds \qquad (56) \\
- \quad & (-1)^{j}(j+2)\int_{-\infty}^{\infty} Q^{(j)}(s)\ s^{j+2}\ ds. \qquad (57)
\end{aligned}$$

Repeatedly integrating (56) and (57) by parts, we obtain

$$\begin{aligned}
& (j+2)!\left[\frac{2\lambda\{1+o(1)\}}{nh}+\int_{-\infty}^{\infty} Q(s)\ s^2\ ds\ \right] \\
= \quad & \int_{-\infty}^{\infty} Q(s)\ s^2\ ds\ (j+3)!/2 \\
- \quad & (j+2)\int_{-\infty}^{\infty} Q(s)\ s^2\ ds\ (j+2)!/2,
\end{aligned}$$

which yields

$$\frac{4\lambda\{1+o(1)\}}{nh}+2\int_{-\infty}^{\infty} Q(s)\ s^2\ ds=\int_{-\infty}^{\infty} Q(s)\ s^2\ ds. \qquad (58)$$

Condition 3.1 follows from relation (58). ■

Proof of Theorem 4.1. Let $\beta_0>0$ and $L_0>0$ be arbitrary. Put

$$\delta_n := (\log n)^{\xi-1}, \quad \psi_n := (\log n)^{\xi-1}, \quad \gamma_n := 2(\log n)^{\xi-1}, \quad \text{and } \beta_n := \beta_0\ n^{2(\delta_n+\gamma_n)}.$$

Recalling that $L_n = 3^{M_n}$, choose M_n to satisfy

$$1 \le \frac{L_n}{L_0\ \beta_n^{1/2} n^{\psi_n}} \le 3.$$

Let

$$I_{1n}(x) := \{a \in I_n : |a-x| < n^{-\gamma_n}\}, \quad I_{2n}(x) := \{a \in I_n : |a-x| < \beta_n^{-1/2}\},$$

$$L_a(x) := \sum_{k=0}^{4j} \left[\widetilde{c}_{k;a} - \mu^{(k)}(a)/k!\right](x-a)^k, \quad U_a(x) := \sum_{k=4j+1}^{\infty} \left[\widetilde{c}_{k;a} - \mu^{(k)}(a)/k!\right](x-a)^k,$$

and $E_a(x) := L_a(x) + U_a(x)$. Note that

$$\inf_{x\in I} \sum_{c\in I_n} \exp[-\beta_n(x-c)^2] \geq \inf_{x\in I} \sum_{c\in I_{2n}(x)} \exp[-1] \geq C_1\ n^{\psi_n} \tag{59}$$

for some constant $C_1 > 0$.

Put $W_{a,n}(x) := \exp\left[-\beta_n(x-a)^2\right] / \sum_{c\in I_n} \exp\left[-\beta_n(x-c)^2\right]$ for each $a \in I_n$. Noting that $\mu(x) = \sum_{a\in I_n} W_{a,n}(x)\mu(x)$, we need to show that

$$\sup_{x\in I} \left| \frac{d^j}{dx^j} \sum_{a\in I_n} W_{a,n}(x) E_a(x) \right| = O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right).$$

We do so by employing the partition

$$\frac{d^j}{dx^j} \sum_{a\in I_n} W_{a,n}(x) E_a(x) = \sum_{a\in \bar{I}_{1n}(x)} \frac{d^j}{dx^j}\{W_{a,n}(x)E_a(x)\} + \sum_{a\in I_{1n}(x)} \frac{d^j}{dx^j}\{W_{a,n}(x)E_a(x)\}. \tag{60}$$

To handle the first piece in (60), note there exists a constant $C_2 > 0$ such that, for any $a \in \bar{I}_{1n}(x)$ and $0 \leq k \leq j$,

$$\left| \frac{d^k}{dx^k} W_{a,n}(x) \right| \leq \frac{C_2\ \beta_n^k \exp\left[-\beta_0\ n^{2\delta_n}\right]}{\sum_{c\in I_n} \exp[-\beta_n(x-c)^2]}.$$

Moreover, by virtue of assumptions (19) and (21), there exists a constant $C_3 > 0$ such that

$$\left| \frac{d^{j-k}}{dx^{j-k}} E_a(x) \right| \leq C_3$$

for any $a \in \bar{I}_{1n}(x)$ and $0 \leq k \leq j$. We then obtain

$$\begin{aligned} \sup_{x\in I} \left| \sum_{a\in \bar{I}_{1n}(x)} \frac{d^k}{dx^k} W_{a,n}(x) \frac{d^{j-k}}{dx^{j-k}} E_a(x) \right| &\leq \frac{L_n C_3 C_2\ \beta_n^k \exp\left[-\beta_0\ n^{2\delta_n}\right]}{\inf_{x\in I} \sum_{c\in I_n} \exp[-\beta_n(x-c)^2]} \\ &= O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right) \end{aligned} \tag{61}$$

because $\exp\left[-\beta_0\ n^{2\delta_n}\right] = \exp\left[-\beta_0\ n^{2(\log n)^{\xi-1}}\right]$ tends to 0 more quickly than any power of n and the denominator of the right member in (61) is bounded below as in (59). Since finitely many terms of the form $\sum_{a\in \bar{I}_{1n}(x)} \frac{d^k}{dx^k} W_{a,n}(x) \frac{d^{j-k}}{dx^{j-k}} E_a(x)$ constitute $\sum_{a\in \bar{I}_{1n}(x)} \frac{d^j}{dx^j}\{W_{a,n}(x)E_a(x)\}$, the latter must be $O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right)$ uniformly over $x \in I$.

To handle the second piece in (60), note there exists a constant $C_4 > 0$ such that, for any $0 \leq k \leq j$,

$$\sup_{x\in I} \sup_{a\in I_{1n}(x)} \left| \frac{d^k}{dx^k} W_{a,n}(x) \right| \leq \frac{C_4\ \beta_n^k}{\inf_{x\in I} \sum_{c\in I_n} \exp[-\beta_n(x-c)^2]} = O\left(n^{2k(\gamma_n+\delta_n)-\psi_n}\right). \tag{62}$$

For any $\phi > 0$ and $0 \le k \le j$, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{x\in I}\sup_{a\in I_{1n}(x)}\left|\frac{d^{j-k}}{dx^{j-k}}L_a(x)\right| \ge \phi\right) &\le \mathbb{P}\left(\sup_{a\in I_n}\sum_{l=j-k}^{4j}\frac{|\widetilde{c}_{l;a}-\mu^{(l)}(a)/l!|\; n^{-\gamma_n\{l-(j-k)\}}l!}{\{l-(j-k)\}!} \ge \phi\right) \\
&\le \sum_{a\in I_n}\mathbb{P}\left(\sum_{l=j-k}^{4j}\frac{|\widetilde{c}_{l;a}-\mu^{(l)}(a)/l!|\; n^{-\gamma_n\{l-(j-k)\}}l!}{\{l-(j-k)\}!} \ge \phi\right) \qquad (63) \\
&\le \phi^{-2}\sum_{a\in I_n}\mathbb{E}\left[\left(\sum_{l=j-k}^{4j}\frac{|\widetilde{c}_{l;a}-\mu^{(l)}(a)/l!|\; n^{-\gamma_n\{l-(j-k)\}}l!}{\{l-(j-k)\}!}\right)^2\right] \\
&\le \phi^{-2}\; L_n\; C_5\sum_{l=j-k}^{4j}\sum_{m=j-k}^{4j} n^{-\alpha_l}n^{-\alpha_m}n^{-\gamma_n\{l+m-2(j-k)\}} \\
&\le \phi^{-2}\; L_n\; C_6\; n^{-2\alpha_{4j}} \qquad (64)
\end{aligned}$$

for some constants $C_5, C_6 > 0$. Line (63) uses the Bonferroni inequality. The next three lines use the Markov inequality, the Cauchy-Schwarz inequality, and assumption (20). Put $\phi := n^{-(4j+1)\gamma_n}$. Then quantity (64) is dominated by a multiple of

$$n^{(8j+3)\gamma_n+\delta_n+\psi_n-2\alpha_{4j}},$$

which tends to 0. This shows that

$$\sup_{x\in I}\sup_{a\in I_{1n}(x)}\left|\frac{d^{j-k}}{dx^{j-k}}L_a(x)\right| = O_p\left(n^{-(4j+1)\gamma_n}\right). \qquad (65)$$

Combining (62) and (65) with

$$\sup_{x\in I}\sup_{a\in I_{1n}(x)}\left|\frac{d^{j-k}}{dx^{j-k}}U_a(x)\right| = O\left(n^{-\gamma_n(3j+1+k)}\right) \quad \text{and} \quad \sup_{x\in I}\text{card}[I_{1n}(x)] = O\left(n^{\delta_n+\psi_n}\right),$$

where card$[\cdot]$ denotes the cardinality of a set, yields

$$\begin{aligned}
&\sup_{x\in I}\left|\sum_{a\in I_{1n}(x)}\frac{d^k}{dx^k}W_{a,n}(x)\frac{d^{j-k}}{dx^{j-k}}E_a(x)\right| \\
\le\;& O\left(n^{2k(\gamma_n+\delta_n)-\psi_n}\right)\times\left\{O_p\left(n^{-(4j+1)\gamma_n}\right)+O\left(n^{-\gamma_n(3j+1+k)}\right)\right\}\times O\left(n^{\delta_n+\psi_n}\right) \\
=\;& O_p\left(n^{(2j+1)(\delta_n-\gamma_n)}\right) \\
=\;& O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right).
\end{aligned}$$

Since finitely many terms of the form $\sum_{a\in I_{1n}(x)}\frac{d^k}{dx^k}W_{a,n}(x)\frac{d^{j-k}}{dx^{j-k}}E_a(x)$ constitute $\sum_{a\in I_{1n}(x)}\frac{d^j}{dx^j}\{W_{a,n}(x)E_a(x)\}$, the latter must be $O_p\left(n^{-(2j+1)(\log n)^{\xi-1}}\right)$ uniformly over $x \in I$. ■

Caption for Table 1. Columns $CE_{0.3}$, $CE_{0.4}$, and $CE_{0.5}$ pertain to the original compound estimator with nearest neighbors smoothing parameters 0.3, 0.4, and 0.5 for the local regression estimators $\widetilde{c}_{0;a}$ through $\widetilde{c}_{7;a}$. Columns $EXCE_{0.3}$, $EXCE_{0.4}$, and $EXCE_{0.5}$ pertain to the extended compound estimator with nearest neighbors smoothing parameters 0.3, 0.4, and 0.5 for the local regression estimators $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$. Implementation details (e.g., other tuning parameters) are described in Section 5. Entries in rows 0 through 6 are the averages, over 100 simulated data sets of size $n = 125$ or $n = 500$, of quantity (22). Entries in row RSD are the averages, over the 100 simulated data sets, of quantity (23). The data sets had equispaced x_i, independent normally distributed ϵ_i with common variance $\sigma^2 = 0.04$, and mean response $\mu_O(x)$ as defined in Section 5. Method $EXCE_{0.3}$ was substantially better than the nearest competitor at both $n = 125$ and $n = 500$.

Caption for Table 2. Columns $CE_{0.3}$, $CE_{0.4}$, and $CE_{0.5}$ pertain to the original compound estimator with nearest neighbors smoothing parameters 0.3, 0.4, and 0.5 for the local regression estimators $\widetilde{c}_{0;a}$ through $\widetilde{c}_{7;a}$. Columns $EXCE_{0.3}$, $EXCE_{0.4}$, and $EXCE_{0.5}$ pertain to the extended compound estimator with nearest neighbors smoothing parameters 0.3, 0.4, and 0.5 for the local regression estimators $\widetilde{c}_{0;a}$ and $\widetilde{c}_{1;a}$. Implementation details (e.g., other tuning parameters) are described in Section 5. Entries in rows 0 through 6 are the averages, over 100 simulated data sets of size $n = 125$ or $n = 500$, of quantity (22). Entries in row RSD are the averages, over the 100 simulated data sets, of quantity (23). The data sets had equispaced x_i, independent normally distributed ϵ_i with common variance $\sigma^2 = 0.04$, and mean response $\mu_M(x)$ as defined in Section 5. Method $CE_{0.4}$ was slightly better than method $EXCE_{0.3}$ at $n = 125$, but method $EXCE_{0.5}$ was the winner at $n = 500$.

Table 1: Recovering derivatives of the oscillatory test function

$n = 125$	$CE_{0.3}$	$CE_{0.4}$	$CE_{0.5}$	$EXCE_{0.3}$	$EXCE_{0.4}$	$EXCE_{0.5}$
0	0.9914	0.9707	0.9462	0.9894	0.9538	0.9679
1	0.9838	0.9472	0.8939	0.9792	0.9562	0.9518
2	0.9603	0.9086	0.8348	0.9623	0.9360	0.9332
3	0.9312	0.9062	0.8335	0.9594	0.9380	0.9108
4	0.8427	0.8879	0.7790	0.9416	0.9029	0.8304
5	0.4664	0.8657	0.7622	0.9164	0.8266	0.6715
6	-0.8712	0.8189	0.6643	0.8729	0.6171	0.2576
RSD	2.6954	0.6948	1.2861	0.3788	0.8694	1.4769
$n = 500$	$CE_{0.3}$	$CE_{0.4}$	$CE_{0.5}$	$EXCE_{0.3}$	$EXCE_{0.4}$	$EXCE_{0.5}$
0	0.9924	0.9719	0.9464	0.9908	0.9547	0.9675
1	0.9863	0.9491	0.8937	0.9810	0.9569	0.9507
2	0.9670	0.9109	0.8339	0.9640	0.9370	0.9319
3	0.9540	0.9089	0.8325	0.9615	0.9404	0.9117
4	0.9268	0.8924	0.7772	0.9455	0.9090	0.8339
5	0.7949	0.8722	0.7615	0.9234	0.8390	0.6786
6	0.4636	0.8321	0.6611	0.8874	0.6437	0.2703
RSD	0.9149	0.6626	1.2938	0.3465	0.8193	1.4554

Table 2: Recovering derivatives of the monotone test function

$n = 125$	$CE_{0.3}$	$CE_{0.4}$	$CE_{0.5}$	$EXCE_{0.3}$	$EXCE_{0.4}$	$EXCE_{0.5}$
0	0.9981	0.9984	0.9983	0.9985	0.9956	0.9976
1	0.9153	0.9571	0.9574	0.9590	0.9281	0.9435
2	0.1512	0.7344	0.7209	0.7129	0.6130	0.6644
3	-1.0428	0.5491	0.4931	0.5273	0.4704	0.4976
4	-2.8851	0.3541	0.2914	0.3257	0.3233	0.3377
5	-4.5431	0.2230	0.1541	0.1843	0.2024	0.2152
6	-5.8813	0.1264	0.0717	0.0828	0.1116	0.1214
RSD	19.2878	3.0575	3.3130	3.2096	3.3555	3.2225
$n = 500$	$CE_{0.3}$	$CE_{0.4}$	$CE_{0.5}$	$EXCE_{0.3}$	$EXCE_{0.4}$	$EXCE_{0.5}$
0	0.9994	0.9992	0.9990	0.9995	0.9980	0.9990
1	0.9770	0.9760	0.9687	0.9818	0.9660	0.9745
2	0.7883	0.8138	0.7552	0.8395	0.8060	0.8357
3	0.5223	0.6244	0.5166	0.6571	0.6757	0.6943
4	0.1458	0.4272	0.3064	0.4522	0.5094	0.5137
5	-0.1759	0.2704	0.1613	0.2824	0.3426	0.3388
6	-0.4325	0.1578	0.0755	0.1605	0.2051	0.1955
RSD	4.1755	2.7311	3.2172	2.6271	2.4971	2.4484