Oliver B. Linton

Estimating additive nonparametric models by partial L_q norm : the curse of fractionality

Originally published in Econometric theory, 17 (6). pp. 1037-1050 © 2001 Cambridge University Press.

Econometric Theory, **17**, 2001, 1037–1050. Printed in the United States of America.

ESTIMATING ADDITIVE NONPARAMETRIC MODELS BY PARTIAL L_q NORM: THE CURSE OF FRACTIONALITY

Oliver Linton
London School of Economics
and
Yale University

We propose a new method for estimating additive nonparametric regression models based on taking the L_q median of a sample of kernel estimators. We establish the consistency and asymptotic normality of our procedures. The rate of convergence depends on the value of q. For $q > \frac{3}{2}$ one has the usual one-dimensional rate, but if $q \leq \frac{3}{2}$ the rate can be slower.

1. INTRODUCTION

Nonparametric estimation of the conditional mean curve has received much attention in the literature. Estimation of other attributes of the conditional distributions, such as the conditional median curves, has not received quite so much attention in the methodological literature, perhaps as a result of their analytical complexity. For example, Härdle (1990) spends less than 1 page out of 300 on this topic. Nevertheless, they are of as much interest for applications and have been extensively applied in economics following the seminal work of Koenker and Bassett (1978) in parametric quantile regression. One advantage of medians as location measure is that they are still consistent in the presence of a certain amount of censoring or outliers, which can be important for some data sets. They are also equivariant to monotone transformations. Unfortunately, as in the mean regression counterpart, nonparametric estimation of the conditional median suffers from the curse of dimensionality, which is manifested in the slower attainable convergence rates in high dimensions. Assuming that the target function is separable, specifically additive, can alleviate this problem, as was originally shown by Stone (1985). Additive nonparametric regression models provide a powerful tool for exploring relationships between a response vari-

This paper is an abbreviated version of Linton (1999), which itself is based partly on some joint work with Nick Hengartner. I thank two referees, Joel Horowitz, and Peter Phillips for comments and the National Science Foundation and the North Atlantic Treaty Organization for financial support. Address correspondence to: Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK; e-mail: lintono@econ.yale.edu.

 0266-4666/01 $9.50

able Y and multivariate covariates $X \in \mathbb{R}^d$ because the estimates enjoy the flexibility of nonparametric regression although not being subject to the curse of dimensionality. In addition, the individual additive components are easy to interpret.

Tjøstheim and Auestad (1994), Newey (1994), and Linton and Nielsen (1995) have independently proposed a procedure for estimating additive nonparametric regression models. The idea is to integrate an initial consistent estimator with respect to a $d-1$–dimensional probability measure: let $\hat{m}(x)$ be some consistent estimator of a function $m(x)$ and let

$$\hat{\tau}_1(x_1) = \int \hat{m}(x_1, x_2, \ldots, x_d)\, dP(x_2, \ldots, x_d). \tag{1}$$

When the function m is additive, i.e., $m(x) = m_1(x_1) + \cdots + m_d(x_d)$, $\hat{\tau}_1(x_1)$ consistently estimates $m_1(x_1)$, up to an additive constant. Typically, P is the empirical distribution of the covariates $X_2, \ldots, X_d$, in which case

$$\hat{\tau}_1(x_1) = \frac{1}{n} \sum_{i=1}^{n} \hat{m}(x_1, X_{2i}, \ldots, X_{di}).$$

Consistency of these estimators readily follows from the uniform consistency of the estimate $\hat{m}(x)$. Rather, the statistical issue is whether this estimator circumvents the curse of dimensionality in the sense that its rate of convergence does not depend on the number of covariates d. In fact, $\hat{\tau}_1(x_1)$ can be shown to be asymptotically normal at the same rate as one-dimensional nonparametric regression. Hence this estimator circumvents the curse of dimensionality. Further refinements of this *integration method* are found in Linton (1998) and Horowitz (1999).

The estimator $\hat{\tau}_1(x_1)$ can be interpreted as the expectation (or sample mean) of $\hat{m}(x_1, X_2, \ldots, X_d)$ with respect to some distribution for $X_2, \ldots, X_d$. The mean is only one example of a location measure; the median is an alternative location measure that has found widespread use. Therefore, why not replace the expectation by the median operator? After all, once $\hat{m}(x_1, X_{2i}, \ldots, X_{di})$ has been computed, we just have a list of n numbers. If taking the mean of these n numbers is a sensible operation, then taking the median seems equally sensible. One advantage of taking medians is that the target quantities are well defined for heavy tailed distributions. In this case, the relevant distribution is that of the covariates $X_2, \ldots, X_d$. For example, suppose that $d = 2$ and that $m(x) = x_1 + x_2$, where X_2 is Cauchy distributed. Then, $\tau_1(x_1) = \int m(x)\, dP(x_2)$ is not defined when P is the marginal distribution of X_2. However, the median of $m(x_1, X_2)$ with respect to any continuous distribution is well defined.

In this paper, we investigate a new proposal for estimation in additive nonparametric regression. We shall estimate the individual additive components by

$$\hat{\tau}_1(x_1) = \arg\min_t \int \varrho(\hat{m}(x) - t)\, dP(x_2, \ldots, x_d), \tag{2}$$

where $\hat{m}(x)$ is an estimate of a population function $m(x)$, ϱ is a bowl-shaped loss function, P is some probability measure, and the minimization is taken with respect to $t \in T$ for some set $T \subseteq \mathbb{R}$. Even though our conclusions hold for general loss functions ϱ, we find it useful to specialize our results to the L_q distance, where $\varrho(t) = |t|^q$ with $q \geq 1$. For example, the L_1 distance, which corresponds to taking $\varrho(t) = |t|$, is less sensitive to heavy tails than the L_2 distance. For the L_1 loss function $\varrho(t) = |t|$, the estimator is

$$\hat{\tau}_1(x_1) = \text{median}(\hat{m}(x_1, X_2, \ldots, X_d)),$$

where $X_2, \ldots, X_d$ have joint distribution P. This amounts to replacing the averaging in the integration estimator by medianing. We expect that medianing has similar asymptotic properties to the integration estimator, at least when both population quantities exist. We establish pointwise consistency of the partial L_q method for any $q \geq 1$ under weak assumptions on the "integrating" measure. Specifically, we allow it to have unbounded support. The result is basically a consequence of uniform consistency of $\hat{m}$ over increasing sets, a result we found in Andrews (1995) for regression. The asymptotic distribution theory for $\hat{\tau}_1(x_1)$ depends on the smoothness of the loss function ϱ (which is measured by q in our case). Estimators defined through smooth loss functions have essentially the same behavior as the integration estimator, whereas lack of smoothness leads to a slower rate of convergence for the estimate. The main conclusion of this paper is that the partial L_q estimator with $q > \frac{3}{2}$ reduces a d-dimensional problem to a one-dimensional problem; i.e., the asymptotic distribution of $\hat{\tau}_1(x_1)$ is normal with one-dimensional rate, whereas if $q \leq \frac{3}{2}$ the rate of convergence is slower. Specifically, the partial L_1 estimator has the convergence rate of two-dimensional nonparametric regression, whereas the partial L_q estimator with $1 < q < \frac{3}{2}$ has an intermediate fractional rate of convergence, hence the title. These three cases correspond well with Arcones (1996).

2. ESTIMATION

Let $(X, Y) \in \mathbb{R}^{d+1}$ be a random vector and let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent and identically distributed (i.i.d.) sample from this population. For each direction x_k, we partition $x = (x_k, x_{-k})$, where x_k is scalar whereas x_{-k} is of dimensions $d - 1$; likewise let $X_i = (X_{ki}, X_{-ki})$. Let f be the Lebesgue density of X and let f_{-k} denote the density of X_{-k}. Let P be an absolutely continuous distribution with density p on $\mathbb{R}^d$ whose support is included in the support of the distribution of X. It will be convenient to denote $p_{-k}(x_{-k}) = \int p(x)\,dx_k$ and $p_k(x_k) = \int p(x)\,dx_{-k}$.

We write

$$Y = m(X) + \varepsilon,$$

where the function m is identified by a conditional location restriction on the disturbance ε—specifically, we suppose that the L_q median of ε given X is zero. The two leading examples would be $E(\varepsilon|X) = 0$ and $\text{median}(\varepsilon|X) = 0$.

In the first case, $m(x)$ is the conditional mean function, whereas in the second case, $m(x)$ is the conditional median. We then suppose that there is a unique solution, which we call $\tau_k(x_k)$, to the population minimization problem $\int \varrho(m(x) - t)p_{-k}(x_{-k})dx_{-k}$. We also suppose that $m(x) = m_1(x_1) + \cdots + m_d(x_d)$, for some functions $m_j(\cdot)$, $j = 1,\ldots,d$, in which case, $\tau_k(x_k)$ is $m_k(x_k)$ up to an additive constant; i.e., we can write $\tau_k(x_k) = m_k(x_k) + \mu$ for some μ. The constant μ is defined through some restriction on $m_k(\cdot)$; e.g., we might suppose that $\int \varrho(m_k(x_k) - t)p_k(x_k)dx_k$ is uniquely minimized at $t = 0$. However, because we only concern ourselves with estimation of $m_k(\cdot)$ up to a constant in this paper, we do not need to be precise on this issue.

Let $\hat{m}(x)$ be a consistent estimator of $m(x)$, such as a local polynomial kernel median or mean. Now consider the *partial criterion function*

$$Q_n(t) = \int_{A_n} \varrho(\hat{m}(x) - t)p_{-k}(x_{-k})dx_{-k} \tag{3}$$

and denote by $\hat{\tau}_k(x_k)$ the minimizer of $Q_n(t)$ with respect to t for every fixed x_k. Then $\hat{\tau}_k(x_j)$ estimates $m_k(x_k) + \mu$. We actually work with solutions to the first-order condition $G_n(t) = 0$, where

$$G_n(t) = \int_{A_n} \psi(\hat{m}(x) - t)p_{-k}(x_{-k})dx_{-k}, \tag{4}$$

where ψ is the (generalized) derivative of the function ϱ. The integration in (3) and (4) is over a set $A_n \subseteq A = \text{supp}(p_{-k}) \subset \text{supp}(f_{-k})$, where we shall allow the set A_n to increase with sample size where necessary. We introduce this additional generality because we wish to show consistency in the case where p_{-k} has support $\mathbb{R}^{d-1}$. The integration can be done numerically when $d \leq 4$; in higher dimensions, it is necessary to replace the integrals by sums.

3. ASYMPTOTICS

We shall restrict our attention to the class of L_q criterion functions with $\varrho(t) = |t|^q$, where $q \geq 1$ and hence $\psi(t) = \text{sign}(t)|t|^{q-1}$.

3.1. Consistency

We first establish the pointwise consistency of the proposed estimator under high level conditions on the pilot estimator.

THEOREM 1. *Suppose that* $q \geq 1$ *and that*

$$\sup_{x_{-k} \in A_n} |\hat{m}(x) - m(x)| \to 0 \tag{5}$$

with probability one for some increasing sequence of sets A_n. Suppose also that there exists a unique solution, $\tau_k(x_k)$, to $\min_{t\in T}\int \varrho(m(x)-t)p_{-k}(x_{-k})dx_{-k}$. Then, the corresponding estimator $\hat{\tau}_k(x_k)$ is strongly consistent, i.e., with probability one

$$|\hat{\tau}_k(x_k)-\tau_k(x_k)|\to 0.$$

This generalizes Lemma 1 of Linton, Chen, Wang, and Härdle (1997). In the case that A_n is uniformly bounded, the result (5) has been shown in Masry (1996) for local polynomial regression and in Chaudhuri (1991a, 1991b) for local polynomial quantile regression. Andrews (1995) extended the mean regression result to increasing sets. The rate at which A_n is allowed to grow depends on the marginal density f of X, specifically, on the quantity $\alpha_n = \inf_{x_{-k}\in A_n} f(x)$; i.e., the faster α_n decreases to zero, the more slowly A_n is permitted to expand. The preceding result gives conditions under which $\hat{\tau}_k(x_k)$ is consistent when the support of p_{-k} is infinite and is the first result of this kind that we are aware of.

3.2. Asymptotic Normality

We shall now suppose that the integration in $G_n(t)$ is carried out over a fixed set $A = \mathrm{supp}(p_{-k})$, which is assumed to be compact. The results divide according to three cases: the "smooth" case where $q > \frac{3}{2}$, the "partly smooth" case $1 < q \le \frac{3}{2}$, and the unsmooth case where $q = 1$. In the partly smooth and unsmooth cases, we shall need some results on fractional integration and generalized functions that can be found in Linton (1999).

We shall require some additional structure on the unrestricted estimation error $\hat{m}(x)-m(x)$. We will suppose that the estimator and its derivatives satisfy the following Bahadur representation:

$$D^{(\boldsymbol{\nu})}\hat{m}(x) - D^{(\boldsymbol{\nu})}m(x) = \sum_{i=1}^{n} W_{n,i}^{(\nu)}(x)\eta_i + \sum_{k=1}^{d}\chi_{nk}^{(\nu)}(x_k) + R_n^{(\nu)}(x) \tag{6}$$

for vectors $\nu = (\nu_1,\ldots,\nu_d)$ to be determined subsequently and for some functions a_ν

$$W_{n,i}^{(\nu)}(x) = \frac{1}{nh^{d+|\nu|}}\, a_\nu(x)\prod_{j=1}^{d} K^{(\nu_j)}\left(\frac{x_j - X_{ji}}{h}\right), \tag{7}$$

where K is a univariate kernel function and $h = h(n)$ is a bandwidth sequence. Here, $K^{(\nu_j)}(t)$ denotes $d^{\nu_j}K(t)/dt^{\nu_j}$. When $\hat{m}(x)$ is a Nadaraya–Watson conditional mean estimator, $\eta_i = Y_i - m(X_i)$, whereas if $\hat{m}(x)$ is a conditional median estimator, $\eta_i = \mathrm{sign}(Y_i - m(X_i))$. Assumptions A, which follow, are needed in the smooth case, whereas in the unsmooth cases we shall need in addition Assumptions B, which are presented subsequently.

Assumption A.

(a) The random sample $\{Z_i \equiv (Y_i, X_i')', Y_i \in \mathbb{R}, X_i \in \mathbb{R}^d\}_{i=1}^n$ is i.i.d. The covariates X have distribution that is absolutely continuous with respect to Lebesgue measure with density $f(\cdot)$.
(b) The density function p_{-k} has support A that is strictly contained in the support of the density function f. The density f is bounded away from zero on A. The functions f and p_{-k} are Lipschitz continuous on A; i.e., there exists a constant c such that $|f(x_k, x'_{-k}) - f(x_k, x'_{-k})| \leq c\|x_{-k} - x'_{-k}\|$, $|p_{-k}(x_{-k}) - p_{-k}(x'_{-k})| \leq c\|x_{-k} - x'_{-k}\|$ for all $x_{-k}, x'_{-k} \in A$.
(c) The function $\psi(t) = \text{sign}(t)|t|^{q-1}$. The set T is compact, and $\tau_k(x_k)$ lies in the interior of T.
(d) $\eta_1, \ldots, \eta_n$ are i.i.d. with $E(\eta_i|X_i) = 0$ and $E(\eta_i^2) < \infty$. Let $\sigma^2(x) = \text{var}(\eta_i|X_i = x)$.
(e) The kernel function $K(\cdot)$ is bounded, symmetric about zero, compactly supported, has Lipschitz continuous derivative, and integrates to one. Let $\|K\|_2^2 = \int K^2(u)du < \infty$, $\mu_j(K) = \int u^j K(u)du < \infty$.
(f) The functions $a = a_{(0)}$ and σ^2 are Lipschitz continuous $|a(x_k, x_{-k}) - a(x_k, x'_{-k})| \leq c\|x_{-k} - x'_{-k}\|$ and $|\sigma^2(x_k, x'_{-k}) - \sigma^2(x_k, x'_{-k})| \leq c\|x_{-k} - x'_{-k}\|$ for all $x_{-k}, x'_{-k} \in A$. Furthermore, the function σ^2 is bounded away from zero on A.
(g) The functions $\chi_{nk}^{(0)}(x_k)$ satisfy $\lim_{n\to\infty} h^{-r} \sum_{k=1}^d \chi_{nk}^{(0)}(x_k) = \chi^{(0)}(x)$, where the function $\chi^{(0)}(x)$ is bounded.
(h) The remainder terms $R_n^{(0)}(x)$ are such that with probability one, $\sup_{x_{-k}\in A} \times |R_n^{(0)}(x)| = o(n^{-r/(2r+1)})$.

Assumption B1. There exists some neighborhood of the set $\mathcal{M}_k = \{x_{-k}: \sum_{\ell\neq k}^d m_\ell(x_\ell) = 0\}$ on which $m_j'(x_j) \neq 0$ for some $j \neq k$. The set $\mathcal{M}_k \subseteq \mathbb{R}^{d-1}$ is of Lebesgue measure zero; furthermore, $f_m(\sum_{\ell\neq k}^d m_\ell(x_\ell)) > 0$ for all x_{-k} in some neighborhood of $\mathcal{M}_k$, where f_m is the Lebesgue density of the random variable $\sum_{\ell\neq k}^d m_\ell(X_{\ell i})$.

Assumption B2.

(a) The density p_{-k} is twice continuously differentiable on A, and it and its first partial derivatives are zero on the boundary of A.
(b) For all ν with $|\nu| \leq 3$, the functions $\chi_{nk}^{(\nu)}(x_k)$ satisfy $\lim_{n\to\infty} h^{-(r-|\nu|)} \times \sum_{k=1}^d \chi_{nk}^{(\nu)}(x_k) = \chi^{(\nu)}(x)$, where the function $\chi^{(\nu)}(x)$ is bounded.
(c) For all ν with $|\nu| \leq 3$, the remainder terms $R_n^{(\nu)}(x)$ are such that with probability one, $\sup_{x_{-k}\in A} |R_n^{(\nu)}(x)| = o(n^{-(r-|\nu|)/(2r+1)})$.
(d) The functions a_ν with $|\nu| \leq 3$ are Lipschitz continuous, and the kernel derivatives $K^{(j)}$ with $j = 1,2,3$ are bounded, symmetric about zero, compactly supported, and Lipschitz continuous.

Under our assumptions, the pilot estimator $\hat{m}$ is uniformly consistent on the compact set A with rate $\sqrt{\log n/nh^d} + h^r$. Many of the usual multivariate regression estimators satisfy this expansion, including local polynomial mean and median regression estimators.

The first part of condition B1 requires that at least one of the component functions be strictly monotonic in some neighborhood of the set $\mathcal{M}_k$. Note that the function $\sum_{\ell \neq k} m_\ell(x_\ell)$ must take on both positive and negative signs, which implies that there must be some singularity points; i.e., there exists $x^*_{-k} \in A$ such that $\sum_{\ell \neq k} m_\ell(x^*_\ell) = 0$. Hence, the set $\mathcal{M}_k$ is nonempty. In general, $\mathcal{M}_k$ will determine a region of dimensions $d - 2$ in $\mathbb{R}^{d-1}$. For example, $m_j(x_j) = x_j$ and $m_l(x_l) = 0$, $l \neq k, j$, where $x_l \in [-1,1]$. In this example, $\mathcal{M}_k = \{x_{-k} : x_j = 0\}$, which is a linear subspace of dimensions $d - 2$. Furthermore, $m'_j(x_j) = 1$ but $m'_l(x_l) = 0$, $l \neq j, k$. Suppose instead that $m_l(x_l) = x_l^2 - \frac{1}{2}$, where $x_l \in [0,1]$, $l = 2, \ldots, d$. The set $\mathcal{M}_k$ is the surface of a $d - 1$–dimensional sphere of diameter $(d-1)/2$, whereas $m'_l(x_l) \neq 0$ for all $x_l \neq 0$. Regarding the condition on the density f_m, this condition is likely to hold in many examples. It is necessary because we have to deal with integrals of the form

$$\int_A \frac{g(x)\,dx_{-k}}{\left|\sum_{j \neq k}^{d} m_j(x_j)\right|^{\alpha}} \tag{8}$$

for constants $\alpha > 0$ and bounded continuous functions g. This assumption helps us to pin down for which values of α integrals such as (8) exist; specifically, it makes the existence of (8) equivalent to the existence of the univariate integral $\int_0^1 |y|^{-\alpha} dy$. It is a weak assumption in this context: if f_m were zero on $\mathcal{M}_k$, we might obtain better results regarding existence of (8) (which might imply better rates of convergence for our estimator).

Condition B2(a) is important because we use integration by parts to borrow the smoothness of p_{-k}. Conditions B2(b)–(d) ensure that the first three derivatives of $\hat{m}$ are uniformly consistent on A. There are further restrictions on r, q, d, which are necessary to make the remainder terms of smaller order than the leading terms; these are given in the statement of Theorem 2, which follows.

THEOREM 2.

(a) ($q > \frac{3}{2}$) *Suppose that the pilot estimator $\hat{m}(x)$ satisfies the linear expansion (6) and (7) and that assumptions A hold. Suppose also that the bandwidth sequence satisfies $h = \gamma n^{-1/(2r+1)}$ for some γ with $0 < \gamma < \infty$ and that $r > \max\{2, d-1\}$. Then there exists an increasing sequence δ_n and some finite constants $\mu_k(x_k), s_k(x_k)$ such that*

$$\delta_n\{\hat{\tau}_k(x_k) - \tau_k(x_k)\} \to N[\mu_k(x_k), s_k(x_k)], \tag{9}$$

where $\delta_n = n^{r/(2r+1)}$.

(b) ($1 < q \leq \frac{3}{2}$) *Suppose that the pilot estimator $\hat{m}(x)$ satisfies the linear expansion (6) and (7) and that Assumptions A and B hold. Furthermore, suppose that $r > \max\{3, d+1\}$. Suppose also that the bandwidth sequence satisfies $h = \gamma(\log n/n)^{1/(2r+4-2q)}$ for some γ with $0 < \gamma < \infty$. Then (9) holds with $\delta_n = n^{r/(2r+4-2q)}/\log n$.*

(c) ($q = 1$) *Suppose that the pilot estimator* $\hat{m}(x)$ *satisfies the linear expansion in (6) and (7) and that Assumptions A and B hold. Furthermore, suppose that* $r > \max\{3, d + 1\}$. *Suppose also that the bandwidth sequence satisfies* $h = \gamma n^{-1/(2r+2)}$ *for some* γ *with* $0 < \gamma < \infty$. *Then (9) holds with* $\delta_n = n^{r/(2r+2)}$.

The constant terms $\mu_k(x_k), s_k(x_k)$ are given in Linton (1999).[1]

4. CONCLUSION

We have one positive result and one negative result about the partial L_q estimator. First, it is well defined in cases where the original marginal integration estimator is not, and it is consistent in such cases. Second, there is a reduced rate of convergence in general, so that whereas the marginal integration estimator has the one-dimensional convergence rate, the partial median estimator has a two-dimensional convergence rate.

NOTE

1. The restrictions on r, d are phrased in this way because we are using a given bandwidth. For comparison, Andrews (1994, p. 2271) requires $r > d/2$ for a stochastic equicontinuity result useful for semiparametric estimation, whereas Linton and Härdle (1996) require $r > d - 1$. In the second case we require at least three derivatives, as already discussed.

REFERENCES

Andrews, D.W.K. (1994) Empirical process methods in econometrics. In R.F. Engle & D.F. MacFadden (eds.), *The Handbook of Econometrics*, vol. IV, pp 2247–2294.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Arcones, M.A. (1996) The Bahadur–Kiefer representation of L_p regression estimator. *Econometric Theory* 12, 257–283.

Chaudhuri, P. (1991a) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760–777.

Chaudhuri, P. (1991b) Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.

Chow, Y.S. & H. Teicher (1997) *Probability Theory: Independence, Interchangeability, and Martingales*, 3rd ed. Berlin: Springer.

Gel'fand, I.M. & G.E. Shilov (1964) *Generalized Functions*. London: Academic Press.

Härdle, W. (1990) *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Horowitz, J. (1999) Nonparametric estimation of a Generalised Additive Model with unknown link function. *Econometrica*, forthcoming.

Koenker, R. & G. Bassett, Jr. (1978) Regression quantiles. *Econometrica* 46, 33–50.

Linton, O.B. (2000) Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory* 16, 502–523.

Linton, O.B. (1999) Estimating Additive Nonparametric Models by Partial L_q-Mediannning: The Curse of Fractionality. Available at http://econ.lse.ac.uk/staff/olinton/.

Linton, O.B., R. Chen, N. Wang, & W. Härdle (1997) An analysis of transformations for additive nonparametric regression. *Journal of the American Statistical Association* 92, 1512–1521.

Linton, O.B. & W. Härdle (1996) Estimation of additive regression models with known links. *Biometrika* 83, 529–540.

Linton, O.B. & J.P. Nielsen. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.
Masry, E. (1996) Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.
Müller, H.G. (1988) *Nonparametric Regression Analysis of Longitudinal Data*. Lecture Notes in Statistics, vol. 46. Heidelberg/New York: Springer-Verlag.
Newey, W.K. (1994) Kernel estimation of partial means. *Econometric Theory* 10, 233–253.
Phillips, P.C.B. (1991) A shortcut to LAD estimator asymptotics. *Econometric Theory* 7, 450–463.
Spivak, M. (1965) *Calculus on Manifolds*. Reading, MA: Addison-Wesley.
Stone, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 685–705.
Tjøstheim, D. & B. Auestad (1994) Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association* 89, 1398–1409.

APPENDIX

Proof of Theorem 1. Let $\bar{G}(t) = \int_{A_n} \psi(m(x) - t) p_{-k}(x_{-k}) dx_{-k}$. Then $|G_n(t) - G(t)| \le |G_n(t) - \bar{G}(t)| + |\bar{G}(t) - G(t)|$, where with probability one

$$\sup_{t\in T} |\bar{G}(t) - G(t)| = \sup_{t\in T} \left| \int_{A_n^c} \psi(m(x) - t) p_{-k}(x_{-k}) dx_{-k} \right| = o(1)$$

by dominated convergence, provided $\sup_{t\in T} \int |\psi(m(x) - t)| p_{-k}(x_{-k}) dx_{-k} < \infty$ and $\lambda(A_n^c) \to 0$ as $n \to \infty$, where λ denotes Lebesgue measure. It remains to show that $\sup_{t\in T} |G_n(t) - \bar{G}(t)| = o(1)$ with probability one. By Arcones (1996, Lemma 4), we have

$$|\psi(a+b) - \psi(a)| \le \begin{cases} 2^{2-q}|b| & \text{if } 1 < q \le 2 \\ 2\{|a| \le |b|\} & \text{if } q = 1. \end{cases}$$

Therefore, when $1 < q \le 2$, we have (taking $a = m(x) - t$ and $b = \hat{m}(x) - m(x)$)

$$\begin{aligned} |G_n(t) - G(t)| &= \left| \int_{A_n} \{\psi(\hat{m}(x) - t) - \psi(m(x) - t)\} p_{-k}(x_{-k}) dx_{-k} \right| \\ &\le \int_{A_n} |\psi(\hat{m}(x) - t) - \psi(m(x) - t)| p_{-k}(x_{-k}) dx_{-k} \\ &\le 2^{2-q} \int_{A_n} |\hat{m}(x) - m(x)| p_{-k}(x_{-k}) dx_{-k} \\ &\le 2^{2-q} \sup_{x_{-k}\in A_n} |\hat{m}(x) - m(x)| \\ &\to 0 \end{aligned}$$

with probability one. When $q = 1$, we have

$$|G_n(t) - G(t)| \le 2\int_{A_n} \{|m(x) - t| \le |\hat{m}(x) - m(x)|\} p_{-k}(x_{-k}) dx_{-k}$$
$$\le 2\int_{A_n} \left\{ |m(x) - t| \le \sup_{x_{-k}\in A_n} |\hat{m}(x) - m(x)| \right\} p_{-k}(x_{-k}) dx_{-k}. \qquad \textbf{(A.1)}$$

Now, for any $\epsilon > 0$ there exists an $\eta > 0$ such that $\sup_{t\in T} \lambda(A_{nt}(\eta)) < \epsilon$, where $A_{nt}(\eta) = \{x_{-k}: |m(x) - t| \le \eta\}$. Therefore, for any given $\epsilon > 0$ we can bound the right-hand side of (A.1) by

$$2\int_{A_{nt}(\eta)} \{|m(x) - t| \le Y_n\} p_{-k}(x_{-k}) dx_{-k} + 2\int_{A^c_{nt}(\eta)} \{\eta \le Y_n\} p_{-k}(x_{-k}) dx_{-k}$$
$$\le 2\int_{A_{nt}(\eta)} \{|m(x) - t| \le Y_n\} p_{-k}(x_{-k}) dx_{-k} + o(1)$$
$$\le 2\epsilon + o(1),$$

with probability one, where $Y_n = \sup_{x_{-k}\in A_n} |\hat{m}(x) - m(x)|$. Therefore, because ϵ is arbitrary and independent of t, we have $\sup_{t\in T} |G_n(t) - G(t)| \to 0$ with probability one. Combining these facts with the unique minimizing condition, we conclude that if $\tau_k(x_k)$ is the unique minimizer of $G(t)$, then $\hat{\tau}_k(x_k) \to \tau_k(x_k)$ with probability one. ■

Proof of Theorem 2. First, we define a linearized version $\mathcal{G}_n(t)$ of $G_n(t)$ and provide a central limit theorem for $\mathcal{G}_n(t)$. We then prove that $\mathcal{G}_n(t)$ and $G_n(t)$ are uniformly close to each other. Finally, a simple mean value expansion gives the asymptotic distribution of zeros of $\mathcal{G}_n(t)$, which we have already shown are close to zeros of $G_n(t)$. See Phillips (1991) for an accessible treatment of the asymptotics for least absolute deviations (LAD) estimators in the parametric case.

Case $Q > 1$. Define $G(t) = \int_A \mathrm{sign}(m(x) - t)|m(x) - t|^{q-1} p_{-k}(x_{-k}) dx_{-k}$ and let

$$\mathcal{G}_n(t) = G(t) + (q - 1)\int_A |m(x) - \tau_k(x_k)|^{q-2} \{\hat{m}(x) - m(x)\} p_{-k}(x_{-k}) dx_{-k}$$
$$\equiv G(t) + \Delta_n.$$

Note that the function $\psi'(u) = (q - 1)|u|^{q-2}$ is continuous everywhere when $q \ge 2$ but is discontinuous at the origin when $1 < q \le 2$. By substituting the stochastic expansion of $\hat{m}(x) - m(x)$ into Δ_n we can write $\Delta_n = \Delta_n^S + \Delta_n^B + \textit{remainder}$. The main difficulty in establishing the asymptotics of $\mathcal{G}_n(t)$ is the random sequence Δ_n—and in particular, the stochastic part of it

$$\Delta_n^S = (q - 1)\int_A |m(x) - \tau_k(x_k)|^{q-2} \left(\sum_{i=1}^n W_{n,i}^{(0)}(x)\eta_i \right) p_{-k}(x_{-k}) dx_{-k}.$$

Case $Q > \frac{3}{2}$. We first approximate Δ_n^S by the random sequence

$$\Delta_n^{S*} = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_k - X_{ki}}{h}\right)\frac{a(x_k, X_{-ki})}{\left|\sum_{\ell\neq k} m_\ell(X_{\ell i})\right|^{2-q}} p_{-k}(X_{-ki})\eta_i, \tag{A.2}$$

which has finite (unconditional) variance and is $O_p(n^{-1/2}h^{-1/2})$ (by the assumption that $f_m > 0$ on a neighborhood of $\mathcal{M}_1$, the approach to singularity occurs at a linear rate; i.e., it suffices to show that the integral $\int_0^1 u^{2q-4}du < \infty$, which it does if and only if $q > \frac{3}{2}$). Furthermore, it is asymptotically normal, i.e., $\sqrt{nh}\Delta_n^{S*} \to N(0, v(x_k))$ for some finite $v(x_k)$ by standard verification of the Lindeberg condition. The bias term $\Delta_n^B = (q-1)\int_A |m(x) - \tau_k(x_k)|^{q-2}(\sum_{k=1}^d \chi_{nk}^{(0)}(x_k))\underline{p_{-k}}(x_{-k})dx_{-k}$ is $O(h^r)$ by Assumption A(g) using standard arguments. Therefore, $\sqrt{nh}\mathcal{G}_n(\tau_k(x_k))$ is asymptotically normal (furthermore, the optimal bandwidth that balances the squared bias against the variance is $h \propto n^{-1/(2r+1)}$, which results in a rate of convergence of $n^{r/(2r+1)}$). Finally, we show that we can restrict attention to the linearization $\mathcal{G}_n(t)$. We have for any sequence ϵ_n converging to zero $\sup_{|t-\tau_k(x_k)|\le\epsilon_n}|G_n(t) - \mathcal{G}_n(t)| = o_p(n^{-1/2}h^{-1/2})$. Furthermore, $\sup_{|t-\tau_k(x_k)|\le\epsilon_n}|\mathcal{G}_n(t) - G(t)| = \Delta_n = O_p(n^{-1/2}h^{-1/2})$. This establishes that the zeros of $\mathcal{G}_n(t)$ and $G_n(t)$ are both $O_p(1/\sqrt{nh})$ distance from the zero of $G(t)$, which is $\tau_k(x_k)$, and moreover that the zero of $\mathcal{G}_n(t)$ is distance $o_p(n^{-1/2}h^{-1/2})$ from the zero of $G_n(t)$. Let $\bar{\tau}_k(x_k)$ be a zero of $\mathcal{G}_n(t)$, in which case, by a Taylor expansion $0 = \mathcal{G}_n(\bar{\tau}_k(x_k)) = G'(\tau_k(x_k))\sqrt{nh}(\bar{\tau}_k(x_k) - \tau_k(x_k)) + \sqrt{nh}\Delta_n + o_p(1)$ using the fact that $G'(t)$ exists and is nonzero in a neighborhood of $\tau_k(x_k)$. This leads to the distribution of $\sqrt{nh}(\bar{\tau}_k(x_k) - \tau_k(x_k))$ and hence $\sqrt{nh}(\hat{\tau}_k(x_k) - \tau_k(x_k))$. We conclude that $n^{r/(2r+1)}(\hat{\tau}_k(x_k) - \tau_k(x_k))$ is asymptotically normal.

Case $1 < Q \le \frac{3}{2}$. In this case, the unconditional variance of Δ_n^{S*} does not exist (unless perhaps $f_m = 0$ on $\mathcal{M}_k$), and we cannot apply a central limit theorem directly to (A.2). In this case, we use fractional integration by parts to approximate Δ_n^S by a statistic with finite variance. There will be a cost to this in terms of the rate of convergence; the corresponding Δ_n^{S*} satisfies $\Delta_n^{S*} = O_p(n^{-1/2}h^{-(4-2q)/2}\log n)$ when $1 < q < \frac{3}{2}$. We will outline the argument in some detail for the special case that $d = 2$, $m_2(x_2) = x_2$, $A = [0,1]$, and $K > 0$. In this case,

$$\begin{aligned}\Delta_n^S &= \frac{1}{nh^2}\sum_{i=1}^n \eta_i K\left(\frac{x_1 - X_{1i}}{h}\right)\int_0^1 \frac{1}{x_2^{2-q}} a(x)K\left(\frac{x_2 - X_{2i}}{h}\right)p_2(x_2)dx_2 \\ &= \frac{1}{nh^2}\sum_{i=1}^n \eta_i K\left(\frac{x_1 - X_{1i}}{h}\right)\int_0^1 D_{x_2}^{-\alpha}\left(\frac{a(x)p_2(x_2)}{x_2^{2-q}}\right)D_{x_2}^{\alpha}\left(K\left(\frac{x_2 - X_{2i}}{h}\right)\right)dx_2 \\ &= \frac{1}{nh^{2+\alpha}}\sum_{i=1}^n \eta_i K\left(\frac{x_1 - X_{1i}}{h}\right)\int_0^1 D_{x_2}^{-\alpha}\left(\frac{a(x)p_2(x_2)}{x_2^{2-q}}\right)K_\alpha\left(\frac{x_2 - X_{2i}}{h}\right)dx_2\end{aligned}$$

for any α with $0 < \alpha < 1$. Here, for any function K, $K_\alpha(a(x+b)) \equiv a^{-\alpha}D_x^\alpha(K(a(x+b)))$, where $\lim\sup_a K_\alpha(a(x+b)) < \infty$.

We now make a change of variables $x_2 \mapsto u = (x_2 - X_{2i})/h$ to obtain $\Delta_n^S = \Delta_n^{S*}\{1 + o_p(1)\}$, where

$$\Delta_n^{S*} = \frac{1}{nh^{1+\alpha}} \sum_{i=1}^{n} \eta_i K\left(\frac{x_1 - X_{1i}}{h}\right) M_{-\alpha}(x_1, X_{2i}) \mu_0(K_\alpha), \tag{A.3}$$

where $M_{-\alpha}(x_1,x_2) = D_{x_2}^{-\alpha}(a(x)p_2(x_2)x_2^{q-2})$. Note that $|M_{-\alpha}(x_1,x_2)| \le \{\sup|a(x) \times p_2(x_2)|\}|D_{x_2}^{-\alpha}(x_2^{q-2})| = O(x_2^{\alpha+q-2})$, as $x_2 \to 0$, because a, p_2, and x_2 are positive. We therefore take $\alpha = (3-2q)/2$, in which case $M_{-((3-2q)/2)}(x_1,x_2) = O(x_2^{-1/2})$ as $x_2 \to 0$. Although the unconditional variance of Δ_n^{S*} does not exist in this case, the conditional variance does with probability one, and we can show that

$$\begin{aligned} \operatorname{var}[\Delta_n^{S*}|X_1,\dots,X_n] &= \frac{1}{n^2h^{5-2q}} \sum_{i=1}^{n} \sigma^2(X_i) K^2\left(\frac{x_1 - X_{1i}}{h}\right) M_i^2 \mu_0^2(K_{(3-2q)/2}) \\ &= O_p(n^{-1}h^{2q-4}\log n), \end{aligned} \tag{A.4}$$

where $M_i = M_{-((3-2q)/2)}(x_1, X_{2i})$. In fact, Δ_n^{S*} is asymptotically normal provided the correct standardization is used, as we now show. Note that provided $f_2(x_2) > 0$, where f_2 is the marginal density of X_2, we have $\min_{i\le n} X_{2i} = O_p(n^{-1})$. Therefore, we can approximate $\sqrt{nh^{4-2q}/\log n}\,\Delta_n^{S*}$ by

$$\frac{1}{\sqrt{nh\log n}} \sum_{i=1}^{n} K\left(\frac{x_1 - X_{1i}}{h}\right)\eta_i M_i \times \mathbf{1}\left(X_{2i} \ge \frac{1}{n\log n}\right)\mu_0(K_{(3-2q)/2}) \equiv \sum_{i=1}^{n} Z_{ni}, \tag{A.5}$$

where Z_{ni} are mean zero and independent random variables with finite second moments. We then apply the Lindeberg central limit theorem for triangular arrays (Chow and Teicher, 1997, p. 351), which in this case requires only that (a) $nEZ_{ni}^2 \to v$ for some finite positive v and (b) $nEZ_{ni}^2\mathbf{1}(|Z_{ni}| \ge \epsilon) \to 0$ for all $\epsilon > 0$. The conclusion is that $\sum_{i=1}^n Z_{ni}$ (and hence $\sqrt{nh^{4-2q}/\log n}\,\Delta_n^{S*}$) is asymptotically normal with mean zero and variance v. We show why (a) is plausible. We have

$$\begin{aligned} nEZ_{ni}^2 &= \frac{\mu_0^2(K_{(3-2q)/2})}{h\log n}\int K^2\left(\frac{x_1 - X_1}{h}\right)\sigma^2(X) M_{-((3-2q)/2)}^2(X) \\ &\quad \times \mathbf{1}\left(X_2 \ge \frac{1}{n\log n}\right) f(X)\,dX \\ &= \mu_0^2(K_{(3-2q)/2})\|K\|^2 \frac{1}{\log n}\int \sigma^2(x) M_{-((3-2q)/2)}(x) \\ &\quad \times \mathbf{1}\left(x_2 \ge \frac{1}{n\log n}\right) f(x)\,dx_2\{1 + o(1)\} \\ &\le \mu_0^2(K_{(3-2q)/2})\|K\|^2 \times c \times \frac{1}{\log n}\int_{\frac{1}{n\log n}}^{1} \frac{dx_2}{x_2} \\ &= \mu_0^2(K_{(3-2q)/2})\|K\|^2 \times c \times \frac{-\log(1/n\log n)}{\log n} \\ &= \mu_0^2(K_{(3-2q)/2})\|K\|^2 \times c \times \left(1 + \frac{\operatorname{loglog} n}{\log n}\right), \end{aligned}$$

where c is some finite constant. The second line follows from a change of variables, and the inequality uses the fact that σ^2 and f are bounded functions and our earlier arguments. Similarly, we can bound nEZ_{ni}^2 away from zero. As for condition (b), we first show that for any subsequence $\{n_k\}_{k=1}^{\infty}$, $(^*)$ $\max_{1\le i\le n_k} w_{n_k i} \to_p 0$, where $w_{ni} = K((x_1 - X_{1i})/h)X_{2i}^{-1/2}/\sqrt{nh\log n} > 0$. By independence and the law of iterated expectation, we have for any $\epsilon > 0$,

$$\Pr\left(\max_{1\le i\le n_k} w_{n_k i} \le \epsilon\right) = [\Pr(w_{n_k i} \le \epsilon)]^{n_k} = \left[1 - E\left\{F_{2|1}\left(\frac{K\left(\dfrac{x_1 - X_{1i}}{h(n_k)}\right)}{\epsilon \times n_k h(n_k)\log n_k}\right)\right\}\right]^{n_k},$$

where $F_{2|1}$ denotes the conditional cumulative distribution function of X_{2i} given X_{1i}. Now, because $K((x_1 - X_{1i})/h(n_k))/\epsilon \times n_k h(n_k)\log n_k \to 0$ and $F_{2|1}(0) = 0$, we have by a Taylor expansion and change of variables that

$$\begin{aligned}\Pr\left(\max_{1\le i\le n_k} w_{n_k i} \le \epsilon\right) &\simeq \left[1 - f_{2|1}(0)\,\frac{E\left\{K\left(\dfrac{x_1 - X_{1i}}{h(n_k)}\right)\right\}}{\epsilon \times n_k h(n_k)\log n_k}\right]^{n_k} \\ &= \left[1 - \frac{f_{2|1}(0)\displaystyle\int K(u) f_1(x_1 - uh(n_k))\,du}{\epsilon \times n_k \log n_k}\right]^{n_k} \\ &\to 1,\end{aligned}$$

where $f_{2|1}$ is the conditional density function of X_{2i} given X_{1i}, whereas f_1 is the marginal density of X_{1i}, which establishes $(^*)$. This in turn implies that $\max_{1\le i\le n_k} w^*_{n_k i} \to 0$ in probability, where $w^*_{ni} = K((x_1 - X_{1i})/h)X_{2i}^{-1/2}\sigma(X_i)\mathbf{1}(X_{2i} \ge 1/n\log n)/\sqrt{nh\log n}$, because $\sigma(X_i)$ is bounded. Therefore, $\max_{1\le i\le n_k} w_{n_{k_l} i} \to 0$ with probability one along some subsubsequence $\{n_{k_l}\}_{l=1}^{\infty}$, which implies that condition (4.14) of Müller (1988, p. 31) is satisfied, and hence the Lindeberg condition (b) is satisfied along these subsequences conditional on $X_1, \ldots, X_{n_{k_l}}$ with probability one. In conclusion, $\sum_{i=1}^{n_{k_l}} Z_{n_{k_l} i} \to N(0, v)$ with probability one conditional on $X_1, \ldots, X_{n_{k_l}}$. Because the limit distribution does not depend on $X_1, \ldots, X_{n_{k_l}}$, the weak convergence holds unconditionally. Finally, we have $\sum_{i=1}^{n} Z_{ni} \to N(0, v)$ by the sequential compactness property of real sequences.

The bias term Δ_n^B is of order h^r as before. This means that the optimal bandwidth is $h \propto (\log n/n)^{1/(2r+4-2q)}$ and the optimal rate of convergence is $O_p(n^{-r/(4-2q+2r)} \times (\log n)^{1/2})$. In conclusion, $\sqrt{nh^{4-2q}/\log n}\,\mathcal{G}_n(\tau_k(x_k))$ is asymptotically normal. Finally, we can show that the linearization error is $o_p(n^{-r/(4-2q+2r)}\log n)$.

Case $Q = 1$. In this case, the preceding methods do not apply because $\psi(u) = \mathrm{sign}(u)$ is discontinuous with a nonremovable singularity. However, the functions $G_n(t) = \int_A \mathrm{sign}(\hat{m}(x) - t)p_{-k}(x_{-k})\,dx_{-k}$ and $G(t) = \int_A \mathrm{sign}(m(x) - t)p_{-k}(x_{-k})\,dx_{-k}$ are continuously differentiable in t, and the derivative of G, e.g., can be represented as $\partial G(t)/\partial t = 2\int \delta_0(m(x) - t)p_{-k}(x_{-k})\,dx_{-k}$, where $\delta_x(\cdot)$ is the Dirac delta (generalized) function, which is defined through its integrals; i.e., $\int \delta_x(t)g(t)\,dt = g(x)$ for any univariate function $g(\cdot)$ that is continuous at x (see Gel'fand and Shilov, 1964). In this case, let

$$\mathcal{G}_n(t) = \int_A \operatorname{sign}\{m(x) - t\} p_{-k}(x_{-k}) dx_{-k}$$
$$+ \int_A \delta_0(m(x) - \tau_k(x_k))\{\hat{m}(x) - m(x)\} p_{-k}(x_{-k}) dx_{-k} \equiv G(t) + \Delta_n.$$

We show for any sequence ϵ_n converging to zero that $\sup_{|t-\tau_k(x_k)|\le\epsilon_n} |\mathcal{G}_n(t) - G_n(t)| = o_p(n^{-1/2}h^{-1})$, and $\sqrt{nh^2}\mathcal{G}_n(\tau_k(x_k)) = \sqrt{nh^2}\int_A \delta_0(m(x) - \tau_k(x_k))\{\hat{m}(x) - m(x)\}p_{-k}(x_{-k})dx_{-k} \to N(b(x_k), v(x_k))$ for some $b(x_k), v(x_k)$. Then, provided $\Gamma = G'(\tau_k(x_k)) = \int \delta_0(m(x) - \tau_k)p_{-k}(x_{-k})dx_{-k} \neq 0$, we have the limiting distribution of $\hat{\tau}_k(x_k)$ by the same arguments used earlier. By Theorem 5.1 of Spivak (1965), $\mathcal{M}_k$ is a manifold provided the vector $(m'_\ell(x_\ell), \ell \neq k) \neq 0$ on some open neighborhood of $\mathcal{M}_k$ in $\mathbb{R}^{d-1}$. In this case, these "delta integrals" can be represented as ordinary Lebesgue integrals over $\mathcal{M}_k$, i.e., $\int \delta_0(m(x) - \tau_k(x_k))p_{-k}(x_{-k})dx_{-k} = \int_{\mathcal{M}_k} p_{-k}(x_{-k})dx_{-k}$ and $\int \delta_0(m(x) - \tau_k(x_k))\{\hat{m}(x) - m(x)\}p_{-k}(x_{-k})dx_{-k} = \int_{\mathcal{M}_k}\{\hat{m}(x) - m(x)\}p_{-k}(x_{-k})dx_{-k}$ (see Gel'fand and Shilov, 1964, Ch. III). Because $\mathcal{M}_k$ generally determines a region of dimensions $d - 2$, the random sequence in $\sqrt{nh^2}\mathcal{G}_n(\tau_k(x_k))$ behaves like the integral of a d-dimensional smoother with respect to a $d - 2$–dimensional integrator; i.e., it should behave like a two-dimensional regression smoother. In the special case that $m_\ell(x_\ell) = x_\ell$ for some $\ell \neq k$, and $m_j(x_j) = 0$ for all $j \neq \ell, k$, this interpretation is exact because then $\mathcal{M}_k = \{x_{-k} : x_\ell = 0\}$ and

$$\int_{\mathcal{M}_k} \{\hat{m}(x) - m(x)\}p_{-k}(x_{-k})dx_{-k} = \int \{\hat{m}(x_{-k}, 0) - m(x_{-k}, 0)\}p_{-k}(x_{-\ell,k}, 0)dx_{-\ell,k},$$

where $x_{-\ell,k}$ is the $(d - 2) \times 1$ vector excluding both x_ℓ and x_k. Note that asymptotic normality follows from an application of the Lindeberg central limit theorem. In the general case, we must apply a change of variables argument to return to ordinary integrals, but we end up with the same qualitative behavior. ■