Estimation and Inference for Generalized Full and Partial Means and Average Derivatives*

Guido W. Imbens† and Geert Ridder‡

First Draft: May 2006
This version: March 2009

Abstract

We propose new semiparametric estimators for parameters that depend on the derivatives (up to any finite order) of unknown conditional expectations and densities. We consider two cases. In the first we average over all (conditioning) variables in the conditional expectation or density. In the second case we average over a strict subset of the conditioning variables. The unknown conditional expectations and densities are estimated by a first step kernel estimator. The kernel estimator has a boundary correction that makes it uniformly consistent if the distribution of the covariates has bounded support. The partial and full mean estimators therefore do not require trimming (asymptotic or fixed) as in the estimators developed by Newey (1994) and Powell, Stock, and Stoker (1989) and in many applications to specific settings. We provide a general formula for the influence function and the asymptotic variance for both full and partial averaging. We also specify a general set of regularity conditions that contains a new restriction on the kernel function to avoid bias in the case that the parameter depends on the derivatives of the conditional expectation or density.

JEL Classification: C14, C21, C52 Keywords: *Nonparametric Density Estimation, Uniform Convergence, Compact Support, Partial Means, Average Derivatives, Trimming*

*Financial support for this research was generously provided through NSF grant SES 0452590. We thank Matias Cattaneo, Richard Crump, Bryan Graham, Michael Jansson, and Jim Powell for comments.

†Department of Economics, Harvard University, M-24 Littauer Center, 1850 Cambridge Street, Cambridge, MA 2138, and NBER. E-Mail: imbens@harvard.edu, Web: http://www.economics.harvard.edu/faculty/imbens/imbens.html.

‡Department of Economics, University of Southern California, 310A Kaprielian Hall, Los Angeles, CA 90089. E-Mail: ridder@usc.edu, Web: http://www-rcf.usc.edu/~ridder/

1 Introduction

In this paper we consider estimation and inference for partial and full means. These estimands were introduced in the econometric literature by Newey (1994). We focus on estimands that are weighted averages of smooth functions of (derivatives of) of regression functions. The regression functions are not known and must be estimated nonparametrically. The averaging can be over the full set of regressors (the full mean case), or over a subset of the regressors (the partial mean case). This setting covers many semiparametric models that have been considered in the literature. Examples for the full mean case include average treatment effects in the treatment effect literature, and the partial linear model developed by Robinson (1988), and in the partial mean case the dose-response function studied by Imbens (2000). Examples of full average derivatives include the density weighted average derivative introduced by Powell, Stock and Stoker (1989), and the unweighted average derivative analyzed by Härdle and Stoker (1989).

Newey (1994) considers estimation and inference for partial and full means and analyzes the properties of estimators where in the first step the regression function is estimated using kernel methods. In order to ensure uniform convergence he uses a fixed trimming procedure. We develop a new boundary correction that ensures uniform convergence of the kernel estimator over the entire support even if this support is compact with the density bounded away from zero. The new estimator projects points close to the boundary to the internal region, defined as the subset of the support not affected by boundary bias and then uses a Taylor series expansion of the estimated density at an internal point to estimate the density at the point of interest. We show that this modified kernel estimator can be used to estimate partial and full means, in the case that the support of the covariates is compact. We derive the asymptotic properties of the estimators. In general the estimators are asymptotically normal with no asymptotic bias. In the case of the full mean the convergence rate is the regular parametric rate and the estimator achieves the semiparametric efficiency bound. For this case the results show that asymptotic properties are similar to those obtained for series or sieve estimates. For the partial mean case the convergence rate is slower, depending on the number of covariates that are not averaged over. In Section 2 we discuss generalized partial means and give some examples where estimands of interest may have this form. In Section 3 we discuss the first stage in the estimation procedure, the nonparametric estimation of the regression function. Here we introduce the new boundary correction and derive its properties. In Section 4 we discuss inference for partial and full means. We also report results from a small simulation study to investigate how reliable the results from our proposed methods are for realistic settings (modelled on the lottery data set). Section 6 concludes.

2 Generalized Full and Partial Means

We consider a parameter that can be expressed as the expected value of a nonlinear function of a vector of conditional expectations of an M dimensional random vector Y given an L dimensional covariate vector X. The covariates have a distribution that is absolutely continuous with respect to the Lebesgue measure and that has a compact support $\mathbb{X}$. Let $g(x)$ denote the conditional expectation:

$$g(x) = \mathbb{E}[Y|X = x]. \tag{2.1}$$

In addition define the $M + 1$ dimensional vector of functions

$$h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix} = \begin{pmatrix} f_X(x) \\ f_X(x) \cdot g(x) \end{pmatrix}. \tag{2.2}$$

Although g depends on the full vector X, the expectation is computed over the the distribution of the L_1 dimensional subvector X_1. The remaining $L_2 = L - L_1$ components of X, X_2, are evaluated at a function of X_1, i.e. at $X_2 = t(X_1)$ with $t : \mathbb{R}^{L_1} \to \mathbb{R}^{L_2}$ a known function. In Newey's (1994) partial mean parameter this is a constant function, i.e. $x_2 = t(X_1)$. A limiting case is that $L_1 = L$, i.e. the expectation is over the full covariate vector. We now define the *Generalized Partial Mean* (GPM) as the Q dimensional parameter vector

$$\theta_{\text{gpm}} = \mathbb{E}\left[\omega(X_1)' m\left(h\left(X_1, t(X_1)\right)\right)\right], \tag{2.3}$$

with $m : \mathbb{R}^M \to \mathbb{R}^P$ a known function that depends on X_1 only through g and ω a $Q \times P$ matrix of known weight functions that depend on X_1. If $L_1 = L$ we refer to this parameter as the *Generalized Full Mean* (GFM),

$$\theta_{\text{gfm}} = \mathbb{E}\left[\omega(X)' m\left(h\left(X\right)\right)\right]. \tag{2.4}$$

The data are a random sample from the joint distribution of Y, X: The estimators of θ_{gpm} and θ_{gfm} are

$$\hat{\theta}_{\text{gpm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_{1i})' m\left(\hat{g}_{NIP,s}\left(X_{1i}, t(X_{1i})\right)\right), \tag{2.5}$$

and

$$\hat{\theta}_{\text{gfm}} = \frac{1}{N}\sum_{i=1}^{N} \omega(X_i)' m\left(\hat{g}_{NIP,s}\left(X_i\right)\right), \tag{2.6}$$

with $\hat{g}_{NIP,s}(x)$ a nonparametric, kernel-type, estimator of $g(x)$ that will be defined later.

INCLUDE EXAMPLES FROM PRESENTATIONS. WHAT ABOUT CAUSAL INTERPRETATION IN FULL AND PARTIAL CASE? E.G. WHY DO LONG REGRESSION IF ONLY INTERESTED IN THE EFFECT OF A SINGLE COVARIATE? WHAT DO WE MEAN BY 'CONTROLLING ON AVERAGE'? LINK WITH ASF IN NONSEPARABLE MODELS.

Example 2.1 (Full Mean: Average Treatment Effect an Average Treatment Effect on the Treated)
Suppose units in a population are characterized by two potential outcomes, $Y(0)$ and $Y(1)$,

outcomes given a control and active treatment respectively. Interest is in the average difference between these, $\theta_{ATT} = \mathbb{E}[Y(1) - Y(0)]$. Suppose one is willing to assume that given some covariates X assignment to the treatment, denoted by $W \in \{0, 1\}$, is independent of the potential outcomes, or $W \perp Y(0), Y(1)|X$ (unconfoundedness/selection-on-observables). This setting is widely studied in the program evaluation literature. See for example Rosenbaum and Rubin (1984), Hahn (1998), Heckman, Ichimura and Todd (1998), Hirano, Imbens and Ridder (2003). Then

$$\theta_{ATT} = \mathbb{E}\left[\frac{\mathbb{E}[Y \cdot W|X]}{\mathbb{E}[W|X]} - \frac{\mathbb{E}[Y \cdot (1-W)|X]}{\mathbb{E}[1-W|X]}\right].$$

This fits into the full mean form by setting

$$\omega(x) = 1,$$

$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ g_3(x) \end{pmatrix} = \begin{pmatrix} \mathbb{E}[W|X=x] \\ \mathbb{E}[YW|X=x] \\ \mathbb{E}[Y(1-W)|X=x] \end{pmatrix},$$

and

$$m(g(x)) = g_2(x)/g_1(x) - g_3(x)/(1 - g_1(x)).$$

Hahn (1998) suggests an efficient estimator for θ_{ATT} based on this representation but using a series estimator for $g(x)$. Another treatment effect parameter of interest is the Average Treatment Effect on the Treated (ATET), $\mathbb{E}[Y(1) - Y(0)|W = 1]$. Following Hirano, Imbens and Ridder (2003), the efficient estimator if we know the ratio of the probability of selection and the fraction in the population that is treated $g_0(x)$, is obtained from

$$\theta_{ATET} = \mathbb{E}\left[g_0(X)\left(\frac{\mathbb{E}[Y \cdot W|X]}{\mathbb{E}[W|X]} - \frac{\mathbb{E}[Y \cdot (1-W)|X]}{\mathbb{E}[1-W|X]}\right)\right].$$

The estimator is again a full mean and the estimator is the same as that for the ATT with the only difference that

$$\omega(x) = g_0(x)$$

which is assumed to be known. As noted by Hirano, Imbens, and Ridder, the efficient estimator requires estimation of the (known) regression function in g.$\square$

Example 2.2 (Full Mean: Robinson Partial Linear Model)
Robinson (1988) is interested in estimating β in the partial linear model for the scalar dependent variable Y:

$$\mathbb{E}[Y|X = z, Z = z] = \beta' x + k(z).$$

Under his assumptions β is equal to

$$\beta = \left(\mathbb{E}\left[(X - \mathbb{E}[X|Z])(X - \mathbb{E}[X|Z])'\right]\right)^{-1} \mathbb{E}\left[(X - \mathbb{E}[X|Z])(Y - \mathbb{E}[Y|Z])\right].$$

All components of the matrix $\mathbb{E}\left[(X - \mathbb{E}[X|Z])\,(Y - \mathbb{E}[Y|Z])\right]$ have the full mean form. In the scalar X and Y case:

$$
\begin{aligned}
&\mathbb{E}\left[(X - \mathbb{E}[X|Z])\,(Y - \mathbb{E}[Y|Z])\right] \\
&\qquad = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X \cdot \mathbb{E}[Y|Z]\right] - \mathbb{E}\left[\mathbb{E}[X|Z] \cdot Y\right] + \mathbb{E}\left[\mathbb{E}[X|Z] \cdot \mathbb{E}[Y|Z]\right] \\
&\qquad = \mathbb{E}\left[\mathbb{E}[XY|Z]\right] - \mathbb{E}\left[\mathbb{E}[X|Z] \cdot \mathbb{E}[Y|Z]\right].
\end{aligned}
$$

This fits into the full mean setting by setting

$$
\begin{aligned}
&\omega(z) = 1, \\
&g(z) = \begin{pmatrix} g_1(z) \\ g_2(z) \\ g_3(z) \end{pmatrix} = \begin{pmatrix} \mathbb{E}[X|Z = z] \\ \mathbb{E}[Y|Z = z] \\ \mathbb{E}[X \cdot Y|Z = z] \end{pmatrix},
\end{aligned}
$$

and

$$
m(g(z)) = g_3(z) - g_1(x) \cdot g_2(x).
$$

Hence β can be written as a smooth function of full means. Robinson (1988) proposes a kernel estimator for $g(z)$, but uses trimming to get around boundary problems, requiring the choice of a second bandwidth (in addition to the bandwidth for the regression function itself). □

Example 2.3 (Partial Mean: Dose Response Function)
Imbens (2000), Hirano and Imbens (2003), and Flores (????) are interested in estimating the dose response function for an assigned value w

$$
\mu(w) = \mathbb{E}[Y(w)],
$$

where the researcher has available a random sample of (Y, W, X), with $Y = Y(W)$, under the assumption that

$$
Y(w) \perp W|X.
$$

If w is continuous we can estimate $\mu(w)$ for fixed w in the partial mean set up by setting

$$
\begin{aligned}
&t(x) = w, \\
&\omega(x) = 1, \\
&g(x, w) = \mathbb{E}[Y|X = x, W = w], \\
&m(g(x, w)) = g(x, w).
\end{aligned}
$$

□

Example 2.4 (Partial Mean: Aggregate Redistributional Effect)
Graham, Imbens and Ridder are interested in estimating the effect of assortive matching where inputs are reallocated. Their estimand can be written as

$$\theta = \mathbb{E}\left[g\left(X, F_W^{-1}\left(F_X(X)\right)\right)\right].$$

We can estimate θ in the partial mean set up by setting

$$t(x) = F_W^{-1}(F_X(x)),$$

$$\omega(x) = 1,$$

$$g(x,w) = \mathbb{E}[Y|X=x, W=w],$$

$$m(g(x,w)) = g(x,w).$$

□

3 Uniform Convergence of Kernel Estimators

In this section we study the problem of estimating $g(x)$. We do this indirectly, first by nonparametrically estimating the probability density function of X, $f_X(x)$, and then by estimating the product of $g(x)$ and $f_X(x)$, which we denote by $h(x)$. The regression function itself will then be estimated by the ratio $\hat{h}(x)/\hat{f}_X(x)$. The key is the development of an estimator for $h(x)$ that is uniformly consistent even if $h(x)$ is bounded away from zero on the compact support of X.

3.1 Notation and Set Up

First we introduce some notation to deal with the case where X is an L-dimensional vector. Let λ denote an L vector of nonnegative integers, with $|\lambda| = \sum_{l=1}^{L} \lambda_l$, and $\lambda! = \prod_{l=1}^{L} \lambda_l!$. For L vectors of nonnegative integers λ and μ let $\mu \leq \lambda$ be equivalent to $\mu_l \leq \lambda_l$ for all $l = 1, \ldots, L$, and define

$$\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \frac{\lambda!}{\mu!(\lambda-\mu)!} = \prod_{l=1}^{L} \frac{\lambda_l!}{\mu_l!(\lambda_l-\mu_l)!} = \prod_{l=1}^{L} \begin{pmatrix} \lambda_l \\ \mu_l \end{pmatrix}.$$

For L vectors λ and x let $x^\lambda = \prod_{l=1}^{L} x_l^{\lambda_l}$. As shorthand for partial derivatives we use $g^{(\lambda)}(x)$:

$$g^{(\lambda)}(x) = \frac{\partial g^{|\lambda|}}{\partial x^\lambda}(x).$$

For matrices we use the matrix norm

$$|A| = \sqrt{\mathrm{tr}(A'A)}.$$

The norm that we use for functions $g : \mathbb{X} \subset \mathbb{R}^L \to \mathbb{R}$ that are at least j times continuously differentiable is the Sobolev norm

$$|g|_j = \sup_{x \in \mathbb{X}, |\lambda| \leq j} \left| \frac{\partial g^{|\lambda|}}{\partial x^\lambda}(x) \right|.$$

Let $K : \mathbb{R}^L \mapsto \mathbb{R}$ denote the kernel function. We will assume that $K(u) = 0$ for $u \notin \mathbb{U}$ with $\mathbb{U}$ compact, and $K(u)$ bounded. The standard Nadaraya-Watson (NW) kernel density estimator, based on the bandwidth b, is:

$$\hat{f}_{X,\mathrm{nw}}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{b^L}K\left(\frac{x - X_i}{b}\right), \tag{3.7}$$

For the bandwidth b define the *internal region* of the support $\mathbb{X}$ as

$$\mathbb{X}_b^I = \left\{ x \in \mathbb{X} \,\middle|\, \left\{ \tilde{x} \in \mathbb{R}^k \,\middle|\, \frac{x - \tilde{x}}{b} \in \mathbb{U} \right\} \subset \mathbb{X} \right\} = \left\{ x \in \mathbb{X} \,|\, \{x - b \cdot u | u \in \mathbb{U}\} \subset \mathbb{X} \right\}. \tag{3.8}$$

This is a compact subset of the interior of $\mathbb{X}$ that contains all points that are sufficiently far away from the boundary that the standard kernel density estimator at those points is not affected by any potential discontinuity in the density at the boundary. In the case with $\mathbb{U} = [-1,1]^L$ and $\mathbb{X} = \bigotimes_{l=1}^{L}[\underline{x}_l, \overline{x}_l]$, we have $\mathbb{X}_b^I = \bigotimes_{l=1}^{L}[\underline{x}_l + b, \overline{x}_l - b]$.[1] Next, we need to develop some notation for Taylor series approximations. Define for a given, $m-1$ times differentiable function $g : \mathbb{R}^L \to \mathbb{R}$, a point $y \in \mathbb{R}^L$ and an integer m, the $m-1$-th order polynomial function $t : \mathbb{R}^L \to \mathbb{R}$ based on the Taylor series expansion of order $m-1$ of $g(\cdot)$ around y:

$$t(x, g(\cdot), y, m) = \sum_{j=0}^{m-1}\sum_{|\lambda|=j}\frac{1}{\lambda!}\frac{\partial^{|\lambda|}}{\partial x^\lambda}g(y) \cdot (x - y)^\lambda. \tag{3.9}$$

If the function $g(x)$ is m times differentiable the remainder term in the Taylor series expansion is

$$g(x) - t(x, g(\cdot), y, m) = \sum_{|\lambda|=m}\frac{1}{\lambda!}\frac{\partial^m}{\partial x^\lambda}g(\overline{y}(x)) \cdot (x - y)^\lambda.$$

with $\overline{y}(x)$ intermediate between x and y. If the m-th order derivative is bounded, this remainder term can be bounded by $C \cdot |x - y|^m$.

3.2 The Nearest Internal Point Estimator

It is well-known that kernel density estimators are biased if the support of the density that is estimated is bounded. Because GPM and GFM estimates require first-stage density estimates, their behavior is affected by this bias. In the literature two types of assumptions are made to avoid this problem.

Newey (1994) shows that if the population probability density function and its derivatives up to order j are zero on the boundary of the support, then

$$\sup_{x \in \mathbb{X}, |\lambda| \le j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{f}_{X,NW}(x) - \frac{\partial^{|\lambda|}}{\partial x^\lambda}f_X(x)\right| = O_p\left(\ln N^{1/2}\left(N \cdot b_N^{L+2j}\right)^{-1/2} + b_N^s\right),$$

[1] The set $[-1,1]^L$ is the set of L vectors with components that are between -1 and 1. The set $\bigotimes_{l=1}^{L}[\underline{x}_l, \overline{x}_l]$ is the set of L vectors with l-th component between $\underline{x}_l$ and $\overline{x}_l$.

If the density is bounded from zero on its support, then the uniform bound holds on a compact subset of the interior of the support. Alternatively one can assume that the support is unbounded. Andrews (1995) gives uniform bounds on the estimation error for this case. Neither the Newey (1994) nor the Andrews (1995) results imply uniform convergence on the full support if the support is compact and the density is bounded away from zero on the support. In fact, in that case it is easy to see that the standard kernel density estimator is not even consistent at the boundary. If X is scalar, the kernel density estimator converges to $1/2$ times the value of the density at the boundary.

Various modifications of the kernel density estimator have been proposed to deal with the boundary bias. Here we discuss two of these, and we show that they do not lead to uniformly consistent estimators of the density and its derivatives. To simplify the discussion we assume that the support of the random variable is $\mathbb{X} = [0, 1]$. The first boundary modification is sometimes referred to as reflection. It consists of adding artificial observations to the data that are mirror images of the existing ones but are outside the support of the distribution. So, for an observation X_i, we add one observation on the other side of the lower boundary zero, $X_i' = 0 - (X_i - 0) = -X_i$ and one observation on the other side of the upper boundary, $X_i'' = 1 + (1 - X_i) = 2 - X_i$. This leads to the estimator

$$\hat{f}_{X,R}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{b}\left[K\left(\frac{x - X_i}{b}\right) + K\left(\frac{x - X_i'}{b}\right) + K\left(\frac{x - X_i''}{b}\right)\right] =$$

$$\frac{1}{N}\sum_{i=1}^{N}\left[K\left(\frac{x - X_i}{b}\right) + K\left(\frac{x + X_i}{b}\right) + K\left(\frac{x - (2 - X_i)}{b}\right)\right).$$

At the boundary it is as if the probability density function is extended by imputing the density at values below the lower boundary point 0 as $f_X(0 - x) = f_X(0 + x)$ for $y > 0$. This removes the bias of the kernel density estimator at the boundary, so that we have uniform convergence over the compact support. However, the kernel density estimator of the derivative of the density now converges to zero at the boundary, and is not consistent for that derivative, so that we cannot have uniform convergence of the derivatives. A second method to remove the boundary bias is a jackknife approach (Gray and Schucany, 1972). It is based on two density estimates that use two different bandwidth sequences b_N and b_N'. It then uses a convex combination of the two estimates to eliminate the first-order bias near the boundary. This cannot restore convergence of the density estimator at the boundary (and hence uniform convergence over a compact support) because at the boundary the probability limit of the estimated density is equal to half times the actual density, irrespective of the bandwidth sequence used.

An estimator that removes the boundary bias should not only ensure uniform convergence of the density estimator and its derivatives, but should also preserve the bias reduction that higher-order kernels deliver. Although this feature of kernel density estimators may not always be important in practice, it is needed to ensure that the bias of the GPM and GFM estimators converges to 0 at an appropriate rate. A final requirement is that the proofs of the asymptotic properties of the GPM and GFM estimators should not be unduly complicated and that no additional smoothing parameters should be involved. Note that the asymptotic trimming arguments that could be used to deal with the boundary bias fail these tests.

The new estimator, the Nearest Internal Point (NIP) estimator, is equal to the NW estimator of the density and its derivatives on the internal region of $\mathbb{X}$ and to a Taylor series expansion of the density and its derivatives on the boundary region. To define the estimator we introduce the transformation $r_b(x)$ indexed by the bandwidth. This transformation maps a point $x \in \mathbb{X}$ to a point in the internal region of $\mathbb{X}$. If $r_b(x) \neq x$, $r_b(x)$ is the point where the Taylor series expansion is taken that replaces the usual kernel density estimator. The usual estimator is used if $r_b(x) = x$.

An obvious choice for $r_b(x)$ is the projection of x onto the set $\mathbb{X}_b^I$:

$$r_b(x) = \text{argmin}_{y \in \mathbb{X}_b^I} |x - y|.$$

Thus, for $x \in \mathbb{X}_b^I$, $r_b(x) = x$ which implies that on the interior region this estimator coincides with the kernel density estimator. If $\mathbb{X} = \bigotimes_{l=1}^2 [\underline{x}_l, \overline{x}_l]$ and $\mathbb{U} = [-1, 1]^2$, the projections are as in Figure 2. Note that for x in the boundary region, moving b in one or both directions will take us out of the support $\mathbb{X}$. For x in the larger rectangles in the boundary set, there is a direction where moving by b keeps us inside the support. The estimator does not use a Taylor series approximation in these "long" directions, but instead uses the kernel density estimator and its derivatives, if derivatives of the density have to be estimated. The NIP estimator of the λ derivative on the boundary region based on this choice of the point of approximation, the NIP estimator is based on a Taylor series expansion of the λ derivative of f_X at $r_b(x)$:

$$\hat{f}^{[\lambda]}_{X,\text{nip},m}(x) = \begin{cases} t\left(x, \frac{\partial^{|\lambda|}}{\partial x^\lambda} \hat{f}_{X,NW}, r_b(x), m\right) & \text{if } x \in \mathbb{X} \\ 0 & \text{elsewhere.} \end{cases} \tag{3.10}$$

Note that in general the nearest interior point estimator of the derivative of the density is not equal to the derivative of the nearest interior point estimator of the density, because the latter depends on the λ derivative of $r_b(x)$ with respect to x. For that reason we use the notation $[\lambda]$ to distinguish this estimator from the λ derivative of the NIP estimator that is indicated by the superscript (λ). Of course the equality $\hat{f}^{[\lambda]}_{X,\text{nip},m}(x) = \hat{f}^{(\lambda)}_{X,\text{nip},m}(x)$ holds on the internal region. Moreover, if $\mathbb{X} = \bigotimes_{l=1}^L [\underline{x}_l, \overline{x}_l]$ and $\mathbb{U} = [-1, 1]^L$ we have for the l-th component of the point of approximation

$$r_b(x)_l = \begin{cases} \underline{x}_l + b & \text{if } x_l < \underline{x}_l + b \\ x_l & \text{if } \underline{x}_l + b \leq x \leq \overline{x}_l - b \\ \overline{x}_l - b & \text{elsewhere.} \end{cases}$$

Note that the derivative of $r_b(x)$ is either 1 or 0, so that if the support is a hyper rectangle $\hat{f}^{[\lambda]}_{X,\text{nip},m}(x) = \hat{f}^{(\lambda)}_{X,\text{nip},m}(x)$ for all $x \in \mathbb{X}$ so that we have uniform convergence in Sobolev norm.

It is essential to bound the distance between x and the point of approximation $r_b(x)$. To obtain the same bound on the bias as Newey (1994), we need the boundary area to be "small", in the sense that $\sup_{x \in \mathbb{X}} |x - r_b(x)|$ is of order $O(b)$. If $\mathbb{X} = \bigotimes_{l=1}^L [\underline{x}_l, \overline{x}_l]$, $\mathbb{X}_b^I = \bigotimes_{l=1}^L [\underline{x}_l + b, \overline{x}_l - b]$, and $r_b(x)$ is the projection of x on the internal set, then the condition is satisfied. An example of a support $\mathbb{X} \subset \mathbb{R}^2$ that has a boundary that is not small is given in Figure 1. This figure shows (part of) a compact support with its internal region $\mathbb{X}_b^I$. Because the support is bounded by the curve $(1-x)^2$, we have that if $r_b(x)$ is the nearest point in the internal set

$$|x_e - r_b(x_e)| = \sqrt{2b+b^2} = O(\sqrt{b}).$$

Hence the shape of the support $\mathbb{X}$ matters.

3.3 Products of Regression and Density Functions

It is useful to look not only at the estimation of densities, but also of a more general class of functions. Let Z be a random vector defined on the same probability space as X, and define

$$h(x) = \mathbb{E}[Z|X=x] \cdot f_X(x).$$

Let

$$h^{(\lambda)}(x) = \frac{\partial^{|\lambda|}}{\partial x^{\lambda}} h(x),$$

be the λ derivative. When we consider partial and full means, we will use a special case where Z is of the form $(1, Y')'$, and so $h(x)$ will be $h(x) = (h_1(x), h_2(x)')' = (f_X(x), f_X(x) \cdot g(x)')'$. Partial and full means depend on h through $g = h_2/h_1$, but for notational ease we formulate and derive the results without this structure. As we shall see later, the expressions for asymptotic variances of estimators simplify, if the dependence is through h_2/h_1.

We consider two estimators for $h(x)$. The first is based on the standard Nadaraya-Watson kernel:

$$\hat{h}_{\mathrm{nw}}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{b^L} \cdot Z_i \cdot K\left(\frac{x - X_i}{b}\right), \tag{3.11}$$

with estimator for the λ derivative equal to the derivative of the estimator:

$$\hat{h}^{(\lambda)}_{\mathrm{nw}}(x) = \frac{\partial^{|\lambda|}}{\partial x^{\lambda}} \hat{h}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{b^{L+|\lambda|}} \cdot Z_i \cdot K^{(\lambda)}\left(\frac{x - X_i}{b}\right). \tag{3.12}$$

The second estimator is based on the NIP density estimator :

$$\hat{h}_{\mathrm{nip},m}(x) = \begin{cases} t\left(x, \hat{h}_{NW}, r_b(x), m\right) & \text{if } x \in \mathbb{X} \\ 0 & \text{elsewhere.} \end{cases} \tag{3.13}$$

As for the estimator of the λ derivative of the density we estimate the λ derivative of h by a Taylor series expansion of the derivative. To emphasize the difference between this estimator and the λ derivative of the NIP estimator of h we use $[\lambda]$ and not (λ).

$$\hat{h}^{[\lambda]}_{\mathrm{nip},m}(x) = \begin{cases} t\left(x, \hat{h}^{(\lambda)}_{NW}, r_b(x), m\right) & \text{if } x \in \mathbb{X} \\ 0 & \text{elsewhere.} \end{cases} \tag{3.14}$$

As noted before we have $\hat{h}^{[\lambda]}_{\mathrm{nip},m}(x) = \hat{h}^{(\lambda)}_{\mathrm{nip},m}(x)$ if $\mathbb{X}$ is a hyper rectangle.

3.4 Assumptions

THE ONLY CHANGE IN MAIN TEXT IS THAT I REPLACED THE ASSUMPTIONS, LEMMAS AND THEOREMS BY THOSE IN THE APPENDIX.

We make three sets of assumptions. The first deals with the joint distribution of (Z_i, X_i), the second with the kernel, and the third with the bandwidth. The assumptions are stronger than necessary for the first set of results, but are formulated in order to be useful in the subsquent discussion of generalized partial means.

Assumption 1 (Distribution)
(i) $(Y_1, X_1), (Y_2, X_2), \ldots,$ *are independent and identically distributed, and one of the components of* Z *is identically equal to 1,*
(ii) the support of X *is* $\mathbb{X} \subset \mathbb{R}^L$, $\mathbb{X} = \bigotimes_{l=1}^{L} [\underline{x}_l, \overline{x}_l]$, $\underline{x}_l < \overline{x}_l$ *for all* $l = 1, \ldots, L$.
(iii) $\sup_{x \in \mathbb{X}} \mathbb{E}[|Y|^p | X = x] < \infty$ *for some* $p > 2$.
(iv) $g(x) = \mathbb{E}[Y|X = x]$ *is* q *times continuously differentiable on the interior of* $\mathbb{X}$ *with the* q*-th derivative bounded,*
(v) $f_X(x)$ *is* q *times continuously differentiable on the interior of* $\mathbb{X}$ *with the* q*-th derivative bounded.*

Definition 3.1 Derivative Order of a Kernel *A Kernel* $K : \mathbb{U} \mapsto \mathbb{R}$ *is of derivative order* d *if for all* u *in the boundary of the set* $\mathbb{U}$ *and all* $|\lambda| \leq d - 1$,

$$\lim_{v \to u} \frac{\partial^\lambda}{\partial u^\lambda} K(v) = 0.$$

Assumption 2 (Kernel)
(i) $K : \mathbb{R}^L \to \mathbb{R}$, *with* $K(u) = \prod_{l=1}^{L} \mathcal{K}(u_l)$,
(ii) $K(u) = 0$ *for* $u \notin \mathbb{U}$, *with* $\mathbb{U} = [-1, 1]^L$,
(iii) K *is* r *times continuously differentiable, with the* r*-th derivative bounded on the interior of* $\mathbb{U}$,
(iv) K *is a kernel of order* s, *so that* $\int_{\mathbb{U}} K(u) du = 1$ *and* $\int_{\mathbb{U}} u^\lambda K(u) du = 0$ *for all* λ *such that* $0 < |\lambda| < s$, *for some* $s \geq 1$,
(v) K *is a kernel of derivative order* d.

We refer to such a kernel as a derivative kernel of order (s, d). Hence a derivative kernel is a higher order kernel that deals with boundary terms that appear in the asymptotically linear expressions for the GFAD and GPAD estimators. If $d = 0$ there is no restriction on the kernel beyond the higher order.

Assumption 3 *The bandwidth* $b_N = N^{-\delta}$ *for some* $\delta > 0$.

We assume that $\mathbb{U} = [-1, 1]^L$ for simplicity. Key is that it has bounded support and that the kernel itself is bounded.

Define $Z_i = (1, Y_i')'$.

3.5 Properties of Standard Kernel Estimators

The first set of results establishes the asymptotic properties of the usual Nadaraya-Watson (NW) and the new Nearest Interior Point (NIP) density estimators. Results on uniform convergence of the NW estimator are given in Newey (1994). The first lemma gives the uniform bias. This result holds for a fixed bandwidth and hence we omit the subscript N on the bandwidth.

Lemma 3.1 (BIAS, NEWEY, 1994)
Suppose Assumptions 1-2 hold, and $q \geq j + s$, then:

$$\sup_{x \in \mathbb{X}_b^I, |\lambda| \leq j} \left| \frac{\partial^{|\lambda|}}{\partial x^\lambda} \mathbb{E}\left[\hat{h}_{\text{nw}}(x)\right] - \frac{\partial^{|\lambda|}}{\partial x^\lambda} h(x) \right| = O\left(b^s\right).$$

i.e. the Sobolev norm (of order j) of the bias of the kernel estimator is $O(b^s)$.

The proofs for the Lemmas and Theorems in the text are given in Appendix C. This Lemma follows directly from Lemma B.2 in Newey (1994), with the one difference that we allow the set $\mathbb{X}_b^I$ to expand with the bandwidth. We give the proof for completeness.

Lemma 3.2 (VARIANCE OF STANDARD KERNEL ESTIMATOR, NEWEY, 1994)
Suppose Assumptions 1-2 hold, $q \geq j$, $r \geq j + 1$, and the bandwidth satisfies $C_1 N^{-\gamma_1} \leq b_N \leq C_2 N^{-\gamma_2}$, for some $0 < \gamma_2 < \gamma_1 < \frac{1-2/p}{2j+L+2}$. Then:

$$\sup_{x \in \mathbb{X}, |\lambda| \leq j} \left| \frac{\partial^{|\lambda|}}{\partial x^\lambda} \hat{h}_{NW}(x) - \frac{\partial^{|\lambda|}}{\partial x^\lambda} \mathbb{E}\left[\hat{h}_{NW}(x)\right] \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^{L+2j}} \right)^{1/2} \right).$$

This follows directly from Lemma B.1 in Newey (1994). We give the proof in the Appendix for completeness. Note that unlike the bias, the variance is not affected by the boundary problem, and the Lemma is valid uniformly over $\mathbb{X}$, not just over $\mathbb{X}_b^I$. We index the bandwidth in this Lemma by the sample size because the variance bound only applies if the bandwidth sequence satisfies the conditions in Assumption.

3.6 Properties of Nearest Internal Point Estimator

THE ONLY CHANGE IN MAIN TEXT IS THAT I REPLACED THE ASSUMPTIONS, LEMMAS AND THEOREMS BY THOSE IN THE APPENDIX.

We use the NIP estimator with the order of the Taylor series expansion equal to $s - 1$ with s the order of the kernel.

The next two lemmas establish the rate of uniform (on $\mathbb{X}$) convergence of the NIP estimator of the λ derivatives of h up to order j. The NIP estimator of order s of $h^{(\lambda)}$ is

$$\hat{h}_{NIP,s}^{(\lambda)}(x) = \sum_{j=0}^{s-1} \sum_{|\mu|=j} \frac{1}{\mu!} \hat{h}_{NW}^{(\lambda+\mu)}(r_b(x))(x - r_b(x))^\mu$$

Lemma 3.3 (Bias)
If Assumptions 1-2 hold, and $q \geq j + 2s - 1$ and $r \geq j + s - 1$, then for all $|\lambda| \leq j$:

$$\sup_{x \in \mathbb{X}} \left| \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right] - h^{(\lambda)}(x) \right| = O\left(b^s\right).$$

Note that the only difference with Lemma 3.1 is that we require $s-1$ additional derivatives of h.

Lemma 3.4 (Variance)
Suppose Assumptions 1-3 hold and $q \geq j + s - 1, r \geq j + s$, and the bandwidth rate δ satisfies $\delta < \frac{1-2/p}{2j+L+2}$. Then:

$$\sup_{x \in \mathbb{X}, |\lambda \leq j} \left| \hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right] \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^{L+2j}} \right)^{1/2} \right).$$

Theorem 3.1 (Uniform Convergence)
If Assumptions 1-3 hold, and $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1$, then:

$$\sup_{x \in \mathbb{X}} \left| \hat{h}^{[\lambda]}_{\text{nip},s}(x) - h^{[\lambda]}(x) \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^{L+2|\lambda|}} \right)^{1/2} + b_N^s \right).$$

If also

$$s > \max\left\{ L + 2|\lambda|, \frac{L + 2|\lambda| + 2}{2 - 4/p} \right\},$$

and

$$\frac{1}{2s} < \delta < \min\left\{ \frac{1 - 2/p}{L + 2|\lambda| + 2}, \frac{1}{2L + 4|\lambda|} \right\},$$

then,
(i)

$$\sup_{x \in \mathbb{X}} \left| \mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(x)\right] - h^{[\lambda]}(x) \right| = o\left(N^{-1/2}\right),$$

and (ii)

$$\sup_{x \in \mathbb{X}} \left| \hat{h}^{[\lambda]}_{\text{nip},s}(x) - h^{[\lambda]}(x) \right| = o_p\left(N^{-1/4}\right).$$

Note that in the case that the support $\mathbb{X}$ is a hyperrectangle Lemmas 3.3, 3.4 and Theorem A.1 imply the same upper bounds for the Sobolev norm, so that the results are directly comparable to those in the previous section.

The results on $\hat{h}_{\text{nip},s}(x)$ will be used for estimation of partial means. In particular, we use the fact that the results imply that under sufficient smoothness conditions the difference between

$\hat{h}_{\text{nip},s}(x)$ and $h(x)$ disappears faster than $N^{-1/4}$, and that the corresponding bias vanishes faster than $N^{-1/2}$, both uniformly over $\mathbb{X}$. These are the well-known minimal rates of convergence of first-stage nonparametric estimators indicated by Newey (1994). The following Lemma makes this precise.

We can always choose the order of the kernel such that the interval for δ is not empty. It is obvious that a higher order kernel is needed. An advantage of the NIP estimator is that it preserves the order of the usual kernel estimator. Note that if $3 \leq p \leq 4$, we have

$$\frac{1}{2s} < \delta < \frac{1-\frac{2}{p}}{L+2}$$

so that the required order of the kernel is less than the number of variables in X.

4 Large Sample Properties of Estimators for Generalized Full and Partial Means and Average Derivatives

4.1 Assumptions

In this section we present the main results of the paper. For four cases, the full mean, the full average derivative, the partial mean and the partial average derivative, we present results on the large sample properties of the estimators. We also present some results for special cases, most importantly for the case where the function $m(h)$ depends only on (derivatives of) the regression function $g = h_2/h_1$. In addition we present estimators for the large sample variances of the four estimators.

Next we make an assumption on the smoothness of $m(\cdot)$ and the weight function $\omega(\cdot)$. Because components of g are continuous functions on the compact set $\mathbb{X}$, we can consider m as a functional on the set $C[\mathbb{X}]$ of continuous (and hence bounded) functions on $\mathbb{X}$. We need to linearize this functional in an open neighborhood of the population value of g. Define this neighborhood by $B(g, \varepsilon) = \{g' \in C[\mathbb{X}] | |g' - g| < \varepsilon\}$ with $\varepsilon > 0$. We also define $\mathbb{G} \subset \mathbb{R}^L$ by $\mathbb{G} = \{g'(x) | x \in \mathbb{X}, g' \in B(g, \varepsilon)\}$. We now give the assumption that ensures that we can linearize m. In this assumption we consider m as a function of M real arguments, i.e. as a function of $g'_1(x), \ldots, g'_M(x)$ with $x \in \mathbb{X}$ and $g' \in B(g, \varepsilon)$. In applications ε is chosen to ensure differentiability of m. An example is the weighting estimator in Example 2.1 where we assume that $g_3(x)$ is bounded from 0 and 1 on $\mathbb{X}$. We choose ε such that all $g' \in B(g, \varepsilon)$ have the same property. If an estimator $\hat{g}$ converges uniformly to g on $\mathbb{X}$ then in large samples $g + \alpha(\hat{g} - g) \in B(g, \varepsilon)$ for all $0 \leq \alpha \leq 1$.

Assumption 4 (Smoothness of m and ω)
(i) The function m is t times continuously differentiable on $\mathbb{H}_\lambda$ with its t-th derivative bounded on this set, and
(ii) the function ω is t times differentiable with bounded t-th derivative on $\mathbb{X}_1 = \bigotimes_{l=1}^{L_1} [\underline{x}_l, \overline{x}_l]$, and $\frac{\partial^\mu \omega}{\partial z^\mu}(z)$ is zero on the boundary of $\mathbb{X}_1$.

Assumption 5 (Smoothness of t)
The function t is v times continuously differentiable on $\mathbb{X}_1$ with its v-th derivative bounded on this set.

It should be noted that this assumption stronger than needed for consistency of the GFM and GPM estimators. For consistency continuity of m on $\mathbb{G}$ suffices.

4.2 Full Means

Here we focus on estimation of the GFM

$$\theta_{\text{gfm}} = \mathbb{E}\left[\omega(X)'m\left(h\left(X\right)\right)\right].$$

The proposed estimator is

$$\hat{\theta}_{\text{gfm}} = \frac{1}{N}\sum_{i=1}^{N}\omega(X_i)'m\left(\hat{h}_{\text{nip},s}\left(X_i\right)\right).$$

Define

$$\psi_{\text{gfm}}(y,x) = \psi_{\text{gfm1}}(y,x) + \psi_{\text{gfm2}}(y,x),$$

where

$$\psi_{\text{gfm1}}(y,x) = \left(\omega(x)'m(h(x)) - \theta_{\text{gfm}}\right) \tag{4.15}$$

and

$$\psi_{\text{gfm2}}(y,x) = f_X(x)\omega(x)'\frac{\partial}{\partial h'}m(h(x))\cdot\begin{pmatrix}1\\y\end{pmatrix} - \mathbb{E}\left[f_X(X)\omega(X)'\frac{\partial}{\partial h'}m(h(X))\cdot\begin{pmatrix}1\\Y\end{pmatrix}\right] \tag{4.16}$$

We will show that $\psi_{\text{gfm}}(y,x)$ is the influence function for the estimator $\hat{\theta}_{\text{gfm}}$. The first term in the influence function, $\psi_{\text{gfm1}}(y,x)$, is the contribution from the averaging of $g(x)$ over the sample, and the second, $\psi_{\text{gfm2}}(y,x)$, captures the uncertainty coming from the estimation of $g(x)$.

Theorem 4.1 (Generalized Full Mean)
Suppose Assumptions 1–4 hold. Then

(i) (Consistency) If $q \geq 2s-1$, $r \geq s-1+L$, $p \geq 3$, $0 < \delta < 1/L$, then

$$\hat{\theta}_{\text{gfm}} \xrightarrow{p} \theta_{\text{gfm}},$$

(ii) (Asymptotic Linearity) If, in addition to the conditions in (i), $t \geq s$, $d \geq s-1$, and

$$\frac{1}{2s} < \delta < \min\left\{\frac{2-\frac{4}{p}}{2L+4}, \frac{1}{2L}\right\},$$

then

$$\sqrt{N}\left(\hat{\theta}_{\text{gfm}} - \theta_{\text{gfm}}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_{\text{gfm}}(Y_i, X_i) + o_p(1),$$

(iii) (Asymptotic Normality) Under the same conditions as in (ii),

$$\sqrt{N}\left(\hat{\theta}_{\text{gfm}} - \theta_{\text{gfm}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_{\text{gfm}}(Y,X)\cdot\psi_{\text{gfm}}(Y,X)'\right]\right).$$

Next we present a result for the special case where $m(\cdot)$ depends on $h(\cdot)$ only through the regression function $h_2(x)/h_1(x)$. Then we can define a function $n : \mathbb{G} \mapsto \mathbb{R}^P$ such that $n(g(x)) = m(f_X(x), g(x)\cdot f_X(x))$ for all $g(x)$. Define

$$\psi_{\text{gfm,reg}}(y,x) = \omega(x)'\frac{\partial}{\partial g'}n(g(x))\cdot(y-g(x)) + \left(\omega(x)'n(g(x)) - \theta\right). \tag{4.17}$$

Corollary 4.1 (Generalized Full Mean of Regression Function)
Suppose Assumptions 1–4 hold with $q \geq 2s-1$, $r \geq s-1+L$, $p \geq 3$, $0 < \delta < 1/L$, $t \geq s$, $d \geq s-1$, and

$$\frac{1}{2s} < \delta < \min\left\{\frac{2-\frac{4}{p}}{2L+4}, \frac{1}{2L}\right\},$$

then (i)

$$\sqrt{N}\left(\hat{\theta}_{\text{gfm}} - \theta_{\text{gfm}}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_{\text{gfm,reg}}(Y_i, X_i) + o_p(1),$$

and (ii)

$$\sqrt{N}\left(\hat{\theta}_{\text{gfm}} - \theta_{\text{gfm}}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\psi_{\text{gfm,reg}}(Y,X)\psi_{\text{gfm,reg}}(Y,X)'\right]\right).$$

Lemma 4.1 Consistent Estimator for Asymptotic Variance for Full Mean

4.3 Full Average Derivatives

Theorem 4.2 (Generalized Full Average Derivative)
Consider the estimator

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\omega(X_i)m(\hat{h}^{[\lambda]}_{NIP,s}(X_i))$$

of the GFM/GFAD

$$\theta = \mathbb{E}\left[\omega(X)m\left(h^{[\lambda]}(X)\right)\right]$$

with $h^{[\lambda]}_{NIP,s}$ the NIP estimator of $h^{[\lambda]}$, then

(i) (Consistency) If Assumptions 1 and 4 hold[2], $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1 + L$, , $p \geq 3$, and $b_N = N^{-\delta}$ with

$$0 < \delta < \frac{1}{L + 2|\lambda|}$$

then

$$\hat{\theta} \xrightarrow{p} \theta.$$

[2] The function n need not be differentiable. Continuity on $\mathbb{H}_\lambda$ suffices.

(ii) (Asymptotic Linearity) If Assumptions 1-4 hold, $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1 + L$,$t \geq |\lambda| + s$, $p \geq 3$, $d \geq \max\{\lambda_1, \dots, \lambda_L\} + s - 1$, *all* $\mu \leq \lambda$, $|\mu| \leq |\lambda| - 1$, *and*

$$\frac{1}{2s} < \delta < \min\left\{\frac{2-\frac{4}{p}}{2L+4\max\{1,|\lambda|\}}, \frac{1}{2L+4|\lambda|}\right\}$$

then the estimator is asymptotically linear with

$$\sqrt{N}(\hat{\theta}-\theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega(X_i)m(h_0^{[\lambda]}(X_i)) - \mathbb{E}\left[\omega(X)m\left(h^{[\lambda]}(X)\right)\right]\right)$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\left(\alpha_{\kappa m}^{(\kappa)}(X_i)Z_{im} - \mathbb{E}[\alpha_{\kappa m}^{(\kappa)}(X)Z_m]\right)\right) + o_p(1)$$

with $Z_{i1} = 1, Z_{i2} = Y_i$ *and for* $m = 1, 2$

$$\alpha_{\kappa m}^{(\kappa)}(x) = f_X(x)\omega(x)\frac{\partial n}{\partial h_m^{(\kappa)}(x)}(h_0^{[\lambda]}(x)).$$

(iii) (Asymptotic Normality) Under the same assumptions as in (ii),

$$\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

with

$$V = \mathrm{Var}\left(\omega(X)n\left(h_0^{[\lambda]}(X)\right)\right) + \sum_{\kappa_1\leq\lambda}\sum_{\kappa_2\leq\lambda}(-1)^{|\kappa_1|+|\kappa_2|}\mathbb{E}\left[\alpha_{\kappa_1,2}^{(\kappa_1)}(X)\alpha_{\kappa_2,2}^{(\kappa_2)}(X)\,Var(Y|X)\right]$$

$$+\sum_{\kappa_1\leq\lambda}\sum_{\kappa_2\leq\lambda}\sum_{m_1=1}^{2}\sum_{m_2=1}^{2}(-1)^{|\kappa_1|+|\kappa_2|}Cov\left(\alpha_{\kappa_1,m_1}^{(\kappa_1)}(X)\mathbb{E}(Z_{m_1}|X)\ ,\ \alpha_{\kappa_2,m_2}^{(\kappa_2)}(X)\mathbb{E}(Z_{m_2}|X)\right)$$

$$+2\sum_{\kappa\leq\lambda}\sum_{m=1}^{2}(-1)^{|\kappa|}\mathbb{E}\left[\left(\omega(X)n\left(h_0^{[\lambda]}(X)\right) - \mathbb{E}\left[\omega(X)n\left(h_0^{[\lambda]}(X)\right)\right]\right)\left(\alpha_{\kappa m}^{(\kappa)}(X)\mathbb{E}(Z_m|X) - \mathbb{E}\left[\alpha_{\kappa m}^{(\kappa)}(X)\mathbb{E}(Z_m|X)\right]\right)\right].$$

Lemma 4.2 CONSISTENT ESTIMATOR FOR ASYMPTOTIC VARIANCE FOR FULL DERIVATIVES

4.4 Partial Mean

Recall the definitions of the generalized partial and full means:

$$\theta^{\text{gpm}} = \mathbb{E}\left[\omega(X_1)'m\left(h\left(X_1, t(X_1)\right)\right)\right],$$

and The corresponding estimators are

$$\hat{\theta}^{\text{gpm}} = \frac{1}{N}\sum_{i=1}^{N}\omega(X_{1i})'m\left(\hat{h}_{\text{nip},s}\left(X_{1i}, t(X_{1i})\right)\right),$$

Theorem 4.3 (Generalized Partial Mean)
(i) Suppose Assumptions 1-5 hold with $q \geq 2s-1$, $r \geq s-1+L$, $p \geq 3$, *and* $0 < \delta < 1/L$, *then*

$$\hat{\theta}^{\text{gpm}} \xrightarrow{p} \theta^{\text{gpm}}.$$

(ii) Suppose Assumptions 1-5 hold with $v \geq 2$, $\lambda = 0$, $q \geq 2s-1$, $r \geq s-1+L$, $t \geq s$, $p \geq 4$, $d \geq s-1$, *and*

$$\frac{1}{2s} < \delta < \min\left\{\frac{2-\frac{4}{p}}{2L+4}, \frac{1}{2L}\right\}$$

then $\hat{\theta}^{\text{gpm}}$ *is asymptotically linear with*

$$\sqrt{N} b_N^{L_2/2}\left(\hat{\theta}^{\text{gpm}} - \theta^{\text{gpm}}\right) =$$

$$\frac{1}{{b_N}^{L_2/2}\sqrt{N}} \cdot \sum_{i=1}^{N}\sum_{m=1}^{2}\left(\alpha_m(X_{i1})Z_{im}\int_{\mathbb{U}_1} K\left(u_1, \frac{X_{i2}-t(X_{i1})}{b_N} + \frac{\partial}{\partial x_1'}t(X_{i1})\cdot u_1\right)du_1\right.$$
$$\left. -\mathbb{E}\left[\alpha_m(X_1)Z_m\int_{\mathbb{U}_1} K\left(u_1, \frac{X_2-t(X_1)}{b_N} + \frac{\partial}{\partial x_1'}t(X_1)\cdot u_1\right)du_1\right]\right) + o_p(1)$$

with $Z_{i1} = 1, Z_{i2} = Y_i$, *and for* $m = 1, 2$,

$$\alpha_{\kappa m}(x_1) = f_X(x_1, t(x_1))\omega(x_1)\frac{\partial n}{\partial h_m^{(\kappa)}(x_1, t(x_1))}(h(x_1, t(x_1)))$$

(iii), under the same asssumptions as in (ii),

$$\sqrt{N} b_N^{L_2/2}(\hat{\theta}^{\text{gpm}} - \theta^{\text{gpm}}) \xrightarrow{d} \mathcal{N}\left(0, \sum_{k=1}^{2}\sum_{m=1}^{2} V_{k,m}\right)$$

with

$$V_{k,m} = \int_{\mathbb{X}_1}\mu_{km}(x_1, t(x_1))\alpha_k(x_1)\alpha_m(x_1)\int_{\mathbb{U}_2}\left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)du_1\right)^2 du_2 f_X(x_1, t(x_1))dx_1$$

with $\mu_{km}(x) = \mathbb{E}[Z_k Z_m | X = x]$ *for* $k, m = 1, 2$.

Lemma 4.3 Consistent Estimator for Asymptotic Variance for Partial Mean

4.5 Partial Average Derivative

Theorem 4.4 (Generalized Partial Mean and Average Derivative)

Consider the estimator

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}\omega(X_{i1})n(\hat{h}_{NIP,s}^{[\lambda]}(X_{i1}, t(X_{i1})))$$

of the GFM/GFAD

$$\theta_0 = \mathbb{E}\left[\omega(X) n \left(h_0^{[\lambda]}\left(X_1, t(X_1)\right)\right)\right]$$

with $h_{NIP,s}^{[\lambda]}$ *the NIP estimator of* $h^{(\lambda)}$ *for an boundary set of width* $2b_N$ *that uses the trimmed NW estimator, then*

(i) *Consistency. If Assumptions 1 and 4 hold*[3], $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1 + L$, , $p \geq 3$, *and* $b_N = N^{-\delta}$ *with*

$$0 < \delta < \frac{1}{L + 2|\lambda|}$$

then

$$\hat{\theta} \xrightarrow{p} \theta_0$$

(ii) *Asymptotic distribution. If Assumptions 1, 2 and 4 hold, then*

a. *if* $t(x_1)$ *is twice continuously differentiable on* $\mathbb{X}_1$, $\lambda = 0$, $q \geq 2s - 1$, $r \geq s - 1 + L$, $t \geq s$, $p \geq 4$, $d \geq s - 1$, *and* $b_N = N^{-\delta}$ *with*

$$\frac{1}{2s} < \delta < \min\left\{\frac{2 - \frac{4}{p}}{2L + 4}, \frac{1}{2L}\right\}$$

then the estimator is asymptotically linear with

$$\sqrt{N} b_N^{L_2/2}(\hat{\theta} - \theta_0) = \frac{1}{{b_N}^{L_2/2}\sqrt{N}} \cdot \sum_{i=1}^{N} \sum_{m=1}^{2} \left(\alpha_m(X_{i1}) Z_{im} \int_{\mathbb{U}_1} K\left(u_1, \frac{X_{i2} - t(X_{i1})}{b_N} + \frac{\partial}{\partial x_1'} t(X_{i1}) \cdot u_1\right) du_1 \right.$$
$$\left. - \mathbb{E}\left[\alpha_m(X_1) Z_m \int_{\mathbb{U}_1} K\left(u_1, \frac{X_2 - t(X_1)}{b_N} + \frac{\partial}{\partial x_1'} t(X_1) \cdot u_1\right) du_1\right]\right) + o_p(1)$$

with $Z_{i1} = 1, Z_{i2} = Y_i$ *and for* $m = 1, 2$

$$\alpha_{\kappa m}(x) = f_X(x)\omega(x) \frac{\partial n}{\partial h_m^{(\kappa)}(x)} (h_0^{[\lambda]}(x))$$

Hence

$$\sqrt{N} b_N^{L_2/2}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

with

$$V = \sum_{m=1}^{2} \sum_{m'=1}^{2} \int_{\mathbb{X}_1} \mu_{mm'}(x_1, t(x_1)) \alpha_m(x_1) \alpha_{m'}(x_1) \int_{\mathbb{U}_2} \left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(x_1) u_1 + u_2\right) du_1\right)^2 du_2 f_X(x_1, t(x_1)) dx_1$$

with $\mu_{mm'}(x) = \mathbb{E}[Z_m Z_{m'} | X = x]$ *for* $m, m' = 1, 2$

[3]The function n need not be differentiable. Continuity on $\mathbb{H}_\lambda$ suffices.

b. If $\lambda \geq 0$ and $t(x_1) = x_{02}$, and $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1 + L$,$t \geq |\lambda| + s$, $p \geq 4$, $d \geq \max\{\lambda_1, \ldots, \lambda_L\} + s - 1$, and $b_N = N^{-\delta}$ with[4]

$$\frac{1}{2s} < \delta < \min\left\{\frac{2 - \frac{4}{p}}{2L + 4\max\{1, |\lambda|\}}, \frac{1}{2L + 4|\lambda|}\right\}$$

then the estimator is asymptotically linear with

$$\sqrt{N} b_N{}^{L_2/2+|\lambda_2|}(\hat{\theta} - \theta_0) =$$

$$\frac{1}{b_N{}^{L_2/2}\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \sum_{m=1}^{2} \left(\alpha^{(\kappa_1)}_{\kappa_1\lambda_1,m}(X_{i1}) Z_{im} K^{(\lambda_2)}\left(\frac{x_{02} - X_{i2}}{b_N}\right) - \mathbb{E}\left[\alpha^{(\kappa_1)}_{\kappa_1\lambda_1,m}(X_1) Z_m K^{(\lambda_2)}\left(\frac{x_{02} - X_2}{b_N}\right.\right.\right.\right.$$

with $Z_{i1} = 1, Z_{i2} = Y_i$ and for $m = 1, 2$

$$\alpha_{\kappa m}(x) = f_X(x)\omega(x)\frac{\partial n}{\partial h^{(\kappa)}_m(x)}(h^{[\lambda]}_0(x))$$

Hence

$$\sqrt{N} b_N{}^{L_2/2+|\lambda_2|}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$

with

$$V = \sum_{m=1}^{2}\sum_{m'=1}^{2} \int_{\mathbb{X}_1} \mu_{mm'}(x_1, t(x_1)) \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1\lambda_2,m}(x_1) \right) \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1\lambda_2,m'}(x_1) \right)$$

$$\cdot \int_{\mathbb{U}_2} \left(K^{(\lambda_2)}(u_2) \right)^2 du_2 f_X(x_1, t(x_1)) dx_1$$

Lemma 4.4 Consistent Estimator for Asymptotic Variance for Partial Derivative

5 Applications

IN THIS SECTION WE SPECIALIZE THE GENERAL RESULT TO INTERESTING CASES: FULL AND PARTIAL MEAN, FULL AVERAGE DERIVATIVES (WEIGHTED AND UNWEIGHTED), PARTIAL AVERAGE DERIVATIVES (UNWEIGHTED?), FULL AND PARTIAL CROSS DERIVATIVES, FULL AND PARTIAL SECOND DERIVATIVES. NOTHING TO PROVE, BUT JUST THE INFLUENCE FUNCTIONS AND ASYMPTOTIC VARIANCE.

6 Conclusion

To be added

[4]These are the bounds on δ if $|\lambda| < 2L$ which is usually true.

References

Ai, C., (1997). "A Semiparametric Maximum Likelihood Estimator," *Econometrica*, Vol 65(4), 933-963.

Altonji, J., and R. Matzkin, (2003). "Cross-section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," ,*Econometrica*, Vol 73(4), 1053-1102.

Andrews, D., (1995). "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, Vol 11, 560-596.

Cheng, M., J. Fan, and J. Marron, (1997). "On Automatic Boundary Corrections," *Annals of Statistics*,Vol 25, 1691-1708.

Graham, B., G. Imbens, and G. Ridder,(2006a). "Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis," *Mimeo.*

Guerre, E., I. Perrigne, and Q. Vuong (2000). "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, Vol 68(3), 525-574.

Härdle, W., and T. Stoker, (1989). "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, Vol 84(408) 986-995.

Jones, M., (1993). "Simple boundary correction for kernel density estimation," *Statistics and Computing*, Vol 3(3), 135-146.

Jones, M., and D. Signorini, (1997). "A Comparison of Higher-order Bias Kernel Density Estimators," *Journal of the American Statistical Association*, Vol 92, No 439 1063-1973.

Li, Q., and Racine, J., (2006). *Nonparametric Econometrics*, Princeton University Press, Princeton, New Jersey.

Marron, J., and D. Ruppert, (1994). "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society*, Series B, Vol 56(4), 653-671.

Newey, W., (1994). "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, Vol 10, 233-253.

Newey, W., and D. McFadden, (1994). "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Engle and McFadden (eds.), Vol 4, Chapter 36, 2113-2245.

Pagan, A., and A. Ullah, (1999). *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Powell, J., J. Stock, and T. Stoker, (1989). "Semiparametric Estimation of Index Coefficients," *Econometrica*, Vol 57(6), 1403-1430.

Powell, J., J. Stock, and T. Stoker, (1989). "Semiparametric Estimation of Index Coefficients," *Econometrica*, Vol 57(6), 1403-1430.

Robinson, P., (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 67: 645-662.

Serfling, R., (1980). *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York.

Silverman, B., (1986). *Density Estimation for Statistics and Data Analysis*, London, Chapman and Hall.

Appendix A: Additional Lemmas and Theorems

Lemma A.1 *Under the assumptions of Lemma 3.1 and in addition $r \geq j$, we have that for $|\lambda| \leq j$, and $x \in \mathbb{X}$,*

$$\mathbb{E}\left[\hat{h}_{\text{nw}}^{(\lambda)}(x)\right] = \frac{\partial^{|\lambda|}}{\partial x^{\lambda}} \mathbb{E}\left[\hat{h}_{\text{nw}}(x)\right].$$

Lemma A.2 *Under the assumptions of Lemma 3.2, we have that for $|\lambda| \leq j$*

$$\sup_{x \in \mathbb{X}} \left| \hat{h}_{NW}^{(\lambda)}(x)` - \mathbb{E}\left[\hat{h}_{NW}^{(\lambda)}(x)\right]\right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^{L+2j}}\right)^{1/2}\right).$$

Here we introduce some additional notation to keep track of derivatives. We do this part only for scalar Y. For a L-vector of nonnegative integers λ, define $\hat{h}_{\text{nip},s}^{[\lambda]}$ as the vector of NIP estimators of the derivatives of h up to order λ. Note that $h = (h_1 \; h_2)'$ with

$$h_1 = f_X \ , \quad h_2 = f_X \cdot g,$$

so that $h_1 : \mathbb{X} \mapsto \mathbb{R}$ and $h_2 : \mathbb{X} \mapsto \mathbb{R}$. Hence $h^{[\lambda]}$ is a 2Λ vector of derivatives of h_1 and h_2 with $\Lambda = \prod_{l=1}^{L}(1 + \lambda_l)$ and the first Λ components the derivatives of h_1, i.e. $h_1^{[\lambda]}$ and the second Λ components the derivatives of h_2, i.e. $h_2^{[\lambda]}$. For example, if $\lambda = (1, 2)$, then $\Lambda = (1 + 1) \cdot (1 + 2) = 6$, and

$$h_1^{[\lambda]}(x) = \begin{pmatrix} h_1(x) \\ \frac{\partial}{\partial x_1} h_1(x) \\ \frac{\partial}{\partial x_2} h_1(x) \\ \frac{\partial^2}{\partial x_1 \partial x_2} h_1(x) \\ \frac{\partial^2}{\partial x_2^2} h_1(x) \\ \frac{\partial^3}{\partial x_1 \partial x_2^2} h_1(x) \end{pmatrix}.$$

Also define the $\Lambda \times L$ matrix $[\lambda]$ with rows equal to the order of differentation in $[\lambda]$, so that for $\lambda = (1, 2)'$,

$$[\lambda] = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 2 \\ 1 & 2 \end{pmatrix},$$

and $|[\lambda]|$ to be the Λ vector with the row sums, so that in the same example with $\lambda = (1, 2)'$,

$$|[\lambda]| = [\lambda]\iota_L = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \\ 2 \\ 3 \end{pmatrix},$$

where ι_L is the L-vector with all elements equal to one. Note that for a scalar μ, $h^{[\mu]}$ contains all derivatives of h_1, h_2 up to order μ. Thus for $\lambda = (1, 2)$, the vector of functions $h^{[|\lambda|]}$ contains all derivatives up to order 3:

$$h_1^{[|\lambda|]}(x) = \begin{pmatrix} h_1(x) \\ \frac{\partial}{\partial x_1} h_1(x) \\ \frac{\partial}{\partial x_2} h_1(x) \\ \frac{\partial^2}{\partial x_1^2} h_1(x) \\ \frac{\partial^2}{\partial x_1 \partial x_2} h_1(x) \\ \frac{\partial^2}{\partial x_2^2} h_1(x) \\ \frac{\partial^3}{\partial x_1^3} h_1(x) \\ \frac{\partial^3}{\partial x_1^2 \partial x_2} h_1(x) \\ \frac{\partial^3}{\partial x_1 \partial x_2^2} h_1(x) \\ \frac{\partial^3}{\partial x_2^3} h_1(x) \end{pmatrix}.$$

Finally $*$ is the component by component multiplication of two vectors with the same number of components, and $|x|_*$ is the vector of absolute values of the components of the vector x.

Theorem A.1 (Uniform Convergence)
If Assumptions 1-3 hold, and $q \geq j + 2s - 1$, $r \geq j + s - 1 + L$, then for all $|\lambda| \leq j$:

$$\sup_{x \in \mathbb{X}} \left| \hat{h}^{(\lambda)}_{\mathrm{nip},s}(x) - h^{(\lambda)}(x) \right| = O_p\left(\left(\frac{\log N}{N \cdot b_N^{L+2j}} \right)^{1/2} + b_N^s \right).$$

Note that this implies that if $\delta < 1/(L + 2|\lambda|)$, we have convergence of $\hat{h}^{(\lambda)}_{\mathrm{nip},s}(x)$ to $h^{(\lambda)}(x)$ uniform in x.

Lemma A.3 (Estimation of $h(x)$)
Suppose that Assumptions 1-3 hold, with

$$s > \max\left\{ L + 2j, \frac{L + 2j + 2}{2 - 4/p} \right\},$$
$$q \geq j + 2s - 1,$$
$$r \geq j + s - 1,$$

and

$$\frac{1}{2s} < \delta < \min\left\{ \frac{1 - 2/p}{L + 2j + 2}, \frac{1}{2L + 4j} \right\}.$$

Then, for $|\lambda| \leq j$,
(i)

$$\sup_{x \in \mathbb{X}} \left| \mathbb{E}\left[\hat{h}^{(\lambda)}_{\mathrm{nip},s}(x) \right] - h^{(\lambda)}(x) \right| = o\left(N^{-1/2} \right),$$

and (ii)

$$\sup_{x \in \mathbb{X}} \left| \hat{h}^{(\lambda)}_{\mathrm{nip},s}(x) - h^{(\lambda)}(x) \right| = o_p\left(N^{-1/4} \right).$$

The results in Lemmas 3.3, 3.4, A.3 and Theorem A.1 with $j = |\lambda|$ can now be read as results on a uniform bound on the bias and variance of $\hat{h}^{[\lambda]}_{\mathrm{nip},s}$ as an estimator of the vector of derivatives $h^{[\lambda]}$.
The next set of Lemmas A.4-A.11 do the groundwork for the proofs of Theorems 4.1-**??**. They provide the tools for the analysis of the correction term that accounts for the effect of the estimation of g and/or its derivatives on the asymptotic distribution of the Generalized Full Mean (GFM), the Generalized Full Average Derivative (GFAD), the Generalized Partial Mean (GPM), and the Generalized Partial Average Derivative (GPAD). The lemmas show that the properties of the GPM and GFM estimators follow from the properties of the NIP estimator. Those lemmas that apply to both the full and partial mean cases are stated for partial means with the understanding that they apply to full means as well.
The estimand of the GFM and GFAD is

$$\theta = \mathbb{E}\left[\omega(X)' m\left(h^{[\lambda]}(X) \right) \right],$$

and that of the GPM and GPAD is

$$\theta = \mathbb{E}\left[\omega(X_1)' m\left(h^{[\lambda]}(X_1, t(X_1)) \right) \right].$$

Here $h^{[\lambda]}$ denotes the vector of derivatives of h up to λ.
We then consider a number of special cases that are obtained by particular choices of λ, e.g. $\lambda = 0$ or a unit vector which corresponds to an average partial derivative.
Because components of $h^{[\lambda]}$ are continuous functions on the compact set $\mathbb{X}$, we can consider m as a functional on the set $C[\mathbb{X}]$ of continuous (and hence bounded) functions on $\mathbb{X}$. We need to linearize this functional in an open neighborhood of the population value of $h^{[\lambda]}$. Denoting the sup norm on $\mathbb{X}$ by

$$\left| \tilde{h}^{[\lambda]} - h^{[\lambda]} \right|_0 = \sup_{x \in \mathbb{X}} \left| \tilde{h}^{[\lambda]}(x) - h^{[\lambda]}(x) \right|,$$

we define this neighborhood by

$$B(h^{[\lambda]}, \varepsilon) = \left\{ \tilde{h}^{[\lambda]} \in C[\mathbb{X}] \left| \left| \tilde{h}^{[\lambda]} - h^{[\lambda]} \right|_0 < \varepsilon \right. \right\},$$

with $\varepsilon > 0$. We also define $\mathbb{H}_\lambda \subset \mathbb{R}^{2\Lambda}$ by

$$\mathbb{H}_\lambda = \left\{ \tilde{h}^{[\lambda]}(x) \left| x \in \mathbb{X}, \tilde{h}^{[\lambda]} \in B(h^{[\lambda]}, \varepsilon) \right. \right\}.$$

The derivatives of the function n with respect to $h^{[\lambda]}$ are the partial derivative of $n(h^{[\lambda]}(x))$ with respect to the argument $h^{[\lambda]}(x)$.
The proofs of the lemmas are for the case that m and ω are scalar. The proofs for the case that these are vectors are essentially the same, but with more complex notation.
In the sequel we assume that the support $\mathbb{X} = \prod_{l=1}^{L} [\underline{x}_l, \overline{x}_l]$. For the definition of the NIP we take the interior region as $\mathbb{X}^I_{2b_N}$, i.e. the series expansion is from points that are $2b_N$ from the boundary. When estimating $h^{[\lambda]}(x)$ for $x \in \mathbb{X}^I_{2b_N}$ we trim observations that are in $\mathbb{X} \diagdown \mathbb{X}^I_{b_N}$, i.e. we do not use observations that are less than b_N from the boundary. It should be emphasized that we do not exclude these observations when we average nonparametric estimators over all or some of their arguments. This asymmetric use of observations allows us to deal with the two types of boundary bias and the boundary variance problems that affect the behavior of the estimators. It is possible to generalize the proofs to the case that the support is not a rectangle as long as the boundary region shrinks at the right rate.
The lemmas that lead to the main theorem are organized as follows. Lemma A.4 gives conditions for the linearization of the estimators with respect to the nonparametric estimator of $h^{[\lambda]}$. The linearization remainder is small if the conditions of Theorem 3.1 are met. Lemma A.5 considers the bias of the linearized estimator. As noted, we have to deal with two types of boundary bias: the usual boundary bias of kernel estimators and the boundary bias due to averaging of the nonparametric estimator over a bounded support. The remaining lemmas deal with the variance of the estimator, i.e. the part due to sampling variation in the nonparametric estimator. This done in a number of steps. In the Lemmas A.9 and A.10 the boundary variance problem appears for the full and partial mean case, respectively.

Lemma A.4 (Linearization)
Suppose that Assumptions 1-4 hold, and $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1$, and $\delta < 1/(L + 2|\lambda|)$. Then[5]

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \omega(X_{i1}) \left(m(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1}, t(X_{i1})) - m(h^{[\lambda]}(X_{i1}, t(X_{i1})) \right) \right.$$
$$\left. -\omega(X_{i1}) \frac{\partial}{\partial h^{[\lambda]\prime}} m(h^{[\lambda]}(X_{i1}, t(X_{i1})) \left(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1}, t(X_{i1})) - h^{[\lambda]}(X_{i1}, t(X_{i1})) \right) \right\}$$
$$= O_p\left(\sqrt{N} \left| \hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]} \right|_0^2 \right).$$

Define for $m = 1, 2$ the functions $\nu_m : \mathbb{R}^{L_1} \mapsto \mathbb{R}^{\Lambda}$ by

$$\nu_m(x_1) = \omega(x_1) \frac{\partial}{\partial h^{[\lambda]}_m} m(h^{[\lambda]}(x_1, t(x_1)),$$

and let $\nu = (\nu_1' \ \nu_2')'$ be a 2Λ vector of functions. By Assumption 4 (ii) ν is bounded on $\mathbb{X}_1$. In the sequel we consider derivatives of ν_m. If μ is an L vector of nonnegative integers, then the μ derivative of ν_m exists if ω is $|\mu|$ times differentiable, m is $|\mu|$ times differentiable, and h is $|\lambda + \mu|$ times differentiable. We also use the notation $v(x_1) = (x_1' \ t(x_1)')'$.
The next lemma relates the bias of the GPM, GFM, GAFD, and GAPD estimators to that of the NIP estimator. Remember that the interior region is taken as $\mathbb{X}^I_{2b_N}$ and the set $\mathbb{X}^I_{b_N}$ is considered as the support. The NIP estimator is for all $x \in \mathbb{X}$

$$\hat{h}^{(\lambda)}_{NIP,s}(x) = \sum_{j=0}^{s-1} \sum_{|\mu|=j} \frac{1}{\mu!} \hat{h}^{(\lambda+\mu)}_{NW}(r_{2b}(x))(x - r_{2b}(x))^{\mu}$$

[5] $O_p\left(\left| \hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}_0 \right|_0 \right)$ indicates an expression that is of the same stochastic order as $\left| \hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]} \right|_0$.

with for $x \in \mathbb{X}^I_{2b_N}$ the trimmed NW estimator of $h^{(\lambda)}$

$$\hat{h}^{(\lambda)}_{NW}(x) = \frac{1}{Nb_N^{L+|\lambda|}} \sum_{j=1}^{N} I\left(X_j \in \mathbb{X}^I_{b_N}\right) Z_j K^{(\lambda)}\left(\frac{x - X_j}{b_N}\right)$$

The trimmed NW estimator has bias $O\left(b_N^s\right)$ on $\mathbb{X}^I_{2b_N}$ so that Lemma 3.3 applies $\mathbb{X}^I_{2b_N}$ as interior region and $\mathbb{X}^I_{b_N}$ as support. Note that the uniform bias on $\mathbb{X}$ of the NIP estimator is $O\left((2b_N)^s\right)$, because of the additional extrapolation. In the sequel we will not use special notation for the trimmed NW estimator.

Lemma A.5 *If Assumptions 1, 2, and 4 hold, and $q \geq |\lambda| + 2s - 1$, $r \geq |\lambda| + s - 1$, and $t \geq 1$, then*

$$\frac{1}{N} \sum_{i=1}^{N} \nu(X_{i1})' \left(\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))\right] - h^{[\lambda]}(v(X_{i1}))\right) = O(b^s).$$

The next lemma gives an alternative linear representation of the weighted partial mean of the vector of derivatives of the Nadaraya-Watson kernel estimator $\hat{h}_{\text{nw}}$ up to λ using U-statistic theory. Note that this result is slightly different for the partial and full mean cases. This lemma is used in the proof of a corresponding result for the weighted partial mean of the NIP estimator of the vector of derivatives up to λ. The NW estimator is the trimmed NW estimator.

Lemma A.6 *Suppose Assumptions 2-4 hold, and $r \geq |\lambda|$. Then:*

$$\begin{aligned} &\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nu(X_{i1})' \left(\hat{h}^{[\lambda]}_{\text{nw}}(v(X_{i1})) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nw}}(v(X_{i1}))\right]\right) \\ &\qquad -\sqrt{N}\left(\int_{\mathbb{X}_1} \nu(x_1)' \left(\hat{h}^{[\lambda]}_{\text{nw}}(v(x_1)) - \mathbb{E}[\hat{h}^{[\lambda]}_{\text{nw}}(v(x_1))]\right) f_{X_1}(x_1) dx_1\right) \\ &\qquad\qquad = O_p\left(N^{-1} b_N^{-L/2 - L_1/2 - |\lambda|} + N^{-1/2} b_N^{-L/2 - |\lambda|}\right). \end{aligned}$$

The next Lemma shows that the same result applies to the nip estimator, using a slightly stronger condition on the degree of differentiability of the kernel. Remember that we use the trimmed NW estimator, but that we average over the full support. Also the NIP estimator extrapolates from $r_{2b_N}(x)$.

Lemma A.7 *Suppose that Assumptions 2-4 hold and that $r \geq |\lambda| + s - 1$. Then:*

$$\begin{aligned} &\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nu(X_{i1})' \left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1})) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))\right]\right) \\ &\qquad\qquad -\sqrt{N}\left(\int_{\mathbb{X}_1} \nu(x_1)' \left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1)) - \mathbb{E}[\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))]\right) f_{X_1}(x_1) dx_1\right) \\ &\qquad = O_p\left(N^{-1} b_N^{-L/2 - L_1/2 - |\lambda|} + N^{-1/2} b_N^{-L/2 - |\lambda|}\right). \end{aligned}$$

Lemma A.8 *Suppose Assumption 2 holds, and that $r \geq d$ and $|\lambda| \leq d - 1$. Then for all $\gamma \leq \lambda$,*

$$\int_{\mathbb{U}} u^{\gamma} K^{(\lambda)}(u) du = \begin{cases} 0 & \text{if } \gamma \neq \lambda \\ (-1)^{|\lambda|} \lambda! & \text{if } \gamma = \lambda. \end{cases}$$

Notation We define the Λ vectors $\alpha_1(x_1) = f_{X_1}(x_1)\nu_1(x_1)$ and $\alpha_2(x_1) = f_{X_1}(x_1)\nu_2(x_1)$. Also $\alpha_m^{(\lambda)}$ is the vector of λ derivatives of the components of α_m for $m = 1, 2$. It will be convenient to refer to the components of α_1 and α_2 by the derivative that they refer to. Remember that ν_1 and ν_2 are derivatives of the scalar function n with respect the $h_1^{[\lambda]}(x)$ and $h_2^{[\lambda]}(x)$, the Λ vectors of derivatives of h_1 and h_2 up to order λ. Hence we indicate components of α_1, α_2 by the derivative $h^{(\kappa)}$, e.g. $\alpha_{\kappa 1}$ is the product of f_X and the derivative of n with respect to $h_1^{(\kappa)}$. By Assumptions 1 and 4 these functions have the same properties as ν_1, ν_2. In particular by Assumptions 4 and 1 α_1 and α_2 are q times differentiable and the q-th derivative is bounded. The following lemma refers only to the full mean case with $L_2 = 0$.
The estimator of $h^{(\lambda)}$ is the NIP estimator that uses the trimmed NW estimator. The trimming in combination with the derivative kernel deals with weights that become unbounded if the bandwidth goes to 0, a problem that we refer to as the boundary variance problem.

Lemma A.9 (Full Mean) *Suppose that Assumptions 2-?? and 4 hold, then if* $r \geq |\lambda| + s - 1$, $q \geq 2|\lambda| + s$, $t \geq |\lambda| + s$ $d = \max\{\lambda_1, \ldots, \lambda_L\} + s - 1$

$$\sqrt{N} \sum_{m=1}^{2} \int_{\mathbb{X}} \alpha_m(x)' \left(\hat{h}^{[\lambda]}_{m,\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{m,\text{nip},s}(x) \right] \right) dx - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{\kappa \leq \lambda} (-1)^{|\kappa|} \sum_{m=1}^{2} \left(\alpha^{(\kappa)}_{\kappa m}(X_i) Z_{im} - \mathbb{E}[\alpha^{(\kappa)}_{\kappa m}(X) Z_m] \right) \right)$$
$$= O_p\left(b_N^{\min\{1, L/2\}} \right).$$

The results thus far establish the asymptotic properties of the GFM and GFAD estimators.
Next, we present the results for the partial mean case.

Lemma A.10 *Suppose that Assumptions 2, 1 and 4 hold. If* $\lambda = 0$ *then if* $t(x_1)$ *is twice continuously differentiable on* $\mathbb{X}_1$*, the kernel is of derivative order* $d \geq s - 1$, $p \geq 2$, $q \geq s$, $t \geq s$, *and* $r \geq s - 1$

$$\sqrt{N} \cdot b_N{}^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \alpha_m(x_1) \left(\hat{h}_{m,\text{nip},s}(x_1, t(x_1)) - \mathbb{E}\left[\hat{h}_{m,\text{nip},s}(x_1, t(x_1)) \right] \right) dx_1$$
$$- \frac{1}{b_N{}^{L_2/2}\sqrt{N}} \cdot \sum_{i=1}^{N} \sum_{m=1}^{2} \left(\alpha_m(X_{i1}) Z_{im} \int_{\mathbb{U}_1} K\left(u_1, \frac{X_{i2} - t(X_{i1})}{b_N} + \frac{\partial}{\partial x_1'} t(X_{i1}) \cdot u_1 \right) du_1 \right.$$
$$\left. - \mathbb{E}\left[\alpha_m(X_1) Z_m \int_{\mathbb{U}_1} K\left(u_1, \frac{X_2 - t(X_1)}{b_N} + \frac{\partial}{\partial x_1'} t(X_1) \cdot u_1 \right) du_1 \right] \right) = O_p\left(b_N^{\min\{1, L_1/2\}} \right).$$

If $t(x_1) = x_{02}$ *then for* $\lambda \geq 0$ *with in addition to the listed assumptions a kernel of derivative order* $d \geq s - 1 + \max\{\lambda_1, \ldots, \lambda_L\}$, $p \geq 2$, $q \geq 2|\lambda| + s$, $t \geq |\lambda| + s$, *and* $r \geq s - 1 + \max\{\lambda_{11}, \ldots, \lambda_{1L_1}\}$

$$\sqrt{N} \cdot b_N{}^{L_2/2 + |\lambda_2|} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \alpha_m(x_1)' \left(\hat{h}^{[\lambda]}_{m,\text{nip},s}(x_1, t(x_1)) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{m,\text{nip},s}(x_1, t(x_1)) \right] \right) dx_1$$
$$- \frac{1}{b_N{}^{L_2/2}\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \sum_{m=1}^{2} \left(\alpha^{(\kappa_1)}_{\kappa_1 \lambda_1, m}(X_{i1}) Z_{im} K^{(\lambda_2)}\left(\frac{x_{02} - X_{i2}}{b_N} \right) - \mathbb{E}\left[\alpha^{(\kappa_1)}_{\kappa_1 \lambda_1, m}(X_1) Z_m K^{(\lambda_2)}\left(\frac{x_{02} - X_2}{b_N} \right) \right] \right) \right)$$
$$= O_p\left(b_N^{\min\{1, L_1/2\}} \right).$$

The final lemma establishes that the asymptotic distribution is normal for partial means and partial average derivatives.

Lemma A.11 *Suppose that Assumptions 2, 1 and 4 hold, and that* $t(x_1)$ *is continuously differentiable on* $\mathbb{X}_1$ *and that* $p \geq 4$ *and* $N^{-1} b_N^{-L_2} \to 0$*. Then*

$$N^{-1/2} b_N^{-L_2/2} \sum_{i=1}^{N} \left(\sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_m(X_{i1}) Z_{im} K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1}) u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N} \right) du_1 \right.$$
$$\left. - \mathbb{E}\left[\sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_m(X_1) Z_m K\left(u_1, \frac{\partial t}{\partial x_1}(X_1) u_1 + \frac{t(X_1) - X_2}{b_N} \right) du_1 \right] \right) \xrightarrow{d} \mathcal{N}(0, V),$$

where

$$V = \sum_{m=1}^{2} \sum_{m'=1}^{2} \int_{\mathbb{X}_1} \mu_{mm'}(x_1, t(x_1)) \alpha_m(x_1) \alpha_{m'}(x_1) \int_{\mathbb{U}_2} \left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(x_1) u_1 + u_2 \right) du_1 \right)^2 du_2 f_X(x_1, t(x_1)) dx_1$$

with $\mu_{mm'}(x) = \mathbb{E}[Z_m Z_{m'} | X = x]$ *for* $m, m' = 1, 2$*. If in addition* $r \geq |\lambda_2|$ *and* $q, t \geq |\lambda_1|$

$$N^{-1/2} b_N^{-L_2/2} \sum_{i=1}^{N} \left(\sum_{m=1}^{2} \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1 \lambda_2, m}(X_{i1}) \right) Z_{im} K^{(\lambda_2)}\left(\frac{x_{02} - X_{i2}}{b_N} \right) \right.$$
$$\left. - \mathbb{E}\left[\sum_{m=1}^{2} \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1 \lambda_2, m}(X_1) \right) Z_m K^{(\lambda_2)}\left(\frac{x_{02} - X_2}{b_N} \right) \right] \right) \xrightarrow{d} \mathcal{N}(0, V)$$

with

$$V = \sum_{m=1}^{2} \sum_{m'=1}^{2} \int_{\mathbb{X}_1} \mu_{mm'}(x_1, t(x_1)) \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1 \lambda_2, m}(x_1) \right) \left(\sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \alpha^{(\kappa_1)}_{\kappa_1 \lambda_2, m'}(x_1) \right)$$

$$\cdot \int_{\mathbb{U}_2} \left(K^{(\lambda_2)}(u_2) \right)^2 du_2 f_X(x_1, t(x_1)) dx_1$$

7 Higher Order Kernels

Let u be scalar, and let $\mathcal{K}(u)$ be a symmetric kernel with $\mu_j = \int_u u^j \mathcal{K}(u)du$, $\mu_0 = 1$, $\mu_j = 0$ for j is odd. Then we can construct a kernel of order s, as

$$K(u) = \sum_{j=0}^{s-1} \alpha_j \cdot u^j \cdot \mathcal{K}(u),$$

where the α_j ensure that $\int u^j \cdot K(u)du = 0$ for $j = 1, \ldots, s-1$ and $\int K(u)du = 1$, leading to

$$\alpha = \begin{pmatrix} \mu_0 & \mu_1 & \ldots & \mu_{s-1} \\ \mu_1 & \mu_2 & \ldots & \mu_s \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{s-1} & \mu_s & \ldots & \mu_{2s-2} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We can choose $\mathcal{K}(u) = 1/2$, for $u \in [-1, 1]$, so that $\mu_j = 1/(j+1)$ for j even, and (assuming $s-1$ is even)

$$\alpha = \begin{pmatrix} 1 & 0 & 1/3 & \ldots & 1/s \\ 0 & 1/3 & 0 & \ldots & 0 \\ 1/3 & 0 & 1/5 & \ldots & 1/(s+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/s & 0 & 1/(s+2) & \ldots & 1/(2s-1) \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If we choose the Epanechnikov kernel as the basis, $\mathcal{K}(u) = (3/4)(1-u^2)$ for $-1 < u < 1$, then $\mu_j = 3/((j+1)(j+3))$ for j even and greater than 0, and $\mu_j = 0$ for j odd. Then

$$\alpha = \begin{pmatrix} 1 & 0 & 1/5 & \ldots & 3/(s(s+2)) \\ 0 & 1/5 & 0 & \ldots & 0 \\ 1/5 & 0 & 3/35 & \ldots & 3/((s+2)(s+4)) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 3/(s(s+2) & 0 & 3/((s+2)(s+4)) & \ldots & 3/((2s-1)(2s+1)) \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

According to Pagan and Ullah (1999) the optimal fourth order kernel is (note that the optimal second order kernel is the Epanechnikov kernel):

$$K(u) = \frac{15}{32} \cdot \left(7 \cdot u^4 - 10 \cdot u^2 + 3\right),$$

for $-1 < u < 1$. Let us consider three cases. In each case we look for the least restrictive conditions on the order of the kernel first, then on the number of moments required for Z given X.

1. $L = 1$, $s = 3$, $p > 2s/(s-L-1) = 6$, so $p = 7$, $j = 0$, $J = 1$, $q = 5$, $r = 2$, $1/6 = 1/(2s) < \delta < (1-2/p)/(2L+2) = 5/28$, so $14/84 < \delta < 15/84$.
2. $L = 2$, $s = 4$, $p > 2s/(s-L-1) = 8$, so $p = 9$, $j = 0$, $J = 1$, $q = 7$, $r = 3$, $1/8 = 1/(2s) < \delta < (1-2/p)/(2L+2) = 7/54$, so $27/216 < \delta < 28/216$.
3. $L = 3$, $s = 5$, $p > 2s/(s-L-1) = 10$, so $p = 11$, $j = 0$, $J = 1$, $q = 9$, $r = 4$, $1/10 = 1/(2s) < \delta < (1-2/p)/(2L+2) = 9/88$, so $44/440 < \delta < 45/440$.

Appendix B: Proofs of Additional Lemmas and Theorems

Proof of Lemma A.1
We have

$$\mathbb{E}\left[\hat{h}_{\text{nw}}^{(\lambda)}(x)\right] = \mathbb{E}\left[Z\frac{1}{b^{L+|\lambda|}}K^{(\lambda)}\left(\frac{x-X}{b}\right)\right] = \int_{\mathbb{X}} g(v)\frac{1}{b^{L+|\lambda|}}K^{(\lambda)}\left(\frac{x-v}{b}\right)f_X(v)\mathrm{d}v,$$

for $g(v) = \mathbb{E}[Z|X=v]$. For all $x, v \in \mathbb{X}$ we have by assumption 1,

$$\left|g(v)K^{(\lambda)}\left(\frac{x-v}{b}\right)f_X(v)\right| \leq C\sup_{x,v\in\mathbb{X}}\left|K^{(\lambda)}\left(\frac{x-v}{b}\right)\right| = C\sup_{u\in\mathbb{U}}\left|K^{(\lambda)}(u)\right| < \infty$$

because by Assumption 2 the kernel $K^{(\lambda)}$ is bounded on $\mathbb{U}$. Hence by dominated convergence we can interchange differentiation and integration λ times, so that

$$\mathbb{E}\left[\hat{h}_{\text{nw}}^{(\lambda)}(x)\right] = \frac{\partial^{|\lambda|}}{\partial x^{\lambda}}\mathbb{E}\left[\hat{h}_{\text{nw}}(x)\right].$$

□

Proof of Lemma A.2
The result follows directly from Lemma 3.2, if we can interchange the expectation and λ derivative. This follows from Lemma A.1. □
Proof of Theorem A.1. By the triangle inequality,

$$\sup_{x\in\mathbb{X}}\left|\hat{h}_{\text{nip},s}^{(\lambda)}(x) - h^{(\lambda)}(x)\right|$$
$$\leq \sup_{x\in\mathbb{X}}\left|\mathbb{E}\left[\hat{h}_{\text{nip},s}^{(\lambda)}(x)\right] - h^{(\lambda)}(x)\right| \tag{B.1}$$
$$+\sup_{x\in\mathbb{X}}\left|\hat{h}_{\text{nip},s}^{(\lambda)}(x) - \mathbb{E}\left[\hat{h}_{\text{nip},s}^{(\lambda)}(x)\right]\right|. \tag{B.2}$$

By Lemma 3.3 it follows that (B.1) is $O(b^s)$. By Lemma 3.4 it follows that (B.2) is $O_p\left(\left(\frac{\log N}{N\cdot b_N^{L+2j}}\right)^{1/2}\right)$. Combining these two results implies the result in the Theorem. □
Note that this implies that if $\delta < 1/(L+2|\lambda|)$, we have convergence of $\hat{h}_{\text{nip},s}^{(\lambda)}(x)$ to $h^{(\lambda)}(x)$ uniform in x.
Proof of Lemma A.3. First consider (i). For the upper bound on the bias in (i) we use Lemma 3.3. If $q \geq j+s$ and $r \geq j+s-1$ this Lemma implies that the bias is $O(b_N^s) = O(N^{-\delta s})$. If $\delta > 1/(2s)$, then the bias is $o_p(N^{-1/2})$.
Next, consider (ii). By the triangle inequality we have

$$\sup_{x\in\mathbb{X}}\left|\hat{h}_{\text{nip},s}^{(\lambda)}(x) - h^{(\lambda)}(x)\right| \leq \sup_{x\in\mathbb{X}}\left|\hat{h}_{\text{nip},s}^{(\lambda)}(x) - \mathbb{E}\left[\hat{h}_{\text{nip},s}^{(\lambda)}(x)\right]\right| + \sup_{x\in\mathbb{X}}\left|\mathbb{E}\left[\hat{h}_{\text{nip},s}^{(\lambda)}(x)\right] - h^{(\lambda)}(x)\right|.$$

By the first part of this Lemma the last term is $o(N^{-1/2}) = o_p(N^{-1/4})$. To deal with the first term on the right hand side we use Theorem A.1. We simultaneously need to satisfy the condition on the bandwidth in this Theorem, which requires

$$\delta < \frac{1-2/p}{L+2j+2}$$

and ensure that the upper bound on the variance is of order $N^{-1/4}$, which requires that

$$O\left(\left(\frac{\log N}{N\cdot b_N^{L+2j}}\right)^{1/2}\right) = o\left(N^{-1/4}\right),$$

which requires that

$$\delta < \frac{1}{2L+4j}.$$

Hence a sufficient condition for the variance is that

$$\delta < \min\left\{\frac{1-2/p}{L+2j+2}, \frac{1}{2L+4j}\right\}.$$

Thus both results hold if

$$\frac{1}{2s} < \delta < \min\left\{\frac{1-2/p}{L+2j+2}, \frac{1}{2L+4j}\right\}.$$

Finally, for all L, j, $p \geq 2$, there are s, q, r and δ that satisfy these conditions. □.

Proof of Lemma A.4: By Assumption 4 there is some $M < \infty$ such that

$$\left|\frac{\partial^2}{\partial h^{[\lambda]}\partial h^{[\lambda]\prime}} m(h^{[\lambda]}(x))\right| \leq M,$$

for all $x \in \mathbb{X}$ and $h^{[\lambda]} \in B(h^{[\lambda]}, \varepsilon)$. Then if $\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]}, \varepsilon)$ and $\overline{h^{[\lambda]}}(x)$ intermediate between $h^{[\lambda]}(x)$ and $\hat{h}^{[\lambda]}_{NIP,s}(x)$ we have for all $x \in \mathbb{X}$,

$$\left|\omega(x_1)\left(m(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)) - m(h^{[\lambda]}(x_1,t(x_1)) - \frac{\partial}{\partial h^{[\lambda]\prime}} m(h^{[\lambda]}(x_1,t(x_1))(\hat{h}^{[\lambda]}(x_1,t(x_1)) - h^{[\lambda]}(x_1,t(x_1)))\right)\right|$$

$$= \frac{1}{2}\left|\omega(x_1)(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)) - h^{[\lambda]}(x_1,t(x_1))'\frac{\partial^2}{\partial h^{[\lambda]}\partial h^{[\lambda]\prime}} m(\overline{h^{[\lambda]}}(x_1,t(x_1))(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1) - h^{[\lambda]}(x_1,t(x_1)))\right|$$

$$\leq \frac{1}{2}\sup_{x_1\in\mathbb{X}_1}|\omega(x_1)| \cdot M \cdot \left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2 \leq C \cdot \left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2.$$

Hence if we denote

$$d(x_1) = \omega(x_1)\left(m(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)) - m(h^{[\lambda]}(x_1,t(x_1)) - \frac{\partial}{\partial h^{[\lambda]\prime}} m(h^{[\lambda]}(x_1,t(x_1))(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)) - h^{[\lambda]}_m(x_1,t(x_1)))\right),$$

we have that $\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]}, \varepsilon)$ implies

$$\left|\frac{1}{N}\sum_{i=1}^N d(X_{i1})\right| \leq C\left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2.$$

Therefore, for all $\eta > 0$ and all sequences a_N,

$$\Pr\left(\left|\frac{1}{a_N\sqrt{N}}\sum_{i=1}^N d(X_{i1})\right| > \eta\right) \leq \Pr\left(\left|\frac{1}{a_N\sqrt{N}}\sum_{i=1}^N d(X_{i1})\right| > \eta\middle|\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]},\varepsilon)\right) \cdot \Pr\left(\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]},\varepsilon)\right)$$

$$+ \Pr\left(\hat{h}^{[\lambda]}_{NIP,s} \notin B(h^{[\lambda]},\varepsilon)\right)$$

$$\leq \Pr\left(C \cdot \frac{\sqrt{N}}{a_N} \cdot \left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2 > \eta\middle|\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]},\varepsilon)\right) \cdot \Pr\left(\hat{h}^{[\lambda]}_{NIP,s} \in B(h^{[\lambda]},\varepsilon)\right)$$

$$+ \Pr\left(\hat{h}^{[\lambda]}_{NIP,s} \notin B(h^{[\lambda]},\varepsilon)\right)$$

$$\leq \Pr\left(C \cdot \frac{\sqrt{N}}{a_N} \cdot \left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2 > \eta\right) + \Pr\left(\left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0 > \varepsilon\right).$$

The second term in the upper bound converges to 0, because by the assumptions of the lemma the NIP estimator is uniformly convergent on $\mathbb{X}$. Now if for some sequence a_N the sequence $\frac{\sqrt{N}}{a_N}\left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2$ is stochastically bounded, then so is the sequence $\frac{1}{a_N\sqrt{N}}\sum_{i=1}^N d(X_{i1})$ so that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N d(X_{i1}) = O_p\left(\sqrt{N}\left|\hat{h}^{[\lambda]}_{NIP,s} - h^{[\lambda]}\right|_0^2\right).$$

□

Proof of Lemma A.5:

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\nu(X_{i1})'\left(\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))\right]-h^{[\lambda]}(v(X_{i1}))\right)\right| \\
&\qquad\le \sup_{x_1\in\mathbb{X}_1}\left|\nu(x_1)'\left(\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))\right]-h^{[\lambda]}(v(x_1))\right)\right| \\
&\qquad\le \sup_{x_1\in\mathbb{X}_1}|\nu(x_1)|\cdot\sup_{x_1\in\mathbb{X}_1}\left|\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))\right]-h_0^{[\lambda]}(v(x_1))\right| \\
&\qquad\le \sup_{x_1\in\mathbb{X}_1}|\nu(x_1)|\cdot\sup_{x\in\mathbb{X}}\left|\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(x)\right]-h^{[\lambda]}(x)\right| \\
&\qquad= C\cdot\sup_{x\in\mathbb{X}}\left|\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(x)\right]-h^{[\lambda]}(x)\right| = O\left((2b)^s\right),
\end{aligned}$$

where the last equality follows from Lemma 3.3. □

Proof of Lemma A.6: Define

$$V=\frac{1}{N}\sum_{m=1}^{2}\sum_{i=1}^{N}\nu_m(X_{i1})'\left(\hat{h}^{[\lambda]}_{m,\text{nw}}(v(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{m,\text{nw}}(v(X_{i1}))\right]\right),$$

$$U=\sum_{m=1}^{2}\int_{\mathbb{X}_1}\nu_m(x_1)'\left(\hat{h}^{[\lambda]}_{m,\text{nw}}(v(x_1))-\mathbb{E}[\hat{h}^{[\lambda]}_{\text{nw}}(v(x_1))]\right)f_{X_1}(x_1)\mathrm{d}x_1,$$

$$V_m=\frac{1}{N}\sum_{i=1}^{N}\nu_m(X_{i1})'\left(\hat{h}^{[\lambda]}_{m,\text{nw}}(v(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{m,\text{nw}}(v(X_{i1}))\right]\right),$$

and

$$U_m=\int_{\mathbb{X}_1}\nu_m(x_1)'\left(\hat{h}^{[\lambda]}_{m,\text{nw}}(v(x_1))-\mathbb{E}[\hat{h}^{[\lambda]}_{m,\text{nw}}(v(x_1))]\right)f_{X_1}(x_1)\mathrm{d}x_1.$$

Then the claim in the Lemma is

$$V-U=O_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}+N^{-1}b_N^{-L/2-|\lambda|}\right),$$

and since $V=\sum_{m=1}^{2}V_m$, and $U=\sum_{m=1}^{2}U_m$, it is sufficient to show that

$$V_m-U_m=O_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}+N^{-1}b_N^{-L/2-|\lambda|}\right),$$

for $m=1,2$.
Note that

$$\hat{h}^{[\lambda]}_{m,\text{nw}}(x)=\frac{1}{N}\sum_{j=1}^{N}I\left(X_j\in\mathbb{X}^I_{b_N}\right)Z_{jm}b_N^{-L-|[\lambda]|}*K^{[\lambda]}\left(\frac{X_j-x}{b_N}\right),$$

where the vector $b_N^{-L-|[\lambda]|}$ is the Λ vector with components equal to $b_N^{-L-|\mu|}$ with $\mu\le\lambda$. The Λ vector $K^{[\lambda]}$ has as its components the derivatives of K up to order λ.
Define

$$a_{N,m}(x_{i1},z_j,x_j)=I\left(x_j\in\mathbb{X}^I_{b_N}\right)z_{jm}\nu_m(x_{i1})'\frac{1}{b_N^{L+|[\lambda]|}}*K^{[\lambda]}\left(\frac{x_j-v(x_{i1})}{b_N}\right),$$

$$b_{N,m}(x_{i1},z_j,x_j)=a_{N,m}(x_{i1},z_j,x_j)-\mathbb{E}\left[a_{N,m}(x_{i1},Z,X)\right],$$

and

$$c_{N,m}(z_j,x_j)=\mathbb{E}\left[b_{N,m}(X_1,z_j,x_j)\right].$$

By definition, if $i\neq j$,

$$\mathbb{E}\left[c_{N,m}(Z_j,X_j)^2\right]\le\mathbb{E}\left[b_{N,m}(X_{i1},Z_j,X_j)^2\right]\le\mathbb{E}\left[a_{N,m}(X_{i1},Z_j,X_j)^2\right],\qquad\text{(B.3)}$$

and

$$\mathbb{E}\left[(b_{N,m}(X_{i1},Z_j,X_j)-c_N(Z_j,X_j))^2\right]\leq\mathbb{E}\left[b_{N,m}(X_{i1},Z_j,X_j)^2\right]\leq\mathbb{E}\left[a_{N,m}(X_{i1},Z_j,X_j)^2\right]. \tag{B.4}$$

We need two results on the expectations of $a_N(X_{i1},Z_j,X_j)$. If $i\neq j$, then

$$\mathbb{E}\left[a_{N,m}(X_{i1},Z_j,X_j)^2\right]=O\left(b_N^{-L-2|\lambda|}\right), \tag{B.5}$$

and

$$\mathbb{E}\left[a_{N,m}(X_{i1},Z_i,X_i)^2\right]=O\left(b_N^{-L-L_1-2|\lambda|}\right). \tag{B.6}$$

First, we prove (B.5):

$$\begin{aligned}\mathbb{E}\left[a_{N,m}(X_{i1},Z_j,X_j)^2\right]&=\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left(Z_{jm}\nu_m(X_{i1})'\frac{1}{b_N^{L+|[\lambda]|}}*K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right)^2\right]\\&\leq\frac{1}{b_N^{2L+2|\lambda|}}\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left(Z_{jm}\nu_m(X_{i1})'K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right)^2\right]\\&\leq\frac{1}{b_N^{2L+2|\lambda|}}\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)Z^2_{jm}|\nu_m(X_{i1})|'_*\left|K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|_*\left|K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|'_*|\nu_m(X_{i1})|_*\right]\\&\leq C\frac{1}{b_N^{2L+2|\lambda|}}\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left|K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|'_*\left|K^{[\lambda]}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|_*\right],\end{aligned}$$

where we use the inequality $|x'(y*z)|\leq\max_{i=1,\ldots,I}|y_i||x|'_*|z|_*$ and the fact that $1/b_N^{2L+2|\lambda|}$ is the largest component of $1/b_N^{2L+2|[\lambda]|}$, the Cauchy-Schwartz inequality, and the boundedness of $\mathbb{E}[Y_j^2|X_j=x]$ and $\nu(x_1)$ on $\mathbb{X}$ and $\mathbb{X}_1$, respectively. The absolute value of the expectation is bounded by

$$C\sup_{x_1\in\mathbb{X}_1}\int_{\mathbb{X}^{\mathbb{I}}_{\mathbb{N}}}\left|K^{(\kappa_1)}\left(\frac{z-v(x_1)}{b_N}\right)K^{(\kappa_2)}\left(\frac{z-v(x_1)}{b_N}\right)\right|f_X(z)\mathrm{d}z,$$

for some $\kappa_1,\kappa_2\leq\lambda$. By a change of variables from z to $u=(z-v(x_1))/b_N$ with Jacobian b_N^L this is equal to

$$\begin{aligned}&C\cdot b_N^L\sup_{x_1\in\mathbb{X}_1}\int_{\{u|u=(x-v(x_1))/b_N,x\in\mathbb{X}^I_{b_N}\}}\left|K^{(\kappa_1)}(u)K^{(\kappa_2)}(u)\right|f_X(v(x_1)+ub_N)\mathrm{d}u\\&\leq C\cdot b_N^L\int_{\mathbb{U}}\left|K^{(\kappa_1)}(u)K^{(\kappa_2)}(u)\right|\mathrm{d}u\leq Cb_N^L,\end{aligned}$$

so that

$$\mathbb{E}\left[a_{N,m}(X_{i1},Z_j,X_j)^2\right]=O\left(b_N^{-L-2|\lambda|}\right).$$

Thus (B.5) holds. Next, we prove (B.6):

$$\begin{aligned}\mathbb{E}\left[a_{N,m}(X_{i1},Z_i,X_i)^2\right]&\leq\frac{1}{b_N^{2L+2|\lambda|}}\mathbb{E}\left[I\left(X_i\in\mathbb{X}^I_{b_N}\right)\left(|Z_{im}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|_*\right)^2\right]\\&\leq\frac{1}{b_N^{2L+2|\lambda|}}C\mathbb{E}\left[I\left(X_i\in\mathbb{X}^I_{b_N}\right)\left|K^{[\lambda]}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|'_*\left|K^{[\lambda]}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|_*\right],\end{aligned}$$

The absolute value of the expectation is bounded by

$$C\int_{\mathbb{X}^I_{b_N}}\left|K^{(\kappa_1)}\left(0,\frac{x_2-t(x_1)}{b_N}\right)\right|\left|K^{(\kappa_2)}\left(0,\frac{x_2-t(x_1)}{b_N}\right)\right|f_X(x)\mathrm{d}x$$

By a change of variables $u_2=(x_2-t(x_1))/b_N$ with Jacobian $b_N^{L_2}$ this is

$$Cb_N^{L_2}\int_{\mathbb{X}^I_{1,b_N}}\int_{\{u_2|u_2=(x_2-t(x_1))/b_N,x_2\in\mathbb{X}^I_{2,b_N}\}}\left|K^{(\kappa_1)}(0,u_2)\right|\left|K^{(\kappa_2)}(0,u_2)\right|f_X(x_1,t(x_1)+u_2b_N)\mathrm{d}u_2\mathrm{d}x_1$$

$$\leq Cb_N^{L_2}\int_{\mathbb{X}_1}\int_{\mathbb{U}_2}\left|K^{(\kappa_1)}(0,u_2)\right|\left|K^{(\kappa_2)}(0,u_2)\right|f_X(x_1,t(x_1)+u_2b_N)\mathrm{d}u_2\mathrm{d}x_1\leq Cb_N^{L_2},$$

implying that

$$\mathbb{E}\left[a_{N,m}(X_{i1},Z_i,X_i)^2\right]=O\left(b_N^{-L-L_1-2|\lambda|}\right).$$

Thus (B.6) holds.
By the definitions of $b_{N,m}(x_{i1},z_j,x_j)$ and $c_{N,m}(x_{i1},z_j,x_j)$ it follows that

$$V_m=\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N b_{N,m}(X_{i1},Z_j,X_j),$$

is a V-statistic with (non-symmetrized) kernel $b_{N,m}(x_{i1},z_j,x_j)$, which depends on N, and

$$U_m=\frac{1}{N}\sum_{j=1}^N c_{N,m}(Z_j,X_j).$$

Define

$$V_{m,1}=\frac{1}{N(N-1)}\sum_{i\neq j=1}^N b_{N,m}(X_{i1},Z_j,X_j),$$

$$V_{m,2}=\frac{1}{N^2}\sum_{i=1}^N b_{N,m}(X_{i1},Z_i,X_i),$$

so that

$$V_m-U_m=\frac{(N-1)N}{N^2}(V_{m,1}-U_m)+\left(\frac{(N-1)N}{N^2}-1\right)U_m+V_{m,2}.$$

Hence to prove the lemma it is sufficient to show that

$$\frac{(N-1)N}{N^2}(V_{m,1}-U_m)+\left(\frac{(N-1)N}{N^2}-1\right)U_m+V_{m,2}=O_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}+N^{-1}b_N^{-L/2-|\lambda|}\right).\quad\text{(B.7)}$$

We do this in three steps. First we show that

$$\frac{(N-1)N}{N^2}(V_{m,1}-U_m)=O_p\left(N^{-1}b_N^{-L/2-|\lambda|}\right).\quad\text{(B.8)}$$

Second, we show that

$$\left(\frac{(N-1)N}{N^2}-1\right)U_m=O_p\left(N^{-3/2}b_N^{-L/2-|\lambda|}\right)=o_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}\right).\quad\text{(B.9)}$$

Third, we show that

$$V_{m,2}=O_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}\right).\quad\text{(B.10)}$$

First consider (B.8). We have

$$V_{m,1}-U_m=\frac{1}{N(N-1)}\sum_{i\neq j=1}^N\left(b_{N,m}(X_{i1},Z_j,X_j)-c_{N,m}(Z_j,X_j)\right),$$

so that

$$\mathbb{E}\left[(V_{m,1}-U_m)^2\right]=\frac{1}{N^2(N-1)^2}\sum_{i\neq j=1}^N\sum_{k\neq l=1}^N\mathbb{E}\left[\left(b_{N,m}(X_{i1},Z_j,X_j)-c_{N,m}(Z_j,X_j)\right)\left(b_{N,m}(X_{k1},Z_l,X_l)-c_{N,m}(Z_l,X_l)\right)\right].$$

We consider

$$\mathbb{E}\left[\left(b_{N,m}(X_{i1},Z_j,X_j)-c_{N,m}(Z_j,X_j)\right)\left(b_{N,m}(X_{k1},Z_l,X_l)-c_{N,m}(Z_l,X_l)\right)\right].$$

There are four cases: (i), $i \neq k$, $j \neq l$, (ii), $i \neq k$, $j = l$, (iii), $i = k$, $j \neq l$, and (iv), $i = k$, $j = l$. In case (i), if $i \neq k, j \neq l$, then the expectation is 0. In case (ii), if $i \neq k, j = l$, then

$$\begin{aligned}
&\mathbb{E}\left[\left(b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)\left(b_{N,m}(X_{k1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)\right] \\
&= \mathbb{E}_{Z_j,X_j}\left[\mathbb{E}\left[\left(b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)\left(b_{N,m}(X_{k1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)|Z_j, X_j\right]\right] \\
&= \mathbb{E}_{Z_j,X_j}\left[\mathbb{E}\left[b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)|Z_j, X_j\right]\mathbb{E}\left[b_{N,m}(X_{k1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)|Z_j, X_j\right]\right] = 0,
\end{aligned}$$

because

$$\mathbb{E}\left[b_{N,m}(X_{i1}, Z_j, X_j)|Z_j, X_j\right] = c_{N,m}(Z_j, X_j).$$

In case (iii), with $i = k, j \neq l$

$$\begin{aligned}
&\mathbb{E}\left[\left(b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)\left(b_{N,m}(X_{i1}, Z_l, X_l) - c_{N,m}(Z_l, X_l)\right)\right] \\
&= \mathbb{E}_{X_{i1}}\left[\mathbb{E}\left[\left(b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)\left(b_{N,m}(X_{i1}, Z_l, X_l) - c_{N,m}(Z_l, X_l)\right)|X_{i1}\right]\right] \\
&= \mathbb{E}_{X_{i1}}\left[\mathbb{E}\left[b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)|X_{i1}\right]\mathbb{E}\left[b_{N,m}(X_{i1}, Z_l, X_l) - c_{N,m}(Z_l, X_l)|X_{i1}\right]\right] = 0,
\end{aligned}$$

because

$$\mathbb{E}\left[b_{N,m}(X_{i1}, Z_j, X_j)|X_{i1}\right] = 0, \quad \text{and} \quad \mathbb{E}\left[c_{N,m}(Z_j, X_j)|X_{i1}\right] = 0.$$

This leaves only case (iv) where $i = k$ and $j = l$, so that because of (B.4) and(B.3),

$$\begin{aligned}
\mathbb{E}\left[(V_{m,1} - U_m)^2\right] &= \frac{1}{N^2(N-1)^2}\sum_{i \neq j = 1}^{N}\mathbb{E}\left[\left(b_{N,m}(X_{i1}, Z_j, X_j) - c_{N,m}(Z_j, X_j)\right)^2\right] \\
&\leq \frac{1}{N^2(N-1)^2}\sum_{i \neq j = 1}^{N}\mathbb{E}\left[b_{N,m}(X_{i1}, Z_j, X_j)^2)\right] \\
&\leq \frac{1}{N^2(N-1)^2}\sum_{i \neq j = 1}^{N}\mathbb{E}\left[a_{N,m}(X_{i1}, Z_j, X_j)^2)\right] \leq \frac{C}{N(N-1)}b_N^{-L-2|\lambda|}.
\end{aligned}$$

Thus, by the Markov inequality we find that

$$V_{m,1} - U_m = O_p\left(N^{-1}b_N^{-L/2-|\lambda|}\right).$$

Next, consider (B.9). Because $\mathbb{E}[(c_{N,m}(Z, X))^2] \leq \mathbb{E}[(a_{N,m}(X_i, Z_j, X_j))^2] = O(b_N^{-L-2|\lambda|})$, we find that

$$\begin{aligned}
\mathbb{E}[U_m^2] &= \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}c_{N,m}(Z_i, X_i) \cdot c_{N,m}(Z_j, X_j)\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^{N}c_{N,m}(Z_i, X_i)^2\right] \\
&\leq \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^{N}a_{N,m}(Z_i, X_i)^2\right] \leq \frac{C}{N}b_N^{-L-2|\lambda|},
\end{aligned}$$

so that

$$U_m = O_p\left(N^{-1/2}b_N^{-L/2-|\lambda|}\right),$$

and

$$\left(\frac{(N-1)N}{N^2} - 1\right)U_m = -\frac{1}{N}U_m = O_p\left(N^{-3/2}b_N^{-L/2-|\lambda|}\right).$$

Finally, consider (B.10).

$$\mathbb{E}\left[V_{m,2}^2\right] = \frac{1}{N^4}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[b_{N,m}(X_{i1}, Z_i, X_i) \cdot b_{N,m}(X_{j1}, Z_j, X_j)\right]$$

$$= \frac{1}{N^4} \sum_{i=1}^{N} \mathbb{E}\left[b_{N,m}(X_{i1}, Z_i, X_i)^2\right]$$

$$\leq \frac{1}{N^4} \sum_{i=1}^{N} \mathbb{E}\left[a_{N,m}(X_{i1}, Z_i, X_i)^2\right] \leq \frac{C}{N^3} b_N^{-L-L_1-2|\lambda|},$$

so that

$$V_{m,2} = O_p\left(N^{-3/2} b_N^{-L/2-L_1/2-|\lambda|}\right).$$

Combining the results we have that

$$V_m - U_m = O_p\left(N^{-3/2} b_N^{-L/2-L_1/2-|\lambda|} + N^{-1} b_N^{-L/2-|\lambda|}\right),$$

proving (B.7), and thus the lemma. □

Proof of Lemma A.7 The proof is analogous to that of the previous lemma. Define

$$V = \frac{1}{N} \sum_{i=1}^{N} \sum_{m=1}^{2} \nu_m(X_{i1})' \left(\hat{h}^{[\lambda]}_{m,NIP,s}(v(X_{i1})) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{m,NIP,s}(v(X_{i1}))\right]\right),$$

$$U = \sum_{m=1}^{2} \int_{\mathbb{X}_1} \nu_m(x_1)' \left(\hat{h}^{[\lambda]}_{m,\text{nip},s}(v(x_1)) - \mathbb{E}[\hat{h}^{[\lambda]}_{m,\text{nip},s}(v(x_1))]\right) f_{X_1}(x_1) \mathrm{d}x_1,$$

$$V_{\mu,m} = \frac{1}{N} \sum_{i=1}^{N} \nu_m(X_{i1})' \left(\hat{h}^{[\lambda]+(\mu)}_{m,\text{nw}}(r_{2b_N}(v(X_{i1})) - \mathbb{E}\left[\hat{h}^{[\lambda]+(\mu)}_{m,\text{nw}}(r_{2b_N}(v(X_{i1}))\right]\right)(v(X_{i1}) - r_{2b_N}(v(X_{i1})))^{\mu},$$

and

$$U_{\mu,m} = \int_{\mathbb{X}_1} \nu_m(x_1)' \left(\hat{h}^{[\lambda]+(\mu)}_{m,\text{nw},s}(v(x_1)) - \mathbb{E}[\hat{h}^{[\lambda]+(\mu)}_{m,\text{nw},s}(v(x_1))]\right)(v(x_1) - r_{2b_N}(v(x_1)))^{\mu} f_{X_1}(x_1) \mathrm{d}x_1.$$

Here $h_m^{[\lambda]+(\mu)}$ is the Λ vector of μ derivatives of the components of the vector $h_m^{[\lambda]}$ that are the derivatives of h_m up to λ. Then

$$V = \sum_{m=1}^{2} \sum_{k=0}^{s-1} \sum_{|\mu|=k} \frac{1}{\mu!} V_{\mu,m}, \qquad \text{and} \quad U = \sum_{m=1}^{2} \sum_{k=0}^{s-1} \sum_{|\mu|=k} \frac{1}{\mu!} U_{\mu,m}.$$

The statement in the Lemma is equivalent to

$$V - U = O_p\left(N^{-3/2} b_N^{-L/2-L_1/2-|\lambda|} + N^{-1} b_N^{-L/2-|\lambda|}\right).$$

Because there are a finite number of μ, it is sufficient to show that for all μ and m

$$V_{\mu,m} - U_{\mu,m} = O_p\left(N^{-3/2} b_N^{-L/2-L_1/2-|\lambda|} + N^{-1} b_N^{-L/2-|\lambda|}\right). \tag{B.11}$$

Define

$$a_{N,m,\mu}(x_{i1}, z_j, x_j) = I\left(x_j \in \mathbb{X}^I_{b_N}\right) \nu_m(x_{i1})' z_{jm} \frac{1}{b_N^{|[\lambda]|+L+|\mu|}} * K^{[\lambda]+(\mu)}\left(\frac{x_j - r_{2b_N}(v(x_{i1}))}{b_N}\right) \cdot (v(x_{i1}) - r_{2b_N}(v(x_{i1})))^{\mu},$$

$$b_{N,m,\mu}(x_{i1}, z_j, x_j) = a_{N,m,\mu}(x_{i1}, z_j, x_j) - \mathbb{E}\left[a_{N,m,\mu}(x_{i1}, Z_j, X_j)\right],$$

and

$$c_{N,m,\mu}(z, x) = \mathbb{E}\left[b_{N,m,\mu}(X_1, z, x)\right].$$

We will first show that if $i \neq j$

$$\mathbb{E}\left[a_{N,m,\mu}(X_{i1}, Z_j, X_j)^2\right] = O\left(b_N^{-L-2|\lambda|}\right), \tag{B.12}$$

and to cover the case with $i = j$,

$$\mathbb{E}\left[a_{N,m,\mu}(X_{i1}, Z_i, X_i)^2\right] = O\left(b_N^{-L-L_1-2|\lambda|}\right). \tag{B.13}$$

(In these proofs we use the stronger assumption on the degree of differentiability of the kernel.) Then we will show that this is sufficient for the statement in the Lemma.
First, consider (B.12):

$$\mathbb{E}\left[a_{N,m,\mu}(X_{i1},Z_j,X_j)^2\right]$$
$$\leq \frac{1}{b_N^{|2\lambda|+2L+2|\mu|}}\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left(|Z_{jm}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|_*\right)^2(v(X_{i1})-r_{2b_N}(v(X_{i1})))^{2\mu}\right]$$
$$\leq \frac{1}{b_N^{|2\lambda|+2L+2|\mu|}}\sup_{x_1\mathbb{X}_1}|v(x_1)-r_{2b_N}(v(x_{x1}))|^{2\mu}\cdot\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left(|Z_{jm}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|_*\right)^2\right]$$
$$\leq \frac{C}{b_N^{|2\lambda|+2L}}\cdot\mathbb{E}\left[I\left(X_j\in\mathbb{X}^I_{b_N}\right)\left(|Z_{jm}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_j-v(X_{i1})}{b_N}\right)\right|_*\right)^2\right]$$
$$\leq \frac{C}{b_N^{|2\lambda|+2L}}\cdot C_2 b_N^L\leq\frac{C}{b_N^{L+2|\lambda|}},$$

using the same argument as in the proof of (B.5) in the proof of Lemma A.6.
Second, consider (B.13):

$$\mathbb{E}\left[a_{N,m,\mu}(X_{i1},Z_i,X_i)^2\right]$$
$$\leq \frac{1}{b_N^{|2\lambda|+2L+2|\mu|}}\mathbb{E}\left[I\left(X_i\in\mathbb{X}^I_{b_N}\right)\left(|Z_{im}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|_*\right)^2(v(X_{i1})-r_{2b_N}(v(X_{i1})))^{2\mu}\right]$$
$$\leq \frac{1}{b_N^{|2\lambda|+2L+2|\mu|}}\sup_{x_1\in\mathbb{X}^I_{1,b_N}}|v(x_1)-r_{2b_N}(v(x_{x1}))|^{2\mu}\cdot\mathbb{E}\left[I\left(X_i\in\mathbb{X}^I_{b_N}\right)\left(|Z_{im}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|_*\right)^2\right]$$
$$\leq \frac{C}{b_N^{|2\lambda|+2L}}\cdot\mathbb{E}\left[I\left(X_i\in\mathbb{X}^I_{b_N}\right)\left(|Z_{im}||\nu_m(X_{i1})|'_*\left|K^{[\lambda]+(\mu)}\left(\frac{X_i-v(X_{i1})}{b_N}\right)\right|_*\right)^2\right]$$
$$\leq \frac{C}{b_N^{|2\lambda|+2L}}\cdot C_2 b_N^{L_1}\leq\frac{C}{b_N^{2L+2|\lambda|-L_2}},$$

using the same argument as in the proof of (B.6) in the proof of Lemma A.6.
Finally, as in the proof for Lemma A.6, note that

$$V_{\mu,m}=\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N b_{N,m,\mu}(X_{i1},Z_j,X_j),\qquad\text{and}\quad U_{\mu,m}=\frac{1}{N}\sum_{i=1}^N c_{N,m,\mu}(Z_i,X_i).$$

Define

$$V_{\mu,m,1}=\frac{1}{N(N-1)}\sum_{i\neq j} b_{N,m,\mu}(X_{i1},Z_j,X_j),\qquad\text{and}\quad V_{\mu,m,2}=\frac{1}{N^2}\sum_{i=1}^N b_{N,m,\mu}(X_{i1},Z_i,X_i).$$

Then

$$V_{\mu,m}-U_{\mu,m}=\frac{(N-1)N}{N^2}(V_{\mu,m,1}-U_{\mu,m})+\left(\frac{(N-1)N}{N^2}-1\right)U_{\mu,m}+V_{\mu,m,2}.$$

Following the same argument as in the proof of Lemma A.6, it follows that

$$\frac{(N-1)N}{N^2}(V_{\mu,m,1}-U_{\mu,m})=O_p\left(N^{-1}b_N^{-L/2-|\lambda|}\right),$$
$$\left(\frac{(N-1)N}{N^2}-1\right)U_{\mu,m}=O_p\left(N^{-3/2}b_N^{-L/2-|\lambda|}\right)=o_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}\right),$$

and

$$V_{\mu,m,2}=O_p\left(N^{-3/2}b_N^{-L/2-L_1/2-|\lambda|}\right).$$

Combining these results implies that (B.11) holds. □

Proof of Lemma A.8:
Consider

$$\int_{\mathbb{U}} u^{\gamma} K^{(\lambda)}(u)\mathrm{d}u = \prod_{l=1}^{L} \int_{-1}^{1} u_l^{\gamma_l} \mathcal{K}^{(\lambda_l)}(u_l)\mathrm{d}u_l.$$

We focus on

$$\int_{-1}^{1} u_l^{\gamma_l} \mathcal{K}^{(\lambda_l)}(u_l)\mathrm{d}u_l,$$

and, droping the subscripts l for ease of notation, we write this as

$$\int_{-1}^{1} u^{\gamma} \mathcal{K}^{(\lambda)}(u)\mathrm{d}u = \int_{-1}^{1} u^{\gamma} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u.$$

The result in the Lemma now follows from the following claim for the scalar case:

$$\int_{-1}^{1} u^{\gamma} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u = \begin{cases} 0 & \text{if } 0 \leq \gamma < \lambda \\ (-1)^{\lambda}\lambda! & \text{if } \gamma = \lambda. \end{cases} \tag{B.14}$$

To prove (B.14), we use partial integration γ times, leading to

$$\begin{aligned} \int_{-1}^{1} u^{\gamma} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u &= \sum_{j=0}^{\gamma-1} (-1)^j \frac{\gamma!}{(\gamma-j)!} u^{\gamma-j} \mathcal{K}^{(\lambda-j-1)}(u) \Big|_{-1}^{1} + (-1)^{\gamma}\gamma! \int_{-1}^{1} \mathcal{K}^{(\lambda-\gamma)}(u)\mathrm{d}u \\ &= \sum_{j=0}^{\gamma-1} \left((-1)^j \frac{\gamma!}{(\gamma-j)!} \mathcal{K}^{(\lambda-j-1)}(1) - (-1)^{\gamma} \frac{\gamma!}{(\gamma-j)!} \mathcal{K}^{(\lambda-j-1)}(-1) \right) + (-1)^{\gamma}\gamma! \int_{-1}^{1} \mathcal{K}^{(\lambda-\gamma)}(u)\mathrm{d}u. \end{aligned}$$

By the assumption on the derivative order of the kernel we have that

$$\frac{\partial^{\mu}}{\partial u^{\mu}} \mathcal{K}(1) = \frac{\partial^{\mu}}{\partial u^{\mu}} \mathcal{K}(-1) = 0,$$

for $\mu = 0, \ldots, \lambda - 1$, so that for all $0 \leq \gamma \leq \lambda$,

$$\int_{-1}^{1} u^{\gamma} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u = (-1)^{\gamma}\gamma! \int_{-1}^{1} \mathcal{K}^{(\lambda-\gamma)}(u)\mathrm{d}u.$$

If $0 \leq \gamma < \lambda$, then,

$$\int_{-1}^{1} \frac{\partial^{\lambda-\gamma}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u = \frac{\partial^{\lambda-\gamma-1}}{\partial u^{\lambda-\gamma-1}} \mathcal{K}(u) \Big|_{-1}^{1} = \frac{\partial^{\lambda-\gamma-1}}{\partial u^{\lambda-\gamma-1}} \mathcal{K}(1) - \frac{\partial^{\lambda-\gamma-1}}{\partial u^{\lambda-\gamma-1}} \mathcal{K}(-1) = 0.$$

so that in this case with $0 \leq \gamma < \lambda$,

$$\int_{-1}^{1} u^{\gamma} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u = 0.$$

This proves the first part of the claim in (B.14). If $\gamma = \lambda$ then

$$\int_{-1}^{1} u^{\lambda} \frac{\partial^{\lambda}}{\partial u^{\lambda}} \mathcal{K}(u)\mathrm{d}u = (-1)^{\lambda}\lambda! \int_{-1}^{1} \mathcal{K}^{(\lambda-\lambda)}(u)\mathrm{d}u = (-1)^{\lambda}\lambda! \int_{-1}^{1} \mathcal{K}(u)\mathrm{d}u = (-1)^{\lambda}\lambda!,$$

proving the second part of (B.14). □

Proof of Lemma A.9: The Λ vector of the NIP estimator of the derivatives of h_m up to λ with a polynomial approximation in $2b_N$ from the boundary and using the trimmed NW estimator is for $x \in \mathbb{X}$

$$\begin{aligned} \hat{h}^{[\lambda]}_{m,NIP,s}(x) &= \sum_{k=0}^{s-1} \sum_{|\mu|=k} \frac{1}{\mu!} \frac{1}{N} \sum_{j=1}^{N} I\left(X_j \in \mathbb{X}^I_{b_N}\right) Z_{jm} \frac{1}{{b_N}^{L+[|\lambda|]+|\mu|}} * K^{[\lambda]+(\mu)} \left(\frac{x - X_j}{b_N} \right) (x - r_{2b_N}(x))^{\mu} \\ &= \frac{1}{N} \sum_{j=1}^{N} \left(\sum_{k=0}^{s-1} \sum_{|\mu|=k} \frac{1}{\mu!} I\left(X_j \in \mathbb{X}^I_{b_N}\right) Z_{jm} \frac{1}{{b_N}^{L+[|\lambda|]+|\mu|}} * K^{[\lambda]+(\mu)} \left(\frac{x - X_j}{b_N} \right) (x - r_{2b_N}(x))^{\mu} \right). \end{aligned}$$

Define

$$\phi_{j\mu,N} = \sum_{m=1}^{2} \int_{\mathbb{X}} \frac{1}{\mu!} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_m(x)' \frac{1}{{b_N}^{L+|[\lambda]|+|\mu|}} * K^{[\lambda]+(\mu)}\left(\frac{x-X_j}{b_N}\right)(x - r_{2b_N}(x))^{\mu} \mathrm{d}x$$

so that

$$\sqrt{N} \cdot \sum_{m=1}^{2} \int_{\mathbb{X}} \alpha_m(x)' \left(\hat{h}_{m,NIP,s}^{[\lambda]}(x) - \mathbb{E}\left[\hat{h}_{m,NIP,s}^{[\lambda]}(x)\right]\right) \mathrm{d}x = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \sum_{k=0}^{s-1} \sum_{|\mu|=k} \left(\phi_{j\mu,N} - \mathbb{E}[\phi_{j\mu,N}]\right).$$

We will show that for all μ such that $1 \leq |\mu| \leq s-1$,

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left(\phi_{j\mu,N} - \mathbb{E}[\phi_{j\mu,N}]\right). \tag{B.15}$$

converges faster than the term for $\mu = 0$ so that that term dominates.
We start with the case that $\mu = 0$. We have with a change of variables

$$\phi_{j0,N} = \sum_{m=1}^{2} \int_{\mathbb{X}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_m(x)' \frac{1}{b_N^{L+|[\lambda]|}} * K^{[\lambda]}\left(\frac{x-X_j}{b_N}\right) \mathrm{d}x =$$

$$\sum_{m=1}^{2} \int_{\{u|u=(x-X_j)/b_N, x\in\mathbb{X}\}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_m(X_j + b_N u)' \frac{1}{b_N^{|[\lambda]|}} * K^{[\lambda]}(u) \mathrm{d}u$$

Define $\mathbb{U}(X_j) = \{u \in \mathbb{U} | u = (x - X_j)/b_N, x \in \mathbb{X}\}$. By the definition of $\mathbb{X}_{b_N}^I$ for all $u \in \mathbb{U}$ there is an $x \in \mathbb{X}$ with $x = X_j - b_N u$, so that[6] $\mathbb{U} \subset \mathbb{U}(X_j)$ and hence $\mathbb{U}(X_j) = \mathbb{U}$. This is not true if X_j is in the boundary region and for such X_j we cannot control the variance. Hence

$$\phi_{j0,N} = \sum_{m=1}^{2} \int_{\mathbb{U}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_m(X_j + b_N u)' \frac{1}{b_N^{|[\lambda]|}} * K^{[\lambda]}(u) \mathrm{d}u$$

We can write

$$\phi_{j0,N} = \sum_{m=1}^{2} \sum_{\kappa \leq \lambda} \int_{\mathbb{U}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_{\kappa m}(X_j + b_N u) \frac{1}{b_N^{|\kappa|}} K^{(\kappa)}(u) \mathrm{d}u \equiv \sum_{\kappa \leq \lambda} \phi_{j0,N\kappa}$$

We expand $\alpha_{\kappa m}(X_j + b_N u)$ in a Taylor series of order $|\lambda|$. This gives the vector of expansions

$$\alpha_{\kappa m}(X_j + b_N u) = \sum_{k=0}^{|\kappa|} b_N^k \sum_{|\mu|=k} \frac{1}{\mu!} \alpha_{\kappa m}^{(\mu)}(X_j) u^{\mu} + b_N^{|\kappa|+1} \sum_{|\mu|=|\kappa|+1} \frac{1}{\mu!} \alpha_{\kappa m}^{(\mu)}(X_j + b_N \overline{u}) u^{\mu}$$

Substitution gives

$$\phi_{j\kappa 0,N} = \sum_{m=1}^{2} \sum_{k=0}^{|\kappa|} b_N^{k-|\kappa|} \sum_{|\mu|=k} \frac{1}{\mu!} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_{\kappa m}^{(\mu)}(X_j) \int_{\mathbb{U}} u^{\mu} K^{(\kappa)}(u) \mathrm{d}u +$$

$$b_N \sum_{|\mu|=|\kappa|+1} \frac{1}{\mu!} \int_{\mathbb{U}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) Z_{jm} \alpha_{\kappa m}^{(\mu)}(X_j + b_N \overline{u}) u^{\mu} K^{(\kappa)}(u) \mathrm{d}u$$

The expression involves negative powers of the bandwidth. The contribution of these terms must be 0 to avoid a variance that grows without bounds if the bandwidth becomes small. By the assumptions on the kernel and by Lemma A.8

$$\phi_{j\kappa 0,N} = \sum_{m=1}^{2} (-1)^{|\kappa|} I\left(X_j \in \mathbb{X}_{b_N}^I\right) \alpha_{\kappa m}^{(\kappa)}(X_j) Z_{jm} + b_N \sum_{m=1}^{2} \sum_{|\mu|=|\kappa|+1} \frac{1}{\mu!} \int_{\mathbb{U}} I\left(X_j \in \mathbb{X}_{b_N}^I\right) \alpha_{\kappa m}^{(\mu)}(X_j + b_N \overline{u}) Z_{jm} u^{\mu} K^{(\kappa)}(u) \mathrm{d}u$$

[6] Remember that if $u \in \mathbb{U}$, then $-u \in \mathbb{U}$.

Define

$$\psi_{j\kappa} = \sum_{m=1}^{2} I\left(X_j \in \mathbb{X}_{b_N}^I\right)(-1)^{|\kappa|}\alpha_{\kappa m}^{(\kappa)}(X_j)Z_{jm}$$

We have

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(\phi_{j0,N\kappa} - \psi_{j\kappa} - \mathbb{E}\left[\phi_{j0,N\kappa} - \psi_{j\kappa}\right]\right)\right)^2\right]$$

$$= \mathbb{E}\left[\left(\phi_{j0,N} - \psi_{j\kappa} - \mathbb{E}\left[\phi_{j0,N\kappa} - \psi_{j\kappa}\right]\right)^2\right]$$

$$\leq \mathbb{E}\left[(\phi_{j0,N\kappa} - \psi_{j\kappa})^2\right] = b_N^2\mathbb{E}\left[\left(\sum_{m=1}^{2}\sum_{|\mu|=|\kappa|+1}\frac{1}{\mu!}\int_{\mathbb{U}} I\left(X_j \in \mathbb{X}_{b_N}^I\right)\alpha_{\kappa m}^{(\mu)}(X_j + b_N\overline{u})Z_{jm}u^{\mu}K^{(\kappa)}(u)\mathrm{d}u\right)^2\right]$$

The expectation is bounded by a finite sum of terms

$$\mathbb{E}\left[I\left(X_j \in \mathbb{X}_{b_N}^I\right)\int_{\mathbb{U}}\left|\alpha_{\kappa m}^{(\mu_1)}(X_j + b_N\overline{u})\right||Z_{jm}|\left|u^{\mu_1}K^{(\kappa)}(u)\right|\mathrm{d}u\int_{\mathbb{U}}\left|\alpha_{\kappa m}^{(\mu_2)}(X_j + b_N\overline{u})\right||Z_{jm}|\left|u^{\mu_2}K^{(\kappa)}(u)\right|\mathrm{d}u\right]$$

Because $\alpha_{\kappa m}$ has derivatives up to order $|\lambda|+1$ that are bounded on $\mathbb{X}$ and K has derivatives up to order $|\lambda|$ that are bounded on $\mathbb{U}$, the expectation is obviously bounded if $Z_{jm} = 1$. If $Z_{jm} = Y_j$, the expectation is bounded by $\mathbb{E}(Y_j^2)$ which is bounded by Assumption 1. By the Markov inequality we have

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(\phi_{j0,N\kappa} - \psi_{j\kappa} - \mathbb{E}\left[\phi_{j0,N\kappa} - \psi_{j\kappa}\right]\right) = O_p(b_N)$$

We define

$$\psi_j = \sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2} I\left(X_j \in \mathbb{X}_{b_N}^I\right)\alpha_{\kappa m}^{(\kappa)}(X_j)Z_{jm}$$

Because the result is for all $\kappa \leq \lambda$ we have

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(\phi_{j0,N} - \psi_j - \mathbb{E}\left[\phi_{j0,N} - \psi_j\right]\right) = O_p(b_N)$$

Next we consider $\phi_{j\mu,N}$ with $1 \leq |\mu| \leq s-1$ which can be written as a sum of terms $\phi_{j\kappa\mu,N}$

$$\phi_{j\kappa\mu,N} = \sum_{m=1}^{2}\frac{1}{{b_N}^{L+|\kappa|+|\mu|}}\int_{\mathbb{X}}\frac{1}{\mu!}I\left(X_j \in \mathbb{X}_{b_N}^I\right)\alpha_{\kappa m}(x)Z_{jm}K^{(\kappa+\mu)}\left(\frac{x - X_j}{b_N}\right)(x - r_{2b_N}(x))^{\mu}\mathrm{d}x$$

By a change of variables from x to $u = (x - X_j)/b_N$ with Jacobian b_N^L

$$\phi_{j\kappa\mu,N} = \sum_{m=1}^{2}\frac{1}{{b_N}^{|\kappa|+|\mu|}}\int_{\mathbb{U}}\frac{1}{\mu!}I\left(X_j \in \mathbb{X}_{b_N}^I\right)\alpha_{\kappa m}(X_j + b_N u)Z_{jm}K^{(\kappa+\mu)}(u)\left(X_j + b_N u - r_{2b_N}(X_j + b_N u)\right)^{\mu}\mathrm{d}u$$

Consider for $\mathbb{X} = \prod_{l=1}^{L}[\underline{x}_l, \overline{x}_l]$ and $X_j \in \mathbb{X}_{b_N}^I$

$$(X_j + b_N u - r_{2b_N}(X_j + b_N u))^{\mu} = \prod_{l=1}^{L}(X_{jl} + b_N u_l - r_{2b_N}(X_{jl} + b_N u_l))^{\mu_l}$$

By the definition of $\mathbb{X}_{b_N}^I$ we have that $X_j + b_N u \in \mathbb{X}$ for all $u \in \mathbb{U}$. If for some $l = 1, \ldots L$ $\underline{x}_l + 2b_N \leq X_{jl} + b_N u \leq \overline{x}_l - 2b_N$, then the expression above is 0. It is nonzero only if all components of X_j are in the boundary region. This event that we denote by E_N has a probability proportional to $(2b_N)^L$, because the density of X is bounded on its support. If E_N occurs $\phi_{j\kappa\mu,N}$ involves negative powers of b_N that again must have a 0 contribution. As before the assumptions on the kernel together with a Taylor series expansion of $\beta_{\kappa m}(X_j + b_N u) \equiv \alpha_{\kappa m}(X_j + b_N u)(X_j + b_N u - r_{2b_N}(X_j + b_N u))^{\mu}$ up to order $\kappa+\mu$ ensure this. By Assumptions 2 and 4 $\beta_{\kappa m}$ has bounded derivatives up to $|\kappa+\mu|+1$ so that the Taylor series expansion is valid. If the derivative order of the kernel is at least equal to $\max\{\kappa_1+\mu_1,\ldots,\kappa_L+\mu_L\}$, then by Lemma A.8

$$\phi_{j\kappa\mu,N} = \sum_{m=1}^{2} I_{E_N}(X_j) I\left(X_j \in \mathbb{X}_{b_N}^I\right) \frac{Z_{jm}}{\mu!} \left((-1)^{|\kappa+\mu|}\beta_{\kappa m}^{(\kappa+\mu)}(X_j) + b_N \sum_{|\pi|=|\kappa+\mu|+1} \frac{1}{\pi!}\int_{\mathbb{U}} \beta_{\kappa m}^{(\pi)}(X_j + b_n\overline{u})K(u)\mathrm{d}u\right)$$

Hence because the factor between parentheses is bounded for all $X_j \in \mathbb{X}_{b_N}^I$

$$|\phi_{j\kappa\mu,N}| \leq C I_{E_N}(X_j)\left|\sum_{m=1}^{2} Z_{jm}\right|$$

so that because $\mathbb{E}(Y_j^2|X_j=x)$ is bounded

$$\mathbb{E}\left[\phi_{j\kappa\mu,N}^2\right] \leq C\,\Pr(X_j \in E_N)$$

We conclude that

$$\mathbb{E}\left[\phi_{j\kappa\mu,N}^2\right] = O(b_N^L)$$

so that

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\phi_{j\kappa\mu,N} - \mathbb{E}(\phi_{j\kappa\mu,N})) = O_p(b_N^{L/2})$$

and upon summation over κ

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(\phi_{j\mu,N} - \mathbb{E}(\phi_{j\mu,N})) = O_p(b_N^{L/2})$$

and the result follows. □

Proof of Lemma A.10:
We have using the notation of Lemma A.9:

$$\hat{h}^{[\lambda]}_{m,\text{nip},s}(v(x_1)) = \sum_{j=0}^{s-1} \sum_{|\mu|=j} \frac{1}{\mu!} \frac{1}{N} \sum_{i=1}^{N} Z_{im} \frac{1}{b_N^{L+|[\lambda]|+|\mu|}} * K^{[\lambda]+(\mu)} \left(\frac{x_1 - X_{i1}}{b_N}, \frac{t(x_1) - X_{i2}}{b_N} \right) (v(x_1) - r_{b_N}(v(x_1)))^{\mu}.$$

Define

$$\phi_{i\mu,N} = b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \frac{1}{\mu!} Z_{im} \alpha_m(x_1)' \frac{1}{b_N^{L+|[\lambda]|+|\mu|}} * K^{[\lambda]+(\mu)} \left(\frac{x_1 - X_{i1}}{b_N}, \frac{t(x_1) - X_{i2}}{b_N} \right) (v(x_1) - r_{b_N}(v(x_1)))^{\mu} \mathrm{d}x_1$$

so that

$$\sqrt{N} \cdot b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \alpha_m(x_1)' \left(\hat{h}^{[\lambda]}_{m,NIP,s}(v(x_1)) - \mathbb{E}\left[\hat{h}^{[\lambda]}_{m,NIP,s}(v(x_1)) \right] \right) \mathrm{d}x_1$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{j=0}^{s-1} \sum_{|\mu|=j} \left(\phi_{i\mu,N} - \mathbb{E}[\phi_{i\mu,N}] \right).$$

We consider first the term $\mu = 0$ that makes a contribution to the asymptotic distribution

$$\phi_{i0,N} = b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} Z_{im} \alpha_m(x_1)' \frac{1}{b_N^{L+|[\lambda]|}} * K^{[\lambda]} \left(\frac{x_1 - X_{i1}}{b_N}, \frac{t(x_1) - X_{i2}}{b_N} \right) \mathrm{d}x_1$$

This is the sum of terms

$$\phi_{i\kappa 0,N} = b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \alpha_{\kappa m}(x_1) Z_{im} \frac{1}{b_N^{L+|\kappa|}} K^{(\kappa)} \left(\frac{x_1 - X_{i1}}{b_N}, \frac{t(x_1) - X_{i2}}{b_N} \right) \mathrm{d}x_1$$

By a change of variables from x_1 to $u_1 = (x_1 - X_{i1})/b_N$ with Jacobian $b_N^{L_1}$

$$\phi_{i\kappa 0,N} = \frac{1}{b_N^{L_2/2+|\kappa|}} \sum_{m=1}^{2} \int_{\{u_1 | (x_1 - X_{i1})/b_N, x_1 \in \mathbb{X}_1\}} \alpha_{\kappa m}(X_{i1} + b_N u_1) Z_{im} K^{(\kappa)} \left(u_1, \frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

Using the same argument as before we have for almost all x_1 and $N \geq N_0(x_1)$

$$\phi_{i\kappa 0,N} = \frac{1}{b_N^{L_2/2+|\kappa|}} \sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_{\kappa m}(X_{i1} + b_N u_1) Z_{im} K^{(\kappa)} \left(u_1, \frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

To analyze this term we note that κ is an L vector of nonnegative integers. We partition κ according to x_1 and x_2. First, we consider the case that $t(x_1) = x_{02}$ so that

$$\phi_{i\kappa 0,N} = \frac{1}{b_N^{L_2/2+|\kappa|}} \sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_{\kappa m}(X_{i1} + b_N u_1) Z_{im} K^{(\kappa)} \left(u_1, \frac{x_{02} - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

By a Taylor series expansion of $\alpha_{\kappa m}(X_{i1} + b_N u_1)$ up to order $|\kappa_1|$

$$\alpha_{\kappa m}(X_i + b_N u_1) = \sum_{j=0}^{|\kappa_1|} b_N^j \sum_{|\mu_1|=j} \frac{1}{\mu_1!} \alpha_{\kappa m}^{(\mu_1)}(X_{i1}) u_1^{\mu_1} + b_N^{|\kappa_1|+1} \sum_{|\mu_1|=|\kappa_1|+1} \frac{1}{\mu_1!} \alpha_{\kappa m}^{(\mu_1)}(X_{i1} + b_N \overline{u}_1) u^{\mu_1}$$

so that

$$\phi_{i\kappa 0,N} = \sum_{j=0}^{|\kappa_1|} b_N^{j - L_2/2 - |\kappa|} \sum_{m=1}^{2} \sum_{|\mu_1|=j} \frac{1}{\mu_1!} \alpha_{\kappa m}^{(\mu_1)}(X_{i1}) Z_{im} \int_{\mathbb{U}_1} u_1^{\mu_1} K^{(\kappa)} \left(u_1, \frac{x_{02} - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

$$+ \frac{b_N^{1-|\kappa_2|}}{b_N^{L_2/2}} \sum_{m=1}^{2} \sum_{|\mu_1|=|\kappa_1|+1} \frac{1}{\mu_1!} Z_{im} \int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\mu_1)}(X_{i1} + b_N \overline{u}_1) u^{\mu_1} K^{(\kappa)} \left(u_1, \frac{x_{02} - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

Note that if the kernel is a derivative kernel we have

$$\int_{\mathbb{U}_1} u_1^{\mu_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1 = K^{(\kappa_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right) \int_{\mathbb{U}_1} u_1^{\mu_1} K^{(\kappa_1)}(u_1) \mathrm{d}u_1$$

and the final integral is $(-1)^{|\kappa_1|}\kappa_1!$ if $\mu_1 = \kappa_1$ and 0 otherwise. Hence we define

$$\psi_{i\kappa,N} = \sum_{m=1}^{2} (-1)^{|\kappa_1|} b_N^{-L_2/2-|\kappa_2|} \alpha_{\kappa m}^{(\kappa_1)}(X_{i1}) Z_{im} K^{(\kappa_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right)$$

Now

$$\mathbb{E}\left[\left(\frac{b_N^{|\kappa_2|}}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\kappa 0,N} - \mathbb{E}(\phi_{i\kappa 0,N}) - \psi_{i\kappa,N} - \mathbb{E}(\psi_{i\kappa,N}))\right)^2\right] \leq b_N^{2|\kappa_2|} \mathbb{E}\left[(\phi_{i\kappa 0,N} - \psi_{i\kappa,N})^2\right] =$$

$$b_N^{2-L_2} \mathbb{E}\left[\left(\sum_{m=1}^{2} \sum_{|\mu_1|=|\kappa_1|+1} \frac{1}{\mu_1!} Z_{im} \int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\mu_1)}(X_{i1} + b_N \overline{u}_1) u^{\mu_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right)^2\right]$$

The expectation on the right-hand side is bounded by a finite sum of terms

$$\mathbb{E}\left[\mathbb{E}(Z_{im}^2|X_i)\left|\int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\mu_1)}(X_{i1} + b_N \overline{u}_1) u^{\mu_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right|\left|\int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\tilde{\mu}_1)}(X_{i1} + b_N \overline{u}_1) u^{\tilde{\mu}_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right|\right]$$

and

$$\mathbb{E}\left[\mathbb{E}(|Z_{im}||X_i)\left|\int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\mu_1)}(X_{i1} + b_N \overline{u}_1) u^{\mu_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right|\left|\int_{\mathbb{U}_1} \alpha_{\kappa m}^{(\tilde{\mu}_1)}(X_{i1} + b_N \overline{u}_1) u^{\tilde{\mu}_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right|\right]$$

with $|\mu_1| = |\tilde{\mu}_1| = |\kappa_1| + 1$. Because $\mathbb{E}(Z_{im}^2|X_i = x), \alpha_{\kappa m}^{(\mu_1)}, \alpha_{\kappa m}^{(\tilde{\mu}_1)}$ are bounded functions of x and x_1, the expectation is bounded by

$$C\mathbb{E}\left[\left(\int_{\mathbb{U}_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-X_{i2}}{b_N}\right) \mathrm{d}u_1\right)^2\right] = C\int_{\mathbb{X}_2}\left(\int_{\mathbb{U}_1} K^{(\kappa)}\left(u_1, \frac{x_{02}-x_2}{b_N}\right) \mathrm{d}u_1\right)^2 f_{X_2}(x_2)\mathrm{d}x_2 \leq C b_N^{L_2}$$

by a change of variables to $u_2 = (x_{02} - x_2)/b_N$ with Jacobian $b_N^{L_2}$. We conclude that

$$\mathbb{E}\left[\left(\frac{b_N^{|\kappa_2|}}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\kappa 0,N} - \mathbb{E}(\phi_{i\kappa 0,N}) - \psi_{i\kappa N,} + \mathbb{E}(\psi_{i\kappa N,}))\right)^2\right] = O(b_N^2)$$

so that by the Markov inequality

$$\frac{b_N^{|\kappa_2|}}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\kappa 0,N} - \mathbb{E}(\phi_{i\kappa 0,N}) - \psi_{i\kappa,N} + \mathbb{E}(\psi_{i\kappa,N})) = O_p(b_N)$$

Summing over κ and defining

$$\psi_{iN} = \sum_{\kappa \leq \lambda} (-1)^{|\kappa_1|} \sum_{m=1}^{2} \alpha_{\kappa m}^{(\kappa_1)}(X_i) Z_{im} b_N^{-L_2/2-|\kappa_2|} K^{(\kappa_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right)$$

we obtain

$$\frac{b_N^{|\lambda_2|}}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i0,N} - \mathbb{E}(\phi_{i0,N}) - \psi_{iN} + \mathbb{E}(\psi_{iN})) = O_p(b_N)$$

Note that because of the multiplication by $b_N^{|\lambda_2|}$ ψ_{iN} simplifies because terms with $\kappa_2 < \lambda_2$ are asymptotically negligible. Hence

$$\psi_{iN} = \sum_{\kappa_1 \leq \lambda_1} (-1)^{|\kappa_1|} \sum_{m=1}^{2} \alpha_{\kappa_1 \lambda_2, m}^{(\kappa_1)}(X_{i1}) Z_{im} b_N^{-L_2/2-|\lambda_2|} K^{(\lambda_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right)$$

In the case that $t(x_1)$ is unrestricted, but $\lambda = 0$

$$\phi_{i0,N} = b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} Z_{im}\alpha_m(x_1)' \frac{1}{b_N^L} K\left(\frac{x_1 - X_{i1}}{b_N}, \frac{t(x_1) - X_{i2}}{b_N}\right) \mathrm{d}x_1$$

For almost all x_1 and $N \geq N_0(x_1)$ this is equal to

$$\phi_{i0,N} = \frac{1}{b_N^{L_2/2}} \sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_m(X_{i1} + b_N u_1) Z_{im} K\left(u_1, \frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N}\right) \mathrm{d}u_1$$

Define

$$\psi_{iN} = \frac{1}{b_N^{L_2/2}} \sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_m(X_{i1}) Z_{im} K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N}\right) \mathrm{d}u_1$$

with $\frac{\partial t}{\partial x_1}$ the $L_2 \times L_1$ matrix of partial derivatives of t with respect to x_1. After factorizing K into K_1 and K_2 according to x_1 and x_2, we have

$$\mathbb{E}[(\phi_{i0,N} - \psi_{iN})^2] =$$

$$\frac{1}{b_N^{L_2}} \mathbb{E}\left[\left(\sum_{m=1}^{2} Z_{im} \int_{\mathbb{U}_1} K_1(u_1)\left(\alpha_m(X_{i1} + b_N u_1) K_2\left(\frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N}\right) - \alpha_m(X_{i1}) K_2\left(\frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N}\right)\right) \mathrm{d}u_1\right)^2\right]$$

This is bounded by a sum of terms of the form

$$\frac{1}{b_N^{L_2}} \mathbb{E}\left[\mathbb{E}[|Z_m Z_{m'}||X]\left|\int_{\mathbb{U}_1} K_1(u_1)\left(\alpha_m(X_{i1} + b_N u_1) K_2\left(\frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N}\right) - \alpha_m(X_{i1}) K_2\left(\frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N}\right) \mathrm{d}u_1\right)\right|\right.$$

$$\left.\cdot \left|\int_{\mathbb{U}_1} K_1(u_1)\left(\alpha_{m'}(X_{i1} + b_N u_1) K_2\left(\frac{t(X_{i1} + b_N u_1) - X_{i2}}{b_N}\right) - \alpha_{m'}(X_{i1}) K_2\left(\frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N}\right) \mathrm{d}u_1\right)\right|\right]$$

with $m, m' = 1, 2$. After a change of variables to $u_2 = (t(x_1) - x_2)/b_N$ and for almost all x_1, x_2 and $N \geq N_0(x_1, x_2)$ this is bounded by

$$C \int_{\mathbb{X}_1} \int_{\mathbb{U}_2} \left(\int_{\mathbb{U}_1} \left|\alpha_m(x_1 + b_N u_1) K_2\left(\frac{t(x_1 + b_N u_1) - t(x_1)}{b_N} + u_2\right) - \alpha_m(x_1) K_2\left(\frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)\right| \mathrm{d}u_1\right)$$

$$\cdot \left(\int_{\mathbb{U}_1} \left|\alpha_{m'}(x_1 + b_N u_1) K_2\left(\frac{t(x_1 + b_N u_1) - t(x_1)}{b_N} + u_2\right) - \alpha_{m'}(x_1) K_2\left(\frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)\right| \mathrm{d}u_1\right) f_X(x_1, t(x_1) - b_N u_2) \mathrm{d}u_2 \mathrm{d}x_1$$

Now

$$\alpha_m(x_1 + b_N u_1) = \alpha_m(x_1) + b_N \frac{\partial \alpha_m}{\partial x_1}(x_1 + b_N \overline{u}_1) u_1$$

and for some $0 \leq \xi \leq 1$

$$K_2\left(\frac{t(x_1 + b_N u_1) - x_2}{b_N}\right) = K_2\left(\frac{\partial t}{\partial x_1'}(x_1)u_1 + \frac{t(x_1) - x_2}{b_N} + \frac{b_N}{2} u_1' \frac{\partial^2 t}{\partial x_1 \partial x_1'}(x_1 + b_N \overline{u}_1) u_1\right)$$

$$= K_2\left(\frac{\partial t}{\partial x_1'}(x_1)u_1 + \frac{t(x_1) - x_2}{b_N}\right) + \frac{b_N}{2} \frac{\partial K_2}{\partial x_2'}\left(\xi\left(\frac{\partial t}{\partial x_1}(x_1)u_1 + \frac{t(x_1) - x_2}{b_N}\right) + (1 - \xi)\frac{b_N}{2} u_1' \frac{\partial^2 t}{\partial x_1 \partial x_1'}(X_{i1} + b_N \overline{u}_1) u_1\right)$$

$$\cdot u_1' \frac{\partial^2 t}{\partial x_1 \partial x_1'}(x_1 + b_N \overline{u}_1) u_1$$

By substitution we find, because the derivatives of α_m and K_2 and the second derivatives of t are bounded, that $\mathbb{E}[(\phi_{i0,N} - \psi_{iN})^2]$ is bounded by terms that are all $O(b_N^2)$, so that by the Markov inequality

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i0,N} - \mathbb{E}(\phi_{i0,N}) - \psi_{iN} + \mathbb{E}(\psi_{iN})) = O_p(b_N)$$

Now we return to the case that $t(x_1) = x_{02}$. However if we set $\kappa = 0$ and $x_{02} = t(X_{i1})$ the same proof can be used with minor changes. As in Lemma A.9 we show that for $1 \le |\mu| \le s-1$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\mu,N} - \mathbb{E}[\phi_{i\mu,N}])$$

converges faster than for $|\mu| = 0$. To show this we note that $\phi_{i\mu,N}$ is the sum of terms

$$\phi_{i\kappa\mu,N} = b_N^{L_2/2} \sum_{m=1}^{2} \int_{\mathbb{X}_1} \frac{1}{\mu!} Z_{im} \alpha_{\kappa m}(x_1) \frac{1}{b_N^{L+|\kappa|+|\mu|}} K^{(\kappa+\mu)} \left(\frac{x_1 - X_{i1}}{b_N}, \frac{x_{02} - X_{i2}}{b_N} \right) (v(x_1) - r_{b_N}(v(x_1)))^{\mu} \mathrm{d}x_1$$

After a change of variables to $u_1 = (x_1 - X_{i1})/b_N$ we have for almost all x_1 and $N \ge N_0(x_1)$

$$\phi_{i\kappa\mu,N} = b_N^{-L_2/2} \sum_{m=1}^{2} \frac{1}{\mu!} Z_{im} \int_{\mathbb{U}_1} \alpha_{\kappa m}(X_{i1}+b_N u_1) \frac{1}{b_N^{|\kappa|+|\mu|}} K^{(\kappa+\mu)} \left(u_1, \frac{x_{02} - X_{i2}}{b_N} \right) (v(X_{i1}+b_N u_1) - r_{b_N}(v(X_{i1}+b_N u_1)))^{\mu} \mathrm{d}u_1$$

If we partition the L vector μ according to x_1, x_2, then

$$(v(X_{i1} + b_N u_1) - r_{b_N}(v(X_{i1} + b_N u_1)))^{\mu} = (X_{i1} + b_N u_1 - r_{b_N}(X_{i1} + b_N u_1))^{\mu_1} (x_{02} - r_{b_N}(x_{02}))^{\mu_2}$$

This is 0 unless all components of X_{i1} are in the boundary region. This event that we call E_{1N} has a probability proportional to $b_N^{L_1}$. If this event occurs, then the L_1 components of $r_{b_N}(X_{i1} + b_N u_1)$ are either equal to $\underline{x}_l$ or to $\overline{x}_l$. Define

$$\beta_{\kappa m}(X_{i1} + b_N u_1) = \alpha_{\kappa m}(X_{i1} + b_N u_1)(X_{i1} + b_N u_1 - r_{b_N}(X_{i1} + b_N u_1))^{\mu_1}$$

A Taylor series expansion of $\beta_{\kappa m}(X_{i1} + b_N u_1)$ up to order $\kappa + \mu$ followed by $\kappa + \mu$ times partial integration of the result gives if the kernel has derivative order $\max\{\kappa_1 + \mu_1, \ldots \kappa_L + \mu_L\}$

$$\phi_{i\kappa\mu,N} = I_{E_{1N}}(X_{i1}) b_N^{-L_2/2} \sum_{m=1}^{2} \frac{1}{\mu!} Z_{im} \left((-1)^{|\kappa+\mu|} \beta_{\kappa m}^{(\kappa+\mu)}(X_{i1}) + b_N \sum_{|\pi|=|\kappa+\mu|+1} \int_{\mathbb{U}_1} \beta_{\kappa m}^{(\pi)}(X_{i1} + b_N \overline{u}_1) K^{(\kappa_1+\mu_1)}(u_1) \mathrm{d}u_1 \right)$$

$$\cdot K^{(\kappa_2+\mu_2)} \left(\frac{x_{02} - X_{i2}}{b_N} \right) (x_{02} - r_{b_N}(x_{02}))^{\mu_2}$$

Note that this term is largest if $\mu_2 = 0$. In that case we have with a change in variables to $u_2 = (x_{02} - x_2)/b_N$

$$\mathbb{E}(\phi_{i\kappa\mu,N}^2) \le C \Pr(X_1 \in E_{1N}) = O(b_N^{L_1})$$

so that by the Markov inequality

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\kappa\mu,N} - \mathbb{E}(\phi_{i\kappa\mu,N})) = O_p(b_N^{L_1/2})$$

and upon summation over κ

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\phi_{i\mu,N} - \mathbb{E}(\phi_{i\mu,N})) = O_p(b_N^{L_1/2})$$

and the result follows. □

Proof of Lemma A.11: If $t(x_1)$ is unrestricted and $\lambda = 0$ we define

$$\psi_{iN} = \frac{1}{b_N^{L_2/2}} \sum_{m=1}^{2} \int_{\mathbb{U}_1} \alpha_m(X_{i1}) Z_{im} K \left(u_1, \frac{\partial t}{\partial x_1}(X_{i1}) u_1 + \frac{t(X_{i1}) - X_{i2}}{b_N} \right) \mathrm{d}u_1$$

so that we have to prove

$$N^{-1/2} \cdot \sum_{i=1}^{N} (\psi_{iN} - \mathbb{E}[\psi_{iN}]) \xrightarrow{d} \mathcal{N}(0, V).$$

We first show that the variance of ψ_{iN} is finite. We have

$$\mathbb{V}(\psi_{iN}) \leq \mathbb{E}(\psi_{iN}^2) = \frac{1}{b_N^{L_2}}\mathbb{E}\left[\left(\sum_{m=1}^{2}\int_{\mathbb{U}_1}\alpha_m(X_{i1})Z_{im}K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1})-X_{i2}}{b_N}\right)\mathrm{d}u_1\right)^2\right]$$

This is bounded by a sum of terms of the form

$$\frac{1}{b_N^{L_2}}\int_{\mathbb{X}}\mathbb{E}(|Z_{im}Z_{im'}||X=x)|\alpha_m(x_1)||\alpha_{m'}(x_1)|\left(\int_{\mathbb{U}_1}\left|K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + \frac{t(x_1)-x_2}{b_N}\right)\right|\mathrm{d}u_1\right)^2 f_X(x)\mathrm{d}x$$

By a change of variables to $u_2 = (t(x_1) - x_2)/b_N$ with Jacobian $b_N^{L_2}$ the upper bound is (the integration region for u_2 is a subset of $\mathbb{U}_2$ and $\mathbb{E}(|Z_{im}Z_{im'}||X = x)$ is a bounded function) bounded by

$$\int_{\mathbb{X}_1}\int_{\mathbb{U}_2}|\alpha_m(x_1)||\alpha_{m'}(x_1)|\left(\int_{\mathbb{U}_1}\left|K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)\right|\mathrm{d}u_1\right)^2 f_X(x_1, t(x_1) - b_N u_2)\mathrm{d}u_2\mathrm{d}x_1 < \infty$$

because all functions in the integrand are bounded and so are the sets $\mathbb{U}_1$ and $\mathbb{U}_2$. In the same way we can show that $\mathbb{E}(|\psi_{iN}|) = O\left(b_N^{L_2/2}\right)$, so that

$$\lim_{N\to\infty}\mathbb{V}(\psi_{iN}) = \lim_{N\to\infty}\mathbb{E}(\psi_{iN}^2)$$

and

$$\mathbb{E}(\psi_{iN}^2) = \frac{1}{b_N^{L_2}}\sum_{m=1}^{2}\sum_{m'=1}^{2}\mathbb{E}\left[Z_{im}Z_{im'}\alpha_m(X_{i1})\alpha_{m'}(X_{i1})\left(\int_{\mathbb{U}_1}K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1})-X_{i2}}{b_N}\right)\mathrm{d}u_1\right)^2\right]$$

$$= \sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}_1}\int_{\{u_2|u_2=(t(x_1)-x_2)/b_N, x_2\in\mathbb{X}_2\}}\mathbb{E}(Z_{im}Z_{im'}|X_1 = x_1, X_2 = t(x_1) - b_N u_2)\alpha_m(x_1)\alpha_{m'}(x_1)$$

$$\cdot\left(\int_{\mathbb{U}_1}K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)\mathrm{d}u_1\right)^2 f_X(x_1, t(x_1) - b_N u_2)\mathrm{d}u_2\mathrm{d}x_1$$

If we define

$$\mu_{mm'}(x) = \mathbb{E}(Z_{im}Z_{im'}|X = x)$$

then by dominated convergence

$$\lim_{N\to\infty}\mathbb{E}(\psi_{iN}^2) = \sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}_1}\mu_{mm'}(x_1, t(x_1))\alpha_m(x_1)\alpha_{m'}(x_1)\int_{\mathbb{U}_2}\left(\int_{\mathbb{U}_1}K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)\mathrm{d}u_1\right)^2\mathrm{d}u_2 f_X(x_1, t(x_1))\mathrm{d}x_1 \equiv V$$

The final step is to check the Lyapunov condition, so that we can apply the Lyapunov Central Limit Theorem. The Lyapunov condition for a triangular array of random variables is

$$\frac{N\mathbb{E}\left[|(\psi_{iN} - \mathbb{E}[\psi_{iN}])^{2+\delta}|\right]}{N^{1+\delta/2}\mathbb{V}(\psi_{iN})^{1+\delta/2}} \to 0$$

for some $\delta > 0$. Because $\mathbb{V}(\psi_{iN}) < \infty$, this is equivalent to

$$N^{-\delta/2}\mathbb{E}\left[\left|(\psi_{iN} - \mathbb{E}[\psi_{iN}])^{2+\delta}\right|\right] \to 0$$

We take $\delta = 2$ and note that because $\mathbb{E}[\psi_{iN}] \to 0$ we only need to check

$$N^{-1}\mathbb{E}\left[\psi_{iN}^4\right] \to 0$$

Now

$$\mathbb{E}\left[\psi_{iN}^4\right] = \frac{1}{b_N^{2L_2}}\mathbb{E}\left[\left(\sum_{m=1}^{2}\alpha_m(X_{i1})Z_{im}\right)^4\left(\int_{\mathbb{U}_1}K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1})-X_{i2}}{b_N}\right)\mathrm{d}u_1\right)^4\right]$$

$$\leq \frac{1}{b_N^{2L_2}} \mathbb{E}\left[\left(\sum_{k_m+k_{m'}=4, k_m, k_{m'}\geq 0} (|\alpha_m(X_{i1})||Z_{im}|)^{k_m}(|\alpha_{m'}(X_{i1})||Z_{im'}|)^{k_{m'}}\right)\left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1})-X_{i2}}{b_N}\right)\mathrm{d}u_1\right)^4\right]$$

$$\leq \frac{C}{b_N^{2L_2}} \mathbb{E}\left[\left(\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(X_{i1})u_1 + \frac{t(X_{i1})-X_{i2}}{b_N}\right)\mathrm{d}u_1\right)^4\right]$$

because $\mathrm{E}(Y^4|X=x)$ and $\alpha_m(x_1)$ are bounded functions on $\mathbb{X}$ and $\mathbb{X}_1$, respectively. By the Hölder inequality the right-hand side is bounded by

$$\frac{C}{b_N^{2L_2}} \int_{\mathbb{X}}\int_{\mathbb{U}_1} K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + \frac{t(x_1)-x_2}{b_N}\right)^4 \mathrm{d}u_1 f_X(x)\mathrm{d}x$$

By a change of variables to $u_2 = (t(x_1)-x_2)/b_N$ with Jacobian $b_N^{L_2}$ we obtain the upper bound (we enlarge the integration region for u_2 to $\mathbb{U}_2$)

$$\frac{C}{b_N^{L_2}} \int_{\mathbb{X}_1}\int_{\mathbb{U}} K\left(u_1, \frac{\partial t}{\partial x_1}(x_1)u_1 + u_2\right)^4 f_X(x_1, t(x_1)-b_N u_2)\mathrm{d}u\mathrm{d}x_1 \leq C b_N^{-L_2}$$

because K is a bounded function and $\mathbb{U}_1$ is a bounded set. Hence the Lyapunov condition holds if

$$N^{-1} b_N^{-L_2} \to 0$$

The proof for the case that $\lambda > 0$ and $t(x_1) = x_{02}$ is the same, if we assume that $K^{(\lambda_2)}$ is bounded on $\mathbb{U}$ and $\alpha_m, m=1,2$ is λ_1 times differentiable with a bounded λ_1 derivative. We consider

$$\psi_{iN} = \frac{1}{b_N^{L_2/2}} \sum_{m=1}^{2}\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m}^{(\kappa_1)}(X_{i1})\right) Z_{im} K^{(\lambda_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right)$$

so that

$$\mathbb{E}(\psi_{iN}^2) = \frac{1}{b_N^{L_2}} \sum_{m=1}^{2}\sum_{m'=1}^{2} \mathbb{E}\left[Z_{im}Z_{im'}\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m}^{(\kappa_1)}(X_{i1})\right)\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m'}^{(\kappa_1)}(X_{i1})\right)\left(K^{(\lambda_2)}\left(\frac{x_{02}-X_{i2}}{b_N}\right)\right)^2\right]$$

$$= \sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}_1}\int_{\{u_2|u_2=(t(x_1)-x_2)/b_N, x_2\in\mathbb{X}_2\}} \mathbb{E}(Z_{im}Z_{im'}|X_1=x_1, X_2=t(x_1)-b_N u_2)\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m}^{(\kappa_1)}(x_1)\right)$$

$$\cdot\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m'}^{(\kappa_1)}(x_1)\right)\left(K^{(\lambda_2)}(u_2)\right)^2 f_X(x_1, t(x_1)-b_N u_2)\mathrm{d}u_2\mathrm{d}x_1 \to V \equiv$$

$$\sum_{m=1}^{2}\sum_{m'=1}^{2}\int_{\mathbb{X}_1}\mu_{mm'}(x_1, t(x_1))\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m}^{(\kappa_1)}(x_1)\right)\left(\sum_{\kappa_1\leq\lambda_1} (-1)^{|\kappa_1|}\alpha_{\kappa_1\lambda_2,m'}^{(\kappa_1)}(x_1)\right)\int_{\mathbb{U}_2}\left(K^{(\lambda_2)}(u_2)\right)^2\mathrm{d}u_2 f_X(x_1, t(x_1))\mathrm{d}x_1$$

□

Appendix C: Proofs of Lemmas and Theorems in Main Text

Proof of Lemma 3.1:
The Nadaraya-Watson estimator of h is

$$\hat{h}_{NW}(x) = \frac{1}{Nb^L}\sum_{i=1}^{N} Z_i \cdot K\left(\frac{x-X_i}{b}\right).$$

In the proofs of the Lemmas 3.1 and 3.2 we omit the subscript NW, because it is understood that we deal with the usual kernel estimator. We have,

$$\begin{aligned}
\mathbb{E}\left[\hat{h}(x)\right] &= \mathbb{E}\left[Z\frac{1}{b^L}K\left(\frac{x-X}{b}\right)\right] = \mathbb{E}\left[\mathbb{E}\left[Z\frac{1}{b^L}K\left(\frac{x-X}{b}\right)\bigg| X\right]\right] \\
&= \mathbb{E}\left[g(X)\frac{1}{b^L}K\left(\frac{x-X}{b}\right)\right] \\
&= \int_{\mathbb{X}} g(v)\frac{1}{b^L}K\left(\frac{x-v}{b}\right) f_X(v)dv \\
&= \int_{\mathbb{X}} h(v)\frac{1}{b^L}K\left(\frac{x-v}{b}\right) dv \\
&= \int_{\mathbb{U}} 1\{x-bu\in\mathbb{X}\}\cdot h(x-bu)K(u)\mathrm{d}u
\end{aligned}$$

where the final equality is obtained by a change of variables from v to $u=(x-v)/b$ with Jacobian b^L. By the definition of $\mathbb{X}_b^I$, $x\in\mathbb{X}_b^I$ implies $x-bu\in\mathbb{X}$ for all $u\in\mathbb{U}$, so that, because $K(u)=0$ if $u\neq\mathbb{U}$, the integration region is $\mathbb{U}$ and the indicator function is identically equal to 1, and thus

$$\mathbb{E}\left[\hat{h}(x)\right] = \int_{\mathbb{U}} K(u)h(x-bu)\mathrm{d}u.$$

By Assumption 2 $\mathbb{U}$ is compact and K is continuous on $\mathbb{U}$, and thus $K(u)$ is bounded on $\mathbb{U}$. By Assumption 1 $f_X(x)$ and $g(x)$ and thus $h(x)=g(x)\cdot f_X(x)$ are continuously differentiable up to order $j+s$. Hence by dominated convergence we can interchange integration and differentiation with respect to x repeatedly, so that for $|\lambda|\leq j$

$$\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right] = \int_{\mathbb{U}} K(u)\,h^{(\lambda)}(x-bu)\mathrm{d}u.$$

Now, again using the continuously differentiability of $h(x)$ up to order $j+s$, we can construct a Taylor series expansion of $h^{(\lambda)}(x-bu)$ around x of order $s-1$, to get

$$\begin{aligned}
&\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right] \\
&\quad= \int_{\mathbb{U}} K(u)\sum_{j=0}^{s-1}\sum_{|\theta|=j}\frac{1}{\theta!}h^{(\theta+\lambda)}(x)u^\theta(-b)^{|\theta|}\mathrm{d}u + \int_{\mathbb{U}} K(u)\sum_{|\theta|=s}\frac{1}{\theta!}h^{(\theta+\lambda)}(\tilde{x})u^\theta(-b)^{|\theta|}du \\
&\quad= \int_{\mathbb{U}} K(u)\,h^{(\lambda)}(x)\mathrm{d}u && \text{(C.1)} \\
&\qquad\quad +\sum_{j=1}^{s-1}\sum_{|\theta|=j}(-b)^{|\theta|}\frac{1}{\theta!}h^{(\theta+\lambda)}(x)\int_{\mathbb{U}} u^\theta K(u)\,\mathrm{d}u && \text{(C.2)} \\
&\qquad\quad +(-b)^s\sum_{|\theta|=s}\int_{\mathbb{U}}\frac{1}{\theta!}h^{(\theta+\lambda)}(\tilde{x})u^\theta K(u)\,du. && \text{(C.3)}
\end{aligned}$$

with $\tilde{x}$ intermediate between x and $x-bu$. If $s=1$, the second term (C.2) does not appear. The first term, (C.1), is equal to $h^{(\lambda)}(x)$. All the terms in (C.2) are equal to 0 because the order of the kernel is s and thus $\int u^\theta K(u)du=0$ if $1\leq|\theta|\leq s-1$. The third term, (C.3), is $O(b^s)$ by the fact that by Assumption 1 $h^{(\theta+\lambda)}(x)$

is continuous and hence bounded for $|\theta+\lambda| \leq j+s$ on the compact set $\mathbb{X}$, and $\tilde{x} \in \mathbb{X}$ for all $x \in \mathbb{X}_b^I$ and $u \in \mathbb{U}$. Hence it follows that

$$\sup_{x\in\mathbb{X}_b^I,|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right]-\frac{\partial^{|\lambda|}}{\partial x^\lambda}h(x)\right| = O\left(b^s\right).$$

□

Proof of Lemma 3.2:
For ease of notation we omit the subscript nw on the kernel estimator, and use $\hat{h}(x)$ as shorthand for $\hat{h}_{\text{nw}}(x)$. Without loss of generality we give the proof for the case with scalar Z. The proof is easy if the support of Z is bounded. To deal with the case of an unbounded support we use the assumption on the conditional moments of Z in Assumption 1. First, for some constant P, and for the $p>2$ in Assumption 1(iii) define

$$Z_{iN} = \begin{cases} -P\cdot N^{1/p} & \text{if } Z_i < -P\cdot N^{1/p}, \\ Z_i & \text{if } -P\cdot N^{1/p} \leq Z_i \leq P\cdot N^{1/p}, \\ P\cdot N^{1/p} & \text{if } P\cdot N^{1/p} < Z_i. \end{cases}$$

Also define the kernel estimator with Z_i replaced by Z_{iN},

$$\tilde{h}(x) = \frac{1}{N}\sum_{i=1}^N Z_{iN}\frac{1}{{b_N}^L}K\left(\frac{x-X_i}{b_N}\right). \tag{C.4}$$

By the triangle inequality,

$$\sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right]\right|$$

$$\leq \sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)\right| \tag{C.5}$$

$$+\sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right]-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\tilde{h}(x)\right]\right| \tag{C.6}$$

$$+\sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\tilde{h}(x)\right]\right|. \tag{C.7}$$

We consider the three terms (C.5)-(C.7) separately. First we show that (C.5) is $o_p\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$. Second we show that (C.6) is $o\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$. Third, we show that the dominant term (C.7) is $O_p\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$.
First consider (C.5). By Assumption 1, the definition of Z_{iN}, and the Markov inequality, it follows that for any $\nu>0$, and all $\epsilon>0$

$$\Pr\left(\sup_{x\in\mathbb{X},|\lambda|\leq j} N^\nu\cdot\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)\right|>\epsilon\right) \leq \Pr\left(\sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)\right|>0\right)$$

$$\leq \sum_{|\lambda|\leq j}\Pr\left(\exists i \text{ with } |Z_i|>PN^{1/p}\right)$$

$$\leq C\sum_{i=1}^N\Pr\left(|Z_i|>PN^{1/p}\right)$$

$$= C\cdot N\cdot\Pr\left(|Z_i|>PN^{1/p}\right)$$

$$= C\cdot N\cdot\Pr\left(|Z_i|^p>P^pN\right)$$

$$\leq C\cdot N\cdot\mathbb{E}\left[|Z_i|^p\right]/(NP^p) = C\mathbb{E}\left[|Z_i|^p\right]/P^p.$$

We can make this probability arbitrarily small by choosing P sufficiently large. Hence

$$\sup_{x\in\mathbb{X},|\lambda|\leq j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)\right| = o_p\left(N^{-\nu}\right).$$

By the condition on the lower bound on the bandwidth there is a ν such that $\lim_{N\to\infty} N^{-\nu}(\log(N))^{-1}Nb_N^{L+2j}=0$, and hence (C.5) is $o_p\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$.
Second, consider (C.6) for fixed $x\in\mathbb{X}$ and $\lambda\le j$. By the boundedness of the λ derivative of the kernel on $\mathbb{U}$, and the boundedness of $\mathbb{E}[Z|X=x]$ and $f_X(x)$ (if Z is not identically equal to 1) on $\mathbb{X}$, we can interchange the expectation and repeated differentiation as in the proof of Corollary A.1, so that using $|Z_i-Z_{iN}|\le 2|Z_i|1\{|Z_i|\ge P\cdot N^{1/p}\}$:

$$
\begin{aligned}
\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right]-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\tilde{h}(x)\right]\right| &= \left|\mathbb{E}\left[\frac{\partial^{|\lambda|}}{\partial x^\lambda}\hat{h}(x)-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\tilde{h}(x)\right]\right| \\
&\le 2\left|{b_N}^{-L-|\lambda|}\cdot\mathbb{E}\left[1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|\cdot K^{(\lambda)}\left(\frac{x-X_i}{b_N}\right)\right]\right| \\
&= 2\left|{b_N}^{-L-|\lambda|}\cdot\mathbb{E}\left[\mathbb{E}\left[1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|\middle| X_i\right]\cdot K^{(\lambda)}\left(\frac{x-X_i}{b_N}\right)\right]\right|. \qquad \text{(C.8)}
\end{aligned}
$$

By the Cauchy-Schwartz and Markov inequalities, and using the definition $\mu_2(x)=\mathbb{E}[Z^2|X=x]$ (finite because of Assumption 1(iii)) we have for any $x\in\mathbb{X}$

$$
\begin{aligned}
&\mathbb{E}\left[1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|\middle| X_i=x\right] \\
&= \mathbb{E}\left[1\{|Z_i|>PN^{1/p}\}\cdot\left(1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|\right)\middle| X_i=x\right] \\
&\le \mathbb{E}\left[\left(1\{|Z_i|>PN^{1/p}\}\right)^2\middle| X_i=x\right]^{1/2}\cdot\mathbb{E}\left[\left(1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|\right)^2\middle| X_i=x\right]^{1/2} \\
&= \Pr\left(|Z_i|>PN^{1/p}\middle| X_i=x\right)^{1/2}\cdot\mathbb{E}\left[1\{|Z_i|>PN^{1/p}\}\cdot|Z_i|^2\middle| X_i=x\right]^{1/2} \\
&\le \Pr\left(|Z_i|>PN^{1/p}\middle| X_i=x\right)^{1/2}\cdot\mathbb{E}\left[|Z_i|^2\middle| X_i=x\right]^{1/2} \\
&= \Pr\left(|Z_i|^p>P^pN\middle| X_i=x\right)^{1/2}\cdot\mu_2(x)^{1/2} \\
&\le \left(\mathbb{E}\left[|Z_i|^p\middle| X_i=x\right]/(P^pN)\right)^{1/2}\cdot\mu_2(x)^{1/2}=C\mu_p(x)^{1/2}\mu_2(x)^{1/2}N^{-1/2}. \qquad \text{(C.9)}
\end{aligned}
$$

where the conditional p-th moment of $|Z|$ is denoted by $\mu_p(x)$. Substituting (C.9) into (C.8), we find by a change of variables in the integral

$$
\begin{aligned}
&\sup_{x\in\mathbb{X},|\lambda|\le j}\left|\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\hat{h}(x)\right]-\frac{\partial^{|\lambda|}}{\partial x^\lambda}\mathbb{E}\left[\tilde{h}(x)\right]\right| \\
&\le C\sup_{x\in\mathbb{X},|\lambda|\le j}{b_N}^{-L-|\lambda|}N^{-1/2}\cdot\mathbb{E}\left[\mu_p(X_i)^{1/2}\mu_2(X_i)^{1/2}\cdot\left|K^{(\lambda)}\left(\frac{x-X_i}{b_N}\right)\right|\right] \\
&\le C\sup_{x\in\mathbb{X},|\lambda|\le j}{b_N}^{-L-|\lambda|}N^{-1/2}\int_{\mathbb{X}}\mu_p(v)^{1/2}\mu_2(v)^{1/2}\cdot\left|K^{(\lambda)}\left(\frac{x-v}{b_N}\right)\right|f_X(v)\mathrm{d}v \\
&\le Cb_N^{-j}\cdot N^{-1/2}\sup_{x\in\mathbb{X}}\sum_{|\lambda|\le j}\int_{\mathbb{U}}1\{x-b_Nu\in\mathbb{X}\}\mu_p(x-b_Nu)^{1/2}\mu_2(x-b_Nu)^{1/2}\left|K^{(\lambda)}(u)\right|f_X(x-b_Nu)\mathrm{d}u \\
&\le Cb_N^{-j}\cdot N^{-1/2}\left(\sup_{x\in\mathbb{X}}\mu_p(x)\mu_2(x)\right)^{1/2}\cdot\sup_{x\in\mathbb{X}}f_X(x)\sum_{|\lambda|\le j}\int_{\mathbb{U}}\left|K^{(\lambda)}(u)\right|\mathrm{d}u\le Cb_N^{-j}\cdot N^{-1/2},
\end{aligned}
$$

by the boundedness of the p-th moment of $|Z|$ given $X=x$ and the boundedness of $f_X(x)$ and $K^{(\lambda)}(u)$, and where we use the condition that $b_N\to 0$, so that $b_N^{-\lambda}\le Cb_N^{-j}$. Finally,

$$
Cb_N^{-j}\cdot N^{-1/2}=O\left(b_N^{-j}\cdot N^{-1/2}\right)=o\left(\left(\frac{\log N}{N\cdot b_N^{L+2j}}\right)^{1/2}\right),
$$

because the ratio

$$
\frac{b_N^{-2j}N^{-1}}{\frac{\ln N}{Nb_N^{2j+L}}}=\frac{b_N^L}{\ln N}\to 0,
$$

by the condition on the upper bound on the bandwidth in the Lemma.
Third, consider (C.7). Because $\mathbb{X}$ is compact, it can be covered by CN^L open balls with radius N^{-1}. Index these balls by $d = 1, \ldots, D$ with $D \leq CN^L$, and their centers by $x_{dc}, d = 1, \ldots, D$. For any $x \in \mathbb{X}$ denote the center of its covering ball (or one of its covering balls if there is more than one) by $x_c(x)$. Hence, by the triangle inequality,

$$\sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\tilde{h}^{(\lambda)}(x) - \mathbb{E}[\tilde{h}^{(\lambda)}(x)]\right|$$

$$\leq \sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\tilde{h}^{(\lambda)}(x_c(x)) - \mathbb{E}[\tilde{h}^{(\lambda)}(x_c(x))]\right| \tag{C.10}$$

$$+ \sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\tilde{h}^{(\lambda)}(x) - \tilde{h}^{(\lambda)}(x_c(x))\right| \tag{C.11}$$

$$+ \sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\mathbb{E}[\tilde{h}^{(\lambda)}(x)] - \mathbb{E}[\tilde{h}^{(\lambda)}(x_c(x))]\right|. \tag{C.12}$$

We show that (C.11) and (C.12) are $o\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$, and that (C.10) is $O_p\left((\log(N)N^{-1}b_N^{-L-2j})^{1/2}\right)$.
First consider (C.11). Because we assume that the λ derivative of the kernel K has bounded derivatives with respect to its arguments, we have

$$\sup_{|\lambda|\leq j} |K^{(\lambda)}(u) - K^{(\lambda)}(u')| \leq \sum_{|\lambda|\leq j} C_\lambda \cdot \|u - u'\| \leq C \cdot \|u - u'\|,$$

for all $u, u' \in \mathbb{U}$. Hence we have, because $r \geq j + 1$,

$$\sup_{\|x-x'\|\leq N^{-1},|\lambda|\leq j} |\tilde{h}^{(\lambda)}(x) - \tilde{h}^{(\lambda)}(x')| \tag{C.13}$$

$$\leq \sup_{\|x-x'\|\leq N^{-1},|\lambda|\leq j} \frac{1}{N}\sum_{i=1}^N \frac{|Z_{iN}|}{b_N^{|\lambda|+L}} \cdot \left|K^{(\lambda)}\left(\frac{x - X_i}{b_N}\right) - K^{(\lambda)}\left(\frac{x' - X_i}{b_N}\right)\right|$$

$$\leq C \sup_{\|x-x'\|\leq N^{-1},|\lambda|\leq j} \frac{1}{N}\sum_{i=1}^N \frac{|Z_{iN}|}{b_N^{|\lambda|+L}} \cdot \|x - x'\| \cdot b_N^{-1}$$

$$\leq CN^{1/p-1}b_N^{-j-L-1}.$$

By (C.13) it follows that (C.11) is $O(N^{1/p-1}b_N^{-j-L-1})$. The next step is to show that

$$C \cdot N^{1/p-1}b_N^{-j-L-1} = o\left(\left(\frac{\ln N}{Nb_N^{2j+L}}\right)^{1/2}\right).$$

To see this, consider the ratio

$$\frac{N^{1/p-1}b_N^{-j-L-1}}{\left(\ln N\Big/\left(Nb_N^{2j+L}\right)\right)^{1/2}} = \left(\frac{N^{2/p-2}b_N^{-2j-2L-2}}{\ln N\Big/\left(Nb_N^{2j+L}\right)}\right)^{1/2} = \left(\frac{N^{2/p-1}b_N^{-L-2}}{\ln N}\right)^{1/2}$$

$$\leq \left(N^{2/p-1}b_N^{-L-2}\right)^{1/2} \to 0,$$

by the condition on the lower bound on the bandwidth in the statement of the Lemma. Thus we have for (C.11):

$$\sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\tilde{h}^{(\lambda)}(x) - \mathbb{E}[\tilde{h}^{(\lambda)}(x)]\right| = o\left(\left(\frac{\ln N}{Nb_N^{2j+L}}\right)^{1/2}\right).$$

Next, consider (C.12). By (C.13) it follows that

$$\sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\mathbb{E}[\tilde{h}^{(\lambda)}(x)] - \mathbb{E}[\tilde{h}^{(\lambda)}(x_c(x))]\right| \leq \mathbb{E}\left[\sup_{x\in\mathbb{X},|\lambda|\leq j} \left|\tilde{h}^{(\lambda)}(x) - \tilde{h}^{(\lambda)}(x_c(x))\right|\right] = O\left(N^{1/p-1}b_N^{-j-L-1}\right) = o\left(\left(\frac{\ln N}{Nb_N^{2j+L}}\right)^{1/2}\right).$$

Finally, consider (C.10):

$$\sup_{x\in\mathbb{X},|\lambda|\le j}\left|\tilde{h}^{(\lambda)}(x_c(x))-\mathbb{E}\left[\tilde{h}^{(\lambda)}(x_c(x))\right]\right|=\max_{d=1,\dots,D,|\lambda|\le j}\left|\tilde{h}^{(\lambda)}(x_{dc})-\mathbb{E}\left[\tilde{h}^{(\lambda)}(x_{dc})\right]\right|.$$

Define $\eta_N=(b_N^{-(L+2j)}N^{-1}\ln N)^{1/2}$. We have $\eta_N^2=b_N^{-(L+2j)}N^{-1}\ln N<b_N^{-(L+2j)}N^{2/p-1}\ln N\to 0$ by the bandwidth conditions in the Lemma so that $\eta_N\to 0$. Let $M>0$ be a constant. By deriving an upper bound on

$$\Pr\left(\sup_{d=1,\dots,D,|\lambda|\le j}\left|\tilde{h}^{(\lambda)}(x_{dc})-\mathbb{E}[\tilde{h}^{(\lambda)}(x_{dc})]\right|>M\eta_N\right), \tag{C.14}$$

that for N sufficiently large goes to 0 as $M\to\infty$, we establish that (C.10) is $O_p(\eta_N)$ and thus by the definition of η_N (C.10) is $O_p((b_N^{-(L+2j)}N^{-1}\ln N)^{1/2})$. We have

$$\Pr\left(\sup_{d=1,\dots,D,|\lambda|\le j}\left|\tilde{h}^{(\lambda)}(x_{dc})-\mathbb{E}[\tilde{h}^{(\lambda)}(x_{dc})]\right|>M\eta_N\right)\le\sum_{|\lambda|\le j}\sum_{d=1}^{D}\Pr\left(\left|\tilde{h}^{(\lambda)}(x_{dc})-\mathbb{E}[\tilde{h}^{(\lambda)}(x_{dc})]\right|>M\eta_N\right).$$

Because $K^{(\lambda)}(u)$ is bounded (with upper bound $\overline{K}_\lambda$), the kernel estimator

$$\tilde{h}^{(\lambda)}(x_{dc})=\frac{1}{N}\sum_{i=1}^{N}\frac{Z_{iN}}{b_N^{L+|\lambda|}}\cdot K^{(\lambda)}\left(\frac{x_{dc}-X_i}{b_N}\right),$$

is an average of bounded random variables (with upper bound $\overline{K}_\lambda\cdot P\cdot N^{1/p}b_N^{-|\lambda|-L}$). Hence by Bernstein's inequality (see e.g. Serfling (1980), p. 95)[7]

$$\Pr\left(\left|\tilde{h}^{(\lambda)}(x_{dc})-\mathbb{E}[\tilde{h}^{(\lambda)}(x_{dc})]\right|>M\eta_N\right)\le 2\exp\left\{-\frac{NM^2\eta_N^2}{2\mathbb{V}\left(\frac{Z_{iN}}{b_N^{|\lambda|+L}}K^{(\lambda)}\left(\frac{x_{dc}-X}{b_N}\right)\right)+\frac{2}{3}\frac{\overline{K}_\lambda PN^{1/p}}{b_N^{|\lambda|+L}}M\eta_N}\right\}.$$

Now, by a change of variables from x to $u=(x_{dc}-x)/b_N$ with Jacobian b_N^L,

$$\begin{aligned}&\mathbb{V}\left(\frac{Z_{iN}}{b_N^{|\lambda|+L}}K^{(\lambda)}\left(\frac{x_{dc}-X}{b_N}\right)\right)\\&\quad\le\frac{1}{b_N^{2|\lambda|+2L}}\mathbb{E}\left[\left(Z_{iN}\cdot K^{(\lambda)}\left(\frac{x_{dc}-X}{b_N}\right)\right)^2\right]\\&\quad\le\frac{1}{b_N^{2|\lambda|+L}}\int_{\mathbb{U}}1\{x_{dc}-b_Nu\in\mathbb{X}\}\mu_2(x_{dc}-b_Nu)K^{(\lambda)}(u)^2f_X(x_{dc}-b_Nu)\mathrm{d}u\\&\quad\le Cb_N^{-2|\lambda|-L},\end{aligned}$$

The last inequality follows because f_X and $\mu_2(x)$ are bounded on $\mathbb{X}$ and so is $K^{(\lambda)}(u)^2$ on $\mathbb{U}$. Hence, using the fact that $D\le CN^L$,

$$\Pr\left(\sup_{d=1,\dots,D,|\lambda|\le j}\left|\hat{h}^{(\lambda)}(x_{dc})-\mathbb{E}[\hat{h}^{(\lambda)}(x_{dc})]\right|>M\eta_N\right)$$

[7] For independent random variables $Z_1,\dots,Z_N$ with variance $\mathbb{V}(Z_i)$ that satisfy $\Pr(|Z_i-\mathbb{E}(Z_i)|\le U)=1$ for all i with $U<\infty$, Bernstein's inequality states that for $z>0$

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i-\mathbb{E}(Z_i))\right|\ge z\right)\le 2\exp\left\{-\frac{N^2z^2}{2\sum_{i=1}^{N}\mathbb{V}(Z_i)+\frac{2}{3}UNz}\right\}.$$

Hence with identically distributed Z_i we have

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i-\mathbb{E}(Z_i))\right|\ge z\right)\le 2\exp\left\{-\frac{Nz^2}{2\mathbb{V}(Z_i)+\frac{2}{3}Uz}\right\}.$$

.

$$\leq C_1 N^L \sum_{|\lambda|\leq j} \exp\left\{-\frac{NM^2\eta_N^2}{2C{b_N}^{-2|\lambda|-L} + \frac{2}{3}\frac{\overline{K}_\lambda P N^{1/p}}{b_N^{|\lambda|+L}} M\eta_N}\right\}$$

$$\leq C_1 N^L \sum_{|\lambda|\leq j} \exp\left\{-\frac{NM^2\eta_N^2 b_N^{2|\lambda|+L}}{C_2 + C_3 b_N^{|\lambda|} N^{1/p} M\eta_N}\right\}$$

$$\leq C_1 \exp\left\{L\log N - \frac{M}{C_2}\frac{N\eta_N^2 b_N^{2j+L}}{1/M + \frac{C_3}{C_2} b_N^j N^{1/p}\eta_N}\right\} \tag{C.15}$$

because for $|\lambda| \leq j$

$$\frac{N\eta_N^2 b_N^{2|\lambda|+L}}{1/M + \frac{C_3}{C_2} b_N^{|\lambda|} M N^{1/p}\eta_N} \geq \frac{N\eta_N^2 b_N^{2j+L}}{1/M + \frac{C_3}{C_2} b_N^{j} M N^{1/p}\eta_N}.$$

The bound in (C.15) will go to zero as M increases for all $N \geq N_0$ for some N_0, if for $N \geq N_0$ and some $C_4 > 0$

$$\frac{N\eta_N^2 b_N^{2j+L}}{1 + \frac{C_3}{C_2} N^{1/p} b_N^j M\eta_N} \geq C_4 \log N. \tag{C.16}$$

First, the numerator in (C.16) is by the definition of η_N equal to

$$N\eta_N^2 b_N^{2j+L} = N\ln N \cdot N^{-1} b_N^{-2j-L} b_N^{2j+L} = \ln N.$$

Because

$$\lim_{N\to\infty} N^{1/p} b_N^j \eta_N = \lim_{N\to\infty} N^{1/p} b_N^j \left(\ln N \cdot N^{-1} b_N^{-2j-L}\right)^{1/2} = \lim_{N\to\infty} \left(\ln N \cdot N^{-1+2/p} b_N^{-L}\right)^{1/2} = 0,$$

by the bandwidth condition in the Lemma, it follows that for $C_4 > 0$ we can find an N_0 and M_0 such that for $N \geq N_0$, and $M \geq M_0$, the denominator in (C.16) is smaller than $1/C_4$ and therefore (C.16) holds. Hence (C.14) goes to zero as $M \to \infty$, and thus (C.10) is $O_p((b_N^{-(L+2j)} N^{-1} \ln N)^{1/2})$. □

Proof of Lemma 3.3: We have for all $x \in \mathbb{X}$,

$$\begin{aligned}
&\left|\mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right] - h^{(\lambda)}(x)\right| \\
&\quad = \left|\mathbb{E}\left[t\left(x, \hat{h}^{(\lambda)}_{NW}, r_b(x), s\right)\right] - h^{(\lambda)}(x)\right| \\
&\quad \leq \left|\mathbb{E}\left[t\left(x, \hat{h}^{(\lambda)}_{NW}, r_b(x), s\right)\right] - t\left(x, h^{(\lambda)}, r_b(x), s\right)\right| &\text{(C.17)}\\
&\qquad\quad + \left|t\left(x, h^{(\lambda)}, r_b(x), s\right) - h^{(\lambda)}(x)\right|, &\text{(C.18)}
\end{aligned}$$

by the triangle inequality. We show that the supremum over $\mathbb{X}$ of (C.17) and (C.18) are both $O(b^s)$. First consider (C.17).

$$\begin{aligned}
&\left|\mathbb{E}\left[t\left(x, \hat{h}^{(\lambda)}_{NW}, r_b(x), s\right)\right] - t\left(x, h^{(\lambda)}, r_b(x), s\right)\right| \\
&\quad = \left|\mathbb{E}\left[\sum_{j=0}^{s-1}\sum_{|\theta|=j}\frac{1}{\theta!}\hat{h}^{(\lambda+\theta)}_{NW}(r_b(x))(x-r_b(x))^\theta\right] - \sum_{j=0}^{s-1}\sum_{|\theta|=j}\frac{1}{\theta!}h^{(\lambda+\theta)}(r_b(x))(x-r_b(x))^\theta\right| \\
&\quad = \left|\sum_{j=0}^{s-1}\sum_{|\theta|=j}\frac{1}{\theta!}\left(\mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{NW}(r_b(x))\right] - h^{(\lambda+\theta)}(r_b(x))\right)(x-r_b(x))^\theta\right| \\
&\quad \leq C \cdot \max_{0\leq|\theta|\leq s-1}\sup_{x\in\mathbb{X}}\left|\mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{NW}(r_b(x))\right] - h^{(\lambda+\theta)}(r_b(x))\right| \\
&\quad = C \cdot \max_{0\leq|\theta|\leq s-1}\sup_{x\in\mathbb{X}_b^I}\left|\mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{NW}(x)\right] - h^{(\lambda+\theta)}(x)\right| = O\left(b^s\right),
\end{aligned}$$

by Lemmas 3.1 and A.1 because $q \geq |\lambda| + 2s - 1$ and $r \geq |\lambda| + s - 1$, which is implied by $q \geq j + 2s - 1$ and $r \geq j + s - 1$.

Now consider (C.18). This is a remainder term of a Taylor series expansion, so we have for all $x \in \mathbb{X}$ and some $\tilde{x}$ intermediate between x and $r_b(x)$

$$\begin{aligned}
&\left|t\left(x, h^{(\lambda)}, r_b(x), s\right) - h^{(\lambda)}(x)\right| \\
&\quad = \left|\sum_{|\theta|=s} h^{(\lambda+\theta)}(\tilde{x})(x - r_b(x))^{\theta}\right| \\
&\quad \leq \sum_{|\theta|=s} \sup_{x\in\mathbb{X}} \left|h^{(\lambda+\theta)}(x)\right| \cdot \left(\sup_{x\in\mathbb{X}} \inf_{y\in\mathbb{X}_b^I} \|x-y\|\right)^s = O(b^s),
\end{aligned}$$

where $\sup_{x\in\mathbb{X}} \inf_{y\in\mathbb{X}_b^I} \|x-y\| = \sup_{x\in\mathbb{X}} \|x - r_b(x)\| = O(b)$ by Assumption 1 and $\sup_{x\in\mathbb{X}} \left|h^{(\lambda+\theta)}(x)\right| < \infty$ because by Assumption 1 $h^{(\lambda+\theta)}(\cdot)$ with $|\theta| = s$ is bounded on $\mathbb{X}$. Hence the supremum over $\mathbb{X}$ of both (C.17) and (C.18) are $O(b^s)$, which proves the result. □

Proof of Lemma 3.4: Without loss of generality we consider the case with scalar h. We have

$$\begin{aligned}
&\sup_{x\in\mathbb{X}} \left|\hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right| \\
&\quad \leq \sup_{x\in\mathbb{X}^I_{b_N}} \left|\hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right| && \text{(C.19)} \\
&\qquad + \sup_{x\in\mathbb{X}^B_{b_N}} \left|\hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right| && \text{(C.20)}
\end{aligned}$$

For (C.19) we have, because in the internal area the estimator $\hat{h}^{(\lambda)}_{\text{nip},s}(x)$ is identical to $\hat{h}^{(\lambda)}_{\text{nw}}(x)$,

$$\begin{aligned}
&\sup_{x\in\mathbb{X}^I_{b_N}} \left|\hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right| = \sup_{x\in\mathbb{X}^I_b} \left|\hat{h}^{(\lambda)}_{\text{nw}}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nw}}(x)\right]\right| \\
&\quad \leq \sup_{x\in\mathbb{X}} \left|\hat{h}^{(\lambda)}_{\text{nw}}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nw}}(x)\right]\right| = O_p\left(\left(\frac{\log N}{N b_N^{L+2\lambda}}\right)^{\frac{1}{2}}\right),
\end{aligned}$$

by Lemmas 3.2 and A.2. The second part, (C.20), corresponds to values in the boundary area. For $x \in \mathbb{X}^B_{b_N}$

$$\begin{aligned}
&\left|\hat{h}^{(\lambda)}_{\text{nip},s}(x) - \mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right| = \left|t\left(x, \hat{h}^{(\lambda)}_{\text{nw}}, r_b(x), s\right) - \mathbb{E}\left[t\left(x, \hat{h}^{(\lambda)}_{\text{nw}}, r_b(x), s\right)\right]\right| \\
&\quad = \left|\sum_{j=0}^{s-1} \sum_{|\theta|=j} \frac{1}{\theta!}\left(\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x)) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))\right]\right)(x - r_{b_N}(x))^{\theta}\right| \\
&\quad \leq \sum_{j=0}^{s-1} \sum_{|\theta|=j} \frac{1}{\theta!}\left|\left(\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x)) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))\right]\right)\right| \cdot \|x - r_{b_N}(x)\|^{|\theta|} \\
&\quad \leq \sum_{j=0}^{s-1} \sum_{|\theta|=j} \frac{1}{\theta!} \sup_{x\in\mathbb{X}}\left|\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x)) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))\right]\right| \cdot \sup_{x\in\mathbb{X}} \|x - r_{b_N}(x)\|^{|\theta|}.
\end{aligned}$$

By Lemmas 3.2 and A.2

$$\begin{aligned}
&\sup_{x\in\mathbb{X}} \left|\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x)) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))\right]\right| \\
&\quad \leq \sup_{x\in\mathbb{X}^I_{b_N}} \left|\hat{h}^{(\lambda+\theta)}_{\text{nw}}(x) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(x)\right]\right| \\
&\quad \leq \sup_{x\in\mathbb{X}} \left|\hat{h}^{(\lambda+\theta)}_{\text{nw}}(x) - \mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(x))\right]\right| = O_p\left(\left(\frac{\log N}{N b_N^{L+2|\theta|+2|\lambda|}}\right)^{\frac{1}{2}}\right).
\end{aligned}$$

Using the fact that $\sup_{x\in\mathbb{X}}\|x-r_{b_N}(x)\|\le Cb_N$

$$\sup_{x\in\mathbb{X}}\left|\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))-\mathbb{E}\left[\hat{h}^{(\lambda+\theta)}_{\text{nw}}(r_{b_N}(x))\right]\right|\cdot\left(\sup_{x\in\mathbb{X}}\|x-r_{b_N}(x)\|\right)^{|\theta|}$$
$$=O_p\left(\left(\frac{\log N}{Nb_N^{L+2|\theta|+2|\lambda|}}\right)^{\frac{1}{2}}b_N^{|\theta|}\right)=O_p\left(\left(\frac{\log N}{Nb_N^{L+2|\lambda|}}\right)^{\frac{1}{2}}\right).$$

We conclude that

$$\sup_{x\in\mathbb{X}}\left|\hat{h}^{(\lambda)}_{\text{nip},s}(x)-\mathbb{E}\left[\hat{h}^{(\lambda)}_{\text{nip},s}(x)\right]\right|=O_p\left(\left(\frac{\log N}{Nb_N^{L+2|\lambda|}}\right)^{\frac{1}{2}}\right)=O_p\left(\left(\frac{\log N}{Nb_N^{L+2j}}\right)^{\frac{1}{2}}\right),$$

because $|\lambda|\le j$. □

Proof of Theorem 3.1:
We can apply Lemmas 3.3, 3.4, A.3 and Theorem A.1 to each row of $[\lambda]$ because for a row μ of $[\lambda]$ we have $|\mu|\le|\lambda|$, so that we can apply the results with $j=|\lambda|$ and $\lambda=\mu$. □

Proof of Theorem ??: First we prove part (i). Consider

$$|\hat{\theta}-\theta|\le\frac{1}{N}\sum_{i=1}^{N}|\omega(X_i)|\left|m(\hat{h}^{[\lambda]}_{NIP,s}(X_i))-m(h^{[\lambda]}(X_i))\right|+\left|\frac{1}{N}\sum_{i=1}^{N}\omega(X_i)m(h^{[\lambda]}(X_i))-\mathbb{E}\left[\omega(X)'\left(h^{[\lambda]}(X)\right)\right]\right|$$
$$\le\frac{1}{N}\sum_{i=1}^{N}|\omega(X_i)|\sup_{x\in\mathbb{X}}\left|m(\hat{h}^{[\lambda]}_{NIP,s}(x))-m(h^{[\lambda]}(x))\right|+\left|\frac{1}{N}\sum_{i=1}^{N}\omega(X_i)m(h^{[\lambda]}(X_i))-\mathbb{E}\left[\omega(X)'m\left(h^{[\lambda]}(X)\right)\right]\right|$$

Hence for all $\eta>0$ a necessary condition for $|\hat{\theta}-\theta|\ge\eta$ is that either

$$\frac{1}{N}\sum_{i=1}^{N}|\omega(X_i)|\sup_{x\in\mathbb{X}}\left|m(\hat{h}^{[\lambda]}_{NIP,s}(x))-m(h^{[\lambda]}(x))\right|\ge\frac{\eta}{2}$$

or

$$\left|\frac{1}{N}\sum_{i=1}^{N}\omega(X_i)m(h^{[\lambda]}(X_i))-\mathbb{E}\left[\omega(X)'m\left(h^{[\lambda]}(X)\right)\right]\right|\ge\frac{\eta}{2}$$

The probability of the latter event goes to 0 if $N\to\infty$ by a law of large numbers. By the law of total probability the probability of the first event is bounded by the sum

$$\text{pr}\left(\frac{1}{N}\sum_{i=1}^{N}|\omega(X_i)|>\mathbb{E}(|\omega(X)|)+\eta\right)+\text{pr}\left(\sup_{x\in\mathbb{X}}\left|m(\hat{h}^{[\lambda]}_{NIP,s}(x))-m(h^{[\lambda]}(x))\right|\ge\frac{\eta}{2|\mathbb{E}(|\omega(X)|)+\eta|}\right)$$

The first probability converges to 0 if $N\to\infty$ by a law of large numbers. Because $m(\cdot)$ is continuous on the bounded set $\mathbb{H}_\lambda$ and hence uniformly continuous on that set, and because the conditions for uniform convergence of the NIP estimator are met, the probability of the second event converges to 0. This finishes the proof for the consistency claim in part (i).
Next, we prove part (ii). Adding and subtracting some terms we can write:

$$\sqrt{N}(\hat{\theta}-\theta)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega(X_i)m(h_0^{[\lambda]}(X_i))-\mathbb{E}\left[\omega(X)m\left(h^{[\lambda]}(X)\right)\right]\right)$$
$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\le\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\left(\alpha^{(\kappa)}_{\kappa m}(X_i)Z_{im}-\mathbb{E}[\alpha^{(\kappa)}_{\kappa m}(X)Z_m]\right)\right)$$
$$=\sqrt{N}(\hat{\theta}-\theta)-\omega(X_{i1})\frac{\partial}{\partial h^{[\lambda]'}}m(h^{[\lambda]}(X_{i1},t(X_{i1}))\left(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))-h^{[\lambda]}(X_{i1},t(X_{i1}))\right)\Bigg\}\quad\text{(C.21)}$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega(X_{i1})\frac{\partial}{\partial h^{[\lambda]'}}m(h^{[\lambda]}(X_{i1},t(X_{i1}))\left(h^{[\lambda]}(X_{i1},t(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))\right]\right)\Bigg\}\quad\text{(C.22)}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nu(X_{i1})'\left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))\right]\right) \tag{C.23}$$

$$-\sqrt{N}\left(\int_{\mathbb{X}_1}\nu(x_1)'\left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))-\mathbb{E}[\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))]\right)f_{X_1}(x_1)\mathrm{d}x_1\right) \tag{C.24}$$

$$\sqrt{N}\sum_{m=1}^{2}\int_{\mathbb{X}}\alpha_m(x)'\left(\hat{h}^{[\lambda]}_{m,\text{nip},s}(x)-\mathbb{E}\left[\hat{h}^{[\lambda]}_{m,\text{nip},s}(x)\right]\right)\mathrm{d}x \tag{C.25}$$

$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\sum_{\kappa\leq\lambda}(-1)^{|\kappa|}\sum_{m=1}^{2}\left(\alpha^{(\kappa)}_{\kappa m}(X_i)Z_{im}-\mathbb{E}[\alpha^{(\kappa)}_{\kappa m}(X)Z_m]\right)\right), \tag{C.26}$$

using the previous definitions of $\nu=(\nu_1'\ \nu_2')'$, where

$$\nu_m(x_1)=\omega(x_1)\frac{\partial}{\partial h^{[\lambda]}_m}m(h^{[\lambda]}(x_1,t(x_1)),$$

and

$$\alpha_1(x_1)=f_{X_1}(x_1)\nu_1(x_1),\qquad \text{and}\quad \alpha_2(x_1)=f_{X_1}(x_1)\nu_2(x_1).$$

The proof consists of showing that (C.21), (C.22), (C.23) plus (C.24) and (C.25) plus (C.26) are all $o_p(1)$ by checking the relevant conditions for Lemmas A.4-A.9 and Theorem A.1.
First consider (C.21). The conditions in the theorem imply that $q\geq|\lambda|+2s-1$, $r\geq|\lambda|+s-1$, $\delta<1/(L+2|\lambda|)$. Hence, by Lemma A.4 it follows that

$$\sqrt{N}(\hat{\theta}-\theta)-\omega(X_{i1})\frac{\partial}{\partial h^{[\lambda]'}}m(h^{[\lambda]}(X_{i1},t(X_{i1}))\left(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))-h^{[\lambda]}(X_{i1},t(X_{i1}))\right)\Big\}$$

$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{\omega(X_{i1})\left(m(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))-m(h^{[\lambda]}(X_{i1},t(X_{i1}))\right)$$

$$-\omega(X_{i1})\frac{\partial}{\partial h^{[\lambda]'}}m(h^{[\lambda]}(X_{i1},t(X_{i1}))\left(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))-h^{[\lambda]}(X_{i1},t(X_{i1}))\right)\Big\}$$

$$=O_p\left(\sqrt{N}\left|\hat{h}^{[\lambda]}_{NIP,s}-h^{[\lambda]}\right|_0^2\right). \tag{C.27}$$

Moreover, because $q\geq|\lambda|+2s-1$, $r\geq|\lambda|+s-1+L$, it follows by Theorem A.1 that

$$\sup_{x\in\mathbb{X}}\left|\hat{h}^{(\lambda)}_{\text{nip},s}(x)-h^{(\lambda)}(x)\right|=O_p\left(\left(\frac{\log N}{N\cdot b_N^{L+2|\lambda|}}\right)^{1/2}+b_N^s\right).$$

Thus

$$=O_p\left(\sqrt{N}\left|\hat{h}^{[\lambda]}_{NIP,s}-h^{[\lambda]}\right|_0^2\right)=O_p\left(\sqrt{N}\left(\frac{\log N}{N\cdot b_N^{L+2|\lambda|}}+b_N^{2s}\right)\right)$$

$$=O_p\left(\log(N)\cdot N^{-1/2+\delta L+2\delta|\lambda|}+N^{1/2-2s\delta}\right).$$

By assumption $\delta<1/(2L+4|\lambda|)$, so $-1/2+\delta L+2\delta|\lambda|<0$ and thus $\log(N)\cdot N^{-1/2+\delta L+2\delta|\lambda|}\to 0$. In addition, by assumption $\delta>1/(2s)$, so $\delta>1/(4s)$, and thus $N^{1/2-2s\delta}\to 0$. Thus

$$O_p\left(\log(N)\cdot N^{-1/2+\delta L+2\delta|\lambda|}+N^{1/2-2s\delta}\right)=o_p(1),$$

and thus (C.27), and thus (C.21) are $o_p(1)$.
Next, consider (C.22). By assumption, $q\geq|\lambda|+2s-1$, $r\geq|\lambda+s-1$, and $t\geq 1$, so the conditions for Lemma A.5 are satisfied. This implies that

$$\frac{1}{N}\sum_{i=1}^{N}\omega(X_{i1})\frac{\partial}{\partial h^{[\lambda]'}}m(h^{[\lambda]}(X_{i1},t(X_{i1}))\left(h^{[\lambda]}(X_{i1},t(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1}))\right]\right)=O_p(b^s),$$

and thus that (C.22) is $O_p(N^{1/2}b^s)$. By assumption $\delta > 1/(2s)$, and thus (C.22) is $o_p(1)$.
Next, consider the sum of (C.23) and (C.24). By Assumption $r \geq |\lambda|$, and so by Lemma A.7 it follows that the sum of (C.23) and (C.24) is

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nu(X_{i1})'\left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))-\mathbb{E}\left[\hat{h}^{[\lambda]}_{\text{nip},s}(v(X_{i1}))\right]\right)$$

$$-\sqrt{N}\left(\int_{\mathbb{X}_1}\nu(x_1)'\left(\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))-\mathbb{E}[\hat{h}^{[\lambda]}_{\text{nip},s}(v(x_1))]\right)f_{X_1}(x_1)\mathrm{d}x_1\right)$$

$$=O_p\left(N^{-1}b_N^{-L/2-L_1/2-|\lambda|}+N^{-1/2}b_N^{-L/2-|\lambda|}\right).$$

By Assumption 3, this is

$$=O_p\left(N^{-1+\delta L/2+\delta L_1/2+\delta|\lambda|}+N^{-1/2+\delta L/2+\delta|\lambda|}\right).$$

By Assumption, $\delta < (1/(2L+4|\lambda)$, so $\delta < 1/(L+|\lambda|) \leq 1/(L/2+L_1/2+|\lambda|)$, and therefore $O_p\left(N^{-1+\delta L/2+\delta L_1/2+\delta|\lambda|}\right) = o_p(1)$. Also, $\delta < (1/(2L+4|\lambda)$, so $\delta < 1/(L+2|\lambda|)$, and therefore $O_p\left(N^{-1/2+\delta L/2+\delta|\lambda|}\right) = o_p(1)$, and thus the sum of (C.23) and (C.24) is $o_p(1)$.
Next, consider the sum of (C.25) and (C.26). By assumption, $r \geq |\lambda|+s-1$, $q \geq 2|\lambda|+s$, $t \geq |\lambda|+s$ and $d = \max\{\lambda_1,\ldots,\lambda_L\}+s-1$. Therefore A.9 implies that the sum of (C.25) and (C.26) is $O_p\left(b_N^{\min\{1,L/2\}}\right)$. By the assumptions on the bandwidth this is $O_p(N^{-\delta/2} = o_p(1)$, and thus the sum of (C.25) and (C.26) is $o_p(1)$.
This finishes the proof of the asymptotic linearity claim in part (*ii*).
Finally, we prove part (*iii*).
This finishes the proof of part (*iii*), and thus all the claims in Theorem **??**. □
Proof of Theorem 4.4: Consistency follows from

$$|\hat{\theta}-\theta_0| \leq \frac{1}{N}\sum_{i=1}^{N}|\omega(X_{i1})|\left|n(\hat{h}^{[\lambda]}_{NIP,s}(X_{i1},t(X_{i1})))-n(h^{[\lambda]}_0(X_{i1},t(X_{i1})))\right|$$

$$+\left|\frac{1}{N}\sum_{i=1}^{N}\omega(X_{i1})n(h^{[\lambda]}_0(X_{i1},t(X_{i1})))-\mathbb{E}\left[\omega(X_1)'n\left(h^{[\lambda]}_0(X_1,t(X_1))\right)\right]\right| \leq$$

$$\frac{1}{N}\sum_{i=1}^{N}|\omega(X_{i1})|\sup_{x_1\in\mathbb{X}_1}\left|n(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)))-n(h^{[\lambda]}_0(x_1,t(x_1)))\right|+\left|\frac{1}{N}\sum_{i=1}^{N}\omega(X_{i1})n(h^{[\lambda]}_0(X_{i1},t(X_{i1})))-\mathbb{E}\left[\omega(X_1)'n\left(h^{[\lambda]}_0(X_1,t(X_1))\right)\right]\right|$$

and the obvious fact that

$$\sup_{x_1\in\mathbb{X}_1}\left|n(\hat{h}^{[\lambda]}_{NIP,s}(x_1,t(x_1)))-n(h^{[\lambda]}_0(x_1,t(x_1)))\right| \leq \sup_{x\in\mathbb{X}}\left|n(\hat{h}^{[\lambda]}_{NIP,s}(x)-n(h^{[\lambda]}_0(x)\right|$$

so that the result follows as in the proof of Theorem 4.1. The asymptotic distribution is derived directly from Lemmas A.4-A.6, A.10-A.11. □

Notation: (page number indicates where it was first introduced)

(Y_i, X_i), $i = 1, \ldots, N$ is data (page 2)

K is dimension of Y (page 2)

L is dimension of X (page 2) (page 3)