Supplementary Appendix for "Recursive Differencing for Estimating Semiparametric Models"

Chan Shen
Pennsylvania State University

Roger Klein
Rutgers University

1 Proofs of Main Theorems

Proof of Theorem 1. The proof for part a) follows from Lemma 1. Parts b-c) follow from Lemma 5. With $\hat{g}(v;\theta)$ converging to $g(v;\theta)>0$, part d) follows from parts a-c).

In proving Theorem 2 we adopt the averaging notation from Pakes and Pollard (1989) and define $\langle a\rangle \equiv \frac{1}{N}\sum_{i=1}^{N} a_i$.

Proof of Theorem 2. To establish the consistency result in a), recall the objective function $\hat{Q}_1(\theta)$ in D7). Recall also that the trimming function in $\hat{Q}_1(\theta)$ is an indicator for the region where each continuous variable in X is between lower and upper sample quantiles. Define a population trimming function by replacing sample with population quantiles. Bound the absolute difference in estimated and population trimming functions by a smooth function as in Klein and Shen (2010). Employing Taylor series arguments to this bound and the smoothed trimming functions, these functions may be taken as known.

Referring to D3), define $\tau(X_i;q')$ as the indicator trimming function depending on quantiles q'_{1k} and q'_{2k} and let $\tau_{sm}(X_i;q(N))$ be the smooth trimming function on a set expanding to the full support for X that depends on quantiles q_{1k} and q_{2k}. Set q'_{1k} and q'_{2k} as fixed quantiles satisfying: $q_{1k}<q'_{1k}<q'_{2k}<q_{2k}$. Let $\hat{M}_s(V_i(\theta);\theta)\equiv \hat{M}_s(V_i(\theta);\theta,\tau_{sm})$, and redefine $\hat{Q}_1(\theta)$ and associated objective functions under known trimming $\tau_i\equiv\tau(X_i;q')$ as:

$$\begin{aligned}
\hat{Q}_1(\theta) &\equiv \left\langle \tau\left[Y-\hat{M}_s(V(\theta);\theta)\right]^2\right\rangle \\
Q_1(\theta) &\equiv \left\langle \tau\left[Y_i-M_s(V(\theta);\theta)\right]^2\right\rangle \\
\hat{S}_1(\theta) &\equiv \hat{Q}_1(\theta)-Q_1(\theta_0).
\end{aligned}$$

As minimizing $\hat{Q}_1(\theta)$ is equivalent to minimizing $\hat{S}_1(\theta)$, we prove consistency by showing that $\hat{S}_1(\theta)$ is uniformly (in θ) close to $ES_1(\theta)$ with $ES_1(\theta)$ having a unique minimum at θ_0. From Lemma 8, it follows that for θ in a compact set:

$$\sup_\theta\left|\hat{S}_1(\theta)-S_1(\theta)\right|=\sup_\theta\left|\hat{Q}_1(\theta)-Q_1(\theta)\right|\xrightarrow{p}0.$$

Under standard arguments, $S_1(\theta)$ is uniformly close to $ES_1(\theta)$. Under conditions in Ichimura(1993) and Ichimura and Lee(1991) $ES_1(\theta)$ is uniquely minimized at θ_0, which completes the consistency argument.

To establish asymptotic normality, from a standard Taylor series expansion:

$$\sqrt{N}\left(\hat{\theta}_1-\theta_0\right)=-\hat{H}\left(\theta^+\right)^{-1}\sqrt{N}\hat{G}(\theta_0),$$

where $\hat{H}$ is the estimated Hessian matrix, $\hat{G}$ is the estimated gradient to the SLS objective function, and θ^+ is between $\hat{\theta}$ and θ_0. Let H denote the Hessian matrix with all estimated functions replaced by the corresponding true ones. As in the consistency argument, all estimated trimming may be taken as known. From Lemma 11), part a),

$$\sup_\theta \left|\hat{H}(\theta) - H(\theta)\right| = o_p(1).$$

Under standard arguments,

$$\sup_\theta \left|\hat{H}(\theta) - E[H(\theta)]\right| = o_p(1).$$

Therefore, with $\theta^+ \xrightarrow{p} \theta_0$, $\hat{H}\left(\theta^+\right) \xrightarrow{p} E[H(\theta_0)]$. It now remains to show that the gradient is asymptotically distributed as normal. With $M_i \equiv M(V_i)$, $\varepsilon_i \equiv Y_i - M_i$, and $\hat{M}_{si} \equiv \hat{M}_s(V_i)$, write the estimated gradient component as $\hat{G}_A - \hat{G}_B$, where:

$$\hat{G}_A \equiv \left\langle \tau(X_i; \hat{q}')\, \varepsilon \nabla_\theta \hat{M}_s \right\rangle;\ \hat{G}_B \equiv \left\langle \tau(X_i; \hat{q}') \left[\hat{M}_s - M\right] \nabla_\theta \hat{M}_s \right\rangle.$$

Write $\hat{G}_A = \hat{G}_{A_1} + \hat{G}_{A_2} + \hat{G}_{A_3} + \hat{G}_{A_4}$, where with $\Delta_{si} \equiv \hat{g}_2(V_i)\left(\hat{M}_s(V_i) - M_i(V_i)\right)$:

$$\begin{aligned}
\hat{G}_{A_1} &\equiv \langle \tau(X_i; q')\, \varepsilon \nabla_\theta M_s \rangle;\ \hat{G}_{A_2} \equiv \langle [\tau(X_i; \hat{q}') - \tau(X_i; q')]\, \varepsilon \nabla_\theta M_s \rangle \\
\hat{G}_{A_3} &\equiv \left\langle \tau(X_i; q')\, \varepsilon \nabla_\theta \left[\frac{\Delta_s}{\hat{g}_2}\right] \right\rangle; \hat{G}_{A_4} \equiv \left\langle [\tau(X_i; \hat{q}') - \tau(X_i; q')]\, \varepsilon \nabla_\theta \left[\frac{\Delta_s}{\hat{g}_2}\right] \right\rangle.
\end{aligned}$$

To establish asymptotic normality, we will first show that

$$\hat{G}_{A_k} = o_p\left(N^{-1/2}\right),\ k = 2, 3, 4. \tag{1}$$

From Pakes and Pollard (1989; Lemma 1.18), $\hat{G}_{A_2} = o_p(N^{-1/2})$. For $\hat{G}_{A_3}$, referring to Lemma 1, recall the approximating recursion $(\Delta_s^*(V_i))$. In Lemma 6, the approximating derivative $\Delta_s^{1*}(V_i)$ is obtained from $\nabla_\theta^1 \Delta_1(v(w; \theta_0))$ by replacing $\hat{A}^{-1}$ with $A^{-1}[1 + \delta_A]$ and $\frac{1}{\hat{g}_2(v)}$ with $\frac{1}{\bar{g}_2}[1 + \delta_{g_2}(V_i)]$ throughout. Define:

$$\begin{aligned}
\rho_{si} &\equiv \nabla_\theta \left[\frac{\Delta_s(V_i)}{\hat{g}_2}\right] = \frac{\nabla_\theta \Delta_s(V_i)}{\hat{g}_2} - \frac{\Delta_s(V_i) \nabla_\theta \hat{g}_2}{\hat{g}_2^2} \\
\rho_{si}^* &\equiv \frac{\Delta_s^{1*}(V_i)}{\bar{g}_2}[1 + \delta_{g_2}(V_i)] - \frac{\Delta_s^*(V_i) \nabla_\theta \hat{g}_2}{\bar{g}_2^2}[1 + \delta_{g_2}(V_i)]^2.
\end{aligned} \tag{2}$$

Note that from Lemmas 1 and 6, $\Delta_s^*(V_i)$ and $\Delta_s^{1*}(V_i)$ are uniformly within $o_p(N^{-1/2})$ of $\Delta_s(V_i)$ and $\nabla_\theta^1 \Delta_s(V_i)$ respectively. From the expansion in Lemma 1, $\frac{1}{\hat{g}_2}$ is uniformly within $o_p(N^{-1/2})$ of $\frac{[1+\delta_{g_2}(V_i)]}{\bar{g}_2}$. Therefore,

$$\hat{G}_{A_3} = \hat{G}_{A_3}^* + o_p(N^{-1/2}),\ \hat{G}_{A_3}^* = \langle \tau_x \varepsilon \rho_s^* \rangle,$$

where under a Taylor series argument, we may take smooth trimming as known. From a mean-square convergence argument, $\hat{G}_{A_3}^* = o_p(N^{-1/2})$.

As above, for $\hat{G}_{A_4}$ we may take smooth trimming as known under a Taylor series expansion. Replace $\nabla_\theta \left[\frac{\Delta_s(V_i)}{\hat{g}}\right]$ with ρ_{si}^* to obtain $\hat{G}_{A_4}^*$ and note that $\hat{G}_{A_4}$ is uniformly within $o_p(N^{-1/2})$ of $\hat{G}_{A_4}^*$. Write:

$$\left|\hat{G}_{A_4}^*\right| \leq \langle |\tau(X_i; \hat{q}') - \tau(X_i; q')|\, |\varepsilon|\, |\tau_x^* \rho_s| \rangle,$$

where $\tau_{xi}^* = 1$ if either $\tau(X_i; \hat{q}') = 1$ or $\tau(X_i; q') = 1$ and is 0 otherwise. Approximating $|\tau(X_i; \hat{q}') - \tau(X_i; q')|$ by a smooth function as in Klein and Shen (2010), from Cauchy-Schwarz and Lemma 5, $\hat{G}_{A_4}^* = o_p(N^{-1/2})$.

For $\hat{G}_B$, denote $\hat{\tau}_x \equiv \tau(X_i; \hat{q}'), \tau_x \equiv \tau(X_i; q'), \hat{\tau}_{sm} \equiv \tau_{sm}(X_i; \hat{q}(N))$ and $\tau_{sm} \equiv \tau_{sm}(X_i; q(N)$. We make the dependence on estimated smooth trimming explicit and write $\hat{G}_B$ as $\hat{G}_{B_1}(\hat{\tau}_{sm}) + \hat{G}_{B_2}(\hat{\tau}_{sm}) + \hat{G}_{B_3}(\hat{\tau}_{sm}) + \hat{G}_{B_4}(\hat{\tau}_{sm})$, where

$$\begin{aligned}
\hat{G}_{B_1}(\hat{\tau}_{sm}) &\equiv \left\langle \tau_x \left[\hat{M}_s(\hat{\tau}_{sm}) - M\right] \nabla_\theta M_s \right\rangle \\
\hat{G}_{B_2}(\hat{\tau}_{sm}) &\equiv \left\langle \tau_x \left[\hat{M}_s(\hat{\tau}_{sm}) - M\right] \left[\nabla_\theta \hat{M}_s(\hat{\tau}_{sm}) - \nabla_\theta M\right] \right\rangle \\
\hat{G}_{B_3}(\hat{\tau}_{sm}) &\equiv \left\langle [\hat{\tau}_x - \tau_x] \left[\hat{M}_s(\hat{\tau}_{sm}) - M\right] \nabla_\theta M \right\rangle \\
\hat{G}_{B_4}(\hat{\tau}_{sm}) &\equiv \left\langle [\hat{\tau}_x - \tau_x] \left[\hat{M}_s(\hat{\tau}_{sm}) - M\right] \left[\nabla_\theta \hat{M}_s(\hat{\tau}_{sm}) - \nabla_\theta M\right] \right\rangle.
\end{aligned}$$

We begin by showing that smooth trimming can be taken as known in all terms. From Lemma 10:

$$\sqrt{N}\left[\hat{G}_{B_1}(\hat{\tau}_{sm}) - \hat{G}_{B_1}(\tau_{sm})\right] = o_p(1).$$

From Lemma 10 and a Taylor series expansion in sample quantiles applied to $\left[\nabla_\theta \hat{M}_s(\hat{\tau}_{sm}) - \nabla_\theta M\right]$,

$$\sqrt{N}\left[\hat{G}_{B_2}(\hat{\tau}_{sm}) - \hat{G}_{B_2}(\tau_{sm})\right] = o_p(1).$$

For the remaining terms, from Klein and Shen (2010) we may approximate $|\hat{\tau}_x - \tau_x|$ by a smooth bound. Employing a Taylor series for this bound and for the smooth trimming function:

$$\sqrt{N}\left[\hat{G}_{B_j}(\hat{\tau}_{sm}) - \hat{G}_{B_j}(\tau_{sm})\right] = o_p(1),\ j = 3, 4.$$

Next, we may replace recursion elements in the gradients by their approximating counterparts. Referring to Lemma 1 and recalling that $\hat{M}_s(V_i) - M(V_i) = \frac{\Delta_s(V_i)}{\hat{g}_2(V_i)}$, define

$$\left[\hat{M}_s^*(V_i) - M(V_i)\right] \equiv \frac{\Delta_s^*(V_i)}{\bar{g}_2(V_i)}\left[1 + \delta_{g_2}(V_i)\right].$$

Obtain $\hat{G}_{B_j}^*(\tau_{sm})$ from $\hat{G}_{B_j}(\tau_{sm})$ by replacing $\left[\hat{M}_s(V_i) - M(V_i)\right]$ with $\left[\hat{M}_s^*(V_i) - M(V_i)\right]$, and ρ_{si} with ρ_{si}^* from (2). From Lemmas 1 and 6, $\sqrt{N}\left[\hat{G}_{B_j}(\tau_{sm}) - \hat{G}_{B_j}^*(\tau_{sm})\right] = o_p(N^{-1/2})$.

Employing these approximations:

$$\hat{G}_{B_1}^*(\tau_{sm}) = \left\langle \tau_x \frac{\Delta_s^*(V_i)}{\bar{g}_2(V_i)}\left[1 + \delta_{g_2}(V_i)\right] \nabla_\theta M_s \right\rangle.$$

Under Lemmas 4-6, $E\left(\hat{G}_{B_1}^*(\tau_{sm})\right) = o(N^{-1/2})$. Then, under arguments similar to those in Jiang (2021), it can then be shown that $N^{1/2}[\hat{G}_{B_1}^* - U_N] = o_p(1)$ where U_N is a centered U-statistic. It follows from U-statistic projection arguments that

$$\hat{G}_{B_1}^*(\tau_{sm}) = \sqrt{N}\left\langle \varepsilon E\left[\tau_{sm}\delta_i(\theta_0)\,|V\right]\right\rangle + o_p(1) = o_p(1),$$

$\delta_i(\theta_0) \equiv \nabla_\theta \left[M(V(W_i;\theta);\theta)\right]_{\theta_0}$. From Cauchy-Schwarz, Lemmas 5-6, and C1), $\hat{G}_{B_2}^* = o_p(N^{-1/2})$. For the remaining terms, from Klein and Shen (2010) we may bound the absolute value of the difference in indicators by a smooth bound. Employing a Taylor series for this bound, it can readily be shown that $\hat{G}^*_{B_3}$ and $\hat{G}^*_{B_4}$ are each $o_p\left(N^{-1/2}\right)$. The normality result follows from the form for $\sqrt{N}\left[\hat{G}_{A_1} - \hat{G}^*_{B_1}\left(\tau_{sm}\right)\right]$.

To establish b), under index trimming at the second step, consistency follows from an extension of Lemma 4 in Klein and Shen (2010) to d multiple indices. To outline the argument, recall the definition of the adjusted estimator in D8), D9), and D10). From D9), $\hat{M}_{sa}(V_i)$ depends on $\hat{g}_{sa}(V_j)$, $j = 1, ..., N$. Obtain $M_a(V_i)$ from $\hat{M}_{sa}(V_i)$ by replacing $\hat{g}_{sa}(V_j)$ with $g_a(V_j) \equiv g(V_j) + A(V_j)$ throughout and all other components by their probability limits. With $g(v)$ as the density of V_i, notice that $M_a(V_i) = M(V_i)$ if we replace $g_a(V_j)$ with $g(V_j)$. Let $\hat{M}_{sa}$, M_a, and M be vectors with respective i^{th} elements: $\hat{M}_{sa}(V_i)$, $M_a(V_i)$, and $M(V_i)$, let $\tau_{vi} \equiv \tau\left(V_i; q'\right)$ and $\hat{\tau}_{vi} \equiv \tau\left(V_i; \hat{q}'\right)$. From Klein and Shen (2010), we may approximate $|\tau_{vi} - \hat{\tau}_{vi}|$ by a smooth bound. Employing a Taylor expansion for this bound and for the estimated smooth function within $\hat{M}_{sa}$, all trimming can be taken as given in analyzing the objective function. Under known trimming, write:

$$\hat{Q}_a \equiv \left\langle \tau_v \left[Y - \hat{M}_{sa}\right]^2 \right\rangle; Q_a \equiv \left\langle \tau_v \left[Y - M_a\right]^2 \right\rangle; Q \equiv \left\langle \tau_v \left[Y - M\right]^2 \right\rangle.$$

With uniformity taken with respect to θ in a compact set, we show

$$\begin{aligned} i) &: \sup_\theta \left|\hat{Q}_a - Q_a\right| \xrightarrow{p} 0 \\ ii) &: \sup_\theta \left|Q_a - E\left[Q_a\right]\right| \xrightarrow{p} 0 \\ iii) &: \sup_\theta \left|E\left[Q_a\right] - E\left[Q\right]\right| \xrightarrow{p} 0, \end{aligned}$$

with $E\left[Q\right]$ having a unique min at θ_0. Consistency would then follow. Part i) holds from Lemma 9 and standard arguments establish part ii). For iii), write the difference as:

$$\left|E\left[Q_a\right] - E\left[Q\right]\right| \leq 2\left|E\left\langle \tau_v Y\left[M_a - M\right]\right\rangle\right|_{D_1} + \left|E\left\langle \tau_v\left[M_a - M\right]\left[M_a - M\right]\right\rangle\right|_{D_2}.$$

Letting $M_0 \equiv E(Y|X)$, for the first difference D_1:

$$D_1 = 2\left|E\left\langle \tau_v M_0 \left[\frac{f}{g_2}\frac{g_2}{g_{2a}} - \frac{f}{g_2}\right]\right\rangle\right| = 2\left|E\left\langle \tau_v M_0 M \left|\frac{g_2 - g_{2a}}{g_{2a}}\right|\right\rangle\right|.$$

With τ_v, M_0, and M bounded and $A \equiv g_{2a} - g_2$ as the adjustment factor in D8), it suffices to show that

$$\sup_\theta \left|E\left\langle |A/g_{2a}|\right\rangle\right| = o_p(1).$$

Let $\mathfrak{C}_N = \{X_i : g_2(V_i) > N^{-ar/2}\}$ and define the indicator $\tau_g(X_i) \equiv 1\{X_i \, \epsilon \, \mathfrak{C}_N\}$. Then,

$$\sup_\theta \left|E\left\langle |A/g_{2a}|\right\rangle\right| \leq \sup_\theta \left|E\left\langle \tau_g A N^{ar/2}\right\rangle\right| + \sup_\theta \left|E\left\langle 1 - \tau_g\right\rangle\right|.$$

Since $A = o(N^{-ar/2})$, the first component is $o_p(1)$. The second component is also $o_p(1)$ under the vanishing probability on the complement of $\mathfrak{C}_N$. The argument for D_2 is similar to that for D_1, which completes the uniform convergence argument.

With the second stage objective function uniformly converging to the fixed function $E[Q]$, Ichimura (1993) and Ichimura and Lee (1991) provide conditions under which $E[Q]$ is uniquely maximized at θ_0. Consistency for the second stage estimator of Theorem 2 follows.

To establish asymptotic normality, note that the estimator has the following linear form:

$$\sqrt{N}\left(\hat{\theta}_2-\theta_0\right)=-\hat{H}^{-1}\left(\theta^{+}\right)\sqrt{N}\hat{G}\left(\theta_0\right)+o_p(1),\ \theta^{+}\epsilon\left[\hat{\theta}_2,\theta_0\right].$$

Employing arguments very similar to those above, all trimming can be taken as known. Noting that τ_{vi} ensures that adjustment factors vanish rapidly at θ_0, under arguments very similar to those in part a) and employing Lemmas 1, 5, 6, 9, 10, and 11b), we have:

$$\begin{aligned} i) &: \quad \hat{H}^{-1}\left(\theta^{+}\right)\xrightarrow{p}E[H(\theta_0)]^{-1}=E\left[\tau_{vi}\nabla_\theta M_i\nabla_\theta M'\right]^{-1}. \\ ii) &: \quad \sqrt{N}\hat{G}\left(\theta_0\right)=\sqrt{N}\left\langle\tau_v\varepsilon\nabla_\theta M\right\rangle-\sqrt{N}\left\langle\tau_v\left[\hat{M}_s-M\right]\nabla_\theta M\right\rangle+o_p(1).\end{aligned}$$

The proof will then follow if

$$B\equiv\sqrt{N}\left\langle\tau_v\left[\hat{M}_s-M\right]\nabla_\theta M\right\rangle=o_p(1).$$

From Lemma 1:

$$B=B^*+o_p(1),\ B^*\equiv\sqrt{N}\left\langle\tau_v\left[\hat{M}_s^*-M\right]\frac{\hat{g}}{\bar{g}}\left[1+\delta_g\right]\nabla_\theta M\right\rangle.$$

From Newey's residual result, $E\left[B^*\right]=0$. In the single index case, Klein and Shen (2010) show that B^* is a degenerate U-statistic under regular kernels. This argument had been extended to multiple indices under the recursive estimator, with the extension provided in Jiang (2021). The result now follows.

The remainder of the Appendix is organized as follows. Lemmas 1-9 provide results under known trimming, while Lemma 10 provides results on taking the estimated trimming set as known. Lemmas 1-5 develop properties of the estimated M-function that are of independent interest apart from estimating index parameters. The remaining Lemmas 6-11, provide results to establish the asymptotic properties for the index parameter estimators.

In Lemmas 1-5, the conditional expectation of the dependent variable, Y_i, depends on the known d-dimensional vector V_i. The vector V_i may be interpreted as $V_i\left(\theta_0\right)$ in the semiparametric case or as a vector of exogenous variables in the nonparametric case. Since gradients to objective functions are evaluated at true parameter values, this interpretation will be useful when we show that these gradients are asymptotically distributed as normal.

Prior to Lemma 10 which addresses estimated trimming, we take all trimming as known. The results in Lemmas 1-5 hold whether trimming is based on the true index vector or on the continuous variables, X_i. In so doing, we redefine $K_i^*(v)\equiv\tau_{sm}(X_i,q(N))K_i(v)$ or $K_i^*(v)\equiv\tau_{sm}(V_i\left(\theta_0\right),q(N)))K_i(v)$.

2 Intermediate Lemmas

Lemma 1. Recursion Approximation. Recall the definition of the conditional expectation estimator in D6) and kernel functions in D4-5). With $D\equiv diag(K(v))$, $\mathcal{L}$ and $\mathcal{P}$ positive integers, $V_i\equiv V(W_i;\theta_0)$, and $Z_i(v) \equiv \frac{V_i - v}{h}$ as the i^{th} row of Z, define:

$$\begin{aligned}
\hat{A} &\equiv \frac{1}{N}\left[Z(v)' DZ(v)\right];\ A \equiv E(\hat{A});\ \delta_A(v) \equiv \sum_{l=1}^{\mathcal{L}} \left[\left(A - \hat{A}\right) A^{-1}\right]^l \\
\bar{g}_s(v) &= E(\hat{g}_s(v)); \delta_{g_s} \equiv \sum_{p=1}^{\mathcal{P}} \left[\frac{\bar{g}_s(V_i) - \hat{g}_s(V_i)}{\bar{g}_s(V_i)}\right]^p \\
\hat{d}(v) &\equiv \left[Z' DZ\right]^{-1} Z' D \left[Y - \underline{1} \cdot \bar{Y}(v)\right] \\
\hat{d}^*(v) &\equiv A^{-1}[I + \delta_A(v)] \frac{1}{N} \sum_{i=1}^{N} \frac{Z_i(v)\left[Y_i - \bar{Y}(v)\right] \hat{g}_1(v) K_i(v)}{\bar{g}_1(v)} \left[1 + \delta_{g_1}(v)\right].
\end{aligned}$$

With $\Delta_s(v) \equiv \hat{g}_2(v)\left[\hat{M}_s(v) - M(v)\right]$, we have:

$$\begin{aligned}
\Delta_1(v) &= \frac{\hat{g}_2(v)}{\hat{g}_1(v)} \left[\begin{array}{c} \frac{1}{N}\sum_{i=1}^{N}\left[M(V_i) - M(v) + \varepsilon_i\right] K_i(v) \\ -\frac{1}{N}\sum_{i=1}^{N} Z_i(v) K_i(v) \hat{d}(v) \end{array}\right] \\
\Delta_s(v) &\equiv \Delta_{s-1}(v) - \sum_{i=1}^{N} \frac{\Delta_{s-1}(V_i)}{\hat{g}_2(V_i)} K_i^*(v) + \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i K_i^*(v),\ s > 1.
\end{aligned}$$

The approximating recursion is given as:

$$\begin{aligned}
\Delta_1^*(v) &\equiv \frac{\hat{g}_2(v)}{\bar{g}_1(v)}[1 + \delta_{g_{s-1}}(v)] \left[\begin{array}{c} \frac{1}{N}\sum_{i=1}^{N}\left[M_1(V_i) - M_1(v) + \varepsilon_i\right] K_i(v) \\ -\frac{1}{N}\sum_{i=1}^{N} Z_i(v) K_i(v) \hat{d}^*(v) \end{array}\right] \\
\Delta_s^*(v) &\equiv \Delta_{s-1}^*(v) - \sum_{i=1}^{N} \left[\frac{\Delta_{s-1}^*(V_i)}{\bar{g}_2(V_i)}\right] [1 + \delta_{g_2}(V_i)] K_i^*(v) + \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i K_i^*(v), s > 1.
\end{aligned}$$

For v in a compact subset of its support and with $\mathcal{L}$, $\mathcal{P}$ sufficiently large and finite:

$$\sup_v |\Delta_s^*(v) - \Delta_s(v)| = o_p(N^{-1/2}).$$

Proof. For $s = 1$, note that:

$$\begin{aligned}
\hat{A}^{-1} &= A^{-1} + \hat{A}^{-1}\left[A - \hat{A}\right] A^{-1} \\
&= A^{-1}[I + \delta_A] + \hat{A}^{-1}\left[\left(A - \hat{A}\right) A\right]^{\mathcal{L}+1}.
\end{aligned}$$

With a similar expansion holding for $1/\hat{g}(V_i)$, for $\mathcal{L}$, $\mathcal{P}$ sufficiently large:

$$\sup_v \left|\hat{A}^{-1} - A^{-1}[I + \delta_A]\right| = o_p(N^{-1/2}) \tag{3}$$

$$\sup_v \left|\frac{1}{\hat{g}(V_i)} - \frac{1}{\bar{g}(V_i)}\left[1 + \delta_{g_1}(V_i)\right]\right| = o_p(N^{-1/2}). \tag{4}$$

The result now follows for $s = 1$. An induction argument completes the proof. ■

Lemma 2. Stage Characterization. Define

$$KP_0(v) \equiv \frac{1}{N}\sum_i \left[M(V_i) - M(v)\right] K_i(v).$$

Recall the definitions of $\delta_A(v)$ and $\delta_{g_1}(V_i)$ in Lemma 1. With $Z_i \equiv \frac{V_i - v}{h}$, let:

$$\begin{aligned} e_1(v, V_i) &\equiv Z_i K_i(v) A^{-1}(v) \\ e_2(v, V_i) &\equiv Z_i K_i(v) \left[M(V_i) - M(v)\right] \\ e_3(v, V_i) &\equiv Z_i K_i(v) \\ e_4(v, V_i) &\equiv \frac{\left[M(V_i) - M(v)\right] K_i(v)}{\bar{g}_1(v)}. \end{aligned}$$

Define $\bar{e}_t(v) = \frac{1}{N} \sum_i e_t(v, V_i)$ and

$$AKP(v) \equiv \bar{e}_1(v) \left[I + \delta_A(v)\right] \left\{\bar{e}_2(v) - \bar{e}_3(v)\bar{e}_4(v) \left[1 + \delta_{g_1}(v)\right]\right\}.$$

Let U_s be a random variable that has zero expectation conditioned on X and define:

$$\begin{aligned} KP_L(v) &\equiv O\left(\frac{1}{N^L}\right) \sum_{i_1} \cdots \sum_{i_L} T_{L, i_1, \ldots, i_L}, \\ T_{L, i_1, \ldots, i_L} &\equiv \begin{cases} \frac{\Delta_1^*(V_{i_1})}{\bar{g}_s(V_{i_l})} K_{i_L}^*(v) \left[1 + \delta_{g_s}(V_{i_l})\right] & L = 1 \\ \frac{\Delta_1^*(V_{i_1})}{\prod_{l=1}^{L} \bar{g}_s(V_{i_l})} K_{i_L}^*(v) \prod_{l=1}^{L} \left[1 + \delta_{g_s}(V_{i_l})\right] \prod_{l=1}^{L-1} K_{i_l}^*(V_{i_l}) & L > 1 \end{cases}. \end{aligned}$$

Then, there exists integers $C_1, \ldots, C_{s-1}$ such that

$$a) \quad : \quad \Delta_1^*(v) = \left[U_1 + KP_0(v) - AKP(v)\right] \frac{\hat{g}_2(v)}{\bar{g}_1(v)} \left[1 + \delta_{g_1}(v)\right] \tag{5}$$

$$b) \quad : \quad \Delta_s^*(v) - \Delta_1^*(v) = U_s + \sum_{l=1}^{s-1} C_l KP_l(v), \; s > 1. \tag{6}$$

Proof. The proof for a) is immediate from Lemma 1. For b), due to the form of $\Delta_s^*(v) - \Delta_{s-1}^*(v)$, the lemma follows because $\Delta_s^*(v) - \Delta_1^*(v)$ is given as:

$$\left[\Delta_s^*(v) - \Delta_{s-1}^*(v)\right] + \left[\Delta_{s-1}^*(v) - \Delta_{s-2}^*(v)\right] + \ldots + \left[\Delta_2^*(v) - \Delta_1^*(v)\right].$$

■

To study the expectations of the estimators as characterized in Lemma 2, Lemma 3 provides asymptotic conditional independence results for studying the expectation of products of averages, which is the structure of the stage s estimator. Part a) of this lemma is stated in a form more general than is required so as to illustrate a method of proof that applies to all other parts.

Lemma 3. Kernel Products. For V_i let $F(V_i)$ be a bounded function of V_i. For p a positive integer, define:

$$\begin{aligned} \delta(V_i) &\equiv \left\{\frac{1}{N} \sum_j \left[F(V_i) - K_j^*(V_i)\right]\right\}^p \\ \rho_i &\equiv \left[\frac{\Delta_{s-1}^*(V_i)}{g(V_i)} \left[1 + \delta_{g_s}(V_i)\right]\right]. \end{aligned}$$

Referring to the definitions in Lemma 1 and to the definition of $KP_0(v)$ in Lemma 2:

$$a) \quad : \quad E\left[\delta(V_i)|V_i\right] = \left\{\frac{1}{N}\sum_j E\left[F(V_i) - K_j^*\left(V_i\right)|V_i\right]\right\}^p + O\left(\frac{1}{Nh^d}\right).$$

$$b) \quad : \quad E\left[\delta_A(v)\right] \equiv E\left\{\left[\bar{e}_2(v)\right]^l\right\} = \left[E\left(\bar{e}_2(v)\right)\right]^l + O\left(\frac{1}{Nh^d}\right).$$

$$c) \quad : \quad E\left[\bar{e}_1(v)\left[I + \delta_A(v)\right]\left\{\bar{e}_2(v) + \bar{e}_3(v)\ \bar{e}_4(v)\left[1 + \delta_{g_1}(v)\right]\right\}\right] =$$

$$E\left[\bar{e}_1(v)\right]\left[I + E\left[\delta_A(v)\right]\right]\left\{\begin{array}{c} E\left[\bar{e}_2(v)\right] + E\left[\bar{e}_3(v)\right]\ \times \\ E\left[\bar{e}_4(v)\right]\left[1 + E\left[\delta_{g_1}(v)\right]\right] \end{array}\right\}\frac{\hat{g}_2(v)}{\bar{g}_1(v)}\left[1 + E\left[\delta_{g_1}\left(v\right)\right]\right] + O\left(\frac{1}{Nh^d}\right).$$

$$d) \quad : \quad E\left[KP_0(v)\frac{\hat{g}_2(v)}{\bar{g}_1(v)}\left[1 + \delta_{g_1}\left(v\right)\right]\right] = E\left[KP_0(v)\right]\frac{E\left[\hat{g}_2(v)\right]}{\bar{g}_1(v)}\left[1 + E\left[\delta_{g_1}\left(v\right)\right]\right] + O\left(\frac{1}{Nh^d}\right).$$

$$e) \quad : \quad E\left\{\rho_i K_i^*(v)\rho_j K_j^*(v)\right\} = E\left\{E\left[\rho_i K_i^*(v)|V_i\right]E\left[\rho_j K_j^*(v)|V_j\right]\right\} + O\left(\frac{1}{Nh^d}\right).$$

$$f) \quad : \quad E\left\{\frac{\Delta_{s-1}^*(V_i)}{\bar{g}_2(V_i)}\left[1 + \delta_{g_2}(V_i)\right]|V_i\right\} = \frac{E\left[\Delta_{s-1}^*(V_i)|V_i\right]}{\bar{g}_2(V_i)}\left[1 + E\left[\delta_{g_2}(V_i)|V_i\right]\right] + O\left(\frac{1}{Nh^d}\right).$$

With window parameter in D4) satisfying: $0 < r < \frac{1}{2d}$, each $O\left(\frac{1}{Nh^d}\right)$ remainder term is $o(N^{-1/2})$.
Proof. For a), write the expectation of a typical term in $\delta(V_i)$:

$$E\left\{O\left(\frac{1}{N^p}\right)\left[\sum_{j\neq i}F(V_i) - K_j^*\left(V_i\right)\right]^p|V_i\right\} = E\left\{O\left(\frac{1}{N^p}\right)\sum_{j_1}\sum_{j_2}...\sum_{j_p}\prod_{l=1}^{p}\left[F(V_i) - K_{j_l}^*\left(V_i\right)\right]|V_i\right\}.$$

If all of the subscripts are distinct, the result is immediate from independence. Assume there are $m + 1 \geqslant 2$ identical subscripts and reorder terms so that these are at the end. The expectation is then:

$$O\left(\frac{1}{N^m}\right)E\left\{O\left(\frac{1}{N^{p-m}}\right)\left[\sum_{j_1}\sum_{j_2}...\sum_{j_{(p-m)}}\left(\prod_{l=1}^{p-m-1}\left[F\left(V_i\right) - K_{j_l}^*\left(V_i\right)\right]\right)\right]\left[F\left(V_i\right) - K_{j_p}^*\left(V_i\right)\right]^{m+1}|V_i\right\}$$

$$= \quad O\left[\left(\frac{1}{Nh^d}\right)^m\right]\prod_{l=1}^{p-m-1}\left[F\left(V_i\right) - E\left[K_{j_l}^*\left(V_i\right)|V_i\right]\right]E\left\{O(h^{md})\left[F\left(V_i\right) - K_{j_p}^*\left(V_i\right)\right]^{m+1}|V_i\right\}.$$

The first term is $O\left[\left(\frac{1}{Nh^d}\right)^m\right]$, $m \geqslant 1$. It can readily be shown the second component is $O(1)$. For the final component, let z be a $d \times 1$ vector with l^{th} component $\frac{V_i(l) - V_{j_p}(l)}{O(h)}$. With this change of variable, and the integral being multi-dimensional, we can write the third component as:

$$\int O(h^{md})\left[F\left(V_i + O(h)z\right) - \frac{1}{O(h^d)}k(z)\right]^{m+1}O(h^d)g(V_i + O(h)z)dz$$

$$= \quad \int O(h^{(m+1)d})\left\{\sum_{i=1}^{m+1}\left(\begin{array}{c} m+1 \\ i \end{array}\right)F\left(V_i + O(h)z\right)^{m+1-i}\left[\frac{1}{O(h^d)}k(z)\right]^i\right\}g(V_i + O(h)z)dz.$$

With $k(z)^p$ integrable under D4) for all finite p, the above integral is $O(1)$ from which the result follows. The proofs for b-e) are very similar to that for a) in that they depend on matching subscripts as in the above argument. For f), from the stage characterization in Lemma 2 and employing the same matching arguments as in a), the result follows. ■

Lemma 4. Bias Expansions for Kernel Expectations. Replacing $M(v)$ in assumption A3b) with $F(v)$ and letting $g(v)$ be the density for V_i. Assume that $F(v)g(v)$ satisfies A3b). Recall that $q(N)$ is an expanding quantile set that expands from fixed quantiles q_1 and q_2 to the full support of the indices or continuous X-variables for which trimming is applied. With $T_i \equiv X_i$ or V_i, recall from D3) that $\tau_{sm}(T_i; q(N))$ smoothly approximates an indicator on a set that expands from fixed quantiles q_1 and q_2 to the full support of the indices or continuous X-variables. With v such that $q_1 < q_1' < v < q_2' < q_2$, kernel functions defined in D4) and $B_j(v) \equiv \left[\nabla_t^j [F(t)g(t)]\right]_{t=v}$:

$$E[F(V_i)K_i^*(v)] = E[\tau_{sm}(T_i; q(N)) F(V_i)K_i(v)] = F(v)g(v) + \sum_{j=1}^{m} h^{2j} B_{2j}(v) + o(h^{2m}) \; + \; O\left(\frac{1}{N}\right).$$

Proof. We first prove the result under index trimming where $T_i = V_i$. Define:

$$\begin{aligned} \mathfrak{C}_j &\equiv \left\{ V_i : |V_i - q_j(N)| < \frac{1}{\ln(N)} \right\}, j = 1, 2, \\ \mathfrak{A} &\equiv (\mathfrak{C}_1 \cup \mathfrak{C}_2)^c, \end{aligned}$$

where c denotes complement. Write $E[F(V_i)\tau_{sm}(V_i; q) K_i(v)] = E_1 + E_2 + E_3$:

$$\begin{aligned} E_1 &\equiv E[\; F(V_i)[\tau_{sm}(V_i; q(N)) - \tau(V_i; q(N))] 1\{V_i \epsilon \mathfrak{A}\} K_i(v)], \\ E_2 &\equiv E[\; F(V_i)\tau(V_i; q(N)) 1\{V_i \epsilon \mathfrak{A}\} K_i(v)], \\ E_3 &\equiv E[\; F(V_i)\tau_{sm}(V_i; q(N)) 1\{V_i \epsilon \mathfrak{A}^c\} K_i(v)]. \end{aligned}$$

On $\mathfrak{A}$, $\sup_v |\tau(v; q) - \tau_{sm}(v; q)| = O\left(\frac{1}{N}\right) \Rightarrow E_1 = O\left(\frac{1}{N}\right)$. For $E_2 = \int_{\mathfrak{A}} F(v_i)K_i(v)g(v_i)dv_i$, let z be a vector with l^{th} element $\frac{v(l)-v_i(l)}{O(h)}$. Then, making this change of variable and restricting all components of v to a strict subset of a compact set, the expectation up to higher order terms is given as:

$$\int_{\mathcal{R}^d} F(v + hz)k(z)g(v + hz)dz,$$

where k is the standardized kernel in D4). From a standard Taylor expansion in h about 0, the result follows on $\mathfrak{A}$.

On $\mathfrak{A}^c \equiv \mathfrak{C}_1 \cup \mathfrak{C}_2$ we provide the argument for $\mathfrak{C}_1$ as the argument for $\mathfrak{C}_2$ is identical. We have

$$|E_3| \leq \sup_{v \epsilon \mathfrak{C}_1} |F(v)g(v)| \int_{\mathfrak{C}_1} K_i(v)dV_i \equiv \sup_{v \epsilon \mathfrak{C}_1} |F(v)g(v)| \, I(\mathfrak{C}_1).$$

With $\sup_v |F(v)g(v)|$ bounded, it suffices to consider the integral $I(\mathfrak{C}_1)$. With v constrained to an interior subset of $[q_1, q_2]$, the argument follows as $I(\mathfrak{C}_1) = o\left(\frac{1}{N}\right)$ with v not being close to $V_i \epsilon \mathfrak{C}_1$. Noting that the smooth X-trimming function is on a set expanding to the full support of X, similar arguments to those above establish the result under smooth X-trimming. ■

Employing the above lemmas, Lemma 5 obtains uniform bias and variance rates for the proposed estimator.

Lemma 5. Stage Bias and Variance. Let v be in a compact subset of its support. Assuming A3), with $\Delta_s^*(v)$ defined as in Lemma 1:

$$\begin{aligned} a) &\;:\; \sup_v |E[\Delta_s^*(v)]| = O(h^{2s}), \\ b) &\;:\; \sup_v Var[\Delta^*(v)] = O\left(N^{-(1-rd)}\right). \end{aligned}$$

Proof. To prove a), we will show that:

$$E\left[\Delta_s^*(v)\right] = h^{2s}B_s(v) + h^{2(s+1)}B_{s+1}\left(v\right) + o(h^{2(s+1)}),$$

where the B-functions are uniformly bounded in v. For $s = 1$ and with KP_0 and KP_1 defined in Lemma 2:

$$E\left[\Delta_1^*\left(v\right)\right] = E\left[KP_0\left(v\right)\right] - E\left[KP_1\left(v\right)\right].$$

It then follows from Lemmas 3-4 that for uniformly bounded B-functions:

$$\left|E\left[\Delta_1^*(v)\right]\right| = h^2B_1\left(v\right) + h^4B_2\left(v\right) + o(h^4).$$

Continuing with an induction argument for a), with $s > 1$, assume that

$$E\left[\Delta_{s-1}^*(v)\right] = h^{2(s-1)}B_{s-1}\left(v\right) + h^{2s}B_s\left(v\right) + o(h^{2s}). \tag{7}$$

Under Lemma 3f), $E\left[\Delta_s^*(v)\right]$ is given as:

$$\begin{aligned} & E\left[\Delta_{s-1}^*(v)\right] - E\left[\frac{\Delta_{s-1}^*(V_i)}{\bar{g}_2(V_i)}\left[1+\delta_{g_2}(V_i)\right]K_i^*(v)\right] \\ = \quad & E\left[\Delta_{s-1}^*(v)\right] - E\left[\frac{E\left[\Delta_{s-1}^*(V_i)|V_i\right]}{\bar{g}_2(V_i)}\left[1+E\left[\delta_{g_2}(V_i)|V_i\right]\right]K_i^*(v)\right]. \end{aligned}$$

From Lemma 4 it now follows that

$$E\left\{E\left[\Delta_{s-1}^*(V_i)|V_i\right]K_i^*(v)\right\} = h^{2(s-1)}B_{s-1}(v) + h^{2s}B_s\left(v\right) + o(h^{2s}).$$

Part a) now follows because $E\left[\delta_{g_s}(V_i)|V_i\right] = O(h^2)$.

The proof for b) at stage $s = 1$ is immediate. Assuming the result holds for stage $s-1$, we show it holds for stage $s > 1$. Recall that

$$\Delta_s^*(v) \equiv \Delta_{s-1}^*(v) - \frac{1}{N}\sum_i\left[\frac{\Delta_{s-1}^*\left(V_i\right)}{\bar{g}_2(V_i)}\left[1+\delta_{g_2}\left(V_i\right)\right]\right]K_i^*(v) - \frac{1}{N}\sum_i\varepsilon_iK_i^*(v).$$

With $T \equiv \Delta_s^*(v) - E\left[\Delta_s^*(v)\right]$ and $\rho_i \equiv \frac{\Delta_{s-1}^*(V_i)}{\bar{g}_2(V_i)}\left[1+\delta_{g_2}\left(V_i\right)\right]$, from the definition of $\Delta_s^*(v)$, $T = T_1 - T_2 - T_3 - T_4$ where

$$\begin{aligned} T_1 &\equiv \Delta_{s-1}^*(v) - E\left[\Delta_{s-1}^*(v)\right] \\ T_2 &\equiv \frac{1}{N}\sum_i\varepsilon_iK_i^*(v) \\ T_3 &\equiv \frac{1}{N}\sum_i\left\{E\left(\rho_i|V_i\right)K_i^*(v) - E\left[\rho_iK_i^*(v)\right]\right\} \\ T_4 &\equiv \frac{1}{N}\sum_i\left[\rho_i - E\left(\rho_i|V_i\right)\right]K_i^*(v). \end{aligned}$$

Part b) will follow if $E(T_l^2) = O\left(N^{-(1-rd)}\right)$, $l = 1,2,3,4$. The result is immediate for T_1. For T_2 and any stage s:

$$\begin{aligned}\sup_v E(T_2^2) &= \left[\frac{1}{Nh^d}\right]\sup_v \frac{1}{N}E\sum_i \left[\varepsilon_i^2 h^d K_i^{*2}(v)\right] \\ &= \frac{1}{Nh^d}\sigma^2 \sup_v E\left[h^d K_i^{*2}(v)\right] = O\left(\frac{1}{Nh^d}\right).\end{aligned}$$

For T_3, since $E\left\{E\left(\rho_i|V_i\right)K_i^*(v) - E\left[\rho_i K_i^*(v)\right]\right\} = E\left\{E\left(\rho_i|K_i^*(v)V_i\right) - E\left[\rho_i K_i^*(v)\right]\right\} = 0$, from conditional independence, the expectations of cross products vanish, which implies that:

$$\sup_v E(T_3^2) = \left[\frac{1}{Nh^d}\right]\sup_v \frac{1}{N}E\sum_i h^d\left\{E\left(\rho_i|V_i\right)K_i^*(v) - E\left[\rho_i K_i^*(v)\right]\right\}^2 = o\left(\frac{1}{Nh^d}\right).$$

For T_4, write $T_4^2 = ST + CPT$, where

$$\begin{aligned}ST &\equiv \frac{1}{Nh^d}\frac{1}{N}\sum_i\left[\rho_i - E\left(\rho_i|V_i\right)\right]^2 h^d K_i^{*2}(v) \\ CPT &\equiv \frac{1}{N^2}\sum_i\sum_{j\neq i}\left[\rho_i - E\left(\rho_i|V_i\right)\right]K_i^*(v)\left[\rho_j - E\left(\rho_j|V_j\right)\right]K_j^*(v).\end{aligned}$$

For the squared terms (ST):

$$\sup_v E\left[ST\right] = \frac{1}{Nh^d}\sup_v E\left\{\left[\rho_i - E\left(\rho_i|V_i\right)\right]^2 hK_i^{*2}(v)\right\} = O\left(\frac{1}{Nh^d}\right).$$

For the cross-product terms (CPT), from Lemma 3, part e) :

$$\sup_v E\left[CPT\right] = \sup_v E\left\{E\left[\rho_i - E\left(\rho_i|V_i\right)|V_i\right]E\left[\rho_j - E\left(\rho_j|V_i\right)|V_j\right]K_i^*(v)K_j^*(v)\right\} + O\left(\frac{1}{Nh^d}\right).$$

As the first component is 0, part b) of the lemma follows. ■

For the estimated derivatives entering the gradient expressions, we require rates at which the squared bias and variance vanish irrespective of the point at which functions are evaluated. In all remaining lemmas, we will denote $V_i(\theta) \equiv V(W_i, \theta)$ as a parametric vector of indices depending on the parameter vector θ. Lemma 6 below provides the required result.

Lemma 6. Mean-Square Convergence for First Derivatives. For w in a compact subset of its support, there exists $\Delta_s^{1*}(v(w;\theta_0))$ such that:

$$\begin{aligned}a) &: \sup_w\left|\nabla_\theta^1\Delta_s(v(w;\theta_0)) - \Delta_s^{1*}(v(w;\theta_0))\right| = o_p\left(N^{-1/2}\right). \\ b) &: \sup_w E\left[\left(\Delta_s^{1*}(v(w;\theta_0))\right)^2\right] = O\left(N^{-4rs} + N^{-(1-r(d+2))}\right).\end{aligned}$$

Proof. With $\hat{A}$ defined as in Lemma 1, note that

$$\hat{A}\hat{A}^{-1} = I \Rightarrow \nabla_\theta^1\hat{A}^{-1} = -\hat{A}^{-1}\nabla_\theta^1\left(\hat{A}\right)\hat{A}^{-1}.$$

In $\nabla_\theta^1\Delta_1(v(w;\theta_0))$, refer to Lemma 1 and replace $\hat{A}^{-1}$ with $A^{-1}\left[1+\delta_A\right]$. Further, in $\nabla_\theta^1\Delta_s$ replace all density reciprocals by the expansion in Lemma 1. With $\Delta_s^{1*}(v(w;\theta_0))$ as the resulting derivative recursion, part a) follows. Part b) follows from the same term decomposition as in Lemma 5. ■

For lemmas on uniform convergence for averages involving unbounded random variables, we require Markov's inequality for joint events in Lemma 7, part a). We attribute the proof of this result to Markov as the proof follows immediately from his argument for single random variables. We also require a tail dominance result given in part b).

Lemma 7. Markov and Moment Inequalities. Let Y_i, X_{1i}, and X_{mi} be continuous random variables with joint density $g(y, x_l, x_m)$. Assume that $E\left[|Y|^{p_y} |X_l^{p_l}| \, |X_m|^{p_m}\right]$ is finite for $p_j \geqslant 0, j = y, x_l, x_m$. Let U_i be a continuous random variable with density in the tails that is less than that of a t-distribution with $m_u + 1$ degrees of freedom. Letting $\mathcal{A}$ be the set on which $|y| > N^{\delta_y}, |x_l| > N^{\delta_1}, |x_m| > N^{\delta_2}$ and B_i a bounded random variable:

$$\begin{aligned} a) &\quad : \quad \Pr(\mathcal{A}) \leq O\left(N^{-\left[\delta_y p_y + \delta_1 p_1 + \delta_2 p_2\right]}\right). \\ b) &\quad : \quad E(|B_i U_i| \, 1\left\{|U_i| > N^{\delta_u}\right\}) = O(N^{-\delta_u m_u}). \end{aligned}$$

Proof. To establish a), following Markov's argument

$$\begin{aligned} E\left[|Y|^{p_y} |X_l|^{p_l} |X_m|^{p_m}\right] &= \iiint |y|^{p_y} |x_l|^{p_l} |x_m|^{p_m} g(y, x_l, x_m) dy dx_l dx_m \\ &\geqslant N^{\delta_y p_y} N^{\delta_l p_l} N^{\delta_m p_m} \iiint_{\mathcal{A}} g(y, x_l, x_m) dy dx_l dx_m \\ &= N^{\delta_y p_y} N^{\delta_l p_l} N^{\delta_m p_m} \Pr(\mathcal{A}), \end{aligned}$$

and part a) follows. For b), since a t-distribution with df degrees of freedom is proportional to

$$\frac{u}{\left[1 + \frac{u^2}{df}\right]^{(df+1)/2}} < \frac{u}{\left[\frac{u^2}{df}\right]^{(df+1)/2}},$$

the expectation in b) has order given as:

$$O\left(\int_{N^{\delta_u}}^{\infty} u^{-df} du\right) = O\left(N^{-(df-1)\delta_u}\right) = O\left(N^{-m_u \delta_u}\right), m_u \equiv df - 1.$$

Part b) follows. ■

To establish uniform convergence results below, we use a theorem by Bhattacharya (1967) and a modest extension in Klein and Spady (1993). Let $\bar{T}$ denote an i.i.d. average of terms and assume that $N^{-q}\bar{T}$ is bounded. Then, subject to regularity conditions, the theorem essentially says that $\bar{T}$ uniformly converges to its expectation, at a rate that is decreased from the parametric rate by q. In Lemmas 8-9, as results will hold uniformly over the point at which functions are evaluated and over θ, we will denote $M(v; \theta) \equiv E(Y|V(\theta) = v)$ and correspondingly refer to the estimator of this function as $\hat{M}_s(v; \theta)$.

Lemma 8. Uniform Convergence. Let $\nabla^{\lambda} T$ denote the λ^{th} derivative of T with respect to θ, with $\nabla^0 T \equiv T$. For an arbitrarily small $\xi > 0$ and for m_y, m_x referring to the least number of finite absolute moments for Y, X, let:

$$\begin{aligned} p_a(\lambda) &\equiv \left[\frac{m_y}{m_y+1}[1/2-\xi] - r(d+\lambda) - \frac{\lambda}{m_x+1}[1/2-\xi]\right], \\ p_b(\lambda) &\equiv \left[(1/2-\xi) - r(d+\lambda) - \frac{\lambda}{m_x+1}[1/2-\xi]\right]. \end{aligned}$$

With $p_b(\lambda) > p_a(\lambda)$, assume $p_a(\lambda) > 0$. Referring to D5)-D6) and recalling that $V_i \equiv W_i\theta$, define the following components:

$$\begin{aligned} \hat{C}_1(W_i;\theta) &\equiv \hat{g}_1(V_i;\theta)\bar{Y}(V_i;\theta);\ \hat{C}_2(V_i;\theta) \equiv Z'DY/N; \hat{C}_3(W_i;\theta) \equiv \hat{g}_1(V_i;\theta)\bar{Z}(V_i;\theta) \\ \hat{C}_4(V_i;\theta) &\equiv \hat{g}_1(V_i;\theta); \hat{C}_5(V_i;\theta) \equiv Z'DZ/N;\ \hat{C}_6(V_i;\theta) \equiv \hat{g}_2(V_i;\theta). \end{aligned}$$

$$\begin{aligned} C_1(W_i;\theta) &\equiv g(V_i;\theta)M(V_i;\theta);\ C_2(V_i;\theta) \equiv 0; C_3(W_i;\theta) \equiv 0 \\ C_4(V_i;\theta) &\equiv g(V_i;\theta); C_5(V_i;\theta) = g_1(V_i;\theta)I; C_6(V_i;\theta) \equiv g_2(V_i;\theta). \end{aligned}$$

As we will study an upper bound on the sup over observation i, replace W_i with w to obtain $C_j \equiv C_j(w;\theta)$. For w in a compact subset of the support for W, $\lambda = 0,1,2$, for θ in a compact set, and for window parameter r:

$$\begin{aligned} a) &: \sup_{w,\theta}\left|\nabla^{\lambda}\left[\hat{C}_j - C_j\right]\right| = O_p\left(N^{-p_a(\lambda)} + N^{-2rd}\right),\ j=1,2. \\ b) &: \sup_{w,\theta}\left|\nabla^{\lambda}\left[\hat{C}_j - C_j\right]\right| = O_p\left(N^{-p_b(\lambda)} + N^{-2rd}\right), j=3,4,5,6. \\ c) &: \sup_{w,\theta}\left|\nabla^{\lambda}\left[\hat{M}_s(v;\theta) - M(v;\theta)\right]\right| = o_p(1). \end{aligned}$$

Proof. From a standard Taylor series expansion in $h = O(N^{-r})$:

$$\sup_{w,\theta}\left|\nabla^{\lambda}\left[E\left(\hat{C}_j\right) - C_j\right]\right| = O(h^2) = O(N^{-2rd}),\ \text{for } r > 0.$$

Accordingly, it suffices to obtain rates for $\sup_{w,\theta}\left|\nabla^{\lambda}\left[\hat{C}_j - E\left(\hat{C}_j\right)\right]\right|$. Beginning with $\hat{C}_1$, and $\lambda = 0$, split $\hat{C}_1$ into three components $\hat{C}_{11}, \hat{C}_{12}, \hat{C}_{13}$:

$$\hat{C}_1 = \frac{1}{h^d}\langle \tau_y Yk - E[\tau_y Yk]\rangle_{\hat{C}_{11}} + \frac{1}{h^d}\langle [1-\tau_y]Yk\rangle_{\hat{C}12} - \frac{1}{h^d}\langle E[[1-\tau_y]Yk]\rangle_{\hat{C}_{13}},$$

where $\tau_y = 1\left\{|Y| < N^{\delta_y}\right\}, \delta_y > 0$. From standard results in the literature:

$$\sup\left|\hat{C}_{11} - E\left(\hat{C}_{11}\right)\right| = o_p\left(N^{-[1/2-\xi-rd-\delta_y]}\right). \tag{8}$$

Since k is bounded, from A4) and Lemma 7, part b):

$$\sup\left|\hat{C}_{12} - E\left(\hat{C}_{12}\right)\right| = O_p\left(N^{-(m_y\delta_y - rd)}\right).$$

Employing the same argument, $\sup\left|\hat{C}_{13} - E\left(\hat{C}_{13}\right)\right|$ has the same order. Equating this rate to that for $\sup\left|\hat{C}_{11} - E\left(\hat{C}_{11}\right)\right|$ in (8) the result follows.

For the derivative with respect to θ_l :

$$\begin{aligned}\nabla_l^1 \hat{C}_1 &= \frac{1}{h^{d+1}} \left\langle \tau_y \tau_x Y(x_l - X_l) k' \right\rangle_{T_1} + \\ & \quad \frac{1}{h^{d+1}} \left\langle [1-\tau_y\tau_x] \, Y(x_l - X_{lj}) k' \right\rangle_{T_2}.\end{aligned}$$

Since x_l is an element of w that is in a compact set, the order will be determined by the X_{lj} terms. With T_1 and T_2, each containing one such term which we respectively denote as T_1^* and T_2^*, we analyze these below. From Bhattacharya (1967),

$$\sup |T_1^* - E(T_1^*)| \equiv \frac{1}{h^{d+1}} \sup |\langle \tau_y \tau_x Y X_l k' \rangle - E \langle \tau_y \tau_x Y X_l k' \rangle| = O_p\left(N^{-[1/2-\xi-r(d+1)-\delta_y-\delta_x]}\right).$$

For T_2^*, write it as:

$$\begin{aligned}T_2^* &\equiv \left\langle [1-\tau_y\tau_x] \frac{1}{h^{d+1}} Y X_l k' \right\rangle = \left\langle [1-\tau_x] \, \tau_y \frac{1}{h^{d+1}} Y X_l k' \right\rangle_{T_{21}^*} + \\ & \quad \left\langle [1-\tau_y] \, \tau_x \frac{1}{h^{d+1}} Y X_l k' \right\rangle_{T_{22}^*} + \left\langle [1-\tau_y] \, [1-\tau_x] \, \frac{1}{h^{d}} Y_j X_{lj} k_j' \right\rangle_{T_{23}^*}.\end{aligned}$$

From Lemma 7, part b):

$$\sup |T_{21}^* - E(T_{21}^*)| = O_p\left(N^{-[\delta_x m_x - \delta_y - r(d+1)]}\right); \ \sup |T_{22}^* - E(T_{22}^*)| = O_p(N^{-[\delta_y m_y - \delta_x - r(d+1)]}).$$

From Cauchy-Schwarz and Lemma 7, part a):

$$\sup |T_{23}^* - E(T_{23}^*)| = O_p\left(N^{-[[\delta_y m_y + \delta_x m_x]/2 - r(d+1)]}\right).$$

Setting δ_x to equate rates for T_1^* and T_{21}^* terms:

$$\delta_x(m_x+1) = [1/2-\xi] \Rightarrow \delta_x = \frac{1/2-\xi}{m_x+1}. \tag{9}$$

Setting δ_y to equate rates for T_1^* and T_{22}^*:

$$\delta_y(m_y+1) = [1/2-\xi] \Rightarrow \delta_y = \frac{1/2-\xi}{m_y+1}. \tag{10}$$

Employing (9) and (10), $\sup |T_1^* - E(T_1^*)|$ and $\sup |T_2^* - E(T_2^*)|$ are each $O_p(N^{-p_a(1)})$. With δ_y and δ_x set as above, it can be shown that the convergence rate for $\sup |T_{23}^* - E(T_{23}^*)|$ is faster than the other $T_2 - terms$. It then follows from the uniform rate on the bias order being $O(N^{-2rd})$:

$$\sup \left| \nabla^1 \left[\hat{C}_j - C_j \right] \right| = O_p\left(N^{-p_a(1)} + N^{-2rd}\right).$$

For the second derivative of $\hat{C}_1$, let X_{lj} and X_{mj} be the j^{th} observation on any two of the continuous index variables with coefficients θ_l and θ_m respectively. Then, for the cross partial with respect to θ_l and θ_m :

$$\begin{aligned}\nabla_{lm}^2 \hat{C}_1 &= \left\langle \tau_y \tau_{x_l} \tau_{x_m} \frac{1}{h^{d+2}} Y(x_l - X_l)(x_m - X_m) k'' \right\rangle_{T_1} + \\ & \quad \left\langle [1-\tau_y\tau_{x_l}\tau_{x_m}] \, \frac{1}{h^{d+2}} Y(x_l - X_l)(x_m - X_m) k'' \right\rangle_{T_2}.\end{aligned}$$

Similar to the analysis for first derivatives, the uniform convergence rate for T_1 to its expectation is determined by the uniform convergence rate for $T_1^* \equiv \frac{1}{h^{d+2}} \langle \tau_y \tau_{x_l} \tau_{x_m} Y X_l X_m k'' \rangle$ to its expectation:

$$\sup |T_1^* - E(T_1^*)| = O_p(N^{-[1/2-\xi-r(d+2)-\delta_y-2\delta_x]}).$$

For T_2, the convergence rate, as above, will be determined by terms containing $X_l X_m$. With T_2^* denoting these terms, write $T_2^* = T_{21}^* + T_{22}^* + T_{23}^* + T_{24}^*$, where:

$$\begin{aligned}
T_{21}^* &= \frac{1}{h^{d+2}} \langle (1-\tau_y) \tau_{x_l} \tau_{x_m} Y X_l X_m k'' \rangle \\
T_{22}^* &= \frac{1}{h^{d+2}} \langle [(1-\tau_{x_l}) \tau_{x_m} + (1-\tau_{x_m}) \tau_{x_l}] \tau_y Y X_l X_m k'' \rangle \\
T_{23}^* &= \frac{1}{h^{d+2}} \langle (1-\tau_y)(1-\tau_{x_l}) \tau_{x_m} Y X_l X_m k'' \rangle \\
T_{24}^* &= \frac{1}{h^{d+2}} \langle (1-\tau_{xl})(1-\tau_{x_l}) \tau_y Y X_l X_m k'' \rangle \\
T_{25}^* &= \frac{1}{h^{d+2}} \langle (1-\tau_y)(1-\tau_{x_l})(1-\tau_{x_m}) Y X_l X_m k'' \rangle .
\end{aligned}$$

Under Lemma 7, part b):

$$\begin{aligned}
\sup |T_{21}^* - ET_{21}^*| &= O_p\left(N^{-[\delta_y m_y - 2\delta_x - r(d+2)]}\right) \\
\sup |T_{22}^* - ET_{22}^*| &= O_p(N^{-[\delta_x(m_x-1)-\delta_y-r(d+2)]}).
\end{aligned}$$

Equating rates in the T_1^* and T_{21}^* terms:

$$1/2 - \xi - \delta_y = \delta_y m_y \Rightarrow \delta_y = \frac{1}{m_y+1}[1/2-\xi], \delta_y m_y = \frac{m_y}{m_y+1}[1/2-\xi].$$

Equating rates for the T_1^* and T_{22}^* terms:

$$[1/2-\xi] = \delta_x (m_x+1) \Rightarrow \delta_x = \frac{1}{m_x+1}[1/2-\xi].$$

By construction, the convergence rates in T_1^*, T_{21}^*, and T_{22}^* are the same. Employing the definitions of δ_y and δ_x, this rate is $O_p(N^{-p_a(2)})$. For the choices of δ_y and δ_x, it can be shown that convergence rates for T_{23}^*, T_{24}^*, and T_{25}^* are the same or faster than $O_p(N^{-p_a(2)})$. It can also be shown that the convergence rates for second derivatives when $l = m$ is not slower than $O_p(N^{-p_a(2)})$. With the uniform bias being $O(N^{-2rd})$, the proof for part a) with $j = 1$ follows. With $\hat{C}_2$ having the same structure as $\hat{C}_1$, the convergence rate is the same.

With $j = 3$, note that $\hat{C}_3$ has the same kernel structure as the terms in a) with the exception that it does not depend on Y. The convergence rate is then obtained by letting $m_y \to \infty$ in a). As all terms in part b) have the same structure, they all have the same convergence rate. Part b) follows.

Turning to part c), with stage $s = 1$, recall from D6) that:

$$\hat{M}_1(v) \equiv \bar{Y}(v) - \bar{Z}(v) [Z'DZ]^{-1} Z'D \left[Y - \underline{1} \cdot \bar{Y}(v)\right].$$

Note that with ∇ as a derivative operator, for any non-singular matrix A:

$$\nabla \left(AA^{-1}\right) = 0 \Rightarrow \nabla \left(A^{-1}\right) = -A^{-1} \nabla (A) A.$$

Let $\hat{f}_1(v;\theta) \equiv \hat{g}_1(v;\theta)\hat{M}_1(v;\theta)$, $f_1(v;\theta) \equiv g_1(v;\theta)M_1(v;\theta)$, and for $\lambda = 0, 1, 2$:

$$\begin{aligned}\delta(\lambda) &\equiv \nabla^\lambda \left[\hat{g}_1(v;\theta)\left(\hat{M}_1(v;\theta) - M(v;\theta)\right)\right] \\ &\equiv \nabla^\lambda \left[\hat{f}_1(v;\theta) - f_1(v;\theta)\left[\hat{g}_1(v;\theta)/g_1(v;\theta)\right]\right]. \end{aligned} \tag{11}$$

With trimming controlling density denominators, $\hat{g}_1(v;\theta)$ uniformly converges to $g_1(v;\theta)$ with $\inf g_1(v;\theta) = O(N^{-\varepsilon})$ under trimming. It then follows from parts a)-b) that $\delta(\lambda)$ is uniformly $O_p(N^{-p_a(1)}+N^{-2rd})$. For $\lambda = 0$, since $\hat{g}_1(v;\theta) = O_p(1)$ under trimming:

$$D^0) : \sup\left|\hat{M}_1(v;\theta) - M(v;\theta)\right| = O_p\left(N^{-p_a(1)} + N^{-2rd}\right).$$

For $\lambda = 1$:

$$\begin{aligned}&\hat{g}_1(v;\theta)\nabla^1\left[\left(\hat{M}_1(v;\theta) - M(v;\theta)\right)\right] \\ =\ &\delta(1) - \left[\nabla^1\hat{g}_1(v;\theta)\right]\left[\left(\hat{M}_1(v;\theta) - M(v;\theta)\right)\right].\end{aligned}$$

Since $\sup|\delta(1)| = O_p\left(N^{-p_a(1)} + N^{-rd}\right)$ and $\sup\left|\nabla^1\hat{g}_1(v;\theta)\right| = O_p(1)$, it follows from D^0) and trimming that

$$D^1) : \sup\left|\nabla^1\left[\left(\hat{M}_1(v;\theta) - M(v;\theta)\right)\right]\right| = O_p\left(N^{-p_a(1)} + N^{-2rd}\right).$$

Employing results for $\delta(2)$, D^0), and D^1), it can be shown that

$$D^2) : \sup\left|\nabla^2\left[\left(\hat{M}_1(v;\theta) - M(v;\theta)\right)\right]\right| = O_p\left(N^{-p_a(2)} + N^{-2rd}\right).$$

For general stage $s > 1$ and $\lambda = 0$:

$$\begin{aligned}\hat{g}_2(v;\theta)\hat{M}_s(v;\theta) &= \frac{1}{N}\sum_{i=1}^N\left\{Y_i - \left[\hat{M}_{s-1}(V_i;\theta) - \hat{M}_{s-1}(v;\theta)\right]\right\}K_i^*(v) \\ &= \frac{1}{N}\sum_{i=1}^N\left\{\begin{array}{c}\left[\hat{M}_{s-1}(v;\theta) - M(v;\theta)\right] - \\ \left[\hat{M}_{s-1}(V_i;\theta) - M(V_i;\theta)\right]\end{array}\right\}K_i^*(v) \\ &\quad + \frac{1}{N}\sum_{i=1}^N\left\{M(v;\theta) + \left[Y_i - M(V_i;\theta)\right]K_i^*(v)\right\}.\end{aligned}$$

Therefore, $\hat{g}_2(v;\theta)\left[\hat{M}_s(v;\theta) - M(v;\theta)\right] = R_1(s) + R_2$, where:

$$\begin{aligned}R_1(s) &= \frac{1}{N}\sum_{i=1}^N\left\{\begin{array}{c}\left[\hat{M}_{s-1}(v;\theta) - M(v;\theta)\right] - \\ \left[\hat{M}_{s-1}(V_i;\theta) - M(V_i;\theta)\right]\end{array}\right\}K_i^*(v) \\ R_2 &\equiv \frac{1}{N}\sum_{i=1}^N\left\{\left[Y_i - M(V_i;\theta)\right]K_i^*(v)\right\}.\end{aligned} \tag{12}$$

Employing the result for $\hat{C}_6$ in part b) and D^1):

$$\sup|R_1(s)| = O_p\left(N^{-p_a(0)} + N^{-2rd}\right)$$

$$\sup|R_2 - ER_2| = O_p\left(N^{-p_a(0)}\right).$$

With the trimming set expanding to the full support, it can be shown that

$$E(R_2) = E\left[E\left\{\left[Y_i - M\left(V_i;\theta\right)\right]\tau_{sm}\left(X_i, q(N)\right)K_i(v)|V_i\right\}\right] + o(1) = 0.$$

As the second expectation is 0, the result follows. An induction argument then completes the proof for $\lambda = 0$. Similar arguments hold for $\lambda = 1, 2$. ■

Lemma 9 provides the analysis for the adjusted expectations that underlie the estimator $\hat{\theta}_2$. This lemma closely parallels Lemma 8 with the exception of the set for θ over which we seek uniformity for $\hat{M}_s$ and its derivatives. To analyze the Hessian matrix, we only require uniformity over a vanishing neighborhood of θ_0. For the trimming structure employed in Lemma 8, this restriction is not required, but is needed in Lemma 9.

Lemma 9. Uniform Convergence for Adjusted Expectations. Let $\nabla^\lambda T$ denote the λ^{th} derivative of T with respect to θ, with $\nabla^0 T \equiv T$. For X_i the i^{th} observation on the continuous variable vector X_i, assume that X_i is bounded. For $\xi' > 0$, define:

$$\begin{aligned} p'_a(\lambda) &\equiv \frac{m_y}{m_y+1}\left[1/2 - \xi'\right] - r\left(d+\lambda\right), \\ p'_b(\lambda) &= \left(1/2 - \xi'\right) - r\left(d+\lambda\right), \end{aligned}$$

where $p'_a(0) - sar > 0$ with an adjustment parameter a satisfying: $a < 2/s$. Recall the definition of $\hat{C}_j$, $j = 1, 2, 3$ in Lemma 8. From D9) these are the same as in the unadjusted case (e.g. $\hat{C}_1\left(V_i;\theta\right) = \hat{g}_{1a}(V_i;\theta)\bar{Y}_a(V_i;\theta) = \hat{g}(V_i;\theta)\bar{Y}(V_i;\theta)$). Referring to D5)-D6) and D8)-D9), define the following components:

$$\begin{aligned} \hat{C}_{4a}\left(V_i;\theta\right) &\equiv \hat{g}_{1a}(V_i;\theta);\ \hat{C}_{5a}\left(V_i;\theta\right) \equiv Z'D_aZ/N;\ \hat{C}_{6a}\left(V_i;\theta\right) \equiv \hat{g}_{2a}(V_i;\theta) \\ C_{4a}\left(V_i;\theta\right) &\equiv g_{1a}(V_i;\theta); C_{5a}\left(V_i;\theta\right) = g_{2a}(V_i;\theta)I; C_{6a}\left(V_i;\theta\right) \equiv g_{2a}(V_i;\theta). \end{aligned}$$

Let $\mathcal{S}_d$ and $\mathcal{S}_x$ be the support for the discrete and continuous variables respectively. Define $\mathcal{S} \equiv \mathcal{S}_d \cup \mathcal{S}_x$ and let $\mathfrak{N}_p$ be an $o_p(1)$ neighborhood containing θ_0. Then, with $\hat{C}_j, j = 1, 2, 3$ defined as in Lemma 8:

$$\begin{aligned}
a) &: \sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{C}} \left|\nabla^\lambda\left[\hat{C}_j - C_j\right]\right| = O_p(N^{-p'_a(\lambda)} + N^{-2r}), j = 1, 2, 3. \\
b) &: \sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{C}} \left|\nabla^\lambda\left[\hat{C}_{ja} - C_{ja}\right]\right| = O_p(N^{-p'_b(\lambda)} + N^{-2r}), j = 4, 5, 6. \\
c) &: \sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{N}_p} \left|\nabla^\lambda\left[\hat{M}_{sa}\left(v;\theta\right) - M\left(v;\theta\right)\right]\tau(v;q)\right| = O_p\left(N^{-p'_a(\lambda)} + N^{-2r}\right). \\
d) &: \sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{C}} \left|\left[\hat{M}_{sa}\left(v;\theta\right) - M_a\left(v;\theta\right)\right]\right| = O_p\left(N^{-\left[p'_a(0) - sar\right]} + N^{-r[2-sa]}\right).
\end{aligned}$$

Proof. For a)-b), writing sup to mean $\sup\limits_{w;\theta\epsilon\mathfrak{C}}$:

$$\sup\left|\nabla^\lambda\left[\hat{C}_j - C_j\right]\right| \leq \sup\left|\nabla^\lambda\left[\hat{C}_j - E\left(\hat{C}_j\right)\right]\right|_A + \sup\left|\nabla^\lambda\left[E\left(\hat{C}_j\right) - C_j\right]\right|_B.$$

For the A-component, with X having moments of all orders and convergence rates being slowed by at most $N^{-\xi''}, \xi'' > 0$ and arbitrarily small, the rate follows from that in Lemma 8 with ξ replaced by $\xi' \equiv \xi + 2\xi''$ and letting $m_x \to \infty$. It can readily be shown that the bias B-component is $O(N^{-2r})$. Part

a) follows.

For the terms in b), recall the adjustment factor from D8), and consider explicitly the dependence on θ :

$$\hat{A}_s(v;\theta) \equiv \hat{\gamma}_s h^a \left[1 - \tau_\Delta(\hat{g}_s(v;\theta))\right],$$

where $0 < a < 2/s$, γ_s is a lower sample quantile for g_s in D5), and $1 - \tau_\Delta$ component is a smoothed trimming function that is positive in the tails of the index and 0 otherwise. It can be shown that $\sup \left|\nabla_\theta \left[\hat{A}_s(v;\theta) - A_s(v;\theta)\right]\right| = o_p\left(N^{-1/2}\right)$. Employing the same arguments as in Lemma 8 part b), the proof for Lemma 9 part b) follows.

For part c), it suffices to show

$$\sup_{w\epsilon\mathcal{S};,\theta\epsilon\mathfrak{N}_\delta} \left|\nabla^\lambda \left[\hat{M}_{sa}\left(v;\theta\right) - M\left(v;\theta\right)\right] \tau(v\left(\theta_0\right);q)\right| = O_p(N^{-p'_a(\lambda)} + N^{-2r}),$$

where $\mathfrak{N}_\delta = \{\theta : |\theta - \theta_0| < \delta\}$, $\delta = o(1)$.[1] With X_i bounded, $\tau(v\left(\theta_0\right);q)$ restricting $v\left(\theta_0\right)$ to a subset of its support, and $\theta\epsilon\mathfrak{N}_\delta$, then v will be restricted to a compact subset of its support where the index density is bounded away from 0. Then, the result follows from the arguments in Lemma 8, part c).

For part d), let $\hat{f}_{1a}(V_i;\theta) \equiv \hat{g}_{1a}(V_i;\theta)\hat{M}_{1a}\left(v;\theta\right)$, $f_{1a}(V_i;\theta) \equiv g_{1a}(V_i;\theta)M_{1a}\left(v;\theta\right)$, write

$$\begin{aligned}
\delta_a &\equiv \hat{g}_{1a}(V_i;\theta)\left(\hat{M}_{1a}\left(v;\theta\right) - M_a\right) \\
&\equiv \left[\hat{f}_{1a}(V_i;\theta) - f_{1a}(V_i;\theta)\left[\hat{g}_{1a}(V_i;\theta)/g_{1a}(V_i;\theta)\right]\right] \equiv T_1 + T_2, \\
T_1 &\equiv M_a\left[\hat{g}_{1a}(V_i;\theta) - g_{1a}(V_i;\theta)\right];\ T_2 \equiv \hat{f}_{1a}(V_i;\theta) - f_a(V_i;\theta).
\end{aligned}$$

In the first part of the argument, we establish:

$$\sup|\delta_a| = O_p(N^{-p'_a(0)} + N^{-2r}). \tag{13}$$

First, with M_a bounded, from Lemma 9 part b): $\sup|T_1 - E\left(T_1\right)| = O_p(N^{-p'_b(0)})$ and it can be shown that $\sup|E\left(T_1\right)| = O_p(N^{-2r})$. Turning to T_2, it similarly follows that $\sup|T_2| = O_p(N^{-p'_a(0)} + N^{-2r})$. Therefore, for stage $s = 1$,

$$\sup\left|\hat{M}_{1a}\left(v;\theta\right) - M_a\left(v;\theta\right)\right| = O_p\left(N^{-\left[p'_a(0) - ar\right]} + N^{-r(2-a)}\right). \tag{14}$$

From D9), for stage $s = 2$, $\hat{g}_{2a}(v)\hat{M}_{2a}(v;\theta) = T_1 - T_2 + T_3$, where:

$$\begin{aligned}
\hat{g}_{2a}(v)\hat{M}_{2a}(v;\theta) &\equiv \frac{1}{N}\sum_{i=1}^{N}\left\{Y_i - \left[\hat{M}_a(V_i;\theta) - \hat{M}_{1a}\left(v;\theta\right)\right]\right\}K_i^*(v), \\
T_1 &\equiv \frac{1}{N}\sum_{i=1}^{N} Y_i K_i^*(v) \\
T_2 &\equiv \frac{1}{N}\sum_{i=1}^{N}\left\{\left[M_a\left(v;\theta\right) - M_a(V_i;\theta)\right]\right\}K_i^*(v) \\
T_3 &\equiv \frac{1}{N}\sum_{i=1}^{N}\left\{\begin{array}{c}\left[\hat{M}_{1a}\left(v;\theta\right) - M_a\left(v;\theta\right)\right] - \\ \left[\hat{M}_a(V_i;\theta) - M_a(V_i;\theta)\right]\end{array}\right\}K_i^*(v).
\end{aligned}$$

[1] When uniformity holds for $\theta\epsilon\mathfrak{N}_\delta$ for all $\delta = o(1)$, it holds in an $o_p(1)$ neighborhood of θ_0.

With $T_1 = \hat{C}_1$ in part a):

$$\sup |T_1 - E[T_1]| = O_p\left(N^{-p'_a(0)}\right).$$

From a standard Taylor expansion in the kernel window:

$$\sup |E[T_1] - g_2 M(v;\theta)| = O_p\left(N^{-2r}\right).$$

With $M_a(v;\theta) - M_a(V_i;\theta)$ bounded, from the same analysis as in part b):

$$\sup |T_2 - E[T_2]| = O_p\left(N^{-p'_b(0)}\right).$$

From standard arguments, $\sup |E[T_2]| = O_p\left(N^{-2r}\right)$. Finally for stage $s = 2$, from (14), $\sup |T_3| = O_p(N^{-\left[p'_a(0)-2ar\right]} + N^{-r(2-2a)})$. Therefore:

$$\sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{C}} \left|\left[\hat{M}_{2a}(v;\theta) - M_a(v;\theta)\right]\right| = O_p(N^{-\left[p'_a(0)-2ar\right]} + N^{-r[2-2a]}).$$

From an induction argument:

$$\sup_{w\epsilon\mathcal{S};\theta\epsilon\mathfrak{C}} \left|\left[\hat{M}_{sa}(v;\theta) - M_a(v;\theta)\right]\right| = O_p\left(N^{-\left[p'_a(0)-sar\right]} + N^{-r[2-sa]}\right).$$

The lemma follows with $[p'_a(0) - sar] > 0$ and $2 - sa > 0$. ■

With one notable exception, the arguments for taking smooth trimming as given are straightforward and are provided in Theorem 2 and Lemma 11. For one of the gradient components, we will employ the result in Lemma 10 below.

Lemma 10. Smooth Gradient Trimming. Set the stage s, and window parameter r, to satisfy either condition C1) or C2) in Theorem 2. Let $\hat{\tau}_{sm}(i) \equiv \tau_{sm}(X_i;\hat{q}_x)$ or $\tau_{sm}(V_i;\hat{q}_v)$ and correspondingly let $\tau_{sm}(i) \equiv \tau_{sm}(X_i;q_x)$ or $\tau_{sm}(V_i;q_v)$. Refer to D3) and let $\tau \equiv \tau(x, q'_x)$ or $\tau(v, q'_v)$. Let q be either q_x or q_v and q' either q'_x or q'_v. Set quantile vectors $q \equiv (q_1, q_2)$ and $q' \equiv (q'_1, q'_2)$ to satisfy: $q_1 < q'_1 < q'_2 < q_2$. Recalling that the first stage estimator $\hat{M}_1(v;\theta_0)$ does not depend on trimming, for stage $s > 1$:

$$\sup_v \tau \left|\hat{M}_s(v;\theta_0,\hat{\tau}_{sm}) - \hat{M}_s(v;\theta_0,\tau_{sm})\right| = o_p(N^{-1/2}).$$

Proof. For stage $s > 1$:

$$\begin{aligned}
\Delta_s(v,\hat{\tau}_{sm})/\hat{g}_2(v,\hat{\tau}_{sm}) &\equiv \left[\hat{M}_s(v;\theta_0,\hat{\tau}_{sm}) - M(v)\right] \\
\Delta_s(v,\tau_{sm})/\hat{g}_2(v,\tau_{sm}) &\equiv \left[\hat{M}_s(v;\theta_0,\tau_{sm}) - M(v)\right].
\end{aligned}$$

Then, $\left[\hat{M}_s(v;\theta_0,\hat{\tau}_{sm}) - \hat{M}_s(v;\theta_0,\tau_{sm})\right]$ is given as:

$$\left\{\begin{array}{c} \hat{g}_2(v,\tau_{sm})\left[\Delta_s(v,\hat{\tau}_{sm}) - \Delta_s(v,\tau_{sm})\right] - \\ \Delta_s(v,\tau_{sm})\left[\hat{g}_2(v,\hat{\tau}_{sm}) - \hat{g}_2(v,\tau_{sm})\right] \end{array}\right\} \frac{1}{\hat{g}_2(v,\tau_{sm})\,\hat{g}_2(v,\hat{\tau}_{sm})}.$$

The proof will follow by showing that for v in a bounded set:

$$
\begin{aligned}
i) &: \sup_v \tau \left|\Delta_s(v,\hat{\tau}_{sm}) - \Delta_s(v,\tau_{sm})\right| = o_p(N^{-1/2}). \\
ii) &: \sup_v \tau \left|\Delta_s(v,\tau_{sm})\left[\hat{g}_2(v,\hat{\tau}_{sm}) - \hat{g}_2(v,\tau_{sm})\right]\right| = o_p(N^{-1/2}). \\
iii) &: \sup_v \tau \left|\hat{g}_2(v,\hat{\tau}_{sm}) - g_2(v)\right| = o_p(1), \inf_v \tau g_2(v) > 0.
\end{aligned}
$$

Beginning with i) and stage $s=2$, recall that $\hat{\Delta}_2(v,\hat{\tau}_{sm})$ is given as:

$$\tau\left\langle \left[\hat{M}_1(v;\theta_0) - M(v) - \hat{M}_1(V_i;\theta_0) - M(V_i) + \varepsilon_i\right]\hat{\tau}_{sm}K(v)\right\rangle.$$

From D3), and a Taylor series expansion:

$$\hat{\tau}_{sm}(i) = \tau_{sm}(i) + \nabla^1_{q_x}\tau_{sm}(i)\left[\hat{q}-q\right] + \tau rem,$$

where $\sup_v \tau rem = o_p(N^{-1/2})$. The derivative has the replicative structure:

$$\nabla^1_q\tau_{sm}(i) = \left[\ln(N)\right]^2\tau_{sm}(i)B,$$

where B is bounded. Noting that $\left[\ln(N)\right]^2 = o\left(N^{\xi}\right)$, $\xi>0$ and arbitrarily small, the result follows. The proof for i) then follows from an induction argument. The argument for ii) is the same as that above with the convergence of $\Delta_s(v,\tau_{sm})$. From part b) of Lemmas 8 and 9 and trimming, iii) follows, which completes the proof. ■

Lemma 11 Hessian Matrices. Recall the objective functions associated with the estimators in D7) and D10). Define:

$$
\begin{aligned}
\hat{H}_1(\theta) &\equiv \left\langle \tau(X;\hat{q}')\nabla^2_\theta\left\{Y - \hat{M}_s\left[V(\theta);\tau_{sm}(X;\hat{q}(N))\right]\right\}\right\rangle \\
H_1(\theta) &\equiv \left\langle \tau(X;q')\nabla^2_\theta\left\{Y - M\left[V(\theta);\tau_{sm}(X;q(N))\right]\right\}\right\rangle \\
\hat{H}_{2a}(\theta) &\equiv \left\langle \tau\left(V(\hat{\theta}_1);\hat{q}'\right)\nabla^2_\theta\left\{Y - \hat{M}_{sa}\left[V(\theta);\hat{\tau}_{sm}\left(V(\hat{\theta}_1);\hat{q}(N)\right)\right]\right\}\right\rangle \\
H_2(\theta) &\equiv \left\langle \tau\left(V(\theta_0);q'\right)\nabla^2_\theta\left\{Y - M\left[V(\theta);\tau_{sm}\left(V(\theta_0);q(N)\right)\right]\right\}\right\rangle.
\end{aligned}
$$

Referring to Lemma 8, assume

$$\left[\frac{m_y}{m_y+1}\left[1/2-\xi\right] - r(d+2) - \frac{2}{m_x+1}\left[1/2-\xi\right]\right] > 0.$$

Then, with $\mathfrak{N}_p$ an $o_p(1)$ neighborhood of θ_0 :

$$a): \sup_{\theta\xi\mathfrak{N}_p}\left|\hat{H}_1(\theta) - E\left[H_1(\theta)\right]\right| \xrightarrow{p} 0.$$

Assume that the X's are bounded with

$$\left[\frac{m_y}{m_y+1}\left[1/2-\xi\right] - r(d+2)\right] > 0.$$

Then:

$$b): \sup_{\theta\xi\mathfrak{N}_p}\left|\left\langle\hat{H}_{2a}(\theta) - E\left[H_2(\theta)\right]\right\rangle\right| \xrightarrow{p} 0.$$

Proof. For a), employing a similar argument as in Klein and Shen (2010, Proof of Theorem 2), we use a result due to James Powell to provide a smooth upper bound to the difference in indicators. Using a Taylor series for this bound and for the smooth trimming function, it can then be shown that indicator and smooth trimming can be taken as known. The result now follows from Lemma 8, part c).

To establish b), as in a) we may take all trimming as given. The proof then follows from Lemma 9, part c).

■

References

[1] Bhattacharya, P. K. (1967), "Estimation of a Probability Density Function and its Derivatives," *The Indian Journal of Statistics*, Series 4, v. 29, 373-382.

[2] Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single index models," *Journal of Econometrics*, 58, 71-120.

[3] Ichimura, H. and Lee. L. F. (1991), "Semiparametric least squares (SLS) and weighted SLS estimation of multiple index models: Single equation estimation," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics,* ed. W.Barnett, J.Powell and G.Tauchen, Cambridge University Press.

[4] Jiang, Y. (2021), "Semiparametric Estimation of a Corporate Bond Rating Model," *Econometrics*, 9(2): p. 23.

[5] Klein, R. and Shen. C. (2010), "Bias Corrections in Testing and Estimating Semiparametric, Single Index Models," *Econometric Theory,* 1683-1718.

[6] Klein, R. and Spady. R. (1993), "An Efficient Semiparametric Estimator for the Binary Response Model," *Econometrica*, 61, 387-421.

[7] Pakes, A. and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1058.