SUPPLEMENT TO "LOCAL PARTITIONED REGRESSION"

By Norbert Christopeit and Stefan G. N. Hoderlein

IN THIS PAPER we provide additional material on four issues in the main text. The first is an estimator for the covariance matrix Σ in Theorem 3.2, precisely defined in (A.15). The second concerns the claim that under our assumptions, (3.5) and (3.6) are already implied. The third issue is the calculation of the matrix V in Remark 3.3 for general heteroscedasticity. The last issue is the proof of Theorem 3.3.

1. AN ESTIMATOR FOR THE COVARIANCE MATRIX Σ

Let Σ_{12} denote the remaining components of Σ, i.e.,

$$\Sigma_{12} = (\kappa_2 f(x))^{-1} \begin{pmatrix} 1 & \pi_1(x) \\ \pi_1(x) & \pi_2(x) \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \overline{\sigma}^2(x) & \pi_1^{\sigma^2}(x) \\ \pi_1^{\sigma^2}(x) & \pi_2^{\sigma^2}(x) \end{pmatrix} \begin{pmatrix} 1 & \pi_1(x) \\ \pi_1(x) & \pi_2(x) \end{pmatrix}^{-1},$$

with all quantities as defined in Assumptions 1 and 2. To estimate $f(x)$, we propose to use the standard kernel density estimator; to estimate $\pi_v(x) = \mathbb{E}\{P_i^v | X_i = x\}$, $v = 1, 2$, any standard nonparametric regression estimator, e.g., local polynomials, will do. More delicate is the estimation of $\pi_v^{\sigma^2}(x) = \mathbb{E}\{P_i^v \sigma^2(X_i, Z_i) | X_i = x\}$, $v = 1, 2$, and of $\overline{\sigma}^2(x) = \pi_0^{\sigma^2}(x) = \mathbb{E}\{\sigma^2(X_i, Z_i) | X_i = x\}$.

However, estimators can be constructed as in Remark 3.6. Starting out with an estimator for $\sigma^2(X_i, Z_i)$, namely

$$\hat{\sigma}^2(X_j, Z_j) = \left[\sum_{l \neq j} K_l^*(X_j, Z_j) \right]^{-1} \sum_{l \neq j} K_l^*(X_j, Z_j) \check{h}^2 \check{\varepsilon}_l^2,$$

where all quantities are defined in the main text in Remark 3.6, recall that undersmoothing is required. Finally, to obtain estimators for $\pi_v^{\sigma^2}(x)$, $v = 0, 1, 2$,

$$\frac{\sum_{j \neq i} \hat{\sigma}^2(X_j, Z_j) P_j^v L_x(\check{h}^{-1}(X_j - X_i))}{\sum_{j \neq i} L_x(\check{h}^{-1}(X_j - X_i))^2}, \qquad (v = 0, 1, 2),$$

where L_x is a standard, symmetric second order kernel and $\check{h}$ is a pilot bandwidth. Together with the discussion in Remark 3.6, this summarizes the estimation of Σ. In the case of general heteroscedasticity, localization in P is also required.

2. EQUATIONS (3.5) AND (3.6) ARE IMPLIED BY THE ASSUMPTIONS

We establish first the following lemma:

LEMMA S0: *Let* $h(x, z, p)$ *be a nonnegative measurable function such that* $\int h(x, z, p)\, dz\, dp < \infty$ *and* $\int p^2 h(x, z, p)\, dz\, dp < \infty$. *Then if* $h(x, z, p) > 0$ *on a* (z, p) *set of positive Lebesgue measure and* $f(x) > 0$,

$$\text{(S1)} \qquad \frac{\int h(x, z, p)\, dz\, dp}{f(x)} \frac{\int p^2 h(x, z, p)\, dz\, dp}{f(x)} > \left[\frac{\int p h(x, z, p)\, dz\, dp}{f(x)} \right]^2.$$

PROOF: By the Cauchy–Schwarz inequality (in $L^2(dz, dp)$),

$$\text{(S2)} \qquad \left[\int p h(x, z, p)\, dz\, dp \right]^2 = \left[\int p\sqrt{h}\sqrt{h}\, dz\, dp \right]^2$$

$$\leq \int p^2 h\, dz\, dp \int h\, dz\, dp,$$

with equality holding if and only if $p\sqrt{h}$ and $\sqrt{h}$ are linearly dependent, i.e., if there exists a constant λ such that

$$p^2 h(x, z, p) = \lambda h(x, z, p) \quad \text{for a.a. } (z, p).$$

However, this is impossible if $h(x, z, p) > 0$ on a (z, p) set of positive Lebesgue measure, so that strict inequality holds in (S2).   *Q.E.D.*

Special cases—$f(x, z, p)$ continuous density:
   (i) $h(x, z, p) = f(x, z, p) \rightsquigarrow \int f(x, z, p)\, dz\, dp = f(x)$, so that (S1) becomes

$$\frac{\int p^2 f(x, z, p)\, dz\, dp}{f(x)} > \left[\frac{\int p f(x, z, p)\, dz\, dp}{f(x)} \right]^2.$$

Alternatively, with

$$\pi_1(x) = \mathbf{E}\{P | X = x\} = \frac{\int p f(x, p)\, dp}{f(x)},$$

$$\pi_2(x) = \mathbf{E}\{P^2 | X = x\} = \frac{\int p^2 f(x, p)\, dp}{f(x)},$$

the relation

$$\pi_2(x) > \pi_1^2(x)$$

holds (this is the first part of (3.5)).

(ii) $h(x, z, p) = \sigma^2(x, z)f(x, z, p) \rightsquigarrow$ (S1) becomes

$$\frac{\int \sigma^2(x, z)f(x, z)\,dz}{f(x)} \frac{\int p^2\sigma^2(x, z)f(x, z, p)\,dz\,dp}{f(x)}$$

$$> \left[\frac{\int p\sigma^2(x, z)f(x, z, p)\,dz\,dp}{f(x)}\right]^2.$$

Alternatively, with

$$\overline{\sigma}^2(x) = \pi_0^{\sigma^2}(x) = \mathbf{E}\{\sigma^2(X, Z)|X = x\} = \frac{\int \sigma^2(x, z)f(x, z)\,dz}{f(x)},$$

$$\pi_1^{\sigma^2}(x) = \mathbf{E}\{P\sigma^2(X, Z)|X = x\} = \frac{\int p\sigma^2(x, z)f(x, z, p)\,dz\,dp}{f(x)},$$

$$\pi_2^{\sigma^2}(x) = \mathbf{E}\{P^2|X = x\} = \frac{\int p^2\sigma^2(x, z)f(x, z, p)\,dz\,dp}{f(x)},$$

the relation

$$\overline{\sigma}^2(x)\pi_2^{\sigma^2}(x) > \pi_1^{\sigma^2}(x)^2$$

holds (this is the first part of (3.6)).

Second parts of (3.5) *and* (3.6): By Bayes formula, $\overline{\sigma}_P^2(x) = 0$ implies that

$$\int \sigma^2(x, z)[p - \pi_1(x, z)]^2 f(x, z, p)\,dz\,dp = 0$$

or, a forteriori,

$$\sigma^2(x, z)[p - \pi_1(x, z)]^2 = 0$$

for all (z, p) such that $f(x, z, p) > 0$. Hence, if $\sigma^2(x, z) > 0$, $p = \pi_1(x, z)$ for all such (z, p), contradicting the assumption of a joint density $f(x, z, p)$.

3. CALCULATION OF THE V MATRIX FOR GENERAL HETEROSCEDASTICITY

In the case of unrestricted heteroscedasticity, we need the following slight modification of Assumption 2:

ASSUMPTION 2′: We have $\mathbb{E}\{\varepsilon_i|X_i, Z_i, P_i\} = 0$ and $\mathbb{E}\{\varepsilon_i^2|X_i, Z_i, P_i\} = \sigma^2(X_i, Z_i, P_i)$. The function $\sigma^2(x, z, p)$ is positive, continuously differentiable with respect to x in a neighborhood of x_0, and $\sup_{|x-x_0|\leq h/2} \sigma^2(x, z, p) \leq \delta(z, p)$ for some function $\delta(z, p)$ that satisfies $\int \delta(z, p)^2\gamma(z, p)\,dz\,dp < \infty$. Moreover, $\mathbb{E}|\varepsilon_i|^{2\alpha} < \infty$ for some $k > 1$.

To start with, as in Appendix A.4 in the main text, we use the expansion

$$\mathbb{E}\{\sigma^2(X_i, Z_i, P_i)\Phi_i\Phi_i'\}$$
$$= \mathbb{E}\{\sigma^2(x_0, Z_i, P_i)\Phi_i\Phi_i'\} + \mathbb{E}\{\partial_x\sigma^2(X_i^*, Z_i, P_i)(X_i - x_0)\Phi_i\Phi_i'\}$$
$$= V_n + R_n,$$

where $R_n = O(h^2)$.

Calculating the Entries of V_n

The entries of V_n are of the form

(S3) $\qquad h^{2-\mu-\nu}\mathbb{E}\big\{(\xi^{\mu,l} - \overline{\xi^{\mu,l}})(\xi^{\nu,m} - \overline{\xi^{\nu,m}})\sigma^2(x_0, Z, P)\big\}$

(with $\xi^{\mu,l} = W(X - x_0)^\mu P^l$). To calculate these expected value, we need a variant of Corollary A.4 for $\tau(z, p) = \sigma^2(x_0, z, p)$ instead of $\tau(z) = \sigma^2(x_0, z)$. From Lemmas A.2 and A.3 we obtain the expressions (S4)–(S6):

(S4) $\qquad v(\mu, l, \nu, m) \equiv \mathbb{E}\xi^{\mu,l}\xi^{\nu,m}\tau(Z, P) = h^{-1}\mathbb{E}\xi^{\mu+\nu,l+m}\tau(Z, P)$
$$= \begin{cases} h^{\mu+\nu-1}\kappa_{\mu+\nu}A_{l+m} + O(h^{\mu+\nu+1}), & \mu + \nu \text{ even,} \\ O(h^{\mu+\nu}), & \mu + \nu \text{ odd,} \end{cases}$$

with

$$A_{l+m} = \int\int p^{l+m}\tau(z, p)f(x_0, z, p)\,dz\,dp;$$

(S5) $\qquad v(\mu, l, \overline{\nu, m})$
$$\equiv \mathbb{E}\xi^{\mu,l}\overline{\xi^{\nu,m}}\tau(Z, P)$$
$$= \mathbb{E}W(X - x_0)^\mu P^l\tau(Z, P)\overline{W(X - x_0)^\nu P^m}$$
$$= \begin{cases} h^{\mu+\nu-1}\kappa_\mu\kappa_\nu B_{l,m} + o(h^{\mu+\nu-1}), & \text{both } \mu \text{ and } \nu \text{ even,} \\ O(h^{\mu+\nu}), & \mu \text{ or } \nu \text{ odd,} \end{cases}$$

with

$$B_{l,m} = \int \pi_m(x_0, z)\pi_l^{\sigma^2}(x_0, z)f(x_0, z)\,dz,$$

where now $\pi_l^{\sigma^2}(x, z)$ is defined as a (continuous) version of the conditional expectation $\mathbb{E}\{P^l\tau(Z, P)|X = x, Z = z\}$ and $\pi_m(x, z) = \mathbb{E}\{P^l|X = x, Z = z\}$ as in the main text (for $l, m \in \mathbb{N}$) (note that, by definition of $\tau(z, p)$, $\pi_l^{\sigma^2}(x_0, z) =$

$\mathbb{E}\{P^l\sigma^2(X, Z, P)|X = x_0, Z = z\});$

(S6) $v(\overline{\mu, l}, \nu, m)$

$$\equiv \mathbb{E}\overline{\xi^{\mu,l}}\xi^{\nu,m}\tau(Z, P)$$

$$= \mathbb{E}W(X - x_0)^\nu P^m \tau(Z, P)\overline{W(X - x_0)^\mu P^l}$$

$$= \begin{cases} h^{\mu+\nu-1}\kappa_\mu\kappa_\nu B_{m,l} + o(h^{\mu+\nu-1}), & \text{both } \mu \text{ and } \nu \text{ even}, \\ O(h^{\mu+\nu}), & \mu \text{ or } \nu \text{ odd}. \end{cases}$$

By (A.1),

(S7) $\mathbb{E}\{\overline{\xi^{\mu,l}}\xi^{\nu,m}\tau(Z, P)|WZ\} = h\overline{\xi^{\mu,l}}\xi^{\nu,m}\mathbb{E}\{W\tau(Z, P)|WZ\}$

$$= h\overline{\xi^{\mu,l}}\xi^{\nu,m}W\overline{\tau}(Z)$$

$$= \overline{\xi^{\mu,l}}\xi^{\nu,m}\overline{\tau}(Z)$$

with

$$\overline{\tau}(z) = \frac{\int \int_{x_0-h/2}^{x_0+h/2} \tau(z, p)f(x, z, p)\,dx\,dp}{\int_{x_0-h/2}^{x_0+h/2} f(x, z)\,dx}.$$

Whereas

$$\text{numerator} = h \int \int_{-1/2}^{1/2} \tau(z, p)f(x_0 + hs, z, p)\,ds\,dp$$

$$= h \int \tau(z, p)f(x_0, z, p)\,dp + O(h^3),$$

$$\text{denominator} = hf(x_0) + O(h^3),$$

we obtain

$$\overline{\tau}(z) = \underbrace{f(x_0)^{-1} \int \tau(z, p)f(x_0, z, p)\,dp}_{\tau_0(z)} + O(h^2).$$

Inserting into (S7) and taking expectation yields

(S8) $\mathbb{E}\{\overline{\xi^{\mu,l}}\xi^{\nu,m}\tau(Z, P)\} = \mathbb{E}\{\overline{\xi^{\mu,l}}\xi^{\nu,m}[\tau_0(z) + O(h^2)]\}$

$$= \mathbb{E}\{\overline{\xi^{\mu,l}}\xi^{\nu,m}\tau_0(z)\} + O(h^{\mu+\nu+1}),$$

provided the interchange of limit and expectation is justified. Now the first term on the right-hand side of (S8) can be calculated as in Corollary A.4, with

$\tau(z)$ replaced by $\tau_0(z)$:

$$\mathbb{E}\{\overline{\xi^{\mu,l}\xi^{\nu,m}}\tau_0(z)\}$$
$$= \begin{cases} h^{\mu+\nu-1}\kappa_\mu\kappa_\nu B_{l,m}^0 + o(h^{\mu+\nu-1}), & \text{both } \mu \text{ and } \nu \text{ even,} \\ O(h^{\mu+\nu}), & \mu \text{ or } \nu \text{ odd,} \end{cases}$$

with

$$B_{l,m}^0 = \int \tau_0(z)\pi_l(x_0, z)\pi_m(x_0, z)f(x_0, z)\, dz.$$

Hence

(S9) $v(\overline{\mu, l}, \overline{\nu, m})$

$$\equiv \mathbb{E}\{\overline{\xi^{\mu,l}\xi^{\nu,m}}\tau(Z, P)\}$$
$$= \begin{cases} h^{\mu+\nu-1}\kappa_\mu\kappa_\nu B_{l,m}^0 + o(h^{\mu+\nu-1}), & \text{both } \mu \text{ and } \nu \text{ even,} \\ O(h^{\mu+\nu}), & \mu \text{ or } \nu \text{ odd.} \end{cases}$$

Putting (S4)–(S6) and (S9) together, we can calculate

$$w_{\mu l\nu m} = \mathbb{E}(\xi^{\mu,l} - \overline{\xi^{\mu,l}})(\xi^{\nu,m} - \overline{\xi^{\nu,m}})\tau(Z, P),$$

noting that

$$w_{\mu\nu lm} = v(\mu, l, \nu, m) - v(\mu, l, \overline{\nu, m}) - v(\overline{\mu, l}, \nu, m) + v(\overline{\mu, l}, \overline{\nu, m}).$$

When $\mu = \nu = 1$ and $l = m = 0$ (corresponding to entry $(1, 1)$ of V_n),

(S10) $w_{1100} = h\kappa_2 A_0 + O(h^2),$

$$A_0 = \int\int \tau(z, p)f(x_0, z, p)\, dz\, dp;$$

when $\mu = 1$, $\nu = 0$, $l = 0$, and $m = 1$ (entry $(1, 2)$),

$$w_{1001} = O(h);$$

when $\mu = \nu = 1$, $l = 0$, and $m = 1$ (entry $(1, 3)$),

(S11) $w_{1101} = h\kappa_2 A_1 + O(h^2),$

$$A_1 = \int\int p\tau(z, p)f(x_0, z, p)\, dz\, dp;$$

when $\mu = \nu = 0$ and $l = m = 1$ (entry $(2, 2)$),

$$w_{0011} = h^{-1}(A_2 - 2B_{1,1} + B_{1,1}^0) + o(h^{-1})$$

with

(S12) $\quad A_2 - 2B_{1,1} + B_{1,1}^0 = \displaystyle\int\int p^2\tau(z,p)f(x_0,z,p)\,dz\,dp$

$$- 2\int \pi_1(x_0,z)\pi_1^{\sigma^2}(x_0,z)f(x_0,z)\,dz$$

$$+ \int \tau_0(z)\pi_1^2(x_0,z)f(x_0,z)\,dz;$$

when $\mu = 0$, $\nu = 1$, and $l = m = 1$ (entry $(2,3)$),

$$w_{0111} = O(h);$$

when $\mu = \nu = 1$ and $l = m = 1$ (entry $(3,3)$),

(S13) $\quad w_{1111} = h\kappa_2 A_2 + O(h^2),$

$$A_2 = \int\int p^2\tau(z,p)f(x_0,z,p)\,dz\,dp.$$

Taking account of (S3), we find that

$$V_n = \begin{pmatrix} h\kappa_2 A_0 + O(h^2) & O(h^2) & h\kappa_2 A_1 + O(h^2) \\ * & h(A_2 - 2B_{1,1} + B_{1,1}^0) + O(h) & O(h^2) \\ * & * & h\kappa_2 A_2 + O(h^2) \end{pmatrix}$$

and hence $V = (nh)^{-1}\sum_{i=1}^n V_i$ is of the form

(S14) $\quad V = \begin{pmatrix} \kappa_2 A_0 & 0 & \kappa_2 A_1 \\ * & A_2 - 2B_{1,1} + B_{1,1}^0 & 0 \\ * & * & \kappa_2 A_2 \end{pmatrix}.$

Introduce the notation (for $\nu \in \mathbb{N}$)

$$\overline{\sigma}^2(x,z) = \mathbb{E}\{\sigma^2(X,Z,P)|X=x, Z=z\},$$

$$\overline{\sigma}^2(x) = \mathbb{E}\{\sigma^2(X,Z,P)|X=x\},$$

$$\pi_\nu^{\sigma^2}(x) = \mathbb{E}\{P^\nu\sigma^2(X,Z,P)|X=x\},$$

$$\overline{\sigma}_P^2(x) = \pi_2^{\sigma^2}(x_0) - 2\mathbb{E}\{\pi_1(X,Z)\pi_1^{\sigma^2}(X,Z)|X=x_0\}$$

$$+ \mathbb{E}\{\overline{\sigma}^2(X,Z)\pi_1^2(X,Z)|X=x_0\}.$$

Evaluate the constants according to (S10)–(S13), making use of the relations

$$\int\int \varphi(x_0,z,p)f(x_0,z,p)\,dz\,dp = \mathbb{E}\{\varphi(X,Z,P)|X=x_0\}f(x_0),$$

$$\int \varphi(x_0, z, p) f(x_0, z, p)\, dp = \mathbb{E}\{\varphi(X, Z, P)|X = x_0, Z = z\} f(x_0),$$

$$\tau_0(z) = f(x_0)^{-1} \int \tau(z, p) f(x_0, z, p)\, dp$$

$$= \mathbb{E}\{\sigma^2(X, Z, P)|X = x_0, Z = z\}.$$

The result is

$$A_0 = \int \int \tau(z, p) f(x_0, z, p)\, dz\, dp$$

$$= \mathbb{E}\{\sigma^2(X, Z, P)|X = x_0\} f(x_0)$$

$$= \overline{\sigma}^2(x_0) f(x_0),$$

$$A_1 = \int \int p\tau(z, p) f(x_0, z, p)\, dz\, dp$$

$$= \mathbb{E}\{P\sigma^2(X, Z, P)|X = x_0\} f(x_0)$$

$$= \pi_1^{\sigma^2}(x_0) f(x_0),$$

$$A_2 = \int \int p^2 \tau(z, p) f(x_0, z, p)\, dz\, dp$$

$$= \mathbb{E}\{P^2 \sigma^2(X, Z, P)|X = x_0\} f(x_0)$$

$$= \pi_2^{\sigma^2}(x_0) f(x_0),$$

$$B_{11} = \int \pi_1(x_0, z) \pi_1^{\sigma^2}(x_0, z) f(x_0, z)\, dz$$

$$= \mathbb{E}\{\pi_1(X, Z)\pi_1^{\sigma^2}(X, Z)|X = x_0\} f(x_0),$$

$$B_{1,1}^0 = \int \tau_0(z) \pi_1^2(x_0, z) f(x_0, z)\, dz$$

$$= \mathbb{E}\{\tau_0(Z)\pi_1^2(X, Z)|X = x_0\} f(x_0)$$

$$= \mathbb{E}\{\overline{\sigma}^2(X, Z)\pi_1^2(X, Z)|X = x_0\} f(x_0),$$

$$A_2 - 2B_{1,1} + B_{1,1}^0$$

$$= f(x_0)\big[\pi_2^{\sigma^2}(x_0) - 2\mathbb{E}\{\pi_1(X, Z)\pi_1^{\sigma^2}(X, Z)|X = x_0\}$$

$$+ \mathbb{E}\{\overline{\sigma}^2(X, Z)\pi_1^2(X, Z)|X = x_0\}\big]$$

$$= \overline{\sigma}_P^2(x_0).$$

So, finally,

$$(S15) \qquad V = \kappa_2 f(x_0) \begin{pmatrix} \overline{\sigma}^2(x_0) & 0 & \pi_1^{\sigma^2}(x_0) \\ * & \kappa_2^{-1}\overline{\sigma}_P^2(x_0) & 0 \\ * & * & \pi_2^{\sigma^2}(x_0) \end{pmatrix}.$$

This illustrates that the structure of the matrix remains preserved, but with all quantities redefined to take account of the dependence on p. Note that if $\sigma^2(x, z, p)$ is independent of p, all entries of (S15) reduce to the corresponding entries of (A.13) in the main text.

4. PROOF OF THEOREM 3.3

4.1. *General Structure*

In this section we will establish conditions under which the effect of preestimation is of lower order than the leading bias term. In particular, we are interested in the difference between the infeasible estimator that contains unknown conditional expectations,

$$\tilde{\theta}_n = \left[\sum_{i=1}^{n} \Phi_i \Phi_i' \right]^{-1} \sum_{i=1}^{n} \Phi_i V_i,$$

where $\Phi_i = [U_i, Q_{0i}, Q_{1i}]'$ with U_i, Q_{0i}, and Q_{1i} as defined in the main paper, and an estimator that contains preestimators for these conditional expectations, denoted as

$$\hat{\theta}_n = \left[\sum_{i=1}^{n} \widehat{\Phi}_i \widehat{\Phi}_i' \right]^{-1} \sum_{i=1}^{n} \widehat{\Phi}_i \widehat{V}_i.$$

Without further mention, the assumptions of Theorem 3.4 are valid. The structure of the proof is as follows: In the second section, we decompose the difference $\tilde{\theta}_n - \hat{\theta}_n$ into an expression that involves individual differences between infeasible and feasible quantities. For instance, let $\lambda_1 = (nh)^{-1} \sum_{i=1}^{n} U_i V_i$ and $\widehat{\lambda}_1 = (nh)^{-1} \sum_{i=1}^{n} \hat{U}_i \widehat{V}_i$. One individual difference is then

$$\widehat{\lambda}_1 - \lambda_1 = (nh)^{-1} \sum_{i=1}^{n} [\hat{U}_i \widehat{V}_i - U_i V_i].$$

Corresponding to the nine separate entries in $\tilde{\theta}_n$, there will be eight additional differences $\widehat{\lambda}_j - \lambda_j$, $j = 2, \ldots, 9$. In the third section we will derive the rates of convergence for these various differences in detail. First we treat the difference $\widehat{\lambda}_1 - \lambda_1$ in detail. This is done in Lemmas S.1 and S.2. Second, in Lemma S.3

we present the analogous results for the differences $\widehat{\lambda}_j - \lambda_j$, $j = 2, \ldots, 9$. Finally, in the last section which contains Lemmas S.4 and S.5, we synthesize the lemmas to produce the result for the overall estimator.

4.2. *The Structure of the Difference* $\tilde{\theta}_n - \hat{\theta}_n$

In this section we break the estimators into their individual components. To this end, define $\lambda_2 = (nh)^{-1} \sum_{i=1}^{n} Q_{1i} V_i$, $\lambda_3 = (nh)^{-1} \sum_{i=1}^{n} Q_{0i} V_i$, $\lambda_4 = (nh)^{-1} \sum_{i=1}^{n} U_i^2$, $\lambda_5 = (nh)^{-1} \sum_{i=1}^{n} U_i Q_{1i}$, $\lambda_6 = (nh)^{-1} \sum_{i=1}^{n} Q_{1i}^2$, $\lambda_7 = h^{-1} \times (nh)^{-1} \sum_{i=1}^{n} U_i Q_{0i}$, $\lambda_8 = h^{-1} (nh)^{-1} \sum_{i=1}^{n} Q_{0i} Q_{1i}$, and $\lambda_9 = (nh)^{-1} \sum_{i=1}^{n} Q_{0i}^2$. The reason for including h^{-1} in the definition of λ_7 and λ_8 is that only then are both random variables $O_p(1)$ like all the others. With this notation, rewrite $\tilde{\theta}_n$ as

$$\tilde{\theta}_n = \begin{bmatrix} \dfrac{\lambda_1 \lambda_6 - \lambda_2 \lambda_5 + h \lambda_9^{-1} [\lambda_3 \lambda_5 \lambda_8 - \lambda_3 \lambda_6 \lambda_7] + h^2 \lambda_9^{-1} [\lambda_2 \lambda_7 \lambda_8 - \lambda_1 \lambda_8^2]}{\lambda_4 \lambda_6 - \lambda_5^2 + h^2 \lambda_9^{-1} [2\lambda_5 \lambda_7 \lambda_8 - \lambda_4 \lambda_8^2 - \lambda_6 \lambda_7^2]} \\[2ex] \dfrac{\lambda_9^{-1} [\lambda_3 \lambda_4 \lambda_6 - \lambda_3 \lambda_5^2 + h [\lambda_1 (\lambda_5 \lambda_8 - \lambda_6 \lambda_7) + \lambda_2 (\lambda_5 \lambda_7 - \lambda_4 \lambda_8)]]}{\lambda_4 \lambda_6 - \lambda_5^2 + h^2 \lambda_9^{-1} [2\lambda_5 \lambda_7 \lambda_8 - \lambda_4 \lambda_8^2 - \lambda_6 \lambda_7^2]} \\[2ex] \dfrac{\lambda_2 \lambda_4 - \lambda_1 \lambda_5 + h \lambda_9^{-1} [\lambda_3 \lambda_5 \lambda_7 - \lambda_3 \lambda_4 \lambda_8] + h^2 \lambda_9^{-1} [\lambda_1 \lambda_7 \lambda_8 - \lambda_2 \lambda_7^2]}{\lambda_4 \lambda_6 - \lambda_5^2 + h^2 \lambda_9^{-1} [2\lambda_5 \lambda_7 \lambda_8 - \lambda_4 \lambda_8^2 - \lambda_6 \lambda_7^2]} \end{bmatrix}.$$

Now, let $\lambda = (\lambda_1, \ldots, \lambda_9)'$. Then $\tilde{\theta}_n = \varphi(\lambda)^{-1} \psi(\lambda)$, where $\varphi(\lambda)$ is the denominator, i.e., $\lambda_4 \lambda_6 - \lambda_5^2 + h^2 \lambda_9^{-1} [2\lambda_5 \lambda_7 \lambda_8 - \lambda_4 \lambda_8^2 - \lambda_6 \lambda_7^2]$, and $\psi(\lambda)$ is the 3×1 vector of numerators. Finally, let the differences $\widehat{\lambda}_j - \lambda_j$, $j = 1, \ldots, 9$, be defined in the obvious fashion by replacing conditional expectations with suitable estimators. Then

$$\tilde{\theta}_n - \hat{\theta}_n = \varphi(\lambda)^{-1} \psi(\lambda) - \varphi(\widehat{\lambda})^{-1} \psi(\widehat{\lambda})$$
$$= [\varphi(\lambda) \varphi(\widehat{\lambda})]^{-1} [\varphi(\widehat{\lambda}) - \varphi(\lambda)] \psi(\lambda) + \varphi(\widehat{\lambda})^{-1} [\psi(\lambda) - \psi(\widehat{\lambda})].$$

Because $\varphi(\widehat{\lambda}) = \varphi(\lambda) + o_p(1)$ and $\varphi(\lambda)$ as well as $\psi(\lambda)$ are $O_p(1)$, the crucial differences are $\psi(\lambda) - \psi(\widehat{\lambda})$ and $\varphi(\widehat{\lambda}) - \varphi(\lambda)$. Start with the differences for the estimators for the first derivative k', i.e., $\tilde{\theta}_1 - \hat{\theta}_1$, where the subscript denotes the first component. With ψ_1 denoting the first component of ψ, we obtain

$$\psi_1(\lambda) - \psi_1(\widehat{\lambda})$$
$$= (\lambda_1 \lambda_6 - \widehat{\lambda}_1 \widehat{\lambda}_6) - (\lambda_2 \lambda_5 - \widehat{\lambda}_2 \widehat{\lambda}_5)$$
$$+ h(\lambda_9 \widehat{\lambda}_9)^{-1} \{\widehat{\lambda}_9 [\lambda_3 \lambda_5 \lambda_8 - \lambda_3 \lambda_6 \lambda_7] - \lambda_9 [\widehat{\lambda}_3 \widehat{\lambda}_5 \widehat{\lambda}_8 - \widehat{\lambda}_3 \widehat{\lambda}_6 \widehat{\lambda}_7]\}$$
$$+ h^2 (\lambda_9 \widehat{\lambda}_9)^{-1} \{\widehat{\lambda}_9 [\lambda_2 \lambda_7 \lambda_8 - \lambda_1 \lambda_8^2] - \lambda_9 [\widehat{\lambda}_2 \widehat{\lambda}_7 \widehat{\lambda}_8 - \widehat{\lambda}_1 \widehat{\lambda}_8^2]\}.$$

Consider in turn the first difference on the right-hand side, that is,

$$\lambda_1 \lambda_6 - \widehat{\lambda}_1 \widehat{\lambda}_6 = \lambda_1 (\lambda_6 - \widehat{\lambda}_6) + \widehat{\lambda}_6 (\lambda_1 - \widehat{\lambda}_1).$$

Because $\widehat{\lambda}_6 = \lambda_6 + o_p(1)$,

$$\lambda_1 \lambda_6 - \widehat{\lambda}_1 \widehat{\lambda}_6 = O_p(\lambda_1 - \widehat{\lambda}_1) + O_p(\lambda_6 - \widehat{\lambda}_6).$$

Similarly, the second expression in brackets on the right-hand side is of order $O_p(\lambda_2 - \widehat{\lambda}_2) + O_p(\lambda_5 - \widehat{\lambda}_5)$, while the third is of order

$$h\{O_p(\lambda_3 - \widehat{\lambda}_3) + O_p(\lambda_5 - \widehat{\lambda}_5) + O_p(\lambda_8 - \widehat{\lambda}_8) + O_p(\lambda_9 - \widehat{\lambda}_9)\}$$

and the fourth is of order h^2 times the order of the slowest converging component. By similar arguments for the denominator, we obtain that

$$\tilde{\theta}_1 - \hat{\theta}_1 = O_p(\lambda_j - \widehat{\lambda}_j) + O_p(h(\lambda_k - \widehat{\lambda}_k))$$

$$(j = 1, 2, 4, 5, 6, k = 3, 7, 8, 9).$$

At this point, it is imperative to note that j covers only those terms that are free of Q_{0i}, while terms that involve Q_{0i} enter only with an additional h. The same applies for the difference in the estimators of the second derivative g'.

Finally, note that the difference between the estimators of the function $h^{-1}g$, i.e., $\tilde{\theta}_2 - \hat{\theta}_2$, behaves differently. Because $\theta_2 = h^{-1}g$, we obtain

$$\tilde{g} - \hat{g} = O_p(h(\lambda_j - \widehat{\lambda}_j)) + O_p(h^2(\lambda_k - \widehat{\lambda}_k))$$

$$(j = 3, 4, 5, 6, 9, k = 1, 2, 7, 8).$$

4.3. The Behavior of the Differences $\widehat{\lambda}_j - \lambda_j$, $j = 1, \ldots, 9$

4.3.1. The behavior of $\widehat{\lambda}_1 - \lambda_1$

In this subsection we will concentrate on the difference $\widehat{\lambda}_1 - \lambda_1$. This difference may be rewritten as

$$(nh)^{-1}\left(\sum_{i=1}^{n} U_i V_i - \sum_{i=1}^{n} \hat{U}_i \widehat{V}_i\right) = T_{1n} + T_{2n} + T_{3n},$$

where $T_{1n} = (nh)^{-1}\sum_{i=1}^{n} U_i S_i$, $T_{2n} = (nh)^{-1}\sum_{i=1}^{n} V_i G_i$ and $T_{3n} = (nh)^{-1} \times \sum_{i=1}^{n} S_i G_i$, with $S_i = \widehat{\mathbb{E}}[W_i Y_i | W_i Z_i] - \mathbb{E}[W_i Y_i | W_i Z_i]$ and $G_i = \widehat{\mathbb{E}}[W_i X_i | W_i Z_i] - \mathbb{E}[W_i X_i | W_i Z_i]$. We will establish in Lemma S.1 conditions under which T_{3n} converges faster than the leading bias term, i.e., is $o_p(h^2)$. More precisely:

LEMMA S.1: *We have $T_{3n} = o_p(h^2)$ if $h_0^{r_0-1} h_1^{r_1} h^{-2} + (n h_0^{(d+3)/2} h_1^{(d+1)/2})^{-1} h^{-2} \times \ln(n) = o(1)$.*

In the following lemmas we establish that under the same conditions T_{3n} dominates T_{1n} and T_{2n} asymptotically. Intuitively, this is due to the independent and identically distributed structure of the data. The terms U_i and S_i (as well as V_i and G_i) are almost uncorrelated, because only the ith observation contained in $\widehat{\mathbb{E}}[W_i Y_i | W_i Z_i]$ is a potential source of correlation. However, this observation has only influence of order n^{-1}.

LEMMA S.2: *Under the same assumptions, T_{1n} and T_{2n} are $o_p(T_{3n})$.*

4.3.2. *The differences $\widehat{\lambda}_j - \lambda_j$, $j = 2, \ldots, 9$*

In this subsection we focus on the speed of convergence of the remaining differences. By closer inspection, we see that all individual differences have approximately the same structure. In particular, it is true that we may always decompose these expressions into a product of the difference between two estimators that depend on all observations and on an "observation i only" residual. By the same argument as in Lemma S.2, we may hence focus on the terms that involve the product of two estimation errors, i.e.,

$$(\widehat{\mathbb{E}}[W_i A_i | W_i Z_i] - \mathbb{E}[W_i A_i | W_i Z_i])(\widehat{\mathbb{E}}[W_i B_i | W_i Z_i] - \mathbb{E}[W_i B_i | W_i Z_i])$$

for $A_i, B_i \in \{X_i, Y_i, P_i, P_i(X_i - x_0)\}$. The following lemma summarizes the results, which follow similar arguments as in Lemma S.1.

LEMMA S.3: *The difference $\widehat{\lambda}_j - \lambda_j$ is of order $o_p(h^2)$ if*

$$\text{for} \quad j = 2, \quad h_1^{r_1} h_3^{r_3-1} h^{-2} + \left(n h_1^{(d+1)/2} h_3^{(d+3)/2}\right)^{-1} h^{-2} \ln(n) = o(1);$$

$$\text{for} \quad j = 3, \quad h_1^{r_1} h_2^{r_2} h^{-3} + \left(n h_1^{(d+1)/2} h_2^{(d+1)/2}\right)^{-1} h^{-3} \ln(n) = o(1);$$

$$\text{for} \quad j = 4, \quad h_0^{2(r_0-1)} + (n h_0^{(d+3)})^{-1} \ln(n) = o(1);$$

$$\text{for} \quad j = 5, \quad h_0^{r_0-1} h_3^{r_2-1} + \left(n h_0^{(d+3)/2} n h_3^{(d+3)/2}\right)^{-1} \ln(n) = o(1);$$

$$\text{for} \quad j = 6, \quad h_3^{2(r_3-1)} + (n h_3^{(d+3)})^{-1} \ln(n) = o(1);$$

$$\text{for} \quad j = 7, \quad h_0^{r_0-1} h_2^{r_2} h^{-3} + \left(n h_0^{(d+1)/2} h_2^{(d+3)/2}\right)^{-1} h^{-3} \ln(n) = o(1);$$

$$\text{for} \quad j = 8, \quad h_2^{r_2} h_3^{r_3-1} h^{-3} + \left(n h_2^{(d+1)/2} h_3^{(d+3)/2}\right)^{-1} h^{-3} \ln(n) = o(1);$$

$$\text{for} \quad j = 9, \quad h_2^{2r_2} h^{-2} + (n h_2^{(d+3)})^{-1} h^{-2} \ln(n) = o(1).$$

Note that for $j = 4, 5, 6$, these conditions are always fulfilled; only $j = 1, 2, 3, 7, 8, 9$ are restrictive. In particular, $j = 3, 7, 8$ are hard to fulfill but they enter only premultiplied by h and thus have an impact on the overall expression similar to $j = 1, 2, 9$.

4.4. *Synthesizing the Results*

In this section we derive the implications of the speed of convergence of the various individual differences for the components of the estimators. Start with the estimators for the derivatives k' and g':

LEMMA S.4: *We have* $(\tilde{\theta}_1 - \hat{\theta}_1, \tilde{\theta}_3 - \hat{\theta}_3)' = o_p(h^2)$ *if*

$$\sum_{s=0,3}\sum_{t=1,2}\left\{h_s^{r_s-1}h_t^{r_t} + \left(nh_s^{(d+3)/2}h_t^{(d+1)/2}\right)^{-1}\ln(n)\right\} = o(h^2).$$

Similarly, for the estimators of the function g, we can state:

LEMMA S.5: *We have* $\tilde{g} - \hat{g} = o_p(h^2)$ *if* $h_1^{r_1}h_2^{r_2} + (nh_1^{(d+1)/2}h_2^{(d+1)/2})^{-1}\ln(n) = o(h^2)$.

Both proofs are trivial and left to the reader. The only thing to notice in the proofs is that although some of the terms disappear always if mean squared error optimal rates for the choice of bandwidth are assumed (e.g., $j = 4, 5, 6$), others are dominated by the leading terms displayed. This concludes the results.

4.5. *Proofs*

PROOF OF LEMMA S.1: Suppressing the subscript on $\hat{f}_{-i}$, $\widehat{mf}_{-i}$ for ease of notation, rewrite

$$T_{3n} = \int\int h^{-3}J(x)\left[\frac{\int J(t)t\hat{f}(t,z)\,dt}{\int J(t)\hat{f}(t,z)\,dt} - \frac{\int J(t)tf(t,z)\,dt}{\int J(t)f(t,z)\,dt}\right]$$

$$\times\left[\frac{\int J(t)\widehat{mf}(t,z)\,dt}{\int J(t)\hat{f}(t,z)\,dt} - \frac{\int J(t)m(t,z)f(t,z)\,dt}{\int J(t)f(t,z)\,dt}\right]\hat{F}(dx,dz),$$

where we use the shorthand $J(t) = \mathbb{1}_{\{|t-x_0|\leq h/2\}}$ and $\hat{F}$ is the empirical cumulative distribution function of X_i and Z_i. Next, consider

$$\frac{\int J(t)t\hat{f}(t,z)\,dt}{\int J(t)\hat{f}(t,z)\,dt} - \frac{\int J(t)tf(t,z)\,dt}{\int J(t)f(t,z)\,dt}$$

$$= \frac{\int J(t)(t-x_0)[\hat{f}(t,z)-f(t,z)]\,dt\int J(t)f(t,z)\,dt}{(\int J(t)f(t,z)\,dt)^2 + \int J(t)[\hat{f}(t,z)-f(t,z)]\,dt\int J(t)f(t,z)\,dt}$$

$$- \int J(t)(t-x_0)f(t,z)\,dt\int J(t)[\hat{f}(t,z)-f(t,z)]\,dt$$

$$\times \left(\left(\int J(t) f(t,z)\, dt \right)^2 \right.$$

$$\left. + \int J(t)[\hat{f}(t,z) - f(t,z)]\, dt \int J(t) f(t,z)\, dt \right)^{-1}.$$

By standard arguments,

$$\int J(t) f(t,z)\, dt = h f(x_0, z) + h^3 \gamma_1,$$

$$\int J(t)(t - x_0) f(t,z)\, dt = h^3 \partial_x f(x_0, z) + h^4 \gamma_2,$$

where $|\gamma_1| = |\int \psi^2 K(\psi)\, \partial_x^2 f(\psi_r, z)\, d\psi|$ and $|\gamma_2| = |\int \psi^3 K(\psi)\, \partial_x^2 f(\psi_r, z)\, d\psi|$ are bounded by $c \sup_{x,z} |\partial_x^2 f(x,z)|$. Here c is a generic constant. Analogously,

$$\int J(t)[\hat{f}(t,z) - f(t,z)]\, dt = h[\hat{f}(x_0, z) - f(x_0, z)] + h^3 \eta_1,$$

$$\int J(t)(t - x_0)[\hat{f}(t,z) - f(t,z)]\, dt$$

$$= h^3[\partial_x \hat{f}(x_0, z) - \partial_x f(x_0, z)] + h^4 \eta_2,$$

where

$$\eta_1 = \int \psi^2 K(\psi)[\partial_x^2 \hat{f}(\psi_r, z) - \partial_x^2 f(\psi_r, z)]\, d\psi,$$

$$\eta_2 = \int \psi^3 K(\psi)[\partial_x^2 \hat{f}(\psi_r, z) - \partial_x^2 f(\psi_r, z)]\, d\psi,$$

and $|\eta_1|, |\eta_2| \leq c \sup_{x,z} |\partial_x^2 \hat{f} - \partial_x^2 f| = O_p(h_0^{r_0 - 2} + (\ln(n)/n h_0^{d+5})^{1/2})$. Therefore,

$$(S16) \quad \frac{\int J(t) t \hat{f}(t,z)\, dt}{\int J(t) \hat{f}(t,z)\, dt} - \frac{\int J(t) t f(t,z)\, dt}{\int J(t) f(t,z)\, dt}$$

$$= \frac{h^2[\partial_x \hat{f}(x_0, z) - \partial_x f(x_0, z)] f(x_0, z) + \theta}{f(x_0, z)[f(x_0, z) + [\hat{f}(x_0, z) - f(x_0, z)]] + \tau},$$

where $\sup_{x,z} |\theta| = o_p(h^2 \sup_{x,z} |\partial_x \hat{f}(x,z) - \partial_x f(x,z)|)$ and $\sup_{x,z} |\tau| = O_p(h^2)$. Hence, with probability approaching 1, $|\tau| < d$ for a constant $d > 0$. Moreover, $|\hat{f}(x,z) - f(x,z)| \leq b/2$ with probability approaching 1, and $|f(x,z)| \geq b$ due to the assumption of continuously distributed random variables with compact

support. Thus, $|f(x_0, z)[f(x_0, z) + [\hat{f}(x_0, z) - f(x_0, z)]]| \geq b^2/2$, and by choosing $d = b^2/4$, the denominator is bounded from below by $b^2/4$.

By similar derivations,

$$
\text{(S17)} \quad \frac{\int J(t)\widehat{mf}(t, z)\, dt}{\int J(t)\hat{f}(t, z)\, dt} - \frac{\int J(t)m(t, z)f(t, z)\, dt}{\int J(t)f(t, z)\, dt}
$$

$$
= (\widehat{mf}(x_0, z) - mf(x_0, z) + \lambda)\big([f(x_0, z) + h^2\gamma_1]^2
$$

$$
+ [f(x_0, z) + h^2\gamma_1][[\hat{f}(x_0, z) - f(x_0, z)] + h^2\eta_1]\big)^{-1},
$$

where $\sup_{x,z}|\lambda| = o_p(\sup_{x,z}|\widehat{mf}(x, z) - mf(x, z)|)$. Using (S16) and (S17), as well as the boundedness of the denominator, we obtain

$$
|T_{3n}| \leq c \int \int h^{-1} J(x) |\partial_x \hat{f}(x_0, z) - \partial_x f(x_0, z)|
$$

$$
\times |\widehat{mf}(x_0, z) - mf(x_0, z)| \hat{F}(dx, dz) + R_n,
$$

where R_n contains faster converging terms that involve θ and λ. Then

$$
|T_{3n}| \leq c \sup_{x,z}|\partial_x \hat{f} - \partial_x f| \sup_{x,z}|\widehat{mf} - mf| \int \int h^{-1} J(x) \hat{F}(dx, dz) + R_n.
$$

Finally, after change of variables, $\int \int h^{-1} J(x) \hat{F}(dx, dz)$ is bounded by $\widehat{f}(x_0, z)$. Hence, employing standard results, we obtain that $T_{3n} = O_p(h_0^{r_0-1} h_1^{r_1} + (nh_0^{(d+3)/2} h_1^{(d+1)/2})^{-1} \ln(n))$. Thus, $T_{3n} = o_p(h^2)$ if $h_0^{r_0-1} h_1^{r_1} h^{-2} + (nh_0^{(d+3)/2} \times h_1^{(d+1)/2})^{-1} h^{-2} \ln(n) = o(1)$. $\hfill$ Q.E.D.

PROOF OF LEMMA S.2: Consider first T_{1n}. Instead of T_{1n}, we will consider T_{1n}^*, where $T_{1n}^* = (nh)^{-1} \sum_{i=1}^n U_i S_i^*$ and

$$
S_i^* = W_i \left[\frac{\int J(t)\widehat{mf}_{-i}(t, Z_i)\, dt}{\int J(t)\hat{f}_{-i}(t, Z_i)\, dt + n^{-\alpha}} - \frac{\int J(t)m(t, Z_i)f(t, Z_i)\, dt}{\int J(t)f(t, Z_i)\, dt} \right]
$$

contains an additional $n^{-\alpha}$, with $\alpha > 0$. Note that in this case

$$
T_{1n} - T_{1n}^*
$$

$$
= (nh)^{-1} \sum_{i=1}^n U_i[S_i - S_i^*]
$$

$$
= n^{-\alpha}(nh)^{-1} \sum_{i=1}^n U_i \frac{\int J(t)\widehat{mf}_{-i}(t, Z_i)\, dt}{[\int J(t)\hat{f}_{-i}(t, Z_i)\, dt + n^{-\alpha}] \int J(t)\hat{f}_{-i}(t, Z_i)\, dt}
$$

$$
= o_p(T_{1n}),
$$

provided that $h = o(n^\alpha)$. This is done to ensure that in the following calculations all expectations exist.

Note that due to Lemma 3.1, S_i^* is $\mathcal{F}_n$-measurable, where $\mathcal{F}_n = \sigma(W_i\zeta_{-i})$ and $\zeta_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n, Z_1, \ldots, Z_n, Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n)$. The $-i$ indicates that the sigma algebra does not contain Y_i and X_i. Then

$$\mathbb{E}\big\{(W_iX_i - \mathbb{E}[W_iX_i|W_iZ_i])S_i^*\big\} = 0$$

by iterated expectations. Turning to the variance, this is

$$(S18) \quad \mathbb{V}\bigg\{\sum_i \frac{U_iS_i^*}{nh}\bigg\} = \mathbb{E}\bigg\{\sum_{i,j}\frac{U_iS_i^*S_j^*U_j}{n^2h^2}\bigg\}$$

$$= \mathbb{E}\bigg\{\sum_i\frac{U_i^2S_i^{*2}}{n^2h^2}\bigg\} + \mathbb{E}\bigg\{2\sum_{i,j>i}\frac{U_iS_i^*S_j^*U_j}{n^2h^2}\bigg\}.$$

In Step 1, we treat the second right-hand side term in (S18); in Step 2, we treat the first.

Step 1: Rewrite the second term as

$$2(nh^2)^{-2}$$

$$\times \sum_{i,j>i}\mathbb{E}\bigg\{J(i)J(j)\bigg[\frac{\int J(t)\widehat{mf}_{-i}(t, Z_i)\,dt}{\int J(t)\hat{f}_{-i}(t, Z_i)\,dt + n^{-\alpha}} - \frac{\int J(t)mf(t, Z_i)\,dt}{\int J(t)f(t, Z_i)\,dt}\bigg]$$

$$\times\bigg[\frac{\int J(t)\widehat{mf}_{-j}(t, Z_j)\,dt}{\int J(t)\hat{f}_{-j}(t, Z_j)\,dt + n^{-\alpha}} - \frac{\int J(t)mf(t, Z_j)\,dt}{\int J(t)f(t, Z_j)\,dt}\bigg]U_iU_j\bigg\},$$

where, in a slight abuse of notation, we write $J(i) = J(X_i)$. Denote $\mathcal{F}_{-i,n} = \sigma(W_1\zeta_{-i}, \ldots, W_i\zeta_{-i}, \ldots, W_n\zeta_{-i})$. By iterated expectations,

$$(S19) \quad 2(nh^2)^{-2}$$

$$\times \sum_{i,j>i}\mathbb{E}\bigg\{J(i)J(j)\bigg[\frac{\int J(t)\widehat{mf}_{-i}(t, Z_i)\,dt}{\int J(t)\hat{f}_{-i}(t, Z_i)\,dt + n^{-\alpha}} - \frac{\int J(t)mf(t, Z_i)\,dt}{\int J(t)f(t, Z_i)\,dt}\bigg]$$

$$\times\mathbb{E}\bigg[\bigg[\frac{\int J(t)\widehat{mf}_{-j}(t, Z_j)\,dt}{\int J(t)\hat{f}_{-j}(t, Z_j)\,dt + n^{-\alpha}}$$

$$- \frac{\int J(t)mf(t, Z_j)\,dt}{\int J(t)f(t, Z_j)\,dt}\bigg]U_i\bigg|\mathcal{F}_{-i,n}\bigg]U_j\bigg\}.$$

The whole expression would be zero if $\widehat{mf}_{-j}$ and $\hat{f}_{-j}$ were not functions of (X_i, Y_i) and X_i, respectively. To see this, note first that in this case S_i^* would be $\mathcal{F}_{-i,n}$-measurable; second note that

$$\mathbb{E}[U_i|\mathcal{F}_{-i,n}] = \mathbb{E}[W_i X_i - \mathbb{E}[W_i X_i|W_i Z_i]|\mathcal{F}_{-i,n}] = 0,$$

due to independence and identical distribution. Hence, the entire dependence hinges only on $\widehat{mf}_{-j} = (nh_1^{d+1})^{-1}\sum_{s\neq j} K_s^*(t, j)Y_s$, where $K_s^*(t, j) = K(h_1^{-1} \times (X_s - t), h_1^{-1}(Z_s - Z_j))$, depending on X_i and Y_i. Applying the decomposition yields

$$\widehat{mf}_{-j}(t, Z_j) = \widehat{mf}_{-(i,j)}(t, Z_j) + (nh_1^{d+1})^{-1}K_i^*(t, j)\{m(X_i, Z_i) + \varepsilon_i\},$$

where $\widehat{mf}_{-(i,j)}$ is $(nh_1^{d+1})^{-1}\sum_{s\neq j,i} K_s^*(t, j)Y_s$, i.e., does not have normalization $((n-1)h_1^{d+1})^{-1}$. Applying a similar type of decomposition to $\hat{f}_{-j}(t, Z_j)$, then

$$\frac{\int J(t)\widehat{mf}_{-j}(t, Z_i)\,dt}{\int J(t)\hat{f}_{-j}(t, Z_i)\,dt + n^{-\alpha}}$$

$$= \frac{\int J(t)\widehat{mf}_{-(i,j)}(t, Z_i)\,dt}{\int J(t)\hat{f}_{-(i,j)}(t, Z_i)\,dt + n^{-\alpha}} + (nh_0^{d+1})^{-1}[Q_{1i} + Q_{2i} + Q_{3i}],$$

where

$$Q_{1i} = \left[\int J(t)\hat{f}_{-j}(t, Z_i)\,dt + n^{-\alpha}\right]^{-1}\int J(t)K_i^*(t, j)m(X_i, Z_i)\,dt,$$

$$Q_{2i} = \left[\int J(t)\hat{f}_{-j}(t, Z_i)\,dt + n^{-\alpha}\right]^{-1}\int J(t)K_i^*(t, j)\varepsilon_i\,dt,$$

and

$$Q_{3i} = -\left[\left(\int J(t)\hat{f}_{-j}(t, Z_i)\,dt + n^{-\alpha}\right)\right.$$

$$\left. \times \left(\int J(t)\hat{f}_{-(i,j)}(t, Z_i)\,dt + n^{-\alpha}\right)\right]^{-1}$$

$$\times \int J(t)\widehat{mf}_{-(i,j)}(t, Z_i)\,dt \int J(t)K_i^*(t, j)\,dt.$$

Note that all of Q_{1i}, Q_{2i}, and Q_{3i} are $\mathcal{F}_{-j,n} = \sigma(W_1\zeta_{-j}, \ldots, W_i\zeta_{-j}, \ldots, W_n\zeta_{-j})$-measurable, because they do not depend on Y_j and X_j. Because the leading term $\widehat{mf}_{-(i,j)}(t, Z_j)/\hat{f}_{-(i,j)}(t, Z_j)$ is not a function of i anymore, if substituted into the summation (S18), the corresponding expression has expectation zero

as mentioned above. Thus, (S19) becomes

$$(S20) \quad \frac{2}{n^3 h^4 h_1^{d+1}}$$

$$\times \sum_{i,j>i} \mathbb{E}\left\{ J(i)J(j)\left[\frac{\int J(t)\widehat{mf}_{-i}(t, Z_i)\, dt}{\int J(t)\hat{f}_{-i}(t, Z_i)\, dt + n^{-\alpha}} - \frac{\int J(t)mf(t, Z_i)\, dt}{\int J(t)f(t, Z_i)\, dt} \right] \right.$$

$$\left. \times [Q_{1i} + Q_{2i} + Q_{3i}]U_i U_j \right\}.$$

Applying again the law of iterated expectations, but now with $\mathcal{F}_{-j,n}$, yields

$$T_{n1} = 2n^{-4}(h^2 h_1^{d+1})^{-2}$$

$$\times \sum_{i,j>i} \mathbb{E}\left\{ J(i)J(j)[Q_{1i} + Q_{2i} + Q_{3i}][Q_{1j} + Q_{2j} + Q_{3j}]U_i U_j \right\}.$$

All terms that contain ε_i, e.g., $\sum_{i,j>i} \mathbb{E}\{J(i)J(j)Q_{2i}Q_{1j}U_iU_j\}$, can be eliminated by an iterated expectations argument using a sigma algebra that contains all variables other than Y_i. The same holds true for terms that contain ε_j. Hence we are left with the four terms that involve Q_{1k} and Q_{3k}, $k = i, j$, only. Pick a typical term:

$$P_n = 2n^{-4}(h^2 h_1^{d+1})^{-2} \sum_{i,j>i} \mathbb{E}\left\{ J(i)J(j)m(X_i, Z_i)m(X_j, Z_j) \right.$$

$$\left. \times \frac{\int J(t)K_j^*(t, i)\, dt}{\int J(t)\hat{f}_{-i}(t, Z_j)\, dt + n^{-\alpha}} \frac{\int J(t)K_i^*(t, j)\, dt}{\int J(t)\hat{f}_{-j}(t, Z_i)\, dt + n^{-\alpha}} U_i U_j \right\}.$$

Then write $P_n = P_{n1} - P_{n2}$, where

$$P_{n1} = 2n^{-4}(h^2 h_1^{d+1})^{-2} \sum_{i,j>i} \mathbb{E}\left\{ J(i)J(j)m(X_i, Z_i)m(X_j, Z_j) \right.$$

$$\left. \times \frac{\int J(t)K_j^*(t, i)\, dt}{\int J(t)f(t, Z_j)\, dt + n^{-\alpha}} \frac{\int J(t)K_i^*(t, j)\, dt}{\int J(t)f(t, Z_i)\, dt + n^{-\alpha}} U_i U_j \right\}$$

and

$$P_{n2} = 2n^{-4}(h^2 h_1^{d+1})^{-2} \sum_{i,j>i} \mathbb{E}\left\{ J(i)J(j)m(X_i, Z_i)m(X_j, Z_j) \right.$$

$$\times \frac{\int J(t)K_j^*(t, i)\, dt \int J(t)[\hat{f}_{-i}(t, Z_j) - f(t, Z_j)]\, dt}{[\int J(t)\hat{f}_{-i}(t, Z_j)\, dt + n^{-\alpha}] \int J(t)f(t, Z_j)\, dt}$$

$$\left. \times \frac{\int J(t)K_i^*(t, j)\, dt \int J(t)[\hat{f}_{-j}(t, Z_i) - f(t, Z_i)]\, dt}{[\int J(t)\hat{f}_{-j}(t, Z_i)\, dt + n^{-\alpha}] \int J(t)f(t, Z_i)\, dt} U_i U_j \right\}.$$

Assume for simplicity a product kernel, so that $K_i^*(t, j) = K_1(X_i - t) \times K_2(Z_i - Z_j)$, and consider P_{n1},

$$(S21) \quad \mathbb{E}\Big\{ J(i)J(j)m(X_i, Z_i)m(X_j, Z_j)$$

$$\times \frac{\int J(t)K_1(X_j - t)\,dt}{\int J(t)f(t, Z_j)\,dt + n^{-\alpha}} \frac{\int J(t)K_1(X_i - t)\,dt}{\int J(t)f(t, Z_i)\,dt + n^{-\alpha}}$$

$$\times (K_2(Z_i - Z_j))^2 U_i U_j \Big\}$$

$$= h^{-2} \int\int\int \mathbb{1}_{\{|x_j - x_0|\leq h/2\}} m(x_j, z_j)\varphi(x_j, z_j)(x_j - x_0)f(x_j|z_j)\,dx_j$$

$$\times \int \mathbb{1}_{\{|x_i - x_0|\leq h/2\}} m(x_i, z_i)\varphi(x_i, z_i)(x_i - x_0)f(x_i|z_i)\,dx_i$$

$$\times (K_2(z_i - z_j))^2\,dF(z_j)\,dF(z_i) + R_n,$$

where $\varphi(x_j, z_j) = [\int J(t)f(t, Z_j)\,dt + n^{-\alpha}]^{-1} \int J(t)K_1(X_j - t)\,dt$ and R_n contains the $O_p(h^2)$ terms in the approximation, i.e., v_i in $U_i = W_i(X_i - x_0 + h^2 v(X_i))$, with $\sup_{x,z}|v| = c\sup_{x,z}|\partial_x f(x, z)|$. From the right-hand side of (S21), after change of variables $(x_j - x_0) = \psi_j h$, it is tedious but straightforward to show that $P_{n1} = O((nh_1^{d+1})^{-2})$.

Next, consider a typical element $P_{n2}(i, j)$ in $P_{n2} = 2n^{-4}(h^2 h_1^{d+1})^{-2} \times \sum_{i,j>i} P_{n2}(i, j)$. By similar arguments as in Step 1,

$$\int J(t)[\hat{f}_{-j}(t, Z_i) - f(t, Z_i)]\,dt = h[\hat{f}(x_0, Z_i) - f(x_0, Z_i)] + h^3\eta_1,$$

where we suppress the indices as $\hat{f}_{-i} \cong \hat{f}_{-j} \cong \hat{f}$ in large samples. Hence, the leading term is

$$|P_{n2}(i, j)| \leq \mathbb{E}\big\{ J(i)J(j)|m(X_i, Z_i)m(X_j, Z_j)\phi(X_i, X_j, Z_i, Z_j)|$$

$$\times [\hat{f}(x_0, Z_i) - f(x_0, Z_i)][\hat{f}(x_0, Z_j) - f(x_0, Z_j)]|U_i||U_j| \big\},$$

where

$$(S22) \quad \phi(X_i, X_j, Z_i, Z_j)$$

$$= \Big(h^2 \int J(t)K_1(X_j - t)\,dt \int J(t)K_1(X_i - t)\,dt\,[K_2(Z_i - Z_j)]^2 \Big)$$

$$\times \Big(\Big[\int J(t)\hat{f}(t, Z_j)\,dt + n^{-\alpha} \Big] \int J(t)f(t, Z_j)\,dt$$

$$\times \int J(t) f(t, Z_i) \, dt \left[\int J(t) \hat{f}(t, Z_i) \, dt + n^{-\alpha}\right]\Big)^{-1}$$

$$= \left(h^2 \int J(t) K_1(X_j - t) \, dt \int J(t) K_1(X_i - t) \, dt \, [K_2(Z_i - Z_j)]^2\right)$$

$$\times \left(\left[\int J(t) f(t, Z_j) \, dt\right]^2\right.$$

$$+ \left(\int J(t)[\hat{f}(t, Z_j) - f(t, Z_j)] \, dt + n^{-\alpha}\right) \int J(t) f(t, Z_j) \, dt\Big)^{-1}$$

$$\times \left(\left[\int J(t) f(t, Z_i) \, dt\right]^2\right.$$

$$+ \left(\int J(t)[\hat{f}(t, Z_i) - f(t, Z_i)] \, dt + n^{-\alpha}\right) \int J(t) f(t, Z_i) \, dt\Big)^{-1}$$

and we assume again for simplicity $K_i^*(t, j) = K_1(X_i - t) K_2(Z_i - Z_j)$. Applying the Cauchy–Schwarz inequality, we obtain that

$$|P_{n2}(i, j)| \leq \mathbb{E}[G_{ij}^4]^{1/4} \mathbb{E}[\phi_{ij}^2]^{1/4} \mathbb{E}(|\hat{f} - f|^4)^{1/2},$$

where $\phi_{ij} = \phi(X_i, X_j, Z_i, Z_j)$, and

$$\mathbb{E}[G_{ij}^4] = \mathbb{E}\big[J(i) J(j) m(X_i, Z_i)^4 m(X_j, Z_j)^4 U_i^4 U_j^4\big]$$

$$= \int \int \int \mathbb{1}_{\{|x_j - x_0| \leq h/2\}} m(x_j, z_j)^4 (x_j - x_0)^4 f(x_j | z_j) \, dx_j$$

$$\times \int \mathbb{1}_{\{|x_i - x_0| \leq h/2\}} m(x_i, z_i)^4 (x_i - x_0)^4$$

$$\times f(x_i | z_i) \, dx_i \, dF(z_j) \, dF(z_i) + R_n,$$

where R_n contains the higher order terms in the approximation $U_i = W_i(X_i - x_0 + h^2 v(X_i))$. By standard arguments, $\mathbb{E}[G_{ij}^4]^{1/4} = O(h^{5/2})$. Next, the following expression in the denominator of (S22) can be bounded by

$$\left[\left[\int J(t) f(t, Z_i) \, dt\right]^2\right.$$

$$+ \int J(t)[\hat{f}(t, Z_i) - f(t, Z_i)] \, dt \int J(t) f(t, Z_i) \, dt\Big]^{-1}$$

$$= \big[h^2 f(x_0, Z_i)\big[f(x_0, Z_i) + [\hat{f}(x_0, Z_i) - f(x_0, Z_i)]\big] + n^{-\alpha} + \tau_i\big]^{-1}$$

$$\leq n^\alpha.$$

Proceeding analogously for the other terms in (S22), we obtain after change of variables that $|\phi_{ij}| \leq h^4 n^{2\alpha}$. Hence, $\mathbb{E}[\phi_{ij}^2]^{1/4} \leq h^2 n^\alpha$. Thus, $P_{n2} = O(n^\alpha \times (nh_1^{d+1})^{-2} h^{-1/2} \mathbb{E}(|\hat{f} - f|^4)^{1/2})$, and this term converges under general assumptions faster than P_{n1}. Hence, we obtain that $P_n = O((nh_1^{d+1})^{-1})$. Similar arguments apply to all other terms in (S20) and hence the behavior of the second term is clarified.

Step 2: Turning to the first term in (S18),

$$(n^2 h^2)^{-1} \mathbb{E}\left\{ \sum_i U_i^2 S_i^{*2} \right\} = (n^2 h^2)^{-1} \sum_i \mathbb{E}[U_i^2 S_i^{*2}]$$

and

$$\mathbb{E}[U_i^2 S_i^{*2}] \leq \mathbb{E}[U_i^4]^{1/2} \mathbb{E}[S_i^{*4}]^{1/2}.$$

Note that $\mathbb{E}[U_i^4]^{1/2} = O(h^{5/2})$ and $\mathbb{E}[S_i^{*4}]^{1/2} \leq n^{2\alpha} \mathbb{E}[[\widehat{mf} - mf]^4]^{1/2}$ by the same arguments as in Step 1. Consequently,

$$(n^2 h^2)^{-1} \mathbb{E}\left\{ \sum_i U_i^2 S_i^{*2} \right\} = O\big(n^{2\alpha} n^{-1} h^{1/2}(h_1^{2r_1} + (nh_1^{d+1})^{-1})\big).$$

Hence, the first term in (S18) is $T_{n1} = O_p((n^{1/2-\alpha})^{-1} h^{1/4} h_1^{r_1} + h^{1/4}(n^{1-\alpha} h_1^{(d+1)/2})^{-1})$. To see that $T_{n1} = o_p(T_{n3})$, set $n^\alpha \sim h^{-2}$. Then this is the case under our assumptions. Finally, by similar arguments, $T_{n2} = o_p(T_{n3})$ and this completes the second step. *Q.E.D.*