

Asymptotic Trimming for Bounded Density Plug-in Estimators

Arthur Lewbel Boston College

October 2000

Abstract

This paper proposes a form of asymptotic trimming to obtain root n convergence of functions of kernel estimated objects. The trimming is designed to deal with the boundary effects that arise in applications where densities are bounded away from zero.

Arthur Lewbel, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (781)-862-3678, lewbel@bc.edu. I'd like to thank Bo Honore and Shakeeb Khan for helpful suggestions. This research was supported in part by the National Science Foundation, through grant number SES9905010.

1 Introduction

This paper obtains root n convergence of functions of kernel estimated objects using asymptotic trimming. The proposed trimming is designed to deal with the boundary effects that arise in applications where densities are bounded away from zero. The method is demonstrated by an example in which the object to be estimated is a weighted average of the inverse of a kernel estimated conditional density, that is, estimates of functions of the form $E[w/f(v|u)]$ where $f(v|u)$ denotes the conditional density of a scalar random variable v given a random vector u .

The estimator is a special case of a two step estimator with a kernel estimated nonparametric first step. The general theory of such estimators is described in Newey and McFadden (1994), Newey (1994), and Sherman (1994). See also Robinson (1988), Powell, Stock, and Stoker (1989), Hardle and Stoker (1989), Lewbel (1995), and Andrews (1995).

The difficulty with applying generic methods like Newey and McFadden (1994), or the types of trimming employed by Hardle and Stoker (1989) or Sherman (1994), is that, for

root N convergence of a sample average of w/f (where a kernel estimator \hat{f} is substituted in for f), to avoid boundary effects such estimators assume either f or the terms being averaged go to zero on the boundary of the support of v, u . When the average is a form like w/f , this would require that w go to zero in this neighborhood, which is not the case for some applications.

This technicality is resolved here by bounding f away from zero, and introducing an asymptotic trimming function that sets to zero all terms in the average having data within a distance τ of the boundary. The estimator has τ go to zero more slowly than the bandwidth to eliminate boundary effects from kernel estimation, but also assumes $N^{1/2}\tau \rightarrow 0$, which makes the volume of the trimmed space vanish quickly enough to send the trimming induced bias to zero.

Formally, this trimming requires that the support of (v, u) be known.. In practice, this support could be estimated, or trimming might be accomplished more easily by simply dropping out a few of the most extreme observations of the data, e.g., observations where the estimated density is particularly small.

Based on Rice (1986), for one dimensional densities Hong and White (2000) use jack-knife boundary kernels to deal with this same problem of boundary bias. Their technique (which also requires known support) could be generalized to higher dimensions as an alternative to the trimming proposed here.

The inverse density example derived here is directly applicable to Lewbel (2000) and Honore and Lewbel (2000). Other estimation problems involving nonparametric density estimators in the denominator to which the technique can likely be applied include Stoker (1991), Robinson (1991), Newey and Ruud (1994), Granger and Lin (1994), and Hong and White (2000).

For ease of notation, the theorem is stated and proved assuming all of the elements of v, u are continuously distributed. The results can be readily extended to include discretely distributed variables as well, either by applying the density estimator separately to each discrete cell of data and averaging the results, or by smoothing across cells using kernels (essentially treating the discrete variables as if they were continuous). See, e.g., Racine and Li (2000) for details, or see the examples in Lewbel (2000) or Honore and Lewbel (2000).

ASSUMPTION A.1: The data consist of N observations (w_i, v_i, u_i) , $i = 1, \dots, N$, which are assumed to be i.i.d. Here u_i is a $k - 1$ vector and v_i is a scalar, u_i and (v_i, u_i) are drawn from distributions that are absolutely continuous with respect to some Lebesgue measures, with Radon-Nikodym densities $f_u(u)$ and $f_{vu}(v, u)$, and have supports denoted Ω_u and Ω_{vu} . Denote kernel estimators of these densities \hat{f}_{vu} and \hat{f}_u , with kernel functions K_{vu} and K_u . Let h be a kernel bandwidth and τ be a density trimming parameter. Assume Ω_{vu} is known, and define the trimming function $I_\tau(v, u)$ to equal zero if (v, u) is within a distance τ of the boundary of Ω_{vu} , otherwise, $I_\tau(v, u)$ equals one. Define

$$h_{\tau i} = w_i \frac{f_u(u_i)}{f_{vu}(v_i, u_i)} I_\tau(v, u)$$

$$\begin{aligned}
h_i &= w_i \frac{f_u(u_i)}{f_{vu}(v_i, u_i)} \\
\eta &= E(h_i) \\
q_i &= h_i + E(h_i|u_i) - E(h_i|v_i, u_i) \\
\widehat{f}_{vu}(v, u) &= \frac{1}{h^k N} \sum_{j=1}^N K_{vu} \left(\frac{v_j - v}{h}, \frac{u_j - u}{h} \right) \\
\widehat{f}_u(u) &= \frac{1}{h^{k-1} N} \sum_{j=1}^N K_u \left(\frac{u_j - u}{h} \right) \\
\widehat{\eta} &= \frac{1}{N} \sum_{i=1}^N w_i \frac{\widehat{f}_u(u_i)}{f_{vu}(v_i, u_i)} I_\tau(v, u)
\end{aligned}$$

ASSUMPTION A.2: The functions $f_{vu}(v, u)$, $f_u(u)$, $\pi_{\tau vu}(v, u) = E(h_\tau|v, u)$, and $\pi_{\tau u}(u) = E(h_\tau|u)$ exist and are continuous in the components of (v, u) for all $(v, u) \in \Omega_{vu}$, and are continuously differentiable in the components of (v, u) for all $(v, u) \in \overline{\Omega}_{vu}$, where $\overline{\Omega}_{vu}$ differs from Ω_{vu} by a set of measure zero. Assume w , Ω_{vu} , and $f_{vu}(v, u)$ are bounded, and $f_{vu}(v, u)$ is bounded away from zero.

ASSUMPTION A.3: There exists some functions m_{vu} and m_u such that the following local Lipschitz conditions hold for some v_u and some (v_v, v_u) in an open neighborhood of zero, for all $\tau \geq 0$:

$$\begin{aligned}
\| f_{vu}(v + v_v, u + v_u) - f_{vu}(v, u) \| &\leq m_{vu}(v, u) \| (v_v, v_u) \| \\
\| f_u(u + v_u) - f_u(u) \| &\leq m_u(u) \| v_u \| \\
\| \pi_{\tau vu}(v + v_v, u + v_u) - \pi_{\tau vu}(v, u) \| &\leq m_{vu}(v, u) \| (v_v, v_u) \| \\
\| \pi_{\tau u}(u + v_u) - \pi_{\tau u}(u) \| &\leq m_u(u) \| v_u \|
\end{aligned}$$

$E[h_\tau^2 f_u(u)^{-2}|u]$ and $E[h_\tau^2 f_{vu}(v, u)^{-2}|v, u]$ exist and are continuous, and the following objects exist

$$\begin{aligned}
&\sup_{\tau \geq 0, (v, u) \in \Omega_{vu}} E[h_\tau^2 f_{vu}(v, u)^{-2}|v, u] \\
&\sup_{\tau \geq 0, u \in \Omega_u} E[h_\tau^2 f_u(u)^{-2}|u] \\
&\sup_{\tau \geq 0} E \left[(|h_\tau / f_{vu}(v, u)| + 1) m_{vu}(v, u)^2 \right] \\
&\sup_{\tau \geq 0} E \left[(|h_\tau / f_u(u)| + 1) m_u(u)^2 \right]
\end{aligned}$$

ASSUMPTION A.4: The support of the kernel function $K(u)$ is $\mathbb{R}^{\dim(u)}$. $K(u) = 0$ for all u on the boundary of, and outside of, a convex bounded subset of \mathbb{R}^k . This subset has a nonempty interior and has the origin as an interior point. $K(u)$ is a bounded differentiable, symmetric function. $\int K(u)du = 1$. The kernel function $K_u(u)$ has order $p > 1$. All partial derivatives of $f_u(u)$ of order p exist. For all $0 \leq \rho \leq p$, for $l_1 + \dots + l_k = \rho$, $\sup_{\tau \geq 0} \int_{u \in \Omega_u} \pi_{\tau u}(x) [\partial^\rho f_u(u) / \partial^{l_1} u_1 \dots \partial^{l_k} u_k] du$ exist. The kernel function $K_{vu}(v, u)$ has all the same properties, replacing u with v, u .

Theorem A: Let Assumptions A.1 to A.4 hold. If $Nh^k \rightarrow \infty$, $Nh^{2p} \rightarrow 0$, $h/\tau \rightarrow 0$, and $N\tau^2 \rightarrow 0$, then the following equations hold

$$\begin{aligned}\sqrt{N}(\hat{\eta} - \eta) &= [N^{-1/2} \sum_{i=1}^N q_i - E(q_i)] + o_p(1) \\ \sqrt{N}(\hat{\eta} - \eta) &\rightarrow N[0, \text{var}(q)]\end{aligned}$$

Theorem A is proved in the Appendix.

2 Appendix: Proof

Begin by stating and proving Theorem B below, which supplies a set of regularity conditions for root N convergence of a weighted average of a kernel estimated density. Theorem A will then be proved using repeated application of Theorem B.

Just for theorem B, let x be a continuously distributed random k vector, and let $f(x)$ be the probability density function of x . For some trimming parameter τ , some random vector s , and some function $w_\tau = w_\tau(x, s)$, let $\mu_\tau = E[w_\tau f(x)]$ and define estimators

$$\begin{aligned}\hat{f}(x) &= \frac{1}{h^k N} \sum_{j=1}^N K\left(\frac{x_j - x}{h}\right) \\ \hat{\mu} &= N^{-1} \sum_{i=1}^N w_{\tau i} \hat{f}(x_i)\end{aligned}$$

where h is a bandwidth and K is a kernel function. Note that h and τ are implicitly functions of N

ASSUMPTION B.1: The data consist of N observations (x_i, s_i) , $i = 1, \dots, N$, which is assumed to be an i.i.d. random sample. The k vector x_i is drawn from a distribution that is absolutely continuous with respect to some Lebesgue measure on \mathbb{R}^k , with Radon-Nikodym density $f(x)$ having bounded support denoted Ω_x . The underlying measure ν

can be written in product form as $\nu = \nu_x \times \nu_s$. The trimming parameter τ is an element of a set Ω_τ .

ASSUMPTION B.2: The density function $f(x)$ and the function $\pi_\tau(x) = E(w_\tau|x)f(x)$ exist and are continuous in the components of x for all $x \in \Omega_x$, and are continuously differentiable in the components of x for all $x \in \overline{\Omega}_x$, where $\overline{\Omega}_x$ differs from Ω_x by a set of measure zero.

ASSUMPTION B.3: For some ν in an open neighborhood of zero there exists some functions m_f and m_{π_τ} such that the following local Lipschitz conditions hold:

$$\begin{aligned} \| f(x + \nu) - f(x) \| &\leq m_f(x) \| \nu \| \\ \| \pi_\tau(x + \nu) - \pi_\tau(x) \| &\leq m_{\pi_\tau}(x) \| \nu \| \end{aligned}$$

For all $\tau \in \Omega_\tau$, $E(w_\tau^2|x)$ is continuous in x . Also, $\sup_{\tau \in \Omega_\tau, x \in \Omega_x} E(w_\tau^2|x)$ and $\sup_{\tau \in \Omega_\tau} E \left[(|w_\tau| m_f(x) + m_{\pi_\tau}(x))^2 \right]$ exist.

ASSUMPTION B.4: The support of $K(u)$ is \mathbb{R}^k . $K(u) = 0$ for all u on the boundary of, and outside of, a convex bounded subset of \mathbb{R}^k . This subset has a nonempty interior and has the origin as an interior point. $K(u)$ is a bounded differentiable, symmetric function. $\int K(u)du = 1$ and $\int u K(u)du = 0$.

ASSUMPTION B.5: The kernel $K(u)$ has order p . All partial derivatives of $f(x)$ of order p exist. For all $0 \leq \rho \leq p$, for $l_1 + \dots + l_k = \rho$, $\sup_{\tau \in \Omega_\tau} \int_{x \in \Omega_x} \pi_\tau(x) [\partial^\rho f(x) / \partial^{l_1} x_1 \dots \partial^{l_k} x_k] dx$ exists. The equality $w_\tau(x, s)f(x) = 0$ holds for all x within a distance τ of the boundary of Ω_x .

Theorem B: Let Assumptions B.1 to B.5 hold. Define $r_{\tau i}$ by

$$r_{\tau i} = [w_{\tau i} + E(w_{\tau i}|x_i)] f(x_i) = f(x_i)w_{\tau i} + \pi_\tau(x_i)$$

If $Nh^k \rightarrow \infty$, $Nh^{2p} \rightarrow 0$, and $h/\tau \rightarrow 0$ and then,

$$\sqrt{N}(\hat{\mu} - \mu_\tau) = [N^{-1/2} \sum_{i=1}^N r_{\tau i} - E(r_{\tau i})] + o_p(1)$$

Proof of Theorem B: Let $z_{\tau i} = (x_i, w_{\tau i})$, and define $p_N(z_{\tau i}, z_{\tau j})$ by

$$p_N(z_{\tau i}, z_{\tau j}) = \frac{w_{\tau i} + w_{\tau j}}{2h^k} K\left(\frac{x_i - x_j}{h}\right)$$

The average kernel estimator $\hat{\mu}$ is equivalent to a data dependent U-statistic

$$\hat{\mu} - \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^N p_N(z_{\tau i}, z_{\tau j}) = O_p(N^{-2}h^{-k}) = o_p(N^{-1})$$

Following Powell, Stock, and Stoker (1989), first verify that $E[\| p_N(z_{\tau i}, z_{\tau j}) \|^2] = o(N)$.

$$\begin{aligned}
E[\| p_N(z_{\tau i}, z_{\tau j}) \|^2] &= \int \frac{1}{4h^{2k}} f(x_i) f(x_j) \\
&\quad \left[E(w_{\tau i}^2 | x_i) + E(w_{\tau j}^2 | x_j) + 2E(w_{\tau i} | x_i) E(w_{\tau j} | x_j) \right] K\left(\frac{x_i - x_j}{h}\right)^2 dx_i dx_j \\
&= \int \frac{1}{4h^k} f(x_i) f(x_i + hu) \\
&\quad \left[E(w_{\tau i}^2 | x_i) + E(w_{\tau i}^2 | x_i + hu) + 2E(w_{\tau i} | x_i) E(w_{\tau i} | x_i + hu) \right] K(u)^2 dx_i du \\
&\leq \frac{1}{h^k} \left[\sup_{\tau \in \Omega_\tau, x \in \Omega_x} E(w_\tau^2 | x) \right] \int f(x_i) f(x_i + hu) K(u)^2 dx_i du = O(h^{-k}) = O[N(Nh^k)^{-1}] = o(N)
\end{aligned}$$

using the change of variables from (x_i, x_j) to $(x_i, u = (x_j - x_i)/h)$ with jacobian h^k .

Let $r_{Ni} = 2E[p_N(z_{\tau i}, z_{\tau j}) | z_{\tau i}]$. It follows from Lemma 3.1 in Powell, Stock, and Stoker (1989) that $N^{1/2}[\widehat{\mu} - E(\widehat{\mu})] = N^{-1/2} \sum_{i=1}^N r_{Ni} - E(r_{Ni}) + o_p(1)$. Next, define t_{Ni} by

$$\begin{aligned}
t_{Ni} &= r_{Ni} - r_{\tau i} = E[p_N(z_{\tau i}, z_{\tau j}) | z_{\tau i}] - r_{\tau i} \\
&= \int \frac{1}{h^k} [w_{\tau i} + E(w_{\tau j} | x_j)] K\left(\frac{x_i - x_j}{h}\right) f(x_j) dx_j - r_{\tau i} \\
&= \int \frac{1}{h^k} [w_{\tau i} f(x_j) + \pi_\tau(x_j)] K\left(\frac{x_i - x_j}{h}\right) dx_j - r_{\tau i} \\
&= \int [w_{\tau i} f(x_i + hu) + \pi_\tau(x_i + hu)] K(u) du - [w_{\tau i} f(x_i) + \pi_\tau(x_i)] \\
&= \int [w_{\tau i} [f(x_i + hu) - f(x_i)] + [\pi_\tau(x_i + hu) - \pi_\tau(x_i)]] K(u) du
\end{aligned}$$

Note that given the properties of $K(u)$, having $w_\tau(x, s)f(x) = 0$ hold for all x within a distance τ of the boundary of Ω_x , and having $h/\tau \rightarrow 0$, ensures that boundary effects do not interfere with the change of variables from (x_i, x_j) to $(x_i, u = (x_j - x_i)/h)$ above. To illustrate this point, suppose that K is a product kernel, the first element of which is non zero only over the range $[-c, c]$, and suppose that the first element of x_i has support given by the interval $[L_-, L^+]$. Then, after the change of variables, the first element of u will be evaluated over the range $[(L_- - x_i)/h, (L^+ - x_i)/h]$, and so boundary effects regarding this first element can only arise at observations i in which the interval $[(L_- - x_i)/h, (L^+ - x_i)/h]$ does not contain the interval $[-c, c]$, which is equivalent to when x_i is within a distance ch of the boundary of $[L_-, L^+]$. However, $w_\tau(x, s) = 0$, and hence the integral equals zero, when x_i is within a distance τ of the boundary, so having $h/\tau \rightarrow 0$, ensures that once the sample size is sufficiently large (for this element, once h becomes smaller than τ/c), the integral will equal zero at all observations i for which boundary effects may arise.

Given the above, we have

$$N^{1/2}[\widehat{\mu} - E(\widehat{\mu})] = \left[N^{-1/2} \sum_{i=1}^N r_{\tau i} - E(r_{\tau i}) \right] + \left[N^{-1/2} \sum_{i=1}^N t_{Ni} - E(t_{Ni}) \right]$$

and, using the local Lipschitz conditions

$$|t_{Ni}| \leq h[|w_{\tau i}|m_f(x_i) + m_{\pi_\tau}(x_i)] \int \|u\| K(u) du$$

$$E(t_{Ni}^2) \leq h^2 \sup_{\tau \in \Omega_\tau} E \left[(|w_\tau|m_f(x) + m_{\pi_\tau}(x))^2 \right] \left[\left(\int \|u\| K(u) du \right)^2 \right] = O(h^2) = o(1)$$

Now $E(t_{Ni}^2) = o(1)$ implies that $N^{-1/2} \sum_{i=1}^N t_{Ni} - E(t_{Ni}) + o_p(1)$, and therefore $N^{1/2}[\widehat{\mu} - E(\widehat{\mu})] = N^{-1/2} \sum_{i=1}^N 2[r_{\tau i} - E(r_{\tau i})] + o_p(1)$.

Write $E(\widehat{\mu})$ as

$$\begin{aligned} E(\widehat{\mu}) &= \frac{1}{h^k} E \left[w_{\tau i} K \left(\frac{x_i - x_j}{h} \right) \right] \\ &= \frac{1}{h^k} \int E(w_{\tau i}|x_i) K \left(\frac{x_i - x_j}{h} \right) f(x_i) f(x_j) dx_i dx_j \\ &= \int \pi_\tau(x_i) K(u) f(x_i + hu) dx_i du \end{aligned}$$

by Assumption B.5, we can substitute into this expression a p 'th order Taylor expansion of $f(x_i + hu)$ around $f(x)$. Then, using $\mu_\tau = \int \pi_\tau(x_i) f(x_i) dx_i$, the existence of $\sup_{\tau \in \Omega_\tau} \int_{x \in \Omega_x} \pi_\tau(x) [\partial^\rho f(x) / \partial^{l_1} x_1 \dots \partial^{l_k} x_k] dx$, the fact that $w_\tau(x, s) f(x)$ and hence $\pi_\tau(x) f(x)$ vanishes in a sufficiently large neighborhood of the boundary of x , and the assumption that $K(u)$ is a p 'th order kernel, gives $E(\widehat{\mu}) = \mu_\tau + O(h^p)$, so $N^{1/2}[E(\widehat{\mu}) - \mu_\tau] = O(N^{1/2}h^p) = o(1)$, which finishes the proof of Theorem B.

Theorem B does not require that the trimming parameter τ equal distance to the boundary, although that is how it will be used in the proof of Theorem A. An extension to Theorem B (which might be useful in other applications), is that if we let τ be any function of N , and let $d(\tau)$ denote distance to the boundary, then Theorem B still holds as long as we replace $h/\tau \rightarrow 0$ with $h/d(\tau) \rightarrow 0$, and replace the last sentence of Assumption B.5 with $w_\tau(x, s) f(x) = 0$ holding for all x within a distance $d(\tau)$ of the boundary of Ω_x .

Before going on to prove Theorem A, the Assumptions A.1 to A.4 are restated in a different form below as Assumptions C.1 to C.4 below, which are more convenient for the proof.

ASSUMPTION C.1: The data consist of N observations (w_i, v_i, u_i) , $i = 1, \dots, N$, which are assumed to be i.i.d. Here w_i is a vector, u_i is a $k - 1$ vector and v_i is a scalar.

Let f_{vu} denote the joint density function of an observation of v and u , and let f_u denote the marginal density function of an observation of u . Denote kernel estimators of these densities \widehat{f}_{vu} and \widehat{f}_u , where the kernels have bandwidth h and order p . Let Ω_{vu} equal the support of (v, u) , which is assumed to be known. Define the trimming function $I_\tau(v, u)$ to equal zero if (v, u) is within a distance τ of the boundary of Ω_{vu} , otherwise, $I_\tau(v, u)$ equals one. Define

$$\begin{aligned} h_{\tau i} &= w_i \frac{f_u(u_i)}{f_{vu}(v_i, u_i)} I_\tau(v_i, u_i) \\ h_i &= w_i \frac{f_u(u_i)}{f_{vu}(v_i, u_i)} \\ \eta &= E(h_i) \\ q_i &= h_i + E(h_i|u_i) - E(h_i|v_i, u_i) \\ \widehat{\eta} &= \frac{1}{N} \sum_{i=1}^N w_i \frac{\widehat{f}_u(u_i)}{\widehat{f}_{vu}(v_i, u_i)} I_\tau(v_i, u_i) \end{aligned}$$

ASSUMPTION C.2: Assumptions B.1 to B.5 hold, with $x = (v, u)$, $f = f_{vu}$, $\widehat{f} = \widehat{f}_{vu}$, $s = w$, and $w_{\tau i} = h_{\tau i}/f_{vu}(v_i, u_i)$.

ASSUMPTION C.3: Assumptions B.1 to B.5 hold, with $x = u$, $f = f_u$, $\widehat{f} = \widehat{f}_u$, $s = (v, w)$, and $w_{\tau i} = h_{\tau i}/f_u(u_i)$.

ASSUMPTION C.4: Assume w , Ω_{vu} , and $f_{vu}(v, u)$ are bounded, and $f_{vu}(v, u)$ is bounded away from zero.

Proof of Theorem A: Define $\eta_\tau = E(h_{\tau i})$ and $q_{\tau i} = h_{\tau i} + E(h_{\tau i}|u_i) - E(h_{\tau i}|v_i, u_i)$. If $h \rightarrow 0$ then

$$\sup_{(v, u) \in \Omega_{vu}} |\widehat{f}_{vu}(v, u) - f_{vu}(v, u)| = O_p \left[(N^{1-\varepsilon} h^k)^{-1/2} \right]$$

$$\sup_{(v, u) \in \Omega_{vu}} |\widehat{f}_u(u) - f_u(u)| = O_p \left[(N^{1-\varepsilon} h^{k-1})^{-1/2} \right]$$

for any $\varepsilon > 0$. See, e.g., Silverman (1978) and Collomb and Hardle (1986).

To ease notation, let $f_{vui} = f_{vu}(v_i, u_i)$ and $f_{ui} = f_u(u_i)$. Consider a second order Taylor expansion of $\widehat{f}_{vui}/\widehat{f}_{ui}$ around f_{vui}/f_{ui} . The quadratic terms in the expansion involve the second derivatives of f_{vui}/f_{ui} evaluated at \widetilde{f}_{ui} and \widetilde{f}_{vui} , where \widetilde{f}_{ui} lies between \widehat{f}_{ui} and f_{ui} , and similarly \widetilde{f}_{vui} lies between \widehat{f}_{vui} and f_{vui} . Substituting the Taylor expansion of $\widehat{f}_{vui}/\widehat{f}_{ui}$ around f_{vui}/f_{ui} into $\widehat{\eta}$ gives

$$N^{1/2} \widehat{\eta} = R_N + N^{-1/2} \sum_{i=1}^N w_i \left(\frac{f_{ui}}{f_{vui}} + \frac{\widehat{f}_{ui} - f_{ui}}{f_{vui}} - \frac{f_{ui}(\widehat{f}_{vui} - f_{vui})}{f_{vui}^2} \right) I_\tau(v_i, u_i)$$

Where R_N is a remainder term.

One component of R_N is

$$\begin{aligned}
& N^{-1/2} \sum_{i=1}^N \frac{w_i \tilde{f}_{ui}}{\tilde{f}_{vui}^3} (\hat{f}_{vui} - f_{vui})^2 I_\tau(v_i, u_i) \\
& \leq \left(N^{-1/2} \sum_{i=1}^N |w_i| \right) [\sup (f_{vu}^{-3})] (\sup |\hat{f}_{vu} - f_{vu}|)^2 \\
& = O_p(N^{\varepsilon-1} h^{-k}) = o_p(1)
\end{aligned}$$

Similarly, the other components of R_N are $O_p(N^{\varepsilon-1/2} h^{-k-1/2})$ and $O_p(N^{\varepsilon-1/2} h^{-k-1})$, which will also be $o_p(1)$.

We now have

$$\begin{aligned}
N^{1/2} \hat{\eta} &= N^{-1/2} \sum_{i=1}^N w_i \left(\frac{f_{ui}}{f_{vui}} + \frac{\hat{f}_{ui} - f_{ui}}{f_{vui}} - \frac{f_{ui}(\hat{f}_{vui} - f_{vui})}{f_{vui}^2} \right) I_\tau(v_i, u_i) + o_p(1) \\
&= N^{-1/2} \sum_{i=1}^N h_{\tau i} + w_i \left(\frac{\hat{f}_{ui} - f_{ui}}{f_{vui}} - \frac{f_{ui}(\hat{f}_{vui} - f_{vui})}{f_{vui}^2} \right) I_\tau(v_i, u_i) + o_p(1)
\end{aligned}$$

Next, using Assumption C.3 apply Theorem B to obtain

$$N^{-1/2} \sum_{i=1}^N w_i \frac{\hat{f}_{ui} - f_{ui}}{f_{vui}} I_\tau(v_i, u_i) = N^{-1/2} \sum_{i=1}^N r_{\tau i} - E(r_{\tau i}) + o_p(1)$$

where $r_{\tau i}$ is given by

$$\begin{aligned}
r_{\tau i} &= \left[\frac{w_i}{f_{vui}} I_\tau(v_i, u_i) + E \left(\frac{w_i}{f_{vui}} I_\tau(v_i, u_i) \mid u_i \right) \right] f_{ui} \\
&= h_{\tau i} + E(h_{\tau i} \mid u_i)
\end{aligned}$$

so

$$N^{-1/2} \sum_{i=1}^N w_i \frac{\hat{f}_{ui} - f_{ui}}{f_{vui}} I_\tau(v_i, u_i) = N^{-1/2} \sum_{i=1}^N h_{\tau i} + E(h_{\tau i} \mid u_i) - E(h_{\tau i}) + o_p(1)$$

Similarly, using Assumption C.2 and again applying Theorem B gives

$$N^{-1/2} \sum_{i=1}^N w_i \frac{f_{ui}(\hat{f}_{vui} - f_{vui})}{f_{vui}^2} I_\tau(v_i, u_i) = N^{-1/2} \sum_{i=1}^N h_{\tau i} + E(h_{\tau i} \mid v_i, u_i) - E(h_{\tau i}) + o_p(1)$$

substituting in these results gives

$$N^{1/2} \hat{\eta} = N^{-1/2} \sum_{i=1}^N h_{\tau i} + [h_{\tau i} + E(h_{\tau i} \mid u_i) - E(h_{\tau i})] - [h_{\tau i} + E(h_{\tau i} \mid v_i, u_i) - E(h_{\tau i})] + o_p(1)$$

so

$$N^{1/2}(\hat{\eta} - \eta_\tau) = N^{-1/2} \sum_{i=1}^N q_{\tau i} - E(q_{\tau i}) + o_p(1)$$

Next, observe that

$$E \left[\left(N^{-1/2} \sum_{i=1}^N (h_i - h_{\tau i}) \right)^2 \right] = E \left[h_i^2 [1 - I_\tau(v_i, u_i)] \right] + (N-1) [E(h_i [1 - I_\tau(v_i, u_i)])]^2$$

Now $E[h_i^2 [1 - I_\tau(v_i, u_i)]] = o(1)$ because the expectation of h_i^2 exists and $\tau \rightarrow 0$ makes $1 - I_\tau(v_i, u_i) \rightarrow 0$. For the second term above we have

$$(N-1) [E(h_i [1 - I_\tau(v_i, u_i)])]^2 \leq \left[N^{1/2} \sup(h_i) E[1 - I_\tau(v_i, u_i)] \right]^2$$

Now $E[1 - I_\tau(v_i, u_i)]$ equals the probability that (v_i, u_i) is within a distance τ of the boundary of Ω_{vu} , which is less or equal to $\sup(f_{vu})$ times the volume of the space within a distance τ of the boundary of Ω_{vu} . That volume is $O(\tau)$, so given boundedness of h_i and f_{vu} , we have $N^{1/2} \sup(h_i) E[1 - I_\tau(v_i, u_i)] = O(N^{1/2} \tau) = o(1)$.

We therefore have $E \left[\left(N^{-1/2} \sum_{i=1}^N (h_i - h_{\tau i}) \right)^2 \right] = o(1)$, and so $\sqrt{N}(\eta_\tau - \eta) = o(1)$.

The same analysis can be applied to $q_{\tau i}$, resulting in

$$N^{1/2}(\hat{\eta} - \eta) = N^{-1/2} \sum_{i=1}^N q_i - E(q_i) + o_p(1)$$

and $\sqrt{N}(\hat{\eta} - \eta) \rightarrow N[0, \text{var}(q)]$ follows immediately.

References

- [1] ANDREWS, D. W. K., (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596.
- [2] COLLOMB, G. AND W. HARDLE (1986), "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: Kernel Regression Estimation From Dependent Observations," *Stochastic Processes and Their Applications*, 23, 77–89.
- [3] GRANGER, C. W. J. AND J. L. LIN, (1994), "Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models," *Journal of Time Series Analysis*, 15, pp. 371–384.
- [4] HÄRDLE, W. AND T. M. STOKER (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

- [5] HONORÉ, B. E. AND A. LEWBEL (2000), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogeneous Regressors," unpublished manuscript.
- [6] HONG, Y. AND H. WHITE, (2000), "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence" Unpublished Manuscript.
- [7] LEWBEL, A. (1995), "Consistent Nonparametric Tests With an Application to Slutsky Symmetry," *Journal of Econometrics*, 67, 379-401
- [8] LEWBEL, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145-177.
- [9] NEWEY, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.
- [10] NEWEY, W. K. AND D. MCFADDEN (1994), "Large Sample Estimation and Hypothesis Testing," in Handbook of Econometrics, vol. iv, ed. by R. F. Engle and D. L. McFadden, pp. 2111-2245, Amsterdam: Elsevier.
- [11] NEWEY, W. K. AND P. A. RUUD (1994), "Density Weighted Linear Least Squares," University of California at Berkeley working paper.
- [12] POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica* 57, 1403–1430.
- [13] RACINE, J. AND Q. LI (2000), "Nonparametric Estimation of Conditional Distributions with Mixed Categorical and Continuous Data," Unpublished Manuscript.
- [14] RICE, J., (1986), "Boundary Modification for Kernel Regression," *Communications in Statistics*, 12, pp. 1215-1230.
- [15] ROBINSON, P. M., (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, pp. 931-954.
- [16] ROBINSON, P. M., (1991), "Consistent Nonparametric Entropy Testing," *Review of Economic Studies*, 58, pp. 437-453.
- [17] SHERMAN, R. P. (1994), "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator," *Econometric Theory*, 10, 372-395.
- [18] SILVERMAN, B. W. (1978) "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density Function and its Derivatives," *Annals of Statistics*, 6, 177–184.
- [19] STOKER, THOMAS M. (1991), "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, W. A. Barnett, J. Powell, and G. Tauchen, Eds., Cambridge University Press.