WAVELET METHODS FOR ERRATIC REGRESSION MEANS IN THE PRESENCE OF MEASUREMENT ERROR

[1]Peter Hall, [2]Spiridon Penev and [1]Jason Tran

[1]*Univeristy of Melbourne and* [2]*University of New South Wales, Sydney*

Supplementary Material

5. Proof of the Theorem.

Assume, without loss of generality, that the support $[a, b] = [-1, 1]$. The proof includes several steps.

Step 1: First approximation to $S(\alpha)$. The approximation is given at (5.3) below. It follows from the formula for g in (2.1) that

$$\begin{aligned} E\{g(X_i)|W_i\}f_W(W_i) &= \sum_j \alpha_j^0 \int_{-1}^{1} \phi_j(x) f_U(W_i - x) f_X(x)dx \\ &+ \sum_{k=0}^{\infty} \sum_j \alpha_{jk}^0 \int_{-1}^{1} \psi_{jk}(x) f_U(W_i - x) f_X(x)dx. \end{aligned} \tag{5.1}$$

Define too $V_i' = V_i \hat{f}_W(W_i)$, $\Delta_{W_i} = g(X_i)\{\hat{f}_W(W_i) - f_W(W_i)\}$.

$$\Delta_{X_{i,\phi}} = \sum_j \alpha_j \int_{-1}^{1} \phi_j(x) f_U(W_i - x)\{\hat{f}_X(x) - f_X(x)\}dx,$$

$$\Delta_{X_{i,\psi}} = \sum_{k=0}^{m} \sum_j \alpha_{jk} \int_{-1}^{1} \psi_{jk}(x) f_U(W_i - x)\{\hat{f}_X(x) - f_X(x)\}dx,$$

$$g(x|\alpha, m) = \sum_j \alpha_j \phi_j(x) + \sum_{k=0}^{m} \sum_j \alpha_{jk} \psi_{jk}(x).$$

Then,

$$E\{g(X|\alpha,m)|W=w\} = \frac{1}{f_W(w)}\Big\{\sum_j \alpha_j \int_{-1}^{1} \phi_j(x) f_U(w-x) f_X(x)dx + \sum_{k=0}^{m}\sum_j \alpha_{jk}\int_{-1}^{1}\psi_{jk}(x) f_U(w-x) f_X(x)dx\Big\}. \qquad (5.2)$$

Since $Y_i = g(X_i) + V_i$ then, in view of the definition of $S(\alpha)$, (5.1) and (5.2),

$$\begin{aligned} S(\alpha) = \sum_{i=1}^{n} \Big[& g(X_i) f_W(W_i) + \Delta_{W_i} + V_i' - \Delta_{X_{i,\phi}} - \Delta_{X_{i,\psi}} \\ & - \sum_j \alpha_j \int_{-1}^{1} \phi_j(x) f_U(W_i - x) f_X(x)dx - \sum_{k=0}^{m}\sum_j \alpha_{jk}\int_{-1}^{1}\psi_{jk}(x) f_U(W_i-x) f_X(x)dx\Big]^2 \\ = \sum_{i=1}^{n} \Big(& \big[g(X_i) - E\{g(X_i|\alpha,m)|W_i\}\big] f_W(W_i) + \Delta_{W_i} + V_i' - \Delta_{X_{i,\phi}} - \Delta_{X_{i,\psi}}\Big)^2 \end{aligned}$$

Therefore, defining

$$S_1(\alpha) = \sum_{i=1}^{n} \Big(\big[g(X_i) - E\{g(X_i|\alpha,m)|W_i\}\big] f_W(W_i) + \Delta_{W_i} + V_i'\Big)^2,$$

$$S_2(\alpha) = \sum_{i=1}^{n} \Delta^2_{X_{i,\phi}}, \quad S_3(\alpha) = \sum_{i=1}^{n} \Delta^2_{X_{i,\psi}},$$

we have:

$$|S(\alpha) - S_1(\alpha)| \le 2\big[2S_1(\alpha)\{S_2(\alpha) + S_3(\alpha)\}\big]^{1/2} + 2\{S_2(\alpha) + S_3(\alpha)\}. \qquad (5.3)$$

Note that *Assumptions A1(a) and A1(b)* together imply that f_W is bounded.

Step 2: Second approximation to $S(\alpha)$. The approximation is given at (5.7). To derive it, we put $s_U = \sup_{u\in\mathbb{R}} f_U(u)$, and let $\beta = \hat f_X - f_X$ and $\Delta^2_\beta = \int_{-1}^{1} \beta^2$. Then note that the support of β is $[-1,1]$. Without loss of generality, the supports of ϕ and ψ are also confined to $[-1,1]$. Then, since $\phi_j(u) = \rho^{1/2}\phi(\rho u - j)$ and $\psi_{jk}(u) = \rho_k^{1/2}\phi(\rho_k u - j)$, the integrals

$$\int_{-1}^{1} \phi_j(x) f_U(w-x)\beta(x)dx, \quad \int_{-1}^{1} \psi_{jk}(x) f_U(w-x)\beta(x)dx$$

vanish, regardless of the value of w, unless, for some $u \in [-1,1], -1+\rho u \le j \le 1+\rho u$ or $-1+\rho_k u \le j \le 1+\rho_k u$, respectively. In particular, if it is not true that $|j| \le \rho+1$ or $|j| \le \rho_k+1$ then the respective integral vanishes. Let ν and ν_k denote the integer parts of $\rho+1$ and ρ_k+1, respectively. Since ϕ_j and ψ_{jk} are orthonormal then, writing χ for either ϕ_j or ψ_{jk}, and using the Cauchy-Schwarz inequality for integrals, we obtain:

$$\sup_{w\in\mathbb{R}} \left\{ \int_{-1}^{1} \chi(x) f_U(w-x)\beta(x)dx \right\}^2 \le s_U^2 \Delta_\beta^2. \tag{5.4}$$

Employing (5.4) and the Cauchy-Schwartz inequality for series we see that

$$\begin{aligned} S_2(\alpha) &= \sum_{i=1}^{n} \left(\sum_{j=-\infty}^{\infty} \alpha_j \int_{-1}^{1} \phi_j(x) f_U(W_i - x)\{\hat{f}_X(x) - f_X(x)\}dx \right)^2 \\ &\le \sum_{i=1}^{n} \left(\sum_{j=-\nu}^{\nu} \alpha_j^2 \right) \left(\sum_{j=-\nu}^{\nu} \sup_{w\in\mathbb{R}} \left[\int_{-1}^{1} \phi_j(x) f_U(W_i - x)\{\hat{f}_X(x) - f_X(x)\}dx \right]^2 \right) \\ &\le n(2\nu+1) s_U^2 \Delta_\beta^2 \sum_j \alpha_j^2 = S_{21}(\alpha), \end{aligned} \tag{5.5}$$

say. Similarly,

$$\begin{aligned} S_3(\alpha) &= \sum_{i=1}^{n} \left(\sum_{k=0}^{m} \sum_{j=-\nu_k}^{\nu_k} \alpha_{jk} \int_{-1}^{1} \psi_{jk}(x) f_U(W_i - x)\{\hat{f}_X(x) - f_X(x)\}dx \right)^2 \\ &\le \sum_{i=1}^{n} \left(\sum_{k=0}^{m} \sum_{j=-\nu_k}^{\nu_k} \alpha_{jk}^2 \right) \left(\sum_{k=0}^{m} \sum_{j=-\nu_k}^{\nu_k} \sup_{w\in\mathbb{R}} \left[\int_{-1}^{1} \psi_{jk}(x) f_U(w - x)\{\hat{f}_X(x) - f_X(x)\}dx \right]^2 \right) \\ &\le n s_U^2 \Delta_\beta^2 \left\{ \sum_{k=0}^{m} (2\nu_k + 1) \right\} \left(\sum_{k=0}^{m} \sum_j \alpha_{jk}^2 \right) \\ &= S_{31}(\alpha). \end{aligned} \tag{5.6}$$

Combining (5.3)-(5.6) we deduce that

$$|S(\alpha) - S_1(\alpha)| \le 2[2S_1(\alpha)\{S_{21}(\alpha) + S_{31}(\alpha)\}]^{1/2} + 2\{S_{21}(\alpha) + S_{31}(\alpha)\}. \tag{5.7}$$

Step 3: Third approximation to $S(\alpha)$. The approximation is given at (5.11). To establish it, define $S_6 = \sum_i (\Delta_{W_i} + V_i')^2$, not depending on α, and

$$S_4(\alpha) = \sum_{i=1}^{n} \Big(\big[g(X_i) - E\{g(X_i|\alpha, m)|W_i\}\big] f_W(W_i) \Big)^2, \tag{5.8}$$

$$S_5(\alpha) = \sum_{i=1}^{n} \Big(\big[g(X_i) - E\{g(X_i|\alpha, m)|W_i\}\big] f_W(W_i) \Big) (\Delta_{W_i} + V_i').$$

Then

$$S_1(\alpha) = S_4(\alpha) + 2S_5(\alpha) + S_6. \tag{5.9}$$

Using a lattice argument it can be proved that, for all $\epsilon > 0$,

$$\sup_{\alpha \in \mathcal{A}_m} |S_4(\alpha) - ES_4(\alpha)| = O_p\Big(n^{\epsilon+(1+r)/2}\Big), \quad \sup_{\alpha \in \mathcal{A}_m} |S_5(\alpha)| = O_p\Big(n^{\epsilon+(1+r)/2}\Big). \tag{5.10}$$

We shall outline the arguments in the next paragraph. Combining (5.7), (5.9) and (5.10) we deduce that, uniformly in $\alpha \in \mathcal{A}_m$, and for all $\epsilon > 0$,

$$\begin{aligned} |S(\alpha) - \{ES_4(\alpha) + S_6\}| \quad &\leq \quad 2[2\{ES_4(\alpha) + S_6 + O_p(n^{\epsilon+(1+r)/2})\}\{S_{21}(\alpha) + S_{31}(\alpha)\}]^{1/2} \\ &+ \quad 2\{S_{21}(\alpha) + S_{31}(\alpha)\} + O_p(n^{\epsilon+(1+r)/2}), \end{aligned} \tag{5.11}$$

where r is as in *Assumption A2.*

Finally in this step we derive (5.10). Observe from the definition of $\mathcal{A}_m$ (see (4.1) and (4.2)) that

$$\sum_j \alpha_j^2 + \sum_{k=0}^{m} \sum_j \alpha_{jk}^2 = \int_{\boldsymbol{c}}^{\boldsymbol{d}} g(x|\alpha, m)^2 dx \leq (d-c)B_3^2 = C_7,$$

say, where c and d are as in (4.1). Note too that $\nu \leq \rho + 1$ and $\nu_k \leq 2^k \rho + 1$. Then, $\sup_{j\in\mathbb{N}} |\alpha_j| \leq C_7^{1/2}$ and $\sup_{k\in\{0,\ldots,m\}} \sup_{j\in\mathbb{N}} |\alpha_{jk}| \leq C_7^{1/2}$ for $m \leq m_0$.

Given a constant $C_1 > r$, let $\mathcal{A}_m^*$ denote the set of all $\alpha \in \mathcal{A}_m$ for which each α_j and α_{jk} lies among the points $\{0, \pm n^{-C_1}, ..., Kn^{-C_1}\}$, where K is the smallest integer such that $Kn^{-C_1} \geq C_7^{1/2}$. If $\alpha \in \mathcal{A}_m$, with components α_j and α_{jk}, let α^*, with respective components α_j^* and α_{jk}^*, denote an element of $\mathcal{A}_m^*$ that has the property that α_j^* is as close as possible to α_j, and α_{jk}^* is as close as possible to α_{jk}, for each α_j and α_{jk}, that is, $\sup_{j\in\mathbb{N}} |\alpha_j - \alpha_j^*| \leq n^{-C_1}$ and $\sup_{k\in\{0,\ldots,m\}} \sup_{j\in\mathbb{N}} |\alpha_{jk} - \alpha_{jk}^*| \leq n^{-C_1}$. Then, recalling the definitions of $\nu < \rho + 1$ and $\nu_k \leq \rho_k + 1$,

$$\begin{aligned}
&\max_{m\le m_0}\sup_{\alpha\in\mathcal{A}_m}\sup_{w\in\mathbb{R}}\big|E\{g(X|\alpha,m)|W=w\}-E\{g(X|\alpha^*,m)|W=w\}\big|f_W(w)\\
&=\max_{m\le m_0}\sup_{\alpha\in\mathcal{A}_m}\sup_{w\in\mathbb{R}}\Big|\sum_j(\alpha_j-\alpha_j^*)\int_{-1}^1\phi_j(x)f_U(w-x)f_X(x)dx\\
&\quad+\sum_{k=0}^m\sum_j(\alpha_{jk}-\alpha_{jk}^*)\int_{-1}^1\psi_{jk}(x)f_U(w-x)f_X(x)dx\Big|\\
&\le\max_{m\le m_0}\sup_{\alpha\in\mathcal{A}_m}\left\{\sum_{j=-\nu}^{\nu}|\alpha_j-\alpha_j^*|\sup_{w\in\mathbb{R}}\Big|\int_{-1}^1\phi_j(x)f_U(w-x)f_X(x)dx\Big|\right.\\
&\quad\left.+\sum_{k=0}^m\sum_{j=-\nu_k}^{\nu_k}|\alpha_{jk}-\alpha_{jk}^*|\sup_{w\in\mathbb{R}}\Big|\int_{-1}^1\psi_{jk}(x)f_U(w-x)f_X(x)dx\Big|\right\}\\
&\le n^{-C_1}s_U\|f_X\|_2\Big\{(2\nu+1)+\sum_{k=0}^{m_0}(2\nu_k+1)\Big\}\\
&\le n^{-C_1}s_U\|f_X\|_2\Big\{(2\rho+3)+\sum_{k=0}^{m_0}(2^{k+1}\rho+3)\Big\}\\
&=n^{-C_1}s_U\|f_X\|_2\{(2\rho+3)+2(2^{m_0+1}-1)+3m_0\}\\
&\le C_3n^{r-C_1}
\end{aligned}\tag{5.12}$$

for large enough constant $0<C_3<\infty$, by *Assumption A2.*

Therefore, by (5.12),

$$\max_{m\le m_0}\sup_{\alpha\in\mathcal{A}_m}\sup_{w\in\mathbb{R}}\big|E\{g(X|\alpha,m)|W=w\}-E\{g(X|\alpha^*,m)|W=w\}\big|f_W(w)\le C_3n^{r-C_1}.$$

Hence, noting the definition of $S_4(\alpha)$ at (5.8), we see that if $C_1>r+1$ then

$$P\left\{\sup_{\alpha\in\mathcal{A}_m}|S_4(\alpha)-S_4(\alpha^*)|\le C_4n^{r+1-C_1}\right\}=1,\tag{5.13}$$

for some constant $0<C_4<\infty$. Similarly, for some constant $0<C_5<\infty$, Bernstein's inequality can be used to prove that for $0\le t\le n^{1/2}$,

$$\sup_{\alpha\in\mathcal{A}_m}P\{|S_4(\alpha)-ES_4(\alpha)|>n^{1/2}t\}\le 2\exp(-C_5t^2).\tag{5.14}$$

The number of elements of $\mathcal{A}_m^*$ equals

$$O(n^{C_6 2^{m_0(n)}}) = O\{\exp(C_6 n^r \log n)\},$$

where $0 < r < \frac{1}{2}$ (see *Assumption A2(a)*) and where we also used *Assumption A2(b)* to derive the above identity. Therefore, by (5.14), if $\frac{1}{2}r < u \le \frac{1}{2}$,

$$P\left\{\sup_{\alpha\in\mathcal{A}_m^*} |S_4(\alpha) - ES_4(\alpha)| > n^{(1/2)+u}\right\} = O\{\exp(-C_5 n^{2u})\exp(n^r \log n)\} \to 0. \qquad (5.15)$$

In particular, (5.15) implies that, for all $\epsilon > 0$,

$$n^{-1} \sup_{\alpha\in\mathcal{A}_m^*} |S_4(\alpha) - ES_4(\alpha)| = O_p(n^{\epsilon-(1-r)/2}). \qquad (5.16)$$

Taking C_1 sufficiently large in (5.13) we deduce the first part of (5.10) from (5.13) and (5.16). The second part can be derived similarly. (It is here that *Assumption A1(f)* is used.)

Step 4: Fourth approximation to $S(\alpha)$. Here we prove that

$$n^{-1} \max_{m\le m_0(n)} \sup_{\alpha\in\mathcal{A}_m} |S(\alpha) - \{ES_4(\alpha) + S_6\}| = o_p(1). \qquad (5.17)$$

By(5.5) and (5.6):

$$n^{-1}S_{21}(\alpha) \le C_8 \int_{-1}^{1} (\hat{f}_X - f_X)^2, \quad n^{-1}S_{31}(\alpha\} \le C_8' 2^m \int_{-1}^{1} (\hat{f}_X - f_X)^2, \qquad (5.18)$$

where $C_8 = s_U^2(2\rho+3)C_7$ and $C_8' = (4\rho+3)C_7 s_U^2$. The right-hand sides of the two inequalities in (5.18) do not depend on α, and *Assumption 2(a)* implies that they equal $O_p(n^{-2r})$ and $O_p(2^m n^{-2r})$, respectively. Moreover, *Assumption 2(b)* asserts that $m \le m_0(n)$ where $2^{m_0(n)} = o(n^r)$, and therefore the right-hand sides of the inequalities in (5.18) both equal $o_p(1)$, uniformly in $\alpha \in \mathcal{A}_m$ and $m \le m_0$. Hence, $n^{-1}S_{21}(\alpha)$ and $n^{-1}S_{31}(\alpha)$ both equal $o_p(1)$, uniformly in $\alpha \in \mathcal{A}_m$ and $m \le m_0(n)$:

$$n^{-1} \max_{m\le m_0(n)} \sup_{\alpha\in\mathcal{A}_m} \{S_{21}(\alpha) + S_{31}(\alpha)\} \to 0, \qquad (5.19)$$

where the convergence is in probability. By *Assumption 1(a)* and *Assumption 1(d)*,

$$n^{-1}E\{S_4(\alpha)\} \leq 4B_3^2 \int f_W(w)^3 dw < \infty, \tag{5.20}$$

where B_3 is as in (4.2).(Note that *Assumption 1(b)* implies that f_W is bounded.) It is straightforward to show that

$$n^{-1}S_6 = O_p(1); \tag{5.21}$$

recall that S_6 does not depend on α. Combining (5.11) and (5.19)-(5.21) we deduce that (5.17) holds.

Step 5: Completion of proof of theorem. First, we note that by Step 4, $n^{-1}S(\alpha) = n^{-1}ES_4(\alpha) + n^{-1}S_6 + o_p(1)$, uniformly in $\alpha \in \mathcal{A}_m$. Then

$$\begin{aligned} n^{-1}ES_4(\alpha) &= E\Big(\big[g(X) - E\{g(X|\alpha,m)|W\}\big]f_W(W)\Big)^2 \\ &= E\Big(\big[g(X) - E\{g(X)|W\} + E\{g(X)\} - E\{g(X|\alpha,m)|W\}\big]f_W(W)\Big)^2 \\ &= S_{41} + 2S_{42}(\alpha) + S_{43}(\alpha) \end{aligned} \tag{5.22}$$

where

$$\begin{aligned} S_{41} &= E\left(\left[g(X) - E\{g(X)|W\}\right] f_W(W)\right)^2 \\ S_{42}(\alpha) &= E\left(\left[g(X) - E\{g(X)|W\}\right]\left[E\{g(X)|W\} - E\{g(X|\alpha,m)|W\}\right] f_W(W)^2\right) \\ S_{43}(\alpha) &= E\left(\left[E\{g(X)|W\} - E\{g(X|\alpha,m)|W\}\right]^2 f_W(W)^2\right). \end{aligned} \tag{5.23}$$

Using the total expectation property $E\{h(X)\} = E\left[E\{h(X)|W\}\right]$ for any measurable function h we see that $S_{42}(\alpha) = 0$ holds for all $\alpha \in \mathcal{A}_m$. Therefore

$$n^{-1}ES_4(\alpha) = S_{41} + S_{43}(\alpha).$$

Define the functional $\kappa_1(h) = E\left(\left[E\{g(X)|W\} - E\{h(X)|W\}\right]^2 f_W(W)^2\right)$ and observe that with $g_{\alpha,m}(x) = g(x|\alpha,m)$ we have that $\kappa_1(g_{\alpha,m}) = S_{43}(\alpha)$. Since $\kappa_1 \geq 0$ and $\kappa_1 = 0$ when $h = g$

we see that $h = g$ is a minimizer of κ_1. This minimizer is unique. Indeed, suppose that $\kappa_1(h) = E\left([E\{g(X)|W\} - E\{h(X)|W\}]^2 f_W(W)^2\right) = 0$, then $E\{g(X)|W\} = E\{h(X)|W\}$ almost surely. By *Assumption A1 e)*, we have that $h(X) = g(X)$ almost surely. Therefore, κ_1 is uniquely minimized at g.

Noting that $n^{-1}S(\alpha) = n^{-1}ES_4(\alpha) + n^{-1}S_6 + o_p(1) = S_{41} + \kappa_1(g_{\alpha,m}) + n^{-1}S_6 + o_p(1)$ uniformly in $\mathcal{A}_m$, and that $S(\alpha)$ is minimized at $\alpha = \hat{\alpha}$, we have that

$$\kappa_1(g_{\hat{\alpha},m}) \xrightarrow{p} \kappa_1(g) = 0 \tag{5.24}$$

as $n \to \infty$.

We will now show that $\int_{-1}^{1}\{g_{\hat{\alpha},m}(x) - g(x)\}^2 dx \xrightarrow{p} 0$ by showing that for each subsequence n_k of n there exists a subsubsequence $n_{k(s)}$ of n_k such that $\int_{-1}^{1}\{g_{\hat{\alpha},m}(x) - g(x)\}^2 dx \xrightarrow{p} 0$ as $s \to \infty$.

Let n_k be an arbitrary subsequence of n. Let $\mathcal{C}(B_3)$ denote the class of functions h that satisfy (4.1) and (4.2). Then $\mathcal{C}(B_3)$ includes $g_{\hat{\alpha},m}$ by the definition of $\hat{\alpha}$. Functions in $\mathcal{C}(B_3)$ can be approximated uniformly and arbitrarily closely in L_2 on $[-1,1]$ by a lattice laid down in $[-1,1]$. Specifically, for any $\epsilon \in (0,1)$ there exists an integer $l = l(\epsilon) > 0$ such that for each $h \in \mathcal{C}(B_3)$ there is a right continuous step function h^* defined on (the subintervals created by) a regular l-point lattice in $[-1,1]$, and taking values only in $[-B_3, B_3]$ (where B_3 is as in (4.2)), such that

$$\int_{-1}^{1}(h - h^*)^2 \leq \epsilon.$$

Taking $h = g_{\hat{\alpha},m}$ we obtain the step-function approximation $g^*_{\hat{\alpha},m}$:

$$P\left\{\int_{-1}^{1}(g_{\hat{\alpha},m} - g^*_{\hat{\alpha},m})^2 \leq \epsilon\right\} = 1. \tag{5.25}$$

In this paragraph we keep $\epsilon > 0$ fixed, and define c_j, $j = 1, ..., l(\epsilon)$, to be the steps of $g^*_{\hat{\alpha},m}$ (we suppress the dependency of c_j on ϵ and n). Define $\mathcal{E} = (\epsilon_1, \epsilon_2, \dots)$ to be a positive sequence decaying to zero. Then, along a subsequence $n_{k(s)}$ by the Prohorov - Varadajan theorem (see, e.g., p. 303 in Athreya and Lahiri (2006)), c_j converges in distribution to some random variable c_j^∞. By the Kolmogorov extension theorem, there exists a probability space $(\Omega, \Sigma, \mathbb{P})$ such that $c_{j,n_{k(s)}}$ is defined on $(\Omega, \Sigma, \mathbb{P})$ such that $c_j = c_{j,n_{k(s)}} + o_p(1)$, so that $\mathcal{G}_{\epsilon,n_{k(s)}}$, $\mathcal{G}_\epsilon$ are step functions defined by the values $c_{j,n_{k(s)}}$ and c_j^∞ for $j = 1, ..., l(\epsilon)$ respectively, and they satisfy (4.1) and (4.2). Then by construction,

$$\int_{-1}^{1} (g^*_{\tilde{\alpha},m} - \mathcal{G}_{\epsilon_i,n_{k(s)}})^2 \xrightarrow{p} 0 \tag{5.26}$$

as $s \to \infty$ for a fixed natural number i. Combining (5.24)-(5.26) we see that we can construct a sequence $\epsilon_i \equiv \epsilon_{n_{k(s)}}$ converging to zero sufficiently slowly enough so that the corresponding sequence $\mathcal{G}_{\epsilon_i}$, satisfies

$$\kappa_1(\mathcal{G}_{\epsilon_i}) \to \kappa_1(g) = 0 \tag{5.27}$$

in probability as $i \to \infty$. Here the value of $l = l_i$ will diverge as ϵ_i decreases, and without loss of generality it diverges dyadically: $l_i = 2^i$ for $i \geq 1$. Express $\mathcal{G}_{\epsilon_i}$ using the Haar wavelet basis rescaled to $[-1, 1]$. Then, for any $\delta > 0$, we can approximate $\mathcal{G}_{\epsilon_i}$ to within δ, in L_2 and for all i, using at most the first N_δ terms in a complete sequence of orthonormal functions $\chi_1, \chi_2, \ldots$ representing an enumeration of the Haar basis. Since each $\mathcal{G}_{\epsilon_i} \in \mathcal{C}(B_3)$ then neither N_δ nor our ordering of the functions χ_j need depend on i, and so for each value of that index,

$$P\left\{ \left\| \mathcal{G}_{\epsilon_i} - \sum_{j=1}^{N_\delta} \Big(\int_{-1}^{1} \mathcal{G}_{\epsilon_i}\chi_j \Big) \chi_j \right\|_2 \leq \delta \right\} = 1. \tag{5.28}$$

Take i_k to be a subsequence of i. Define $\mathcal{G}^\delta_{\epsilon_i} = \sum_{j=1}^{N_\delta} \big(\int_{-1}^{1} \mathcal{G}_{\epsilon_i}\chi_j \big) \chi_j$. Define $\mathcal{D} = (\delta_1, \delta_2, \ldots)$ to be a sequence decaying to zero. Note that $\mathcal{G}^{\delta_K}_{\epsilon_{ik}}$ is defined by the uniformly bounded finite dimensional vector $\big(\mathcal{G}^{\delta_K}_{\epsilon_{ik}}\big)_{1 \leq j \leq N_\delta}$, and hence by the Prohorov-Varadajan Theorem, we can construct a further subsequence $\epsilon_{i_k(s)}$ such that $\mathcal{G}^{\delta_K}_{\epsilon_{ik(s)}}$ converges in distribution to a random function $\mathcal{G}_0^\delta$. Then by the continuous mapping theorem,

$$\int_{-1}^{1} \mathcal{G}_{\epsilon_{i_{k(s)}}}^{\delta_K} \chi_j = \int_{-1}^{1} \mathcal{G}_{\epsilon_{i_{k(s)}}} \chi_j \xrightarrow{D} \int_{-1}^{1} \mathcal{G}_0^{\delta_K} \chi_j,$$

for each fixed $K \geq 1, j \leq N_{\delta_K}$. Since the left hand side does not depend on δ_K, we have that $\mathcal{G}_0^{\delta_k} \equiv \mathcal{G}_0$.

Hence, by (5.27) and by choosing $\delta \equiv \delta(i)$ to converge slowly to 0, $\kappa_1(\mathcal{G}_0) = \kappa_1(g) = 0$ almost surely. Note that *Assumption A1 e)* implies that if $\kappa_1(h) = 0$ then $h(X) = g(X)$ almost surely, and therefore, $\mathcal{G}_0 = g$ almost surely. Hence, $\mathcal{G}_{e_{i_{k(s)}}}$ converges weakly to g. Since this holds for any subsubsequence $i_{k(s)}$ of any arbitrary subsequence i_k, we have that $\mathcal{G}_{\epsilon_i}$ converges weakly to g. Armed with this result and (5.28), an argument by contradiction can now be used to prove that as $\epsilon_i = \epsilon_i(n) \to 0$, $\int_{-1}^{1}(\mathcal{G}_{\epsilon_i, n_{k_s}} - g)^2 \xrightarrow{p} 0$. Hence, by (5.25) and (5.26),

$$\int_{-1}^{1} (g_{\hat{\alpha},m} - g)^2 \to 0 \tag{5.29}$$

in probability as $n_{k_s} \to \infty$. Since we have found a subsubsequence n_{k_s} for any arbitrary subsequence n_k of n such that (5.29) holds, then (5.29) holds as $n \to \infty$. Since, by construction, $g(x|\hat{\alpha}, m)$ and g are bounded then (5.29), which is equivalent to (4.3) in the case $q = 2$, implies (4.3) for all $q \in (0, \infty)$.

References

Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory.* Springer.