A note on "Convergence rates and asymptotic normality for series estimators": uniform convergence rates

Robert M. de Jong*
Michigan State University

This revision, January 2002

Abstract

This paper establishes improved uniform convergence rates for series estimators. Series estimators are least-squares fits of a regression function where the number of regressors depends on sample size. I will specialize my results to the cases of polynomials and regression splines. These results improve upon results obtained earlier by Newey, yet fail to attain the optimal rates of convergence.

JEL classification codes: C14; C21.

Keywords and phrases: series estimation, uniform consistency, nonparametric regression.

---

*Current address: Department of Economics, Michigan State University, 217 Marshall Hall, East Lansing, MI 48824, USA, Email dejongr@msu.edu, phone (517)-353-9933. I thank an anonymous referee for truly excellent comments and for pointing out an error in an earlier version of this paper.

1 Introduction

In a recent and interesting contribution, Newey (1997) establishes consistency in L_2-norm, uniform consistency, and asymptotic normality of series estimators. Newey (1997) improves several results from the existing literature on this subject. In this note, I show that the argument in Newey (1997) can be extended to show that his uniform convergence rates can be improved. The mathematical machinery that is used for establishing parts of this result is from Pollard (1984).

2 Definitions and main results

In this note, I will use notation as in Newey (1997). $g_0(x) = E(y|x)$ denotes the true conditional expectations function that is to be estimated. Let (y_i, x_i), $i = 1, \ldots, n$ be a random sample from a distribution $F(y, x)$. The vectors of approximating functions will be denoted as

$$p^K(x) = (p_{1K}(x), \ldots, p_{KK}(x))', \tag{1}$$

and it will be assumed that linear combinations $p^K(x)'\beta$ can approximate g_0. Let $\mathcal{X}$ denote the support of the x. Define $\partial^\lambda h(x) = \partial^\lambda h(x)/\partial x_1^{\lambda_1} \ldots \partial x_r^{\lambda_r}$, where λ denotes a vector of nonnegative elements such that $|\lambda| = \sum_{j=1}^r \lambda_j$, and let $|g|_d$ denote

$$\max_{|\lambda| \leq d} \sup_{x \in \mathcal{X}} |\partial^{|\lambda|} g(x)|. \tag{2}$$

Let $F_0(x)$ denote the probability distribution of the x_i. Below, as in Newey (1997), I will sometimes drop the x argument from expressions such as $g_0(x)$ and $\hat{g}(x)$. The series estimator for g_0 is $\hat{g}(x) = p^K(x)'\hat{\beta}$, where $\hat{\beta} = (P'P)^{-1}P'Y$, $P = [p^K(x_1), \ldots, p^K(x_n)]'$ and $Y = (y_1, \ldots, y_n)'$. From the assumptions, it follows that $P'P/n$ is asymptotically nonsingular, which makes it valid to use the above notation. Note that the dimension K of $p^K(x)$ is assumed to grow with n. For establishing his results, Newey (1997) makes three assumptions (Assumptions 1,2 and 3 below). I will need the following four assumptions:

Assumption 1 *$(y_1, x_1), \ldots, (y_n, x_n)$ are i.i.d. and* $\mathrm{Var}(y|x)$ *is bounded.*

Assumption 2 *For every K there is a nonsingular constant matrix B such that for $P^K(x) = Bp^K(x)$; (i) the smallest eigenvalue of $E[P^K(x_i)P^K(x_i)']$ is bounded away from zero uniformly in K and; (ii) $K = K(n)$ and $\zeta_0^2(K)K/n \to 0$ as $n \to \infty$.*

Assumption 3 *For an integer $d \geq 0$ there are α, β such that $|g_0 - \beta' p^K(x)|_d = O(K^{-\alpha})$ as $K \to \infty$.*

Assumption 4 *$\mathcal{X}$ is a compact subset of $\mathbb{R}^r$; $E(\varepsilon_i^4|x_i) \leq C$ almost surely for some constant C; for some $\varphi > 0$,*

$$\limsup_{n\to\infty} n^{-\varphi}\zeta_{d+1}(K)K^{\alpha}(K^{1/2}n^{-1/2} + K^{-\alpha}) < \infty; \tag{3}$$

and $\sup_{n\geq 1} K^{1/2}n^{-1/2}\zeta_d(K) < \infty$, where for any nonnegative integer d,

$$\zeta_d(K) = \max_{|\lambda|\leq d} \sup_{x\in\mathcal{X}} \| \partial^{\lambda} P^K(x) \| . \tag{4}$$

The fourth conditional moment condition is potentially restrictive; however, I was unable to remove this condition. The condition of Equation (3) is hardly a condition; it essentially imposes that $\zeta_{d+1}(K)$ and K can be bounded by polynomials in K and n respectively. Note that Assumptions 1, 2 and 3 are the assumptions from Newey (1997), while Assumption 4 is an assumption that is not needed in Newey (1997). The assumption of compactness of $\mathcal{X}$ is not needed for Theorem 1 of Newey (1997), but for many series estimators - such as the ones discussed in Newey (1997) - this assumption is natural. Alternatively, total boundedness of $\mathcal{X}$ can be assumed.

The central results of this paper are the following:

Theorem 1 *Under Assumptions 1, 2, 3, and 4,*

$$|\hat{g} - g_0|_d = O_P(K^{-\alpha}\zeta_d(K) + \zeta_d(K)(\log(n)/n)^{1/2}). \tag{5}$$

This theorem is an alternative to the uniform convergence rate of Newey's (1997) Theorem 1, where a rate of $K^{-\alpha}\zeta_d(K) + \zeta_d(K)K^{1/2}n^{-1/2}$ is established. Clearly if $K/\log(n) \to \infty$, the result of the above Theorem 1 will provide a sharper uniform convergence rate.

The following result characterizes the rate of convergence. Let $\tilde{\beta} = Q^{-1}Ep^K(x_i)g_0(x_i)$, where $Q = Ep^K(x_i)p^K(x_i)'$.

Theorem 2 *Under Assumptions 1, 2, 3, and 4,*

$$|\hat{g} - g_0|_d = O_P(K^{-\alpha} + (\log(n)/n)^{1/2}\zeta_d(K) + |(\tilde{\beta} - \beta)'p^K(x)|_d). \quad (6)$$

The problem of characterizing the asymptotic behavior of the last term (essentially the difference between the best L_2- and uniform approximation) is difficult, and I was unable to find an appropriate upper bound for the general case. I do not know whether the problem of bounding $|(\beta - \tilde{\beta})'p^K(x)|_d$ for any particular series $p^K(.)$ can be solved in such a way as to attain the optimal convergence rates as determined by Stone (1982). It may be that the presence of this term, and its asymptotic relevance, is a generic property of series estimators.

3 Power series

For power series, $\zeta_d(K) = K^{1+2d}$ according to Newey (1997), under the following assumption:

Assumption 5 *$\mathcal{X}$ is a Cartesian product of compact connected intervals on which x has a probability density function that is bounded away from zero.*

For the definition of multivariate power series, see Newey (1997). The following assumption ensures that one can find a useful value for α.

Assumption 6 *$g_0(x) = E(y|x)$ is continuously differentiable of order s on $\mathcal{X}$.*

Newey (1997) now claims that for $d = 0$ we can choose $\alpha = s/r$, while for $r = 1$ the only result is $\alpha = s - d$. Using those values for α, we obtain

Theorem 3 *For power series, if Assumption 1, 5, and 6 hold, $K^3/n \to 0$, $s > d$, and $\sup_{n \geq 1} K^{3+4d}n^{-1} < \infty$, then for $d = 0$*

$$|\hat{g} - g_0|_0 = O_P(K^{-s/r+1+2d} + K^{1+2d}(\log(n)/n)^{1/2}), \quad (7)$$

and for $r = 1$,

$$|\hat{g} - g_0|_d = O_P(K^{d-s+1+2d} + K^{1+2d}(\log(n)/n)^{1/2}). \quad (8)$$

The above theorem is unable to show that the optimal convergence rate for nonparametric estimators can be attained for the case of power series.

4 Regression splines

This section covers regression splines as discussed and defined in Newey (1997). For regression splines, we need the following assumption:

Assumption 7 $\mathcal{X} = [-1, 1]^r$.

For regression splines, Newey notes that $\zeta_d(K) = K^{1/2+d}$. Newey (1997) notes that for regression splines, identically to the case of power series, we can set $\alpha = s/r$ if $d = 0$ and $\alpha = s - d$ if $r = 1$. Combining these values of α with the bound for $\zeta_d(K)$, we obtain

Theorem 4 *For splines, if Assumptions 1, 5, 6, and 7 are satisfied, $K^2/n \to 0$, $s > d$, and $\sup_{n\geq 1} K^{2+2d} n^{-1} < \infty$, then for $d = 0$*

$$|\hat{g} - g_0|_0 = O_P(K^{-s/r+1/2+d} + K^{1/2+d}(\log(n)/n)^{1/2}), \tag{9}$$

and for $r = 1$,

$$|\hat{g} - g_0|_d = O_P(K^{2d-s+1/2} + K^{1/2+d}(\log(n)/n)^{1/2}). \tag{10}$$

5 Mathematical proofs

First, we need to make some definitions. Newey (1997) notes that (under Assumption 2) without loss of generality it can be assumed that $Q = Ep^K(x_i)p^K(x_i)' = I$. Newey also shows that under Assumptions 1 and 2, the minimal eigenvalue of $\hat{Q} = P'P/n$ converges to one. Identically to Newey (1997), let 1_n equal 1 if the minimal eigenvalue of $\hat{Q}$ is greater than 1/2 and 0 otherwise. Clearly $\lim_{n\to\infty} P(1_n = 1) = 1$. For any matrix A, define $\| A \| = (\mathrm{tr}(A'A))^{1/2}$, where "tr" denotes the trace operator, and let $\lambda_{max}(A)$ denotes the maximal eigenvalue of A. Let $G = (g_0(x_1), \ldots, g_0(x_n))'$ and let $\varepsilon = Y - G$.

Before starting the proof of Theorem 1, we first state and prove five lemmas that all assume that Assumptions 1, 2, 3 and 4 hold.

Lemma 1 $\sup_{x\in\mathcal{X}} 1_n |\partial^\lambda p^K(x)'(\hat{Q}^{-1} - I)P'\varepsilon/n| = o_P(n^{-1/2}\zeta_d(K))$.

Proof of Lemma 1:

This follows because

$$E \sup_{x\in\mathcal{X}} 1_n |\partial^\lambda p^K(x)'(\hat{Q}^{-1} - I)P'\varepsilon/n|^2$$

$$= E \sup_{x\in\mathcal{X}} 1_n n^{-2} |\partial^\lambda p^K(x)'(I - \hat{Q})\hat{Q}^{-1}P'\varepsilon|^2$$

$$\leq E n^{-2} 1_n \varepsilon' P \hat{Q}^{-1}(I - \hat{Q})^2 \hat{Q}^{-1} P'\varepsilon \sup_{x\in\mathcal{X}} |\partial^\lambda p^K(x)'\partial^\lambda p^K(x)|$$

$$\leq n^{-2} E 1_n \operatorname{tr}\left((I - \hat{Q})\hat{Q}^{-1}P'E(\varepsilon\varepsilon'|x_1, \ldots, x_n)P\hat{Q}^{-1}(I - \hat{Q})\right) \sup_{x\in\mathcal{X}} |\partial^\lambda p^K(x)'\partial^\lambda p^K(x)|$$

$$\leq C n^{-1}\zeta_d(K)^2 E 1_n \operatorname{tr}\left((I - \hat{Q})\hat{Q}^{-1}(I - \hat{Q})\right)$$

$$\leq 2C n^{-1}\zeta_d(K)^2 E \operatorname{tr}\left((I - \hat{Q})'(I - \hat{Q})\right)$$

$$= 2C n^{-1}\zeta_d(K)^2 E \parallel I - \hat{Q} \parallel^2$$

$$= o(n^{-1}\zeta_d(K)^2) \tag{11}$$

by assumption, where the first inequality is $(a'b)^2 \leq a'ab'b$, and the last equality uses

$$E \parallel I - \hat{Q} \parallel^2 = O(K\zeta_0(K)^2 n^{-1}) = o(1) \tag{12}$$

as established in Newey (1997). ∎

Lemma 2 $\sup_{x\in\mathcal{X}} 1_n |\partial^\lambda p^K(x)'\hat{Q}^{-1}P'(G - P\tilde{\beta})/n|^2 = O_P(n^{-1}\zeta_d(K)^2 K^{1-2\alpha})$.

Proof of Lemma 2:

This follows because

$$
\begin{aligned}
& E\sup_{x\in\mathcal{X}} 1_n|\partial^\lambda p^K(x)'\hat{Q}^{-1}P'(G-P\tilde{\beta})/n|^2 \\
& \leq E1_n n^{-2}(G-P\tilde{\beta})P\hat{Q}^{-1}\hat{Q}^{-1}P'(G-P\tilde{\beta})\sup_{x\in\mathcal{X}}|\partial^\lambda p^K(x)'\partial^\lambda p^K(x)| \\
& \leq 4n^{-2}\zeta_d(K)^2 E(G-P\tilde{\beta})'PP'(G-P\tilde{\beta}) \\
& \leq 4n^{-2}\zeta_d(K)^2 E(\sum_{i=1}^n (g_0(x_i)-p^K(x_i)'\tilde{\beta})p^K(x_i)')(\sum_{l=1}^n (g_0(x_l)-p^K(x_l)'\tilde{\beta})p^K(x_l)) \\
& \leq 4n^{-1}\zeta_d(K)^2 E(g_0(x_i)-p^K(x_i)'\tilde{\beta})^2 p^K(x_i)'p^K(x_i) \\
& = O(n^{-1}\zeta_d(K)^2 K^{1-2\alpha}) \qquad (13)
\end{aligned}
$$

by assumption. Note that the fourth inequality uses the fact that

$$E(g_0(x_i)-p^K(x_i)'\tilde{\beta})p^K(x_i) = 0 \qquad (14)$$

by the definition of $\tilde{\beta}$. The "0" here denotes a zero K-vector. ∎

Lemma 3 *For each x, $E|n^{1/2}\zeta_d(K)^{-1}\partial^\lambda p^K(x)'P'\varepsilon/n|^2 \leq C$ for some constant C not depending on x.*

Proof of Lemma 3:

This follows from noting that for each $x \in \mathcal{X}$,

$$
\begin{aligned}
& n\zeta_d(K)^{-2}E|\partial^\lambda p^K(x)'P'\varepsilon/n|^2 \\
& \leq n\zeta_d(K)^{-2}E(E(|\partial^\lambda p^K(x)P'\varepsilon/n|^2|x_1,\ldots,x_n)) \\
& \leq Cn\zeta_d(K)^{-2}\partial^\lambda p^K(x)'E(P'P/n)\partial^\lambda p^K(x)/n \\
& = Cn\zeta_d(K)^{-2}(\zeta_d(K)^2/n) = C \qquad (15)
\end{aligned}
$$

by assumption. ∎

Lemma 4 *For some constant C, $\sup_{n\geq 1} E\lambda_{max}(n^{-1}\sum_{i=1}^n \varepsilon_i^2 p^K(x_i)p^K(x_i)') \leq C$.*

Proof of Lemma 4:

Let C_1 and C_2 be constants such that $E(\varepsilon_i^2|x_i) \leq C_1$ almost surely and $E(\varepsilon_i^4|x_i) \leq C_2$ almost surely. Then

$$E\lambda_{max}(n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 p^K(x_i)p^K(x_i)')$$

$$\leq \lambda_{max}(E\varepsilon_i^2 p^K(x_i)p^K(x_i)') + E\lambda_{max}(n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 p^K(x_i)p^K(x_i)' - E\varepsilon_i^2 p^K(x_i)p^K(x_i)')$$

$$\leq C_1\lambda_{max}(Ep^K(x_i)p^K(x_i)') + (E \parallel n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 p^K(x_i)p^K(x_i)' - E\varepsilon_i^2 p^K(x_i)p^K(x_i)' \parallel^2)^{1/2}$$

$$\leq C_1 + (\sum_{k=1}^{K}\sum_{l=1}^{K}E(n^{-1}\sum_{i=1}^{n}\varepsilon_i^2 p_{lK}(x_i)p_{kK}(x_i) - E\varepsilon_i^2 p_l(x_i)p_k(x_i))^2)^{1/2}$$

$$\leq C_1 + (\sum_{k=1}^{K}\sum_{l=1}^{K}n^{-1}E\varepsilon_i^4 p_{lK}(x_i)^2 p_{kK}(x_i)^2)^{1/2}$$

$$\leq C_1 + (C_2\sum_{k=1}^{K}\sum_{l=1}^{K}n^{-1}Ep_{lK}(x_i)^2 p_{kK}(x_i)^2)^{1/2}$$

$$\leq C_1 + (C_2\zeta_0(K)^2 K/n)^{1/2}, \tag{16}$$

where the last inequality follows from the reasoning in Newey (1997), page 162. By Assumption 2, the second term converges to zero, and therefore the result follows. ∎

Lemma 5 *Let $\{x^j : j = 1, \ldots, N\}$ denote a set of x-values such that $x^j \in \mathcal{X}$ for all j. If $N = O(n^\gamma)$ some $\gamma > 0$, then*

$$1_n \max_j |\partial^\lambda p^K(x^j)\hat{Q}^{-1}P'\varepsilon/n| = O_P((\log(n))^{1/2}n^{-1/2}\zeta_d(K))). \tag{17}$$

Proof of Lemma 5:

Because of the result from Lemma 1, it suffices to show that

$$\max_j |\partial^\lambda p^K(x^j)P'\varepsilon/n| = O_P((\log(n))^{1/2}n^{-1/2}\zeta_d(K)). \tag{18}$$

Next, note that by Lemma 3, it follows that for all $M > 0$ and for n large enough,

$$\max_j P\left((\log n)^{-1/2}n^{1/2}\zeta_d(K)^{-1}|\partial^\lambda p^K(x^j)'P'\varepsilon/n| \le M\right) \ge 1/2. \tag{19}$$

Therefore, following Pollard (1984), page 15, Equation (11), we have

$$P\left((\log n)^{-1/2}n^{1/2}\zeta_d(K)^{-1}\max_j |\partial^\lambda p^K(x^j)'P'\varepsilon/n| > M\right)$$

$$\le 4P\left((\log n)^{-1/2}n^{1/2}\zeta_d(K)^{-1}\max_j |\partial^\lambda p^K(x^j)'n^{-1}\sum_{i=1}^n p^K(x_i)\varepsilon_i\sigma_i| > M/4\right) = 4P_1 \tag{20}$$

for all $M > 0$, where $\sigma_i = 1$ with probability 1/2 and $\sigma_i = -1$ with probability 1/2 also, and σ_i is independent of ε_l and x_l for all l including i. Note that M will be specified at the end of the proof. Next, note that by conditioning on $X = \{(x_1, \varepsilon_1), \ldots, (x_n, \varepsilon_n)\}$,

$$P_1 = EP\left((\log n)^{-1/2}n^{1/2}\zeta_d(K)^{-1}\max_j |\partial^\lambda p^K(x^j)n^{-1}\sum_{i=1}^n p^K(x_i)\varepsilon_i\sigma_i| > M/4|X\right)$$

$$\le E\max\left(1, N\max_j P\left((\log n)^{-1/2}n^{1/2}\zeta_d(K)^{-1}|\partial^\lambda p^K(x^j)n^{-1}\sum_{i=1}^n \varepsilon_i\sigma_i p^K(x_i)| > M/4|X\right)\right)$$

$$\le E\max(1, 2N\max_j \exp(-2(M/4)^2 n(\log(n))\zeta_d(K)^2/\sum_{i=1}^n(2\partial^\lambda p^K(x^j)'p^K(x_i)\varepsilon_i)^2)$$

$$= P_2 \tag{21}$$

say, where the second inequality is Hoeffding's (see Pollard (1984), Appendix B). Let C be the constant of Lemma 4. Then, note that for all $\eta > 0$,

$$P_2 \le P(\max_j |\sum_{i=1}^n(\partial^\lambda p^K(x^j)'\varepsilon_i p^K(x_i))^2| > \eta^{-1}Cn\zeta_d(K)^2)$$

$$+2N\exp(-2(M/4)^2 4^{-1} n(\log(n))\zeta_d(K)^2/(\eta^{-1}Cn\zeta_d(K)^2))$$

$$\leq \eta C^{-1}\zeta_d(K)^{-2} E\max_j \partial^\lambda p^K(x^j)'(n^{-1}\sum_{i=1}^n \varepsilon_i^2 p^K(x_i)p^K(x_i)')\partial^\lambda p^K(x^j)$$

$$+2N\exp(-\eta C^{-1}M^2(\log(n))/32)$$

$$\leq \eta C^{-1} E\lambda_{max}(n^{-1}\sum_{i=1}^n \varepsilon_i^2 p^K(x_i)p^K(x_i)') + 2N\exp(-\eta C^{-1}M^2(\log(n))/32)$$

$$\leq \eta + 2N\exp(-\eta C^{-1}M^2(\log(n))/32)$$

$$\leq \eta + O(n^\gamma n^{-\eta C^{-1}M^2/32}), \tag{22}$$

where the second inequality is Markov's and the fourth inequality uses the result from Lemma 4. Clearly if we choose $M^2 > 32\gamma C/\eta$, the last term will be $o(1)$. ∎

Proof of Theorems 1 and 2:

Let $\{x^j : j = 1, \ldots, N\}$ be the smallest set of x-values such that for all $x \in \mathcal{X}$ we can find a $j(x)$ such that $\| x - x^j \| \leq b_n = n^{-\varphi}$, where φ is as specified in Assumption 4. Since $x \in \mathbb{R}^r$ and $\mathcal{X}$ is compact, $N = O(b_n^{-r}) = O(n^{\varphi r})$. To start the proof, first notice that

$$|\partial^\lambda(\hat{g}(x) - g_0(x)) - \partial^\lambda(\hat{g}(x^{j(x)}) - g_0(x^{j(x)}))| \leq |\hat{g} - g_0|_{d+1} \| x - x^{j(x)} \|$$

$$\leq |\hat{g} - g_0|_{d+1} b_n \tag{23}$$

if $|\lambda| \leq d$. Therefore,

$$1_n|\hat{g} - g_0|_d$$

$$\leq \max_\lambda \sup_{x\in\mathcal{X}} |\partial^\lambda(\hat{g}(x) - \hat{g}(x^{j(x)}) - g_0(x) + g_0(x^{j(x)}))|$$

$$+1_n \max_j |\hat{g}(x^j) - g_0(x^j)|_d$$

$$\leq |\hat{g} - g_0|_{d+1} b_n + |p^K(x)'\beta - g_0|_d + 1_n \max_\lambda \max_j |\partial^\lambda p^K(x^j)'(\hat{\beta} - \beta)|$$

$$\leq |\hat{g} - g_0|_{d+1} b_n + |p^K(x)'\beta - g_0|_d$$

$$+1_n \max_\lambda \max_j |\partial^\lambda p^K(x^j)'\hat{Q}^{-1} P'(G - P\beta)/n|$$

$$+1_n \max_\lambda \max_j |\partial^\lambda p^K(x^j)'\hat{Q}^{-1} P'\varepsilon/n|. \tag{24}$$

Theorem 1 now follows by noting that the first term is $O_P(b_n \zeta_{d+1}(K)(K^{1/2}n^{-1/2} + K^{-\alpha}))$ by Theorem 1 of Newey (1997), the second term is $O_P(K^{-\alpha})$ by assumption, the third is $O_P(K^{-\alpha}\zeta_d(K))$ by the results in the proof of Theorem 1 of Newey (1997), and the fourth is

$$O_P((\log(n))^{1/2} n^{-1/2} \zeta_d(K)) \tag{25}$$

by Lemma 5. Therefore, combining those results we conclude that

$$|\hat{g}(x) - g_0(x)|_d = O_P\left(n^{-\varphi}\zeta_{d+1}(K)(K^{1/2}n^{-1/2} + K^{-\alpha})\right.$$

$$\left. +K^{-\alpha}\zeta_d(K) + n^{-1/2}K^{1/2-\alpha}\zeta_d(K) + (\log(n))^{1/2} n^{-1/2}\zeta_d(K)\right). \tag{26}$$

The first and third term are $O(K^{-\alpha})$ by Assumption 4, and the result of Theorem 1 follows. The proof of Theorem 2 follows by noting that

$$1_n|\hat{g} - g_0|_d \leq |\hat{g} - g_0|_{d+1} b_n + |p^K(x)'\beta - g_0|_d + |(\beta - \tilde{\beta})'p^K(x)|_d$$

$$+1_n \max_\lambda \max_j |\partial^\lambda p^K(x^j)'\hat{Q}^{-1} P'(G - P\tilde{\beta})/n|$$

$$+1_n \max_\lambda \max_j |\partial^\lambda p^K(x^j)'\hat{Q}^{-1} P'\varepsilon/n|. \tag{27}$$

Using the above results and Lemma 2, it follows that

$$|\hat{g}(x) - g_0(x)|_d = O_P(\left(n^{-\varphi}\zeta_{d+1}(K)(K^{1/2}n^{-1/2} + K^{-\alpha})\right.$$

$$\left. +K^{-\alpha} + |(\beta - \tilde{\beta})'p^K(x)|_d + n^{-1/2}K^{1/2-\alpha}\zeta_d(K) + (\log(n))^{1/2} n^{-1/2}\zeta_d(K)\right). \tag{28}$$

The first and fourth term are now $O(K^{-\alpha})$ by Assumption 4. The result of Theorem 2 now follows. ∎

Proof of Theorem 3 and 4:

Theorems 3 and 4 are obtained by combining the rates for $\zeta_d(K)$ and the values for α with Theorem 1. ∎

References

Newey, W.K., 1997, Convergence rates and asymptotic normality for series estimators, Journal of Econometrics 79, 147-168.

Pollard, D., 1984. Convergence of Stochastic Processes (Springer-Verlag, New York).

Stone, C.J., 1982, Optimal global rates of convergence for nonparametric models, Annals of Statistics 10, 1040-105 3.