Econometric Theory, **24**, 2008, 1463–1499. Printed in the United States of America.
doi:10.1017/S0266466608080584

UNIFORM CONVERGENCE OF SERIES ESTIMATORS OVER FUNCTION SPACES

KYUNGCHUL SONG
University of Pennsylvania

This paper considers a series estimator of $\mathbf{E}[\alpha(Y)|\lambda(X) = \bar{\lambda}]$, $(\alpha,\lambda) \in \mathcal{A} \times \Lambda$, indexed by function spaces, and establishes the estimator's uniform convergence rate over $\bar{\lambda} \in \mathbf{R}$, $\alpha \in \mathcal{A}$, and $\lambda \in \Lambda$, when $\mathcal{A}$ and Λ have a finite integral bracketing entropy. The rate of convergence depends on the bracketing entropies of $\mathcal{A}$ and Λ in general. In particular, we demonstrate that when each $\lambda \in \Lambda$ is locally uniformly $\mathcal{L}_2$-continuous in a parameter from a space of polynomial discrimination and the basis function vector p^K in the series estimator keeps the smallest eigenvalue of $\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))']$ above zero uniformly over $\lambda \in \Lambda$, we can obtain the same convergence rate as that established by de Jong (2002, *Journal of Econometrics* 111, 1–9).

1. INTRODUCTION

This paper studies the convergence rate of a series estimator for $\mathbf{E}[\alpha(Y)|\lambda(X) = \bar{\lambda}]$ indexed by $(\alpha, \lambda) \in \mathcal{A} \times \Lambda$, where the convergence is uniform not only over $\bar{\lambda} \in \mathbf{R}$ but also over $(\alpha, \lambda) \in \mathcal{A} \times \Lambda$. The index spaces $\mathcal{A}$ and Λ can be infinite-dimensional, and their sizes are measured by bracketing numbers. The convergence rate in this paper extends the uniform convergence results of Newey (1997) and de Jong (2002) to uniformity over function spaces.

A uniform convergence rate of a conditional mean estimator over function spaces is useful in numerous situations of estimation and hypothesis testing that involve a semiparametric model. Suppose that a semiparametric model involves a conditional mean function $\mathbf{E}(r(Y;\tau)|X = x)$ where r is a function indexed by $\tau \in \mathcal{T}$, in a finite- or infinite-dimensional parameter space. The restriction $\mathbf{E}(r(Y;\tau)|X = x) = 0$ amounts to a semiparametric conditional moment restriction. This restriction can be used to estimate an unknown structural parameter by minimizing an appropriate norm of the conditional moment restriction. Uniform consistency of the conditional mean estimator $\hat{E}(r(Y;\tau)|X = x)$ uniform over $\tau \in \mathcal{T}$ is useful in this context (e.g., Ai and Chen, 2003).

I am grateful to Xiaohong Chen, a co-editor, and three referees for their valuable comments that led to a substantial improvement of this paper. A referee kindly pointed out an error in a previous version of this paper. All errors are mine. Address correspondence to Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Bldg., 3718 Locust Walk, Philadelphia, PA 19104-6297, USA; e-mail: kysong@sas.upenn.edu.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Uniform convergence is also useful for hypothesis testing that contains a nonparametric conditional mean estimator, in particular, for the testing of a conditional moment restriction that involves a nonparametric function of a conditional mean type. Examples include single-index restrictions, partially parametric regression models, and partially parametric quantile regression models. A test statistic can be constructed as a functional of an empirical process that involves the unknown conditional mean function (e.g., Andrews, 1997; Stute, 1997; Bierens and Ploberger, 1997). To construct a feasible test, one should replace the unknown conditional mean function with its nonparametric estimator, and to obtain the limiting distribution of the test statistic under the null hypothesis, one usually resorts to the weak convergence theory of the empirical process. However, weak convergence of the empirical process is often complicated by the presence of the estimation error from this nonparametric estimator. Combined with stochastic equicontinuity of the empirical process, the uniform consistency of the conditional mean estimator can be used to deal with this nonparametric estimation error (e.g., Andrews, 1994).

Our result applies to a situation even more general than the preceding examples in the sense that we allow the conditioning variable to be indexed by a function space Λ. This is relevant in particular when the conditional mean function is of the form $\mathbf{E}(r(Y;\tau_1)|s(X;\tau_2) = \bar{s})$ where τ_1 and τ_2 belong to finite- or infinite-dimensional spaces. To the best of the author's knowledge, uniform convergence of this type has not been studied in the literature, although its usefulness for inferences in semiparametric models is evident. See motivating examples provided in Section 2.

The uniform convergence over the index space Λ of the conditioning variable introduces a nontrivial problem of degeneracy. That is, when λ is close to a constant function, the smallest eigenvalue of the matrix $\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))']$ (with p^K denoting the vector of basis functions) becomes close to zero. In this paper, we introduce a truncation that is based on the smallest eigenvalue of the sample version of $\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))']$. We obtain a rate of convergence that is a tight extension of de Jong (2002) in the sense that when there is no need for truncation and each $\lambda \in \Lambda$ is locally uniformly $\mathcal{L}_2$-continuous in a parameter from a space of polynomial discrimination, the rate of convergence becomes exactly the same as that of de Jong (2002). This is partly due to our general result that establishes the locally uniform $\mathcal{L}_2$-continuity of conditional mean functions in a parameter of the conditioning variable.

Much of the literature on the uniform convergence of nonparametric estimators has focused on kernel estimation. Early literature includes Parzen (1962), Silverman (1978), and Stute (1982, 1984), to name but a few. The work of Nolan and Pollard (1987) appears to be among the early literature that saw the potential of the empirical process theory for the study of kernel estimators. For a recent reference in this line, see Giné and Guillou (2002). Uniform rate of convergence over function spaces was investigated in particular by Härdle,

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Janssen, and Serfling (1988), who established a strong uniform rate of convergence for kernel estimators of conditional mean functions $\mathbf{E}(r(Y;\tau)|X = x)$, $\tau \in \mathcal{T}$, when $\mathcal{T}$ is a subset of a real line and $r(\cdot,\tau)$ is uniformly Lipschitz in τ. Rather recently, Einmahl and Mason (2000) investigated a similar problem when $\mathcal{T}$ is a Vapnik-Čhervonenkis subgraph(VC subgraph) class using recent developments of the empirical process theory.[1] In the literature of kernel estimation, the study of Andrews (1995) contains a result that is closest to ours in the sense that it deals with the uniform convergence of kernel estimators $\mathbf{E}(\alpha(Y)|\lambda(X) = \lambda(x))$, $(\alpha,\lambda) \in \mathcal{A} \times \Lambda$, via the stochastic equicontinuity of a process involved in the variance part of the estimator.

The uniform convergence rate of series estimators was studied by Cox (1988) and Newey (1994, 1995, 1997) and recently by de Jong (2002). Although the results of Newey (1997) and de Jong (2002) contain improvements over previous results, their uniform convergence rates fail to attain the optimal global rate of convergence prescribed by Stone (1982).

Few studies in the literature seem to have dealt with uniform convergence of series estimators over function spaces, with the marked exception of a recent study by Ai and Chen (2003) (see Lem. A.1 there). Ai and Chen (2003) established a uniform rate of convergence for a series estimator of $\mathbf{E}(r(Y,\tau)|X)$, $\tau \in \mathcal{T}_n$, in a form that is different from this paper.[2]

The rest of the paper is organized as follows. In the next section, we introduce definitions and examples that provide the scope of this paper. Section 3 introduces a series estimator and presents the main results of uniform convergence. Section 4 concludes. Mathematical proofs are relegated to the Appendix.

2. DEFINITIONS AND EXAMPLES

2.1. Definitions

Let $(Y_i, X_i)_{i=1}^n$ be a random sample from a joint distribution $\mathbf{P}$ of a random vector $(Y,X) \in \mathbf{R}^{d_Y+d_X}$ that has marginal distributions, $\mathbf{P}_Y$ and $\mathbf{P}_X$. Let $\lambda(\cdot)\colon \mathbf{R}^{d_X} \to \mathbf{R}$ be a function that belongs to a class of measurable functions denoted by Λ. We introduce another class of measurable functions, $\mathcal{A}$, with a generic element $\alpha(\cdot)\colon \mathbf{R}^{d_Y} \to \mathbf{R}$. We consider a conditional mean function, $g_{\alpha,\lambda}(\bar{\lambda})$, defined by

$$g_{\alpha,\lambda}(\bar{\lambda}) \triangleq \mathbf{E}(\alpha(Y)|\lambda(X) = \bar{\lambda}). \tag{1}$$

The conditional mean function $g_{\alpha,\lambda}$ is indexed by $(\alpha,\lambda) \in \mathcal{A} \times \Lambda$.

We define $\|\cdot\|$ to be the euclidean norm: $\|a\| = \sqrt{\mathrm{tr}(aa')}$ for a vector a and $\|A\| = \sqrt{\mathrm{tr}(AA')}$ for a matrix A. For a symmetric matrix A, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ indicate the smallest and the largest eigenvalues of A. We also define the $\mathcal{L}_p$ norm, $\|\cdot\|_p$, by $\|f\|_p = [\mathbf{E}\|f(X)\|^p]^{1/p}$ and let $\mathcal{L}_p(\mathbf{P})$ be the space of $\mathcal{L}_p$-bounded functions. Let $\mathcal{S}_X \subset \mathbf{R}^{d_X}$ denote a set that is contained in

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

$\text{supp}(X)$, the support of X, i.e., the closure of $\{x : f(x) > 0\}$ where f is the density of $\mathbf{P}_X$ with respect to a dominating σ-finite measure. We introduce a sup norm $\|\cdot\|_\infty$ over the bounded functions on $\mathcal{S}_X$: $\|f\|_\infty = \sup_{x \in \mathcal{S}_X} \|f(x)\|$. The notation $\circ$ indicates the composition of functions. Hence $\|p^K \circ \lambda\|_2$ means $(\mathbf{E}\|p^K(\lambda(X))\|^2)^{1/2}$. The notation λ indicates a function as an element of Λ, and $\bar{\lambda}$ indicates a point in a real line. The notations $\vee$ and $\wedge$ are shorthand for max and min as in $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For a number $h \in \mathbf{R}$, $\lfloor h \rfloor$ denotes a greatest integer that does not exceed h. The notation $\triangleq$ indicates a definitional relation.

To obtain a rate of convergence that is uniform over a class of functions, we need to have a device to measure the size of the class. Throughout this paper, we assume that $\mathcal{A} \subset \mathcal{L}_p(\mathbf{P})$ and $\Lambda \subset \mathcal{L}_q(\mathbf{P})$ for some p and q. (See Assumption 1(b) later in this section and Assumption 5L in Sect. 3.) The bracketing entropy of $\mathcal{A}$ with respect to $\|\cdot\|_p$, denoted by $N_{[]}(\varepsilon, \mathcal{A}, \|\cdot\|_p)$, is defined to be the smallest number N of pairs of functions $(\alpha_j, \Delta_j)_{j=1}^N$ such that $\|\Delta_j\|_p < \varepsilon$, $j = 1, \ldots, N$, and for all $\alpha \in \mathcal{A}$, there exists a pair (α_j, Δ_j), $j \in \{1, \ldots, N\}$, satisfying the inequality $|\alpha_j(y) - \alpha(y)| < \Delta_j(y)$. Similarly we define $N_{[]}(\varepsilon, \Lambda, \|\cdot\|_q)$.

Suppose that $\mathcal{A}$ is of the form $\mathcal{A} = \{\alpha_\tau : \tau \in \mathcal{T}\}$ where $(\mathcal{T}, d)$ is a primitive index space with a pseudometric d and the function α_τ is locally uniformly $\mathcal{L}_p$-continuous in $\tau \in \mathcal{T}$ with $p \geq 2$. That is, there exists $s \in (0, 1]$ such that for each $\tau \in \mathcal{T}$ and $\delta > 0$,

$$\left\{\mathbf{E}\left(\sup_{\tau' \in \mathcal{T}: d(\tau, \tau') \leq \delta} |\alpha_\tau(Y) - \alpha_{\tau'}(Y)|^p\right)\right\}^{1/p} \leq C\delta^s.$$

Then we can compute a bracketing entropy bound for $\mathcal{A}$ in terms of that for $\mathcal{T}$. (See, e.g., Chen, Linton, and van Keilegom, 2003, Thm. 3 and its proof.) As demonstrated by Chen et al. (2003), this local uniform $\mathcal{L}_p$-continuity condition is easy to verify, and it is flexible as it admits even the case of $\mathcal{A}$ depending on τ discontinuously. Let us introduce basic assumptions for this paper.

Assumption 1.

(a) Random vectors (Y_i, X_i), $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) from a distribution $\mathbf{P}$.

(b) $\mathcal{A}$ has an envelope $\bar{\alpha}$ such that $\sup_{x \in \mathcal{S}_X} \mathbf{E}[|\bar{\alpha}(Y)|^p | X = x] < \infty$, $p > 2$, and satisfies the following:

$$\int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{A}, \|\cdot\|_p)}\, d\varepsilon < \infty.$$

The uniform boundedness of the conditional moment of $\bar{\alpha}(Y)$ in (b) is slightly stronger than the condition of Newey (1997), which requires $\text{Var}(Y|X)$ to be bounded. An exact condition (with $p > 2$) was used by Einmahl and Mason (2000) (see cond. (1.7) for Thm. 1 there) for kernel estimation.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Condition (b) implies that $\mathcal{A}$ has a finite bracketing integral entropy, and hence they are P-Donsker. The finite bracketing integral entropy condition is not needed in general for establishing a uniform convergence rate (e.g., Ai and Chen, 2003). The condition is convenient for our purpose. First, the condition simplifies the proofs and is useful for illuminating the results. Indeed, the rate of convergence takes a form that explicitly involves the bracketing entropy integrals, making it transparent how the size of the classes plays a role in determining the uniform rate of convergence. Second, the condition is integral for our purpose of obtaining a rate of convergence as a tight extension of one of the most recent results for the case where $\mathcal{A}$ and Λ are singletons. Third, in the context of hypothesis testing, the conditions of finite bracketing integral entropy are often already needed for weak convergence theory of empirical processes.

Assumption 2. For each $x \in \mathcal{S}_X$, the set $\{\lambda(x) \in \mathbf{R} : \lambda \in \Lambda\}$ is compact in $\mathbf{R}$.

Assumption 2 requires that at each point x in the support of X, the cross section $\{\lambda(x) : \lambda \in \Lambda\}$ of Λ should be compact. This condition holds when corresponding to each $x \in \mathcal{S}_X$, the set $\{\lambda(x) \in \mathbf{R} : \lambda \in \Lambda\}$ is contained in $[-M_x, M_x]$ for some $M_x > 0$ and there exists $\varepsilon > 0$ such that for any $\lambda \in \Lambda$ satisfying $\lambda(x) \in [-M_x + \varepsilon, M_x - \varepsilon]$ and for any $a \in [-\varepsilon, \varepsilon]$, $\lambda + a \in \Lambda$. For example, the class of Hölder functions that are uniformly bounded on $\mathcal{S}_X$ satisfy this condition. When $\lambda(x)$ takes a parametric form, say, $\lambda(x) = \lambda(x;\tau)$, $\tau \in \mathcal{T}_0 \subset \mathbf{R}^k$, Assumption 2 is satisfied if $\lambda(x;\cdot)$ is continuous in τ and $\mathcal{T}_0$ is compact. Note that the condition of Assumption 2 per se does not require that for each $\lambda \in \Lambda$, $\lambda(X)$ be compactly supported.

2.2. Examples

A uniform rate of convergence over function spaces can be useful for inferences on semiparametric models. To illustrate this point, we provide two examples of semiparametric models that involve the conditional mean function of type $\mathbf{E}[\alpha(Y)|\lambda(X) = \bar{\lambda}]$.

2.2.1. Semiparametric Quantile Regressions. Consider the following type of a semiparametric quantile regression model:

$$\mathbf{E}[\pi - 1\{Y - \tau_0(X'\theta_0) < 0\}|X] = 0, \qquad \pi \in (0,1), \qquad \theta_0 \in \Theta \subset \mathbf{R}^d \tag{2}$$

for a nonparametric function $\tau_0 \in \mathcal{T}$ (for related models, see, e.g., Chaudhuri, 1991; Lee, 2003, and references in these two papers). Let the distribution of Y be continuous and let its conditional distribution function given $X = x$ be $F(y|x)$ whose inverse is unique and denoted by F^{-1}. Then $\tau_0(X'\theta_0)$ is identified by

$$\tau_0(X'\theta_0) = \mathbf{E}[F^{-1}(\pi|X)|X'\theta_0],$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

and for inference, we may use $\hat{\tau}(X'\hat{\theta})$, an appropriate conditional mean estimator $\hat{\mathbf{E}}[\hat{F}^{-1}(\pi|X)|X'\hat{\theta}]$.

The asymptotic analysis of statistical inference is facilitated by the stochastic equicontinuity of the empirical process (indexed by x) in $(\tau,\theta) \in \mathcal{T} \times \Theta$:

$$V_{n,x}(\tau,\theta) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i,Y_i;x,\tau,\theta) - \mathbf{E}f(X_i,Y_i;x,\tau,\theta)\}, \tag{3}$$

where $f(X_i,Y_i;x,\tau,\theta) = 1\{X_i \geq x\}(\pi - 1\{Y_i - \tau(X_i'\theta) < 0\})$. The feasible version of the process would involve indicator functions that contain a nonparametric estimator $\hat{\tau}_{\hat{\theta}}(X'\hat{\theta})$ of $\tau_0(X'\theta_0)$. The asymptotic behavior of the feasible empirical process can be analyzed when the conditional mean estimator $\hat{\tau}_\theta(\cdot)$ is uniformly consistent over a shrinking neighborhood of θ_0 in which $\hat{\theta}$ takes values with probability approaching one. (See, e.g., Chen et al., 2003. See also Pollard, 1990, and Andrews, 1994, for a general discussion on the stochastic equicontinuity of empirical processes.) In this case, we can take $\Lambda = \{\lambda_\theta(\cdot) : \theta \in \Theta\}$ with $\lambda_\theta(x) = x'\theta$ and choose $\mathcal{A}$ to be an infinite-dimensional space such that $F^{-1}(\pi|\cdot) \in \mathcal{A}$ and $\mathbf{P}\{\hat{F}^{-1}(\pi|\cdot) \in \mathcal{A}\} \to 1$. The uniform convergence of the estimator $\hat{\mathbf{E}}[\alpha(X)|\lambda_\theta(X)]$ uniform over $\lambda_\theta \in \Lambda$ and over $\alpha \in \mathcal{A}$ is useful in this situation.

2.2.2. Conditional Independence. Conditional independence restriction is often used in nonseparable models (e.g., Altonji and Matzkin, 2005). Consider a model of weakly separable endogenous dummy variables (e.g., Vytlacil and Yildiz, 2007, and references therein) where Y, X, and Z are generated by

$$Y = g(v_1(X,D),\varepsilon) \quad \text{and}$$

$$D = 1\{v_2(Z) \geq u\},$$

where ε and u represent unobserved variables. Without loss of generality, we assume that $Y \in [0,1]$. Although we do not know the form g, suppose that we have consistent estimators for $v_1(X,D)$ and $v_2(X)$. We want to know whether there is endogeneity such that ε and u are dependent conditional on the given information. The null hypothesis of exogeneity is

$$Y \perp D|v_1(X,D),v_2(Z),$$

where the notation $Y \perp D|W$ means conditional independence of Y and D given W.

Let us denote $v(X,Z,D) = (v_1(X,D),v_2(Z))$. The null hypothesis is written as a conditional moment restriction: for all $(y,d) \in [0,1] \times \{0,1\}$,

$$\mathbf{E}[1\{Y \leq y, D \leq d\} - F_Y(y|v(X,Z,D))F_D(d|v(X,Z,D))|v(X,Z,D)] = 0,$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

where $F_Y(\cdot|v(X,Z,D))$ and $F_D(\cdot|v(X,Z,D))$ denote the conditional distribution functions of Y and D given $v(X,Z,D)$. In this case, we take $\mathcal{A} = \{1\{\cdot \leq y\} : y \in [0,1]\}$ and Λ to be a space into which the estimator $\hat{v}(\cdot)$ of $v(\cdot)$ falls with probability approaching one. Then the uniform consistency of the series estimator $\hat{\mathbf{E}}[\alpha(Y)|\lambda(X,Z,D)]$ over $(\alpha,\lambda) \in \mathcal{A} \times \Lambda$ can be useful for dealing with the estimation error in $\hat{v}(\cdot)$ as discussed in the preceding example.

3. UNIFORM CONVERGENCE OF SERIES ESTIMATORS

In this section, we investigate the uniform convergence of series estimators. Imposing uniformity over the index space Λ for the conditioning variable causes nontrivial difficulty for uniform consistency. The difficulty is primarily due to the unavailability of the uniform lower bound of $\lambda_{\min}\{\mathbf{E}[\,p^K(\lambda(X))p^K(\lambda(X))']\}$. Obviously, when Λ is uniformly bounded and contains a nonempty interior (with respect to the sup norm) that includes a constant function, the infimum of this smallest eigenvalue over Λ is zero. Despite this difficulty, we can still obtain useful results of uniform convergence by confining our attention to appropriately truncated subspaces of Λ.

For a class of measurable functions $\mathcal{F}$ equipped with a pseudometric $\|\cdot\|_{(2)}$, let $N(\varepsilon,\mathcal{F},\|\cdot\|_{(2)})$ denote *the covering number* of $\mathcal{F}$: the smallest number of ε-balls with respect to $\|\cdot\|_{(2)}$ that it takes to cover $\mathcal{F}$. Following Pollard (1984), we say that the class $\mathcal{F}$ is *of polynomial discrimination with respect to* $\|\cdot\|_{(2)}$ if for $C_2, C_3 > 0$,

$$N(\varepsilon,\mathcal{F},\|\cdot\|_{(2)}) \leq C_2\varepsilon^{-C_3}.$$

For example, VC classes or euclidean classes are of polynomial discrimination. (For euclidean classes, see Nolan and Pollard, 1987, and Pakes and Pollard, 1989, and for VC classes, see Pollard, 1990, and van der Vaart and Wellner, 1996.)

Assumption 3. There exist $C > 0$ and $s \in (0,1]$ such that for each $\lambda \in \Lambda$, the conditional density function $f_\lambda(y,x|\bar{\lambda})$ of (Y,X) given $\lambda(X) = \bar{\lambda}$ with respect to a σ-finite measure w_λ satisfies that for all $(y,x) \in \mathbf{R}^{d_Y+d_X}$ and for all $\bar{\lambda} \in \mathcal{S}_\lambda$ and for all $\delta > 0$,

$$\sup_{\bar{\lambda}_1 \in \mathcal{S}_\lambda : |\bar{\lambda}-\bar{\lambda}_1|<\delta} |f_\lambda(y,x|\bar{\lambda}) - f_\lambda(y,x|\bar{\lambda}_1)| \leq \varphi_\lambda(y,x)\delta^s,$$

where $\varphi_\lambda(\cdot,\cdot)$ is a real-valued function that satisfies $\sup_{x\in\mathcal{S}_X}\int|\bar{\alpha}(y)|\varphi_\lambda(y,x)w_\lambda(dy) < C$.

Assumption 3 requires that the conditional density function of (Y,X) given $\lambda(X) = \bar{\lambda}$ is Lipschitz continuous in $\bar{\lambda}$, where the Lipschitz coefficient may be

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

different across different λ's but satisfies a certain condition. The exponent s indicates the smoothness of the conditional density function in the conditioning variable. It is worth noting that Assumption 3 does not require that $\lambda(X)$ be continuously distributed. Indeed, it does not presuppose a common measure that dominates all the distributions of $\lambda(X)$, $\lambda \in \Lambda$. For example, when X is a discrete random vector with a finite support and $\lambda(x) = x'\theta$ with $\theta \in \Theta \subset \mathbf{R}^k$, we can take $\mathcal{S}_\lambda$ to be a finite set that contains the support of $X'\theta$ and take w_λ to be a counting measure on $\mathcal{S}_\lambda$. In this case, there is no common dominating measure for all the distributions of $x'\theta$.

Let us define a series estimator. We introduce a vector of basis functions:

$$p^K(\bar{\lambda}) \triangleq (p_{1K}(\bar{\lambda}), \ldots, p_{KK}(\bar{\lambda}))'. \tag{4}$$

Using these basis functions, we approximate $g_{\alpha,\lambda}(\bar{\lambda})$ by $p^K(\bar{\lambda})'\beta_{\alpha,\lambda}$ for an appropriate vector $\beta_{\alpha,\lambda}$. Consider a series estimator

$$\hat{g}_{\alpha,\lambda}(\bar{\lambda}) \triangleq p^K(\bar{\lambda})'\hat{\beta}_{\alpha,\lambda}, \qquad \alpha \in \mathcal{A}, \qquad \lambda \in \Lambda, \tag{5}$$

where $\hat{\beta}_{\alpha,\lambda} \triangleq [P_\lambda' P_\lambda]^- P_\lambda' a_\alpha$,

$$a_\alpha \triangleq \begin{bmatrix} \alpha(Y_1) \\ \vdots \\ \alpha(Y_n) \end{bmatrix} \quad \text{and} \quad P_\lambda \triangleq \begin{bmatrix} p^K(\lambda(X_1))' \\ \vdots \\ p^K(\lambda(X_n))' \end{bmatrix}. \tag{6}$$

Matrix $[P_\lambda' P_\lambda]^-$ denotes the Moore–Penrose inverse of $P_\lambda' P_\lambda$. Vectors $\hat{\beta}_{\alpha,\lambda}$ and a_α are indexed by $(\alpha, \lambda) \in \mathcal{A} \times \Lambda$ and $\alpha \in \mathcal{A}$, and matrix P_λ is indexed by $\lambda \in \Lambda$.

For each $\lambda \in \Lambda$, let $\mathcal{S}_\lambda$ be a subset of $\mathbf{R}$ such that $\operatorname{supp}(\lambda(X)) \subset \mathcal{S}_\lambda$, where $\operatorname{supp}(\lambda(X))$ denotes the support of $\lambda(X)$. For a nonnegative integer ν, we define $|\cdot|_\nu$ as follows. Given a function $g_\lambda : \mathcal{S}_\lambda \to \mathbf{R}$ indexed by $\lambda \in \Lambda$,

$$|g_\lambda|_\nu \triangleq \sup_{\mu \le \nu} \sup_{\bar{\lambda} \in \mathcal{S}_\lambda} \left| \frac{\partial^\mu}{\partial \bar{\lambda}^\mu} g_\lambda(\bar{\lambda}) \right|. \tag{7}$$

Let $K = K(n)$ and define $\zeta_{0,K} \triangleq \sup_{\bar{\lambda} \in \mathbf{R}} \|p^K(\bar{\lambda})\|$ and $\zeta_{1,K} \triangleq \sup_{\mu \le 1} \sup_{\bar{\lambda} \in \mathbf{R}} \|(\partial^\mu / \partial \bar{\lambda}^\mu) p^K(\bar{\lambda})\|$.

Assumption 4. For some $d > 0$ and $\nu \ge 0$, there exist a set of vectors, $\{\beta_{\alpha,\lambda} \in \mathbf{R}^K : (\alpha, \lambda) \in \mathcal{A} \times \Lambda\}$, such that $\sup_{(\alpha,\lambda) \in \mathcal{A} \times \Lambda} |p^{K\prime} \beta_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = O(K^{-d})$.

Assumption 4 is analogous to Assumption 3 in Newey (1997) and condition B(iv) of Lemma A.1 in Ai and Chen (2003). The condition requires that the functional approximation error should vanish at a polynomial rate of K uniformly over $\mathcal{A} \times \Lambda$. The rate $O(K^{-d})$ is often determined by the smoothness of

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

the functions in $\{g_{\alpha,\lambda}:(\alpha,\lambda)\in\mathcal{A}\times\Lambda\}$. For example, Newey (1997) provides this error bound for the examples of regression splines and power series.[3]

For a sequence $c_K\to 0$, we define truncated spaces:

$$\Lambda(c_K)\triangleq\{\lambda\in\Lambda:\lambda_{\min}(\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))'])\geq c_K\}. \tag{8}$$

Our results in Theorem 1L in Section 3.1 concern the uniform convergence rate of the series estimator uniform over $\mathcal{A}\times\Lambda(c_K)$. The relevance of such a set $\Lambda(c_K)$ arises when we are interested in the uniform convergence of the estimator of $\mathbf{E}[\alpha(Y)|\lambda_0(X)=\bar{\lambda}]$ for $\lambda_0\in\Lambda$ unknown, and $\Lambda(c_K)$ contains a shrinking neighborhood around λ_0 that is constituted by potential values of estimators for λ_0. We may force the shrinking neighborhoods to be contained in $\Lambda(c_K)$ for large samples by using an appropriate data-dependent truncation. (See Corollary 1L in Section 3.1.)

3.1. Uniform Convergence with $\mathcal{L}_q$-Bounded Λ

We first consider the case when Λ is $\mathcal{L}_q$-bounded. For this section, we introduce a further structure on Λ. Suppose that

$$\Lambda=\{\lambda(\cdot;\tau):\tau\in\mathcal{T}_0\},$$

where $\lambda(\cdot;\tau)$ is a real-valued function with index τ and the index space $\mathcal{T}_0$ is a totally bounded subset of a pseudometric space $(\mathcal{T},\|\cdot\|_{(1)})$. We continue to use the notation $\lambda(\cdot)$ (with the index suppressed) to denote a generic element of Λ.

Assumption 5L.

(a) Λ has an envelope $\tilde{\lambda}$ such that $\|\tilde{\lambda}\|_q<\infty$, $q\geq 2$, and there exists $r\in(0,1]$ such that for each $\tau_1\in\mathcal{T}_0$,

$$\{\mathbf{E}[\sup_{\tau_2\in\mathcal{T}_0:\|\tau_1-\tau_2\|_{(1)}\leq\delta}|\lambda(X;\tau_1)-\lambda(X;\tau_2)|^q]\}^{1/q}\leq\delta^r\quad\text{for all }\delta>0.$$

(b) $\int_0^1\sqrt{1+\log N(\varepsilon^{(q+2s)/qrs},\mathcal{T}_0,\|\cdot\|_{(1)})}\,d\varepsilon<\infty$.

Assumption 5L(a) indicates that $\lambda(\cdot;\tau)$ is locally uniformly $\mathcal{L}_q$-continuous in τ. This condition immediately gives a bracketing entropy bound of Λ in terms of the covering number of $\mathcal{T}_0$ (see, e.g., Chen et al., 2003, Thm. 3). Condition (b) requires that an integral entropy of $\mathcal{T}_0$ with respect to a covering number should be finite. The exponent $(q+2s)/qrs$ in $\varepsilon^{q+2s/qrs}$ of Assumption 5L(b) appears due to our need to find a bracketing entropy bound for the class of functions $\{g_{\alpha,\lambda}\circ\lambda:(\alpha,\lambda)\in\mathcal{A}\times\Lambda\}$. (See Lemma A2(i) in the Appendix.) The entropy condition in (b) can be checked easily in many cases. For example, the condition is satisfied when $\mathcal{T}_0$ is a VC class or a euclidean class with respect to $\|\cdot\|_{(1)}$. Suppose that $\mathcal{S}_X$ is a bounded, convex subset of $\mathbf{R}^{d_X}$ with nonempty inte-

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

rior and $\mathcal{T}_0$ is a Hölder class, i.e., each $\tau \in \mathcal{T}_0$ has uniformly bounded derivatives up to order h and the highest partial derivatives are Lipschitz of order $h - \lfloor h \rfloor$. If $h > d_X(q + 2s)/(2qrs)$, then condition (b) is satisfied (see, e.g., van der Vaart and Wellner, 1996, Thm. 2.7.1).

As for classes $\mathcal{A}$ and Λ, we introduce real-valued functions $\varrho_{\mathcal{A}}(\cdot)$ and $\varrho_{\mathcal{T}_0,\eta}(\cdot)$ on the positive part of the real line: for $a > 0$,

$$\varrho_{\mathcal{A}}(a) \triangleq a \int_0^{1/a} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{A}, \|\cdot\|_p)}\, d\varepsilon \quad \text{and}$$

$$\varrho_{\mathcal{T}_0,\eta}(a) \triangleq a \int_0^{1/a} \sqrt{1 + \log N(\varepsilon^\eta, \mathcal{T}_0, \|\cdot\|_{(1)})}\, d\varepsilon.$$

We also introduce

$$\varrho^K_{s,r,q}(a) \triangleq \varrho_{\mathcal{A}}(a) + \varrho_{\mathcal{T}_0,((q+2s)/qrs)}(a) + \varrho_{\mathcal{T}_0,(1/r)}(a\zeta_{1,K}). \tag{9}$$

The quantity $\varrho_{\mathcal{A}}(a)$ is used to associate the rate of convergence with the size of $\mathcal{A}$. Suppose that $a_K \to \infty$. When the size of $\mathcal{A}$ is large in terms of the bracketing entropy, $\varrho_{\mathcal{A}}(a_K)$ increases fast, and when the size of $\mathcal{A}$ is small, $\varrho_{\mathcal{A}}(a_K)$ increases slowly. In the extreme case of $\mathcal{A}$ being a singleton, $\varrho_{\mathcal{A}}(a_K) = 1$. Note that the role of $\varrho_{\mathcal{A}}(a_K)$ makes sense only when $\mathcal{A}$ has a finite integral bracketing entropy (see Assumption 1(b)). A similar remark can also be made about $\varrho^K_{s,r,q}(a_K)$.

Assumption 6L.

(a) For an arbitrary small number $\sigma \in (0, 1/p \wedge (p-2)/p)$, $\limsup_{K\geq 1} K^{\sigma/2}/\zeta_{0,K} < \infty$ and

$$n^{-1/2}K^\sigma\{\varrho_{\mathcal{T}_0,(1/r)}(\zeta_{0,K}\zeta_{1,K}/K^\sigma) + \sqrt{K}\} \to 0. \tag{10}$$

(b) For any $m_1, m_2 > 0$, there exists $\varphi > 0$ such that $\limsup_{n\geq 1} n^{-\varphi}\zeta^{m_1}_{\nu+1,K}K^{m_2}/c_K^2 < \infty$.

(c) $\varrho^K_{s,r,q}(Ca_K) = O(\varrho^K_{s,r,q}(a_K))$ for any sequence $a_K \to \infty$ and for any constant $C > 0$.

The conditions in (a) determine the fastest possible rate of K and affect the best rate that can be obtained by the rate of convergence in Theorem 1L, which follows. To facilitate comparison with Newey (1997), suppose that $\mathcal{A}$ and Λ are singletons and $c_K = c$, so that the condition in (10) becomes $n^{-1/2}K^{(1/2)+\sigma} \to 0$. This latter condition is weaker than the condition $n^{-1/2}\zeta_{0,K}\sqrt{K} \to 0$ used by Newey (1997) and de Jong (2002) when $\zeta_{0,K}/K^\sigma \to \infty$.

Assumption 6L(b) is analogous to Assumption 5 of de Jong (2002). As de Jong (2002) noted, this condition is hardly a condition; it only requires that $\zeta_{\nu+1,K}$, K, and c_K are bounded by polynomials in n. Condition (c) is not integral to our procedure but is made only to simplify the notation for constants

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

in the presentation of arguments and results. This condition is mostly satisfied; it suffices that $\varrho^K_{s,r,q}(a_K)$ increases at the rate of a polynomial order of a_K as $a_K \to \infty$.

THEOREM 1L.

(i) Suppose that Assumptions 1–4 and 5L and 6L hold. Then,

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = O_P\left(\frac{\zeta_{\nu,K}}{c_K}\{n^{-1/2}K^{\sigma/2}\{\varrho^K_{s,r,q}(K^{-(\sigma/2)}) + \sqrt{K}\} + \sqrt{c_K}K^{-d}\}\right).$$

(ii) Moreover, suppose that (a) $\mathcal{T}_0$ is of polynomial discrimination with respect to $\|\cdot\|_{(1)}$, (b) for each $\lambda \in \Lambda$, the set $\{\lambda(x): x \in \mathcal{S}_X\}$ is contained in a common compact set in $\mathbf{R}$, (c) $n^{2-p}K^{\sigma p}(\log n)^{p-1} \to 0$ and

$$\limsup_{n\geq 1} n^{-1/2}K^{3\sigma/2}\{\varrho_{\mathcal{T}_0,(1/r)}(\zeta_{0,K}\zeta_{1,K}/K^\sigma) + \sqrt{K}\} \times \{\varrho^K_{s,r,q}(K^{-(\sigma/2)}) + \sqrt{K}\}/c_K < \infty, \text{ and} \quad (11)$$

(d) for ν in Assumption 4, $\zeta_{\nu+1,K} \triangleq |p^K|_{\nu+1}$ exists. Then

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = O_P\left(\frac{\zeta_{\nu,K}}{c_K}\{n^{-1/2}\sqrt{\log n} + \sqrt{c_K}K^{-d}\}\right).$$

The result of Theorem 1L(i) shows that the convergence rate is determined by the size of $\mathcal{A}$ and Λ. It does not require that either $\lambda(X)$ or X is compactly supported and continuously distributed. Theorem 1L(ii) improves upon Theorem 1L(i) when $\mathcal{T}_0$ is of polynomial discrimination. The existence of the common compact set in Theorem 1L(ii) plays a role similar to the assumption used by de Jong (2002) that requires the support of X to be contained in a compact set. When $\mathcal{A}$ and Λ are singletons and $(\log n)^{p-1}/\zeta_{0,K}^{2(p-2)} \to 0$, the condition (c) is weaker (or only slightly stronger) than the condition $n^{-1/2}\zeta_{0,K}\sqrt{K} \to 0$ used by de Jong (2002) when the basis functions are chosen to be orthogonal polynomials (or regression splines).[4]

The results in Theorem 1L make an interesting comparison with Newey (1997) and de Jong (2002). To facilitate the comparison, we assume that $\mathcal{A}$ and Λ are singletons and $c_K = c$. In this case, the convergence rate in (i) is equal to

$$O_P(\zeta_{\nu,K}\{n^{-1/2}K^{(1+\sigma)/2} + K^{-d}\}),$$

which is slightly slower than the rate $O_P(\zeta_{\nu,K}\{n^{-1/2}\sqrt{K} + K^{-d}\})$ established by Newey (1997). The result in (ii) shows that when $\mathcal{T}_0$ is of polynomial dis-

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

crimination, the rate of convergence is exactly the same as that established by de Jong (2002) (when we ignore the truncation parameter c_K.) This result demonstrates that as long as the index space Λ is locally uniformly $\mathcal{L}_2$-continuous in a parameter from a space of polynomial discrimination and there is no need for truncation (i.e., $c_K = c$), we can obtain the same convergence rate of de Jong (2002) even when we impose uniformity over function spaces of $\mathcal{A}$ and Λ upon the convergence. As for the conditions for $\mathcal{A}$, it suffices that $\mathcal{A}$ has a finite integral bracketing entropy with respect to $\|\cdot\|_p$ with $p > 2$. It is worth noting that we do not require the finite fourth moment condition for $\varepsilon_{\alpha,\lambda} \triangleq \alpha(Y) - \mathbf{E}[\alpha(Y)|\lambda(X)]$ as de Jong (2002) assumed. The convergence rates in Theorem 1L do not attain the optimal rate prescribed by Stone (1982) even in the degenerate case of $\mathcal{A}$ and Λ being a singleton, because the best possible rate that the convergence rate in the theorem can achieve is that of de Jong (2002), which is already known to fail to achieve Stone's optimal rate.

Frequently, the space of interest is a sequence of spaces Λ_n that contain a shrinking neighborhood of $\lambda_0 \in \Lambda$. Such interest arises when λ_0 has to be replaced by a nonparametric estimator $\hat{\lambda}$ that lies in Λ_n with probability approaching one. We can apply the result of Theorem 1L to this case of a shrinking sequence of Λ_n. For some $\lambda_0 \in \Lambda$, let

$$\Lambda_n \triangleq \{\lambda \in \Lambda : \|\lambda - \lambda_0\|_\infty \leq n^{-b}\}, \qquad b \in \left(0, \frac{1}{2}\right].$$

The rate n^{-b} can be chosen appropriately, corresponding to the uniform convergence rate of $\hat{\lambda}$. We demonstrate that Theorem 1L immediately yields a useful result of uniform convergence rate over $\mathcal{A} \times \Lambda_n$. Let us make the following assumption.

Assumption 7L.

(a) $n^{-b}\zeta_{0,K}\zeta_{1,K} \to 0$ and
(b) $\lambda_{\min}(\mathbf{E}[p^K(\lambda_0(X))p^K(\lambda_0(X))']) > 0$.

The condition $n^{-b}\zeta_{0,K}\zeta_{1,K} \to 0$ is used to force $\Lambda(c_K)$ to contain Λ_n from some large n on. Condition (b) requires that $\lambda_0(X)$ has a nondegenerate distribution such that the smallest eigenvalue of $\mathbf{E}[p^K(\lambda_0(X))p^K(\lambda_0(X))']$ is positive. This condition is also used by Andrews (1991) and Newey (1997). We can enforce the lower bound condition in (8) by using appropriate data-dependent truncation. Let us define

$$\hat{1}_{\lambda,K} \triangleq 1\{\lambda_{\min}(n^{-1}\boldsymbol{\Sigma}_{i=1}^n p^K(\lambda(X_i))p^K(\lambda(X_i))') > c_K/2\}.$$

Then the following result establishes the uniform convergence of the truncated estimator $(\hat{g}_{\alpha,\lambda} \circ \lambda)\hat{1}_{\lambda,K}$ to $(g_{\alpha,\lambda_0} \circ \lambda_0)$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

COROLLARY 1L.

(i) Suppose that the conditions of Theorem 1L(i) with Λ *replaced by* Λ_n *and Assumption 7L are satisfied. Then we have*

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n} \|(\hat{g}_{\alpha,\lambda}\circ\lambda)\hat{\mathrm{I}}_{\lambda,K} - (g_{\alpha,\lambda_0}\circ\lambda_0)\|_\infty = O_P\left(\frac{\zeta_{0,K}}{c_K}\{n^{-1/2}K^{\sigma/2}\{\varrho^K_{s,r,q}(K^{-(\sigma/2)}) + \sqrt{K}\} + \sqrt{c_K}K^{-d}\} + n^{-bs}\right).$$

(ii) Suppose that the conditions of (i) are satisfied except that Assumption 4 holds with Λ *replaced by* $\{\lambda_0\}$. *Suppose furthermore that* $n^{2-p}K^{(\sigma+1)p-2}(\log n)^{p-1} \to 0$ *and* $\limsup_{n\geq 1}\sqrt{K\log n}/\zeta_{1,K} < \infty$. *Then we have*

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n} \|(\hat{g}_{\alpha,\lambda}\circ\lambda)\hat{\mathrm{I}}_{\lambda,K} - (g_{\alpha,\lambda_0}\circ\lambda_0)\|_\infty = O_P\left(\frac{\zeta_{0,K}}{c_K}\{n^{-b}\zeta_{1,K} + n^{-bs} + \sqrt{c_K}K^{-d}\}\right).$$

Corollary 1L(i) presents the uniform convergence of the truncated series estimator $(\hat{g}_{\alpha,\lambda}\circ\lambda)\hat{\mathrm{I}}_{\lambda,K}$ to the conditional mean function $(g_{\alpha,\lambda_0}\circ\lambda_0)$. As compared to the result of Theorem 1L, this rate has an additional term n^{-bs} that is mainly due to the approximation error of $\lambda - \lambda_0$. Corollary 1L(ii) uses a weaker version of Assumption 4 and yields a convergence rate that is in general slower than (i).[5] This result is convenient when it is cumbersome to establish the uniform approximation rate of $g_{\alpha,\lambda}$ by $p^{K\prime}\beta_{\alpha,\lambda}$ uniform over the parameter of the conditioning variable, $\lambda \in \Lambda_n$. Note that the condition that $\limsup_{n\geq 1}\sqrt{K\log n}/\zeta_{1,K} < \infty$ is satisfied as long as $Kn^{-\eta} \to \infty$ for some small $\eta > 0$.

3.2. Uniform Convergence with Uniformly Bounded Λ

In this section, we investigate the case where Λ is uniformly bounded. The conditions and results are very similar to the case when Λ is $\mathcal{L}_q$-bounded, and hence we briefly summarize the conditions and the results. When Λ is uniformly bounded, we do not need to suppose that each $\lambda \in \Lambda$ is indexed by $\mathcal{T}_0$ and λ is locally uniformly $\mathcal{L}_q$-continuous. We can formulate our results directly in terms of the bracketing entropy of Λ, because of the result of Lemma A2(ii) in the Appendix. As for Λ, we introduce $\varrho_{\Lambda,\eta}(\cdot)$ defined by

$$\varrho_{\Lambda,\eta}(a_K) \triangleq a_K\int_0^{1/a_K}\sqrt{1+\log N_{[]}(\varepsilon^\eta,\Lambda,\|\cdot\|_\infty)}\,d\varepsilon.$$

We also define $\varrho^K_{s,\infty}(a_K) \triangleq \varrho_{\mathcal{A}}(a_K) + \varrho_{\Lambda,1/s}(a_K) + \varrho_{\Lambda,1}(a_K\zeta_{1,K})$. In place of Assumptions 5L and 6L, we make the following assumptions.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Assumption 5U. $\int_0^1 \sqrt{1 + \log N(\varepsilon^{1/s}, \Lambda, \|\cdot\|_\infty)}\, d\varepsilon < \infty$.

Assumption 6U. Assumption 6L holds with $\varrho_{\mathcal{T}_0,1/r}$ and $\varrho^K_{s,r,q}$ replaced by $\varrho_{\Lambda,1}$ and $\varrho^K_{s,\infty}$, respectively.

THEOREM 1U.

(i) Suppose that Assumptions 1–4 and 5U and 6U hold. Then,

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = O_P\left(\frac{\zeta_{\nu,K}}{c_K}\{n^{-1/2}K^{\sigma/2}\{\varrho^K_{s,\infty}(K^{-(\sigma/2)}) + \sqrt{K}\} + \sqrt{c_K}K^{-d}\}\right).$$

(ii) Moreover, suppose that (a) Λ is of polynomial discrimination with respect to $\|\cdot\|_\infty$, (b) for each $\lambda \in \Lambda$, the set $\{\lambda(x) : x \in \mathcal{S}_X\}$ is contained in a common compact set in $\mathbf{R}$*, (c) $n^{2-p}K^{\sigma p}(\log n)^{p-1} \to 0$ and*

$$\text{limsup}_{n\geq 1}\, n^{-1/2}K^{3\sigma/2}\{\varrho_{\Lambda,1}(\zeta_{0,K}\zeta_{1,K}/K^\sigma) + \sqrt{K}\} \times \{\varrho^K_{s,\infty}(K^{-(\sigma/2)}) + \sqrt{K}\}/c_K < \infty, \text{ and}$$

(d) for ν in Assumption 4, $\zeta_{\nu+1,K} \triangleq |p^K|_{\nu+1}$ exists. Then

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = O_P\left(\frac{\zeta_{\nu,K}}{c_K}\{n^{-1/2}\sqrt{\log n} + \sqrt{c_K}K^{-d}\}\right).$$

The results of Theorem 1U are analogous to those of Theorem 1L. Similarly, we can obtain a result analogous to Corollary 1L.

4. CONCLUSION

This paper studies the uniform convergence of a series estimator with uniformity imposed on function spaces. One of the results demonstrates that the convergence rate of de Jong (2002) remains the same when we extend uniformity to function spaces as long as the function of the conditioning variable is locally uniformly $\mathcal{L}_2$-continuous in a parameter from a space of polynomial discrimination and there is no need for truncation. We may extend the framework further to accommodate weak dependence (e.g., β-mixing) of data by employing the maximal inequality in Doukhan, Massart, and Rio (1995) or that of the tail probability in Lemma 1 of Chen and Shen (1998). Research in this direction is relegated to the future.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

NOTES

1. For the definition of a VC subgraph class, see Section 2.6 of van der Vaart and Wellner (1996).

2. Ai and Chen (2003) established the rate of convergence for $\sup_{\alpha\in\mathcal{A}}(1/n)\sum_{i=1}^{n}\{\hat{\mathbf{E}}[\alpha(Y)|X_i]-\mathbf{E}[\alpha(Y)|X_i]\}^2$, whereas the rate of convergence of focus in this paper is for $\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda}\sup_{\bar{\lambda}\in\mathcal{S}_\lambda}|\hat{\mathbf{E}}[\alpha(Y)|\lambda(X)=\bar{\lambda}]-\mathbf{E}[\alpha(Y)|\lambda(X)=\bar{\lambda}]|$, where $\mathcal{S}_\lambda$ is a set that contains the support of $\lambda(X)$.

3. When our focus is on the uniform convergence over shrinking neighborhoods of $\lambda_0\in\Lambda$ with respect to the sup norm, a weaker version of Assumption 4 that does not require uniformity over $\lambda\in\Lambda$ can be employed, and the convergence rate in this case is slower than the one obtained under Assumption 4. (See Corollary 1L(ii) and Sect. 3.1 and a subsequent remark there.)

4. Suppose that $\mathcal{A}$ and Λ are singletons, $(\log n)^{p-1}/\zeta_{0,K}^{2(p-2)}\to 0$. Then the rate condition in (c) of Theorem 1L is weaker than the condition $n^{-1/2}\zeta_{0,K}\sqrt{K}\to 0$ used by de Jong (2002). As for (11), it becomes $\limsup_{n\geq 1}n^{-1/2}K^{1+(3\sigma/2)}<\infty$. When $\zeta_{\nu,K}=K^{1+2\nu}$ (orthogonal polynomials), the condition in (11) is weaker than the condition $n^{-1/2}\zeta_{0,K}\sqrt{K}\to 0$ used by Newey (1997) and de Jong (2002). When $\zeta_{\nu,K}=K^{(1/2)+\nu}$ (regression splines), the condition in (11) is reduced to $\limsup_{n\geq 1}n^{-1/2}\zeta_{0,K}K^{(1/2)+(3\sigma/2)}<\infty$, which is only slightly stronger than the condition $n^{-1/2}\zeta_{0,K}\sqrt{K}\to 0$.

5. To see this, suppose that $\log N_{[]}(\varepsilon,\mathcal{A},\|\cdot\|_2)\leq C\varepsilon^{2(\eta-1)}$ and $\log N_{[]}(\varepsilon,\Lambda,\|\cdot\|_2)\leq C\varepsilon^{2(\eta-1)}$ for $\eta>0$ arbitrarily small. Then $K^{\sigma/2}\varrho_{s,r,q}^{K}(K^{-(\sigma/2)})=O(K^{\sigma\eta/2}\zeta_{1,K}^{1-\eta})=O(\zeta_{1,K}^{1-\eta+(\sigma\eta/2)})$, and $n^{-1/2}\zeta_{1,K}^{1-\eta+(\sigma\eta/2)}$ decreases to zero faster than $n^{-b}\zeta_{1,K}$.

REFERENCES

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

Altonji, J. & R. Matzkin (2005) Cross-section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73, 1053–1102.

Andrews, D.W.K. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric models. *Econometrica* 59, 307–345.

Andrews, D.W.K. (1994) Empirical process method in econometrics. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2247–2294. North-Holland.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Berge, C. (1963) *Topological Spaces*. Macmillan.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment test. *Econometrica* 65, 1129–1151.

Chaudhuri, P. (1991) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760–777.

Chen, X., O. Linton, & I. van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.

Chen, X. & X. Shen (1998) Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.

Cox, D.D. (1988) Approximation of least squares regression on nested subspaces. *Annals of Statistics* 16, 713–732.

de Jong, R.M. (2002) A note on "Convergence rates and asymptotic normality for series estimators": Uniform convergence rates. *Journal of Econometrics* 111, 1–9.

Doukhan, P., P. Massart, & E. Rio (1995) Invariance principles for absolutely regular empirical processes. *Annales de l'Institut Henri Poincaré* 31, 393–427.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Einmahl, U. & D.M. Mason (2000) An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability* 13, 1–37.
Giné, E. & A. Guillou (2002) Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré* 38, 907–921.
Härdle, W., P. Janssen, & M. Serfling (1988) Strong uniform consistency rates for estimators of conditional functionals. *Annals of Statistics* 16, 1428–1449.
Lee, S. (2003) Efficient semiparametric estimation of a partial linear quantile regression model. *Econometric Theory* 19, 1–31.
Newey, W.K. (1994) Series estimation of regression functionals. *Econometric Theory* 10, 1–28.
Newey, W.K. (1995) Convergence rates for series estimators. In G.S. Maddalla, P.C.B. Phillips, & T.N. Srinavasan (eds.), *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao Blackwell*, pp. 254–275. Cambridge.
Newey, W.K. (1997) Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79, 147–168.
Nolan, D. & D. Pollard (1987) *U*-processes: Rates of convergence. *Annals of Statistics* 15, 780–799.
Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
Parzen, E. (1962) On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.
Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.
Pollard, D. (1989) A Maximal Inequality for Sums of Independent Processes under a Bracketing Entropy Condition. Manuscript, Yale University.
Pollard, D. (1990) *Empirical Processes: Theory and Applications*. CBMS Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics.
Silverman, B.W. (1978) Weak and strong uniform consistency of kernel estimate of a density and its derivatives. *Annals of Statistics* 6, 177–184.
Stone, C.J. (1982) Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* 10, 1040–1053.
Stute, W. (1982) A law of the iterated logarithm for kernel density estimators. *Annals of Probability* 10, 414–422.
Stute, W. (1984) Oscillation behavior of empirical processes: The multivariate case. *Annals of Probability* 12, 361–379.
Stute, W. (1997) Nonparametric model checks for regressions. *Annals of Statistics* 25, 613–641.
van der Vaart, A. (1996) New Donsker classes. *Annals of Probability* 24, 2128–2140.
van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer-Verlag.
Vytlacil, E. & N. Yildiz (2007) Dummy endogenous variables in weakly separable models. *Econometrica* 75, 757–779.

APPENDIX: Mathematical Proofs

Throughout the proofs, C denotes a universal constant that may take different values in different places in the proofs. The notation $\mathbf{E}_*$ and $\mathbf{E}^*$ denotes inner and outer integrals, respectively. We also let $\mathbf{P}_*$ and $\mathbf{P}^*$ denote inner and outer probabilities. (For their definitions, see van der Vaart and Wellner, 1996.) For any (possibly nonmeasurable) vector-valued function f on $\mathcal{S}_X$, we define $\|f\|_p \triangleq \{\mathbf{E}^*\|f(X)\|^p\}^{1/p}$. In the proofs, the notation D^μ means a differentiation operator as in $(D^\mu g)(\bar{\lambda}) \triangleq (\partial^\mu/\partial\bar{\lambda}^\mu)g(\bar{\lambda})$. The notation S_K indicates the unit sphere in $\mathbf{R}^K$, i.e.,

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

$S_K \triangleq \{b \in \mathbf{R}^K : \|b\| = 1\}$.

The notation M_σ indicates a positive real number such that $M_\sigma = 1/\sigma$ for σ that appears in Assumption 6L(a). Last we define a $K \times K$ matrix function

$$R_K(\lambda(x)) \triangleq p^K(\lambda(x))p^K(\lambda(x))',$$

as terms involving it frequently appear in the proofs.

A.1. The Proof of the Main Results.

Proof of Theorem 1L(i). We define

$$G_{\alpha,\lambda} \triangleq (\mathbf{E}[\alpha(Y_1)|\lambda(X_1)], \ldots, \mathbf{E}[\alpha(Y_n)|\lambda(X_n)])', \tag{A.1}$$

$Q_\lambda \triangleq \mathbf{E}[R_K(\lambda(X))]$, $\hat{Q}_\lambda \triangleq P_\lambda' P_\lambda/n$, and $\varepsilon_{\alpha,\lambda} \triangleq a_\alpha - G_{\alpha,\lambda}$ for a_α and P_λ defined in (6).

We first rotate p^K by a matrix. For each $b \in S_K$, $b'R_K(\bar{\lambda})^{M_\sigma}b$ is continuous in $\bar{\lambda} \in \mathbf{R}$, and for each $x \in \mathcal{S}_X$, the set $\Lambda_x \triangleq \{\lambda(x) : \lambda \in \Lambda\} \subset \mathbf{R}$ is compact by Assumption 2. Hence the supremum of $\mathcal{Z}_{x,b} \triangleq \{b'R_K(\bar{\lambda})^{M_\sigma}b : \bar{\lambda} \in \Lambda_x\}$ belongs to $\mathcal{Z}_{x,b}$ for each $(x,b) \in \mathcal{S}_X \times S_K$, so that we can denote the supremum by $b'R_K(\lambda_{x,b}(x))^{M_\sigma}b$ for some $\lambda_{x,b} \in \Lambda$. We define

$$B_K(x;b) \triangleq R_K(\lambda_{x,b}(x)).$$

The matrix function $B_K(x;b)$ is constituted by entries that are potentially nonmeasurable functions of $x \in \mathcal{S}_X$ for each $b \in S_K$. By the theorem of the maximum (e.g., Berge, 1963), the function $\varphi(b) \triangleq b'B_K^{M_\sigma}(x;b)b$ is continuous in $b \in S_K$ for each $x \in \mathcal{S}_X$, and hence $b'\mathbf{E}_*[B_K^{M_\sigma}(X;b)]b$ is upper semicontinuous in $b \in S_K$. (To see this, note that for each $b_0 \in S_K$, we have $\lim_{\delta\to 0} \sup_{b\in S_K:\|b-b_0\|\le\delta} b'\mathbf{E}_*[B_K^{M_\sigma}(X;b)]b \le \lim_{\delta\to 0} \mathbf{E}_*[\sup_{b\in S_K:\|b-b_0\|\le\delta} b'B_K^{M_\sigma}(X;b)b] = \mathbf{E}_*[\lim_{\delta\to 0} \sup_{b\in S_K:\|b-b_0\|\le\delta} b'B_K^{M_\sigma}(X;b)b] = b_0'\mathbf{E}_*[B_K^{M_\sigma}(X;b_0)]b_0$. For the leading equality, we used the monotone convergence theorem [e.g., van der Vaart and Wellner, 1996, Problem 1.2.3, p. 13], and for the last equality, we used the continuity of $b'B_K^{M_\sigma}(X;b)b$ in $b \in S_K$.) Hence a vector $b^* \in S_K$ such that

$$b^{*\prime}\mathbf{E}_*[B_K^{M_\sigma}(X;b^*)]b^* = \sup_{b\in S_K} b'\mathbf{E}_*[B_K^{M_\sigma}(X;b)]b \tag{A.2}$$

exists. We fix a constant $C_1 > 0$ and apply the spectral decomposition to the matrix $\mathbf{E}_*[B_K^{M_\sigma}(X;b^*)] + C_1 I$ to obtain

$$\mathbf{E}_*[B_K^{M_\sigma}(X;b^*)] + C_1 I = \tilde{B}_K \Omega \tilde{B}_K', \tag{A.3}$$

where $\tilde{B}_K$ is an orthogonal matrix and Ω is a diagonal matrix with positive diagonal elements. Note that from (A.3),

$$\lambda_{\min}(\tilde{B}_K \Omega \tilde{B}_K') \ge C_1 > 0. \tag{A.4}$$

Now, we rotate p^K by $\tilde{B}_K/\{\lambda_{\max}(\Omega)\}^{1/2M_\sigma}$, yielding $P^K \triangleq \tilde{B}_K' p^K/\{\lambda_{\max}(\Omega)\}^{1/2M_\sigma}$. We investigate the manner in which the postrotation counterpart $\tilde{\zeta}_{\mu,K} \triangleq |P^K|_\mu$ of $\zeta_{\mu,K}$ is related to $\zeta_{\mu,K}$. Observe that by (A.4),

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

$$\tilde{\zeta}^2_{\mu,K} = \frac{|\tilde{B}'_K p^K|^2_\mu}{\{\lambda_{\max}(\Omega)\}^{1/M_\sigma}} \leq \frac{\sup_{\bar{\lambda}\in\mathbf{R}} \operatorname{tr}(D^\mu p^K(\bar{\lambda}) D^\mu p^K(\bar{\lambda})')\lambda_{\max}(\tilde{B}_K \tilde{B}'_K)}{\{\lambda_{\max}(\Omega)\}^{1/M_\sigma}}$$

$$= \frac{\sup_{\bar{\lambda}\in\mathbf{R}} \operatorname{tr}(D^\mu p^K(\bar{\lambda}) D^\mu p^K(\bar{\lambda})')}{\{\lambda_{\max}(\Omega)\}^{1/M_\sigma}} \leq \zeta^2_{\mu,K}/C_1^{1/M_\sigma}$$

and

$$\lambda_{\min}(\mathbf{E}[P^K(\lambda(X))P^K(\lambda(X))']) = \inf_{b\in S_K} \frac{b'\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))']b}{\{\lambda_{\max}(\Omega)\}^{1/M_\sigma}}$$

$$\geq \frac{\lambda_{\min}(\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))'])}{C_1^{1/M_\sigma}}.$$

Conditions involving bounds for $\zeta_{\mu,K}$ also hold for $\tilde{\zeta}_{\mu,K}$ up to a multiplication by a constant, and the smallest eigenvalue of $\mathbf{E}[P^K(\lambda(X))P^K(\lambda(X))']$ is positive when that of $\mathbf{E}[p^K(\lambda(X))p^K(\lambda(X))']$ is positive.

Define the postrotation counterpart of $B_K(x;b): B_{1K}(x;b) \triangleq P^K(\tilde{\lambda}_{x,b}(x))P^K(\tilde{\lambda}_{x,b}(x))'$, where $\tilde{\lambda}_{x,b}(x) \in \Lambda_x$ achieves the supremum of $\{b'\tilde{B}'_K R_K(\bar{\lambda})^{M_\sigma}\tilde{B}_K b/\lambda_{\max}(\Omega) : \bar{\lambda} \in \Lambda_x\}$. Then

$$b'B^{M_\sigma}_{1K}(X;b)b = b'\tilde{B}'_K B^{M_\sigma}_K(X;b)\tilde{B}_K b/\lambda_{\max}(\Omega)$$

because $\tilde{B}_K$ is an orthogonal matrix, i.e., $\tilde{B}_K\tilde{B}'_K = I$. Hence, for all $c \in S_K$,

$$c'\mathbf{E}_*[B^{M_\sigma}_{1K}(X;c)]c = \tilde{b}'\mathbf{E}_*[R_K(\tilde{\lambda}_{X,c}(X))^{M_\sigma}]\tilde{b}/\lambda_{\max}(\Omega), \quad \text{where } \tilde{b} \triangleq \tilde{B}_K c \in S_K \qquad \textbf{(A.5)}$$

$$\leq \tilde{b}'\mathbf{E}_*[B^{M_\sigma}_K(X;\tilde{b})]\tilde{b}/\lambda_{\max}(\Omega) \leq b^{*\prime}\mathbf{E}_*[B^{M_\sigma}_K(X;b^*)]b^*/\lambda_{\max}(\Omega)$$

$$\leq \lambda_{\max}(\mathbf{E}_*[B^{M_\sigma}_K(X;b^*)])/\lambda_{\max}(\Omega) \leq 1.$$

From here on, we assume that we have rotated p^K by $\tilde{B}_K/\{\lambda_{\max}(\Omega)\}^{1/2M_\sigma}$, but we continue to use the same notation p^K for this rotated vector. We also redefine Q_λ, $\hat{Q}_\lambda$, and $\zeta_{\mu,K}$ using this rotated p^K but retain the same notation. We will show the following first.

(A) $\mathbf{P}^*\{\inf_{\lambda\in\Lambda(c_K)} \lambda_{\min}(\hat{Q}_\lambda) > c_K/2\} \to 1$, as $n \to \infty$.

Then we let $1_{n,\lambda} \triangleq 1\{\lambda_{\min}(\hat{Q}_\lambda) > c_K/2\}$. For any sequence ε_n such that $0 < \varepsilon_n \to 0$,

$$\mathbf{P}^*\left\{\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |1_{n,\lambda} - 1|\,|\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu > \varepsilon_n\right\}$$

$$\leq \mathbf{P}^*\{\exists\lambda \in \Lambda(c_K) \quad \text{such that } 1_{n,\lambda} = 0\} \qquad \textbf{(A.6)}$$

$$\leq \mathbf{P}^*\left\{\inf_{\lambda\in\Lambda(c_K)} \lambda_{\min}(\hat{Q}_\lambda) \leq c_K/2\right\} \to 0$$

by (A). From this, we deduce that

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu = \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda}|\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu + O_P(\varepsilon_n), \qquad \textbf{(A.7)}$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

where the sequence ε_n can be chosen to vanish arbitrarily fast as $n \to \infty$. For the leading term, note that $\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda}|\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu$ is bounded by

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda} \sup_{\mu\le\nu} \sup_{\bar{\lambda}\in\mathcal{S}_\lambda} |(D^\mu p^K)'(\bar{\lambda})\{\hat{Q}_\lambda^{-1} P_\lambda' \alpha/n - \beta_{\alpha,\lambda}\}|$$

$$+ \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda}|p^{K\prime}\beta_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu. \tag{A.8}$$

The second term is $O(K^{-d})$ by Assumption 4, and the first term is bounded by

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda} \sup_{\mu\le\nu} \sup_{\bar{\lambda}\in\mathcal{S}_\lambda} |(D^\mu p^K)'(\bar{\lambda})\hat{Q}_\lambda^{-1} P_\lambda'(\varepsilon_{\alpha,\lambda} + \delta_{\alpha,\lambda})/n|$$

$$\le \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda} \sup_{\mu\le\nu} \sup_{\bar{\lambda}\in\mathcal{S}_\lambda} |(D^\mu p^K)'(\bar{\lambda})\hat{Q}_\lambda^{-1} P_\lambda' \varepsilon_{\alpha,\lambda}/n|$$

$$+ \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} 1_{n,\lambda} \sup_{\mu\le\nu} \sup_{\bar{\lambda}\in\mathcal{S}_\lambda} |(D^\mu p^K)'(\bar{\lambda})\hat{Q}_\lambda^{-1} P_\lambda' \delta_{\alpha,\lambda}/n|$$

$$\triangleq B_{1n} + B_{2n}, \quad \text{say,}$$

where $\delta_{\alpha,\lambda} \triangleq G_{\alpha,\lambda} - P_\lambda\beta_{\alpha,\lambda}$, with $G_{\alpha,\lambda}$ defined in (A.1). Then the proof is complete once we show the following:

(B) $B_{1n} = O_P(n^{-1/2}\zeta_{\nu,K}K^{\sigma/2}\{\varrho^K_{s,r,q}(K^{-\sigma/2}) + \sqrt{K}\}/c_K)$.

(C) $B_{2n} = O_P(\zeta_{\nu,K}K^{-d}/\sqrt{c_K})$.

Proof of (A). Note that $|\lambda_{\min}(\hat{Q}_\lambda) - \lambda_{\min}(Q_\lambda)| \le \sup_{b\in S_K}|b'(\hat{Q}_\lambda - Q_\lambda)b|$. (Note that $|\lambda_{\min}(\hat{Q}_\lambda) - \lambda_{\min}(Q_\lambda)|$ is bounded by $|\inf_{b\in S_K}\{b'(\hat{Q}_\lambda - Q_\lambda)b + b'Q_\lambda b\} - \inf_{b\in S_K} b'Q_\lambda b| \le |\inf_{b\in S_K}\{\sup_{c\in S_K}|c'(\hat{Q}_\lambda - Q_\lambda)c| + b'Q_\lambda b\} - \inf_{b\in S_K} b'Q_\lambda b| = \sup_{b\in S_K}|b'(\hat{Q}_\lambda - Q_\lambda)b|$.) Therefore, it suffices to show that $\sup_{(b,\lambda)\in S_K\times\Lambda}|b'(\hat{Q}_\lambda - Q_\lambda)b| = o_P(1)$. Define for each $(b,\lambda) \in S_K \times \Lambda$,

$$\xi_{b,\lambda} \triangleq \left| \frac{1}{n}\sum_{i=1}^n \{\{b'p^K(\lambda(X_i))\}^2 - \mathbf{E}[\{b'p^K(\lambda(X_i))\}^2]\} \right|,$$

so that we have $\sup_{(b,\lambda)\in S_K\times\Lambda}|b'(\hat{Q}_\lambda - Q_\lambda)b| = \sup_{(b,\lambda)\in S_K\times\Lambda}\xi_{b,\lambda}$.

Let $J(\bar{\lambda}) \triangleq [\mathrm{tr}(R_K(\bar{\lambda})^{M_\sigma})]^{1/M_\sigma}$. Because $J(\bar{\lambda})$ is continuous in $\bar{\lambda} \in \mathbf{R}$ and the set $\{\lambda(x) : \lambda \in \Lambda\}$ is compact, the supremum of $\mathcal{J}_x \triangleq \{J(\lambda(x)) : \lambda \in \Lambda\}$ belongs to $\mathcal{J}_x$ for each $x \in \mathbf{R}^{d_X}$, so that we can denote the supremum by $J(\tilde{\lambda}_x(x))$ for some $\tilde{\lambda}_x(\cdot) \in \Lambda$. Let $\mathcal{H}_1 \triangleq \{b'R_K(\lambda(\cdot))b : (b,\lambda) \in S_K \times \Lambda\}$. We take $H_1(x) \triangleq [J(\tilde{\lambda}_x(x))]_*$ to be a maximal measurable minorant of $J(\tilde{\lambda}_x(x))$ and let it serve as an envelope of $\mathcal{H}_1$. (A maximal measurable minorant $[T]_*$ of a [potentially nonmeasurable] map T is defined to be the supremum of measurable maps enveloped by T.)

Let $R_K(\lambda(x)) = U_K'V_KU_K$ be the spectral decomposition of $R_K(\lambda(x))$ in which U_K is an orthogonal matrix and V_K is a diagonal matrix with diagonal elements being equal to the eigenvalues of V_K. Because $R_K(\lambda(x))$ is positive semidefinite, these eigenvalues are nonnegative. Let $b_1 = b'U_K$ and write

$$b'R_K(\lambda(x))b = b'U_K'V_KU_Kb = b_1'V_Kb_1 = \sum_{k=1}^K b_{1k}^2 v_k,$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

where b_{1k} is the k-element of b_1 corresponding to the kth diagonal element v_k in V_K. Because U_K is an orthogonal matrix, $\sum_{k=1}^{K} b_{1k}^2 = \|b_1\|^2 = \|b\|^2 = 1$. By Jensen's inequality, the last sum in the preceding expression is bounded by

$$\left\{\sum_{k=1}^{K} b_{1k}^2 v_k^{M_\sigma}\right\}^{1/M_\sigma} = \{b_1' V_K^{M_\sigma} b_1\}^{1/M_\sigma} = \{b' U_K V_K^{M_\sigma} U_K b\}^{1/M_\sigma} = \{b' R^{M_\sigma}(\lambda(x))b\}^{1/M_\sigma}.$$

This yields the first inequality in the following:

$$\begin{aligned} b' R_K(\lambda(x))b &\le (b' R_K(\lambda(x))^{M_\sigma} b)^{1/M_\sigma} \le (\lambda_{\max}\{R_K(\lambda(x))^{M_\sigma}\})^{1/M_\sigma} \\ &\le (\mathrm{tr}[R_K(\lambda(x))^{M_\sigma}])^{1/M_\sigma} \\ &\le (\mathrm{tr}[R_K(\tilde{\lambda}_x(x))^{M_\sigma}])^{1/M_\sigma} = J(\tilde{\lambda}_x(x)). \end{aligned}$$

The second inequality follows because $R_K(\lambda(x))^{M_\sigma}$ is positive semidefinite. Hence H_1 is a measurable envelope of $\mathcal{H}_1$. Observe that $\|H_1\|_{M_\sigma}^{M_\sigma}$ is equal to $\mathrm{tr}(\mathbf{E}_*\{J(\tilde{\lambda}_X(X))^{M_\sigma}\})$. However, this latter term is bounded by

$$\begin{aligned} K\lambda_{\max}(\mathbf{E}_*\{R_K(\tilde{\lambda}_X(X))^{M_\sigma}\}) &= K \sup_{b\in S_K} b' \mathbf{E}_*\{R_K(\tilde{\lambda}_X(X))^{M_\sigma}\} b \\ &\le K \sup_{b\in S_K} b' \mathbf{E}_*(B_{1K}(X;b)^{M_\sigma}) b \le K, \end{aligned}$$

by (A.5). Hence $\|H_1\|_2 \le \|H_1\|_{M_\sigma} \le K^{1/M_\sigma}$. By the maximal inequality (e.g., Pollard, 1989),

$$\begin{aligned} \mathbf{E}^*\left[\sup_{(b,\lambda)\in S_K\times\Lambda} \sqrt{n}\,\xi_{b,\lambda}\right] &\le \int_0^{\|H_1\|_2} \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}_1, \|\cdot\|_2)}\, d\varepsilon \\ &= O(K^{1/M_\sigma}\{\varrho_{\Lambda,1}(\zeta_{0,K}\zeta_{1,K}/K^{1/M_\sigma}) + \sqrt{K}\}), \end{aligned}$$

by Lemma A1(i) later in this Appendix. (The maximal inequality and its proof there are found in van der Vaart, 1996, Thm. A.2, p. 2136.) Therefore, by Assumption 5L(a),

$$\sup_{(b,\lambda)\in S_K\times\Lambda} |b'(\hat{Q}_\lambda - Q_\lambda)b| = O_P(n^{-1/2}K^\sigma\{\varrho_{\mathcal{T}_0,(1/r)}(\zeta_{0,K}\zeta_{1,K}/K^\sigma) + \sqrt{K}\}). \tag{A.9}$$

(Assumption 5L(a) implies that $\varrho_{\Lambda,1}(\zeta_{0,K}\zeta_{1,K}/K^{1/M_\sigma}) \le \varrho_{\mathcal{T}_0,1/r}(\zeta_{0,K}\zeta_{1,K}/K^{1/M_\sigma})$. For its proof, see the proof of Theorem 3 in Chen et al., 2003.) By Assumption 6L(a), the last term in (A.9) is $o_P(1)$. ■

Proof of (B). Let us define $\hat{d}_\lambda' \triangleq (D^\mu p^K)(\bar{\lambda})'\hat{Q}_\lambda^{-1}/\|(D^\mu p^K)(\bar{\lambda})'\hat{Q}_\lambda^{-1}\|$ and consider

$$1_{n,\lambda}|(D^\mu p^K)'(\bar{\lambda})\hat{Q}_\lambda^{-1} P_\lambda' \varepsilon_{\alpha,\lambda}/n| \le 1_{n,\lambda}(\zeta_{\mu,K}/c_K)|\hat{d}_\lambda' P_\lambda' \varepsilon_{\alpha,\lambda}/n|,$$

where we can write $|\hat{d}_\lambda' P_\lambda' \varepsilon_{\alpha,\lambda}/n| = |(1/n)\sum_{i=1}^n \hat{d}_\lambda' p^K(\lambda(X_i))\{\alpha(Y_i) - \mathbf{E}(\alpha(Y_i)|\lambda(X_i))\}|$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

We define $\mathcal{H}_2 \triangleq \{b'(p^K \circ \lambda)(\alpha - (g_{\alpha,\lambda} \circ \lambda)) : (\alpha, \lambda, b) \in \mathcal{A} \times \Lambda \times S_K\}$ with its envelope denoted by H_2. Note that by the maximal inequality (e.g., Pollard, 1989),

$$\mathbf{E}^*\left[\sup_{(\alpha,\lambda,b)\in\mathcal{A}\times\Lambda\times S_K}\left|\frac{1}{n}\sum_{i=1}^{n} b'p^K(\lambda(X_i))\{\alpha(Y_i) - \mathbf{E}(\alpha(Y_i)|\lambda(X_i))\}\right|\right] \tag{A.10}$$

$$\leq \frac{C}{\sqrt{n}}\left(\int_0^{\|H_2\|_2}\sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{H}_2, \|\cdot\|_2)}\, d\varepsilon\right).$$

Because $J_1(\bar{\lambda}) \triangleq [\mathrm{tr}(R_K(\bar{\lambda})^{M_\sigma})]^{1/2M_\sigma}$ is continuous in $\bar{\lambda} \in \mathbf{R}$ and the set $\{\lambda(x) : \lambda \in \Lambda\}$ is compact, the supremum of $\mathcal{J}_{1,x} \triangleq \{J_1(\lambda(x)) : \lambda \in \Lambda\}$ belongs to $\mathcal{J}_{1,x}$ for each $x \in \mathcal{S}_X$, so that we can denote the supremum by $J_1(\lambda_x(x))$ for some $\lambda_x \in \Lambda$. We define $F_K(x) \triangleq [(\mathrm{tr}\{R_K(\lambda_x(x))^{M_\sigma}\})^{1/2M_\sigma}]_*$ to be a maximal measurable minorant of $[\mathrm{tr}\{R_K(\lambda_x(x))^{M_\sigma}\}]^{1/2M_\sigma}$ and choose the envelope $H_2(x, y)$ to be $F_K(x)\tilde{e}(y)$, where

$$\tilde{e}(y) \triangleq |\bar{\alpha}(y)| + \sup_{x\in\mathcal{S}_X}\mathbf{E}[|\bar{\alpha}(Y)|\,|X = x]. \tag{A.11}$$

Indeed, similarly as before, $|b'p_K(\lambda(x))| \leq [\mathrm{tr}(R_K(\lambda_x(x))^{M_\sigma})]^{1/2M_\sigma}$, and hence $F_K(x)\tilde{e}(y)$ is an envelope of $\mathcal{H}_2$. As for $\tilde{e}(Y)$, note that

$$\sup_{x\in\mathcal{S}_X}\mathbf{E}[\tilde{e}^{\,p}(Y)|X = x] \leq C\sup_{x\in\mathcal{S}_X}\mathbf{E}[|\bar{\alpha}(Y)|^p|X = x]$$

$$+ C\sup_{x\in\mathcal{S}_X}(\mathbf{E}[|\bar{\alpha}(Y)|\,|X = x])^p < \infty \tag{A.12}$$

by Assumption 1(b). Using (A.12), we deduce that

$$\|H_2\|_p^p = \mathbf{E}[F_K^p(X)\tilde{e}(Y)^p] \leq \mathbf{E}[F_K^p(X)\sup_{x\in\mathcal{S}_X}\mathbf{E}[\tilde{e}(Y)^p|X = x]]$$

$$\leq C\mathbf{E}[F_K^p(X)] = C\mathbf{E}_*([\mathrm{tr}\{R_K(\lambda_X(X))^{M_\sigma}\}]^{p/2M_\sigma})$$

$$\leq C\{\mathbf{E}_*[\mathrm{tr}(R_K(\lambda_X(X))^{M_\sigma})]\}^{p/2M_\sigma} \tag{A.13}$$

by noting that $p/(2M_\sigma) = p\sigma/2 < 1$ and applying Jensen's inequality. The inner integral in the last term is bounded by

$$K\lambda_{\max}(\mathbf{E}_*[R_K(\lambda_X(X))^{M_\sigma}]) \leq K\sup_{b\in S_K}\{b'\mathbf{E}_*[B_{1K}^{M_\sigma}(X; b)]b\}^{1/2} \leq K, \tag{A.14}$$

by (A.5). Hence $\|H_2\|_2 \leq \|H_2\|_p \leq CK^{\sigma/2}$. By Lemma A1(ii), we obtain from (A.10) that

$$\mathbf{E}^*\left[\sup_{(\alpha,\lambda,b)\in\mathcal{A}\times\Lambda\times S_K}\left|\frac{1}{n}\sum_{i=1}^{n} b'p^K(\lambda(X_i))\{\alpha(Y_i) - \mathbf{E}(\alpha(Y_i)|\lambda(X_i))\}\right|\right]$$

$$= O(n^{-1/2}K^{\sigma/2}\{\varrho_{s,r,q}^K(K^{-(\sigma/2)}) + \sqrt{K}\}). \tag{A.15}$$

■

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Proof of (C). Because $1_{n,\lambda} = 1$ implies $\lambda_{\min}(P_\lambda' P_\lambda/n) \geq c_K/2$, observe that $1_{n,\lambda}\|\hat{Q}_\lambda^{-1} P_\lambda' \delta_{\alpha,\lambda}/n\|$ is bounded by

$$1_{n,\lambda}\sqrt{\lambda_{\min}(P_\lambda' P_\lambda/n)^{-1}} \times \sqrt{\mathrm{tr}(\{G_{\alpha,\lambda} - P_\lambda \beta_{\alpha,\lambda}\}' P_\lambda (P_\lambda' P_\lambda)^{-} P_\lambda' \{G_{\alpha,\lambda} - P_\lambda \beta_{\alpha,\lambda}\})/n}$$

$$\leq C c_K^{-1/2} \sup_{\alpha,\lambda} \sqrt{\mathrm{tr}(\{G_{\alpha,\lambda} - P_\lambda \beta_{\alpha,\lambda}\}' \{G_{\alpha,\lambda} - P_\lambda \beta_{\alpha,\lambda}\})/n}.$$

By Assumption 4, the last term is $O_p(K^{-d} c_K^{-1/2})$, uniformly over $(\alpha, \lambda) \in \mathcal{A} \times \Lambda$, yielding the wanted result. ■

Proof of Theorem 1L(ii). Let $\mathbf{R}_0$ denote the common compact set that contains the sets $\{\lambda(x) : x \in \mathcal{S}_X\}$, $\lambda \in \Lambda$. We choose $\bar{\lambda}_1, \ldots, \bar{\lambda}_{N_1}$ from $\mathbf{R}_0$ such that for all $\bar{\lambda} \in \mathbf{R}_0$, there exists $j(\bar{\lambda})$ satisfying $\|\bar{\lambda} - \bar{\lambda}_{j(\bar{\lambda})}\| \leq n^{-M_1}$ (note that $N_1 = O(n^{M_1})$). Following the inequality (24) in de Jong (2002, p. 8),

$$\sup_{(\alpha,\lambda) \in \mathcal{A} \times \Lambda(c_K)} 1_{n,\lambda} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_\nu \leq B_{1n}' + B_{2n}' + O_P(K^{-d}),$$

where, with $\varepsilon_{\alpha,\lambda}$ and $\delta_{\alpha,\lambda}$ defined as in the proof of Theorem 1L(i),

$$B_{1n}' \triangleq \sup_{(\alpha,\lambda) \in \mathcal{A} \times \Lambda(c_K)} 1_{n,\lambda} |\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda}|_{\nu+1} n^{-M_1}, \tag{A.16}$$

$$B_{2n}' \triangleq \sup_{(\alpha,\lambda) \in \mathcal{A} \times \Lambda(c_K)} 1_{n,\lambda} \sup_{\mu \leq \nu} \max_{1 \leq j \leq N_1} |(D^\mu p^K)'(\bar{\lambda}_j) \hat{Q}_\lambda^{-1} P_\lambda' (\varepsilon_{\alpha,\lambda} + \delta_{\alpha,\lambda})/n|.$$

As for B_{1n}', note that from (B) and (C) in the proof of Theorem 1L(i),

$$B_{1n}' = O_P(n^{-M_1} \zeta_{\nu+1,K} \{n^{-1/2} K^{\sigma/2} \{\varrho_{s,r,q}^K (K^{-(\sigma/2)}) + \sqrt{K}\}/c_K + K^{-d}/\sqrt{c_K}\}).$$

Because we can take arbitrarily large M_1 and apply Assumption 6L(b), the following completes the proof.

(D) $B_{2n}' = O_P((\zeta_{\nu,K}/c_K)\{n^{-1/2}\sqrt{\log n} + \sqrt{c_K} K^{-d}\})$.

Proof of (D). Consider that $1_{n,\lambda} |D^\mu p^K(\bar{\lambda}_j)' \hat{Q}_\lambda^{-1} P_\lambda' (\varepsilon_{\alpha,\lambda} + \delta_{\alpha,\lambda})/n|$ is bounded by

$$1_{n,\lambda} |D^\mu p^K(\bar{\lambda}_j)' \hat{Q}_\lambda^{-1} P_\lambda' \varepsilon_{\alpha,\lambda}/n| + 1_{n,\lambda} |D^\mu p^K(\bar{\lambda}_j)' \hat{Q}_\lambda^{-1} P_\lambda' \delta_{\alpha,\lambda}/n|. \tag{A.17}$$

The second term is bounded by $\zeta_{\mu,K} 1_{n,\lambda} \|\hat{Q}_\lambda^{-1} P_\lambda' \delta_{\alpha,\lambda}/n\| = O(\zeta_{\mu,K} K^{-d} c_K^{-1/2})$ by (C), and the first term is bounded by

$$1_{n,\lambda} |D^\mu p^K(\bar{\lambda}_j)' (\hat{Q}_\lambda^{-1} - Q_\lambda^{-1}) P_\lambda' \varepsilon_{\alpha,\lambda}/n| + 1_{n,\lambda} |D^\mu p^K(\bar{\lambda}_j)' Q_\lambda^{-1} P_\lambda' \varepsilon_{\alpha,\lambda}/n|$$

$$\triangleq |A_{1j}(\alpha,\lambda)| + |A_{2j}(\alpha,\lambda)|, \text{ say.}$$

As for $|A_{1j}(\alpha,\lambda)|$, we will show the following later.

(D1) $\sup_{(\alpha,\lambda) \in \mathcal{A} \times \Lambda(c_K)} \max_{1 \leq j \leq N_1} |A_{1j}(\alpha,\lambda)| = O_P(n^{-1/2} \zeta_{\mu,K}/c_K)$.

Let us define $\mathcal{T}_0(c_K) \triangleq \{\tau \in \mathcal{T}_0 : \lambda(\cdot;\tau) \in \Lambda(c_K)\}$. Because $\mathcal{T}_0$ is of polynomial discrimination, we can choose any $M_3 > 0$ and $\{\tau_j\}_{j=1}^{N_3}$ from $\mathcal{T}_0(c_K)$ such that $N_3 \leq C_2 n^{C_3 r M_3}$ for some positive constants C_2 and C_3 and for any $\tau \in \mathcal{T}_0(c_K)$, there exists $j \in \{1, \ldots, N_3\}$ satisfying $\|\tau - \tau_j\|_{(1)} < n^{-rM_3}$. (The definition of the covering number

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

of $\mathcal{T}_0(c_K)$ does not require that $\tau_1, \ldots, \tau_{N_3}$ be chosen from $\mathcal{T}_0(c_K)$. However, we can replace τ_j outside $\mathcal{T}_0(c_K)$ by any $\tau_j' \in \mathcal{T}_0(c_K)$ such that $\|\tau_j - \tau_j'\|_{(1)} < \varepsilon$ so that $\tau_1', \ldots, \tau_{N_3}' \in \mathcal{T}_0(c_K)$ become centers of 2ε-balls that cover $\mathcal{T}_0(c_K)$. The factor 2 is subsumed into the constant $C_2 > 0$ in the bound $C_2 n^{C_3 r M_2}$ for the covering number N_3.) Let $\lambda_j(\cdot) \triangleq \lambda(\cdot; \tau_j)$ and $\Delta_j(\cdot) \triangleq \sup_{\|\tau - \tau_j\|_{(1)} < n^{-rM_3}} |\lambda(\cdot; \tau) - \lambda(\cdot; \tau_j)|$. Then the set $\{[\lambda_j - \Delta_j, \lambda_j + \Delta_j]\}_{j=1}^{N_3}$ forms n^{-M_3}-brackets that cover $\Lambda(c_K)$ so that for each $\lambda \in \Lambda$, $|\lambda - \lambda_j| \leq \Delta_j$ and $\|\Delta_j\|_2 < n^{-M_3}$ for some $j \in \{1, \ldots, N_3\}$ by Assumption 5L(a). Observe that

$$\begin{aligned}
&\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} \max_{1\leq j\leq N_1} |A_{2j}(\alpha,\lambda) - A_{2j}(\alpha,\lambda_{m(\lambda)})| \\
&\quad\leq \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} \max_{1\leq j\leq N_1} |D^{\mu}p^K(\bar{\lambda}_j)'\{Q_\lambda^{-1} - Q_{\lambda_{m(\lambda)}}^{-1}\}P_\lambda'\varepsilon_{\alpha,\lambda}/n| \\
&\qquad + \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} \max_{1\leq j\leq N_1} |D^{\mu}p^K(\bar{\lambda}_j)'Q_{\lambda_{m(\lambda)}}^{-1}\{P_\lambda'\varepsilon_{\alpha,\lambda} - P_{\lambda_{m(\lambda)}}'\varepsilon_{\alpha,\lambda_{m(\lambda)}}\}/n|. \qquad \textbf{(A.18)}
\end{aligned}$$

The term following the inequality is bounded by

$$\begin{aligned}
&\zeta_{\nu,K} \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} \|\{Q_\lambda^{-1} - Q_{\lambda_{m(\lambda)}}^{-1}\}P_\lambda'\varepsilon_{\alpha,\lambda}\|/n \\
&\quad\leq \frac{\zeta_{\nu,K}}{c_K^2} \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda(c_K)} \|Q_\lambda - Q_{\lambda_{m(\lambda)}}\| \|P_\lambda'\varepsilon_{\alpha,\lambda}/n\| \\
&\quad\leq \frac{\zeta_{\nu,K}}{c_K^2} \zeta_{0,K}\zeta_{1,K} \max_{1\leq m\leq N_3} \|\Delta_m\| \|P_\lambda'\varepsilon_{\alpha,\lambda}/n\|.
\end{aligned}$$

The last term becomes $O(n^{-CM_3}\zeta_{\nu,K}\zeta_{0,K}\zeta_{1,K}c_K^{-2}) \times O_P(\zeta_{0,K})$. Similarly the last term in (A.18) can be shown to be $O_P(n^{-CM_3}\zeta_{\nu,K}\zeta_{0,K}\zeta_{1,K}c_K^{-1})$ using Lemma A2(i). With sufficiently large M_3 taken, the difference between $A_{2j}(\alpha,\lambda)$ and $A_{2j}(\alpha,\lambda_{m(\lambda)})$ can be made to converge to zero arbitrarily fast uniformly over $(\alpha,\lambda) \in \mathcal{A} \times \Lambda$ and over $j \in \{1, \ldots, N_1\}$, leaving us with term $A_{2j}(\alpha,\lambda_{m(\lambda)})$ to deal with.

Let us define $d_{mj} \triangleq (D^{\mu}p^K)(\bar{\lambda}_j)'Q_{\lambda_m}^{-1}/\|(D^{\mu}p^K)(\bar{\lambda}_j)'Q_{\lambda_m}^{-1}\|$ and consider that

$$\begin{aligned}
&\max_{1\leq m\leq N_3} \sup_{\alpha\in\mathcal{A}} \max_{1\leq j\leq N_1} |A_{2j}(\alpha,\lambda_m)| \\
&\quad\leq \max_{1\leq m\leq N_3} \max_{1\leq j\leq N_1} \|(D^{\mu}p^K)(\bar{\lambda}_j)'Q_{\lambda_m}^{-1}\| \times \max_{1\leq m\leq N_3} \sup_{\alpha\in\mathcal{A}} \max_{1\leq j\leq N_1} |d_{mj}'P_{\lambda_m}'\varepsilon_{\alpha,\lambda_m}/n| \\
&\quad\leq (\zeta_{\mu,K}/c_K) \times \max_{1\leq m\leq N_3} \sup_{\alpha\in\mathcal{A}} \max_{1\leq j\leq N_1} |d_{mj}'P_{\lambda_m}'\varepsilon_{\alpha,\lambda_m}/n|. \qquad \textbf{(A.19)}
\end{aligned}$$

Define $\tau_n \triangleq \{\sqrt{n}K^{\sigma p/2}\}^{1/(p-1)}$ and $1_{n,i} \triangleq 1\{F_K(X_i)\tilde{e}(Y_i) \leq \tau_n\}$, where $F_K(x)$ and $\tilde{e}(y)$ are as defined in the proof of (B). We bound $|d_{mj}'P_{\lambda_m}'\varepsilon_{\alpha,\lambda_m}/n|$ in (A.19) by

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^n d_{mj}'p^K(\lambda_m(X_i))\varepsilon_{\alpha,\lambda_m,i}1_{n,i}\right| + \left|\frac{1}{n}\sum_{i=1}^n d_{mj}'p^K(\lambda_m(X_i))\varepsilon_{\alpha,\lambda_m,i}\{1-1_{n,i}\}\right| \\
&\quad\triangleq |V_{mj,n}^A(\alpha)| + |V_{mj,n}^B(\alpha)|, \quad \text{say,}
\end{aligned}$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

where $\varepsilon_{\alpha,\lambda_m,i} \triangleq \alpha(Y_i) - g_{\alpha,\lambda_m}(\lambda_m(X_i))$. We complete the proof by showing the following.

(D2) $\max_{1\le m\le N_3} \sup_{\alpha\in\mathcal{A}} \max_{1\le j\le N_1} |V^A_{mj,n}(\alpha)| = O_P(n^{-1/2}\sqrt{\log n})$.
(D3) $\max_{1\le m\le N_3} \sup_{\alpha\in\mathcal{A}} \max_{1\le j\le N_1} |V^B_{mj,n}(\alpha)| = O_P(n^{-1/2})$.

Proof of (D1). Let $\tilde{d}'_j \triangleq (D^\mu p^K)(\bar{\lambda}_j)'[\hat{Q}_\lambda^{-1} - Q_\lambda^{-1}]/\|(D^\mu p^K)(\bar{\lambda}_j)'(\hat{Q}_\lambda^{-1} - Q_\lambda^{-1})\|$ and write

$$1_{n,\lambda}|(D^\mu p^K)(\bar{\lambda}_j)'[\hat{Q}_\lambda^{-1} - Q_\lambda^{-1}]P'_\lambda \varepsilon_{\alpha,\lambda}|^2/n^2$$
$$= 1_{n,\lambda}|\tilde{d}'_j P'_\lambda \varepsilon_{\alpha,\lambda}/n|^2 \times \|(D^\mu p^K)(\bar{\lambda}_j)'(\hat{Q}_\lambda^{-1} - Q_\lambda^{-1})\|^2.$$

For a symmetric matrix A such that $A = UVU'$ where U is an orthogonal matrix and V is a diagonal matrix with the eigenvalues of A being its diagonal elements, we define $A_+ \triangleq U'V_+U$ with V_+ being equal to V with its diagonal elements replaced by their absolute values. Observe that $1_{n,\lambda}\|(D^\mu p^K)(\bar{\lambda}_j)'(\hat{Q}_\lambda^{-1} - Q_\lambda^{-1})\|^2$ is bounded by

$$1_{n,\lambda}(D^\mu p^K)(\bar{\lambda}_j)'\hat{Q}_\lambda^{-1}(Q_\lambda - \hat{Q}_\lambda)Q_\lambda^{-2}(Q_\lambda - \hat{Q}_\lambda)\hat{Q}_\lambda^{-1}(D^\mu p^K)(\bar{\lambda}_j)$$
$$= 1_{n,\lambda}\mathrm{tr}\{Q_\lambda^{-2}(Q_\lambda - \hat{Q}_\lambda)\hat{Q}_\lambda^{-1}(D^\mu p^K)(\bar{\lambda}_j)(D^\mu p^K)(\bar{\lambda}_j)'\hat{Q}_\lambda^{-1}(Q_\lambda - \hat{Q}_\lambda)\}$$
$$\le (C/c_K^2)1_{n,\lambda}\mathrm{tr}\{(Q_\lambda - \hat{Q}_\lambda)_+(Q_\lambda - \hat{Q}_\lambda)_+^{1/2}\hat{Q}_\lambda^{-1}(D^\mu p^K)(\bar{\lambda}_j)$$
$$\times (D^\mu p^K)(\bar{\lambda}_j)'\hat{Q}_\lambda^{-1}(Q_\lambda - \hat{Q}_\lambda)_+^{1/2}\}$$
$$\le (C/c_K^2)\lambda_{\max}(Q_\lambda - \hat{Q}_\lambda)_+ 1_{n,\lambda}\mathrm{tr}\{(Q_\lambda - \hat{Q}_\lambda)_+^{1/2}\hat{Q}_\lambda^{-1}(D^\mu p^K)(\bar{\lambda}_j)$$
$$\times (D^\mu p^K)(\bar{\lambda}_j)'\hat{Q}_\lambda^{-1}(Q_\lambda - \hat{Q}_\lambda)_+^{1/2}\}$$
$$\le C(\zeta_{\mu,K}^2/c_K^4)\{\lambda_{\max}(Q_\lambda - \hat{Q}_\lambda)_+\}^2.$$

The last term is equal to

$$\frac{C\zeta_{\mu,K}^2}{c_K^4}\left\{\sup_{b\in S_K}|b'(Q_\lambda - \hat{Q}_\lambda)b|\right\}^2$$
$$= O_P\left(\frac{\zeta_{\mu,K}^2}{c_K^4}\{n^{-1/2}K^\sigma\{\varrho_{\mathcal{T}_0,1/r}(\zeta_{0,K}\zeta_{1,K}/K^\sigma) + \sqrt{K}\}\}^2\right) \quad \textbf{(A.20)}$$

by (A.9). On the other hand, by (A.15),

$$\sup_{(\alpha,\lambda,b)\in\mathcal{A}\times\Lambda\times S_K}|b'P'_\lambda\varepsilon_{\alpha,\lambda}/n| = O_P(n^{-1/2}K^{\sigma/2}\{\varrho^K_{s,r,q}(K^{-(\sigma/2)}) + \sqrt{K}\}). \quad \textbf{(A.21)}$$

Combining (A.20) and (A.21) and using condition (11) of this theorem, we obtain the wanted result. ■

Proof of (D2). For fixed $(\lambda_m, d_{mj}) \in \Lambda \times S_K$, let $\mathcal{H}_{3mj} \triangleq \{d'_{mj}(p^K \circ \lambda_m)\{\alpha - (g_{\alpha,\lambda_m} \circ \lambda_m)\} : \alpha \in \mathcal{A}\}$ and let $\sigma^2_{mj} \triangleq \sup_{\varphi\in\mathcal{H}_{3mj}}\|\varphi\|_2^2$ and $F_{mj}(y,x) \triangleq \sqrt{d'_{mj}(p^K \circ \lambda_m)(x)(p^K \circ \lambda_m)(x)'d_{mj}}\tilde{e}(y)$. Hence F_{mj} is an envelope of $\mathcal{H}_{3mj}$. Furthermore, observe that by (A.12) and (A.5),

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

$$\begin{aligned}\|F_{mj}\|_2 &= (\mathbf{E}[d'_{mj}R_K(\lambda_m(X))d_{mj}\mathbf{E}[\tilde{e}(Y)^2|X]])^{1/2}\\ &\le C(\mathbf{E}[d'_{mj}R_K(\lambda_m(X))d_{mj}])^{1/2} \le C(\mathbf{E}[\{d'_{mj}R_K(\lambda_m(X))d_{mj}\}^{M_\sigma}])^{1/2M_\sigma}\\ &\le C(\mathbf{E}[d'_{mj}R_K(\lambda_m(X))^{M_\sigma}d_{mj}])^{1/2M_\sigma} \le (b^{*\prime}\mathbf{E}_*[B_{1K}^{M_\sigma}(X;b^*)]b^*)^{1/2M_\sigma} \le C. \quad \textbf{(A.22)}\end{aligned}$$

Therefore, $\sigma_{mj}^2 \le C$.

Let $\mu_n \triangleq \max_{1\le m\le N_3}\max_{1\le j\le N_1}\mathbf{E}^*[\sup_{\alpha\in\mathcal{A}}|\sqrt{n}V_{mj,n}^A(\alpha)|]$. Now, observe that

$$\begin{aligned}&\mathbf{P}^*\left\{\max_{1\le m\le N_3}\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N_1}|\sqrt{n}V_{mj,n}^A(\alpha)| > C(\mu_n + t\sqrt{\log n})\right\}\\ &\quad\le N_3N_1\max_{1\le m\le N_3}\max_{1\le j\le N_1}\mathbf{P}^*\left\{\sup_{\alpha\in\mathcal{A}}|\sqrt{n}V_{mj,n}^A(\alpha)| > C(\mu_n + t\sqrt{\log n})\right\} \qquad \textbf{(A.23)}\end{aligned}$$

and apply Theorem 2.14.25 of van der Vaart and Wellner (1996) to the preceding probability to bound it by

$$\exp\left[-C\left(\frac{t^2\log n}{\sigma_{mj}^2}\wedge\frac{t\sqrt{\log n}\sqrt{n}}{\tau_n}\right)\right] \le \exp\left[-C\left(\frac{t^2\log n}{C}\wedge\frac{t\sqrt{\log n}}{n^{-1/2}\tau_n}\right)\right].$$

By condition (c) of this theorem, $n^{-1/2}\tau_n\sqrt{\log n} = n^{-(1/2)+(1/2(p-1))}K^{\sigma p/2(p-1)}\sqrt{\log n} \to 0$. Hence for sufficiently large n, the bound following the inequality in (A.23) is less than or equal to

$$\begin{aligned}CN_3N_1\exp\left[-C\left(\frac{t^2\log n}{C}\right)\right] &= n^{C(M_3+M_1)}\exp\left[-C\left(\frac{t^2\log n}{C}\right)\right]\\ &\le n^{C(M_3+M_1)-Ct^2}\to 0\end{aligned}$$

by taking sufficiently large t.

As for μ_n, we will show that $\limsup_{n\ge1}\mu_n/\sqrt{\log n} < C$ shortly. In view of the preceding developments, this proves that

$$\max_{1\le m\le N_3}\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N_1}|V_{mj,n}^A(\alpha)| = O_P(n^{-1/2}\sqrt{\log n}),$$

giving the wanted result of (D2).

Let us turn to μ_n. By the maximal inequality and Lemma A1(iii) and by (A.22),

$$\begin{aligned}\mathbf{E}^*\left[\sup_{\alpha\in\mathcal{A}}|V_{mj,n}^A(\alpha)|\right] &\le \frac{1}{\sqrt{n}}\mathbf{E}^*\left[\sup_{\alpha\in\mathcal{A}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n d'_{mj}p^K(\lambda_m(X_i))\varepsilon_{\alpha,\lambda_m,i}1_{n,i}\right|\right]\\ &\le \frac{1}{\sqrt{n}}\int_0^{\|F_{mj}\|_2}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{H}_{3mj},\|\cdot\|_2)}\,d\varepsilon\\ &\le \frac{1}{\sqrt{n}}\int_0^{C}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{H}_{3mj},\|\cdot\|_2)}\,d\varepsilon = O(n^{-1/2}),\end{aligned}$$

leading to the claim that $\limsup_{n\ge1}\mu_n/\sqrt{\log n} < C$. ■

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Proof of (D3). First observe that

$$\mathbf{E}^*\left[\max_{1\le m\le N_3}\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N_1}|V^B_{mj,n}(\alpha)|\right]$$

$$\le \mathbf{E}^*\left[\max_{1\le m\le N_3}\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N_1}|d'_{mj}p^K(\lambda_m(X_i))\varepsilon_{\alpha,\lambda_m,i}(1-1_{n,i})|\right].$$

The last term is bounded by $C\mathbf{E}[F_K(X)\tilde{e}(Y)|1-1_{n,i}|]$ with F_K as defined in the proof of (B). Using the Hölder inequality and Markov inequality sequentially, we bound this latter quantity by

$$C(\mathbf{E}[F_K^p(X)\tilde{e}^p(Y)])^{1/p}(\mathbf{P}\{F_K(X)\tilde{e}(Y)\ge\tau_n\})^{(1-(1/p))}$$

$$\le C(\mathbf{E}[F_K^p(X)])^{1/p}(\mathbf{E}[F_K^p(X)])^{1-(1/p)}/\tau_n^{p-1}\le CK^{p/2M_\sigma}/\tau_n^{p-1}=O(n^{-1/2}),$$

by (A.13) and (A.14), and by our choice of $\tau_n=\{\sqrt{n}K^{p/2M_\sigma}\}^{1/(p-1)}$. ∎

Proof of Corollary 1L. (i) First, observe that

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|(\hat{g}_{\alpha,\lambda}\circ\lambda)\hat{1}_{\lambda,K}-(g_{\alpha,\lambda_0}\circ\lambda_0)\|_\infty$$

$$\le \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|(\hat{g}_{\alpha,\lambda}\circ\lambda)\hat{1}_{\lambda,K}-(g_{\alpha,\lambda}\circ\lambda)\|_\infty$$

$$+\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|(g_{\alpha,\lambda}\circ\lambda)-(g_{\alpha,\lambda_0}\circ\lambda_0)\|_\infty. \tag{A.24}$$

By Lemma A2(ii), the uniform convergence rate for the second term is $O(n^{-bs})$. For the first term, note that $\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|\hat{g}_{\alpha,\lambda}\hat{1}_{\lambda,K}-g_{\alpha,\lambda}\|_\infty$ is bounded by

$$\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|\hat{g}_{\alpha,\lambda}\hat{1}_{\lambda,K}-g_{\alpha,\lambda}1_{\lambda,K}\|_\infty+\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}\|g_{\alpha,\lambda}1_{\lambda,K}-g_{\alpha,\lambda}\|_\infty. \tag{A.25}$$

Write $\hat{g}_{\alpha,\lambda}\hat{1}_{\lambda,K}-g_{\alpha,\lambda}1_{\lambda,K}=\hat{g}_{\alpha,\lambda}(\hat{1}_{\lambda,K}-1_{\lambda,K})+\hat{g}_{\alpha,\lambda}1_{\lambda,K}-g_{\alpha,\lambda}1_{\lambda,K}$. For the first term, observe that for any arbitrary sequence $0<\varepsilon_n\to 0$, we can bound $\mathbf{P}^*\{\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}|\hat{g}_{\alpha,\lambda}(\hat{1}_{\lambda,K}-1_{\lambda,K})|>\varepsilon_n\}$ by

$$\mathbf{P}^*\left\{\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}|\hat{g}_{\alpha,\lambda}\hat{1}_{\lambda,K}(1-1_{\lambda,K})|>\varepsilon_n/2\right\}$$

$$+\mathbf{P}^*\left\{\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}|\hat{g}_{\alpha,\lambda}(\hat{1}_{\lambda,K}-1_{\lambda,K})|1_{\lambda,K}>\varepsilon_n/2\right\}. \tag{A.26}$$

For the first term, we bound $|1-1_{\lambda,K}|$ as follows:

$$|1-1_{\lambda,K}|\le 1\{\lambda_{\min}(\mathbf{E}[R_K(\lambda_0(X))])-\Delta_n\le c_K\},$$

where $\Delta_n\triangleq\sup_{\lambda\in\Lambda_n}|\lambda_{\min}(\mathbf{E}[R_K(\lambda(X))])-\lambda_{\min}(\mathbf{E}[R_K(\lambda_0(X))])|$. As for the term Δ_n, observe that (see the beginning of the proof of (A))

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

$$\begin{aligned}\Delta_n &\le \sup_{\lambda\in\Lambda_n}\sup_{b\in S_K}|b'\mathbf{E}[R_K(\lambda(X)) - R_K(\lambda_0(X))]b| \\ &\le \sup_{\lambda\in\Lambda_n}\sup_{b\in S_K}|b'\mathbf{E}[\{p^K(\lambda(X)) - p^K(\lambda_0(X))\}'p^K(\lambda(x))]b| \\ &\quad + \sup_{\lambda\in\Lambda_n}\sup_{b\in S_K}|b'\mathbf{E}[p^K(\lambda_0(x))'\{p^K(\lambda(X)) - p^K(\lambda_0(X))\}]b| \\ &\le 2\zeta_{0,K}\zeta_{1,K}\sup_{\lambda\in\Lambda_n}\mathbf{E}[|\lambda(X) - \lambda_0(X)|] \\ &\le 2\zeta_{0,K}\zeta_{1,K}\sup_{\lambda\in\Lambda_n}\|\lambda - \lambda_0\|_2 = O(\zeta_{0,K}\zeta_{1,K}n^{-b}).\end{aligned}$$

The third inequality is due to the mean-value expansion applied to $R_K(\lambda(X)) - R_K(\lambda_0(X))$ and the definitions of ζ_0 and ζ_1. By Assumption 7L(a), the last term in the preceding expression is $o(1)$. From the fact that $c_K \to 0$ and from Assumption 7L(b), we deduce that from some large K on, $\sup_{\lambda\in\Lambda_n}|1 - 1_{\lambda,K}| = 0$. Hence from some large K on, the first term in (A.26) is equal to zero and $\Lambda_n \subset \Lambda(c_K)$. The second term in (A.26) is bounded by $\mathbf{P}^*\{\inf_{\lambda\in\Lambda(c_K)}\lambda_{\min}(\hat{Q}_\lambda) \le c_K/2\}$, which is equal to

$$1 - \mathbf{P}^*\left\{\inf_{\lambda\in\Lambda(c_K)}\lambda_{\min}(\hat{Q}_\lambda) > c_K/2\right\} \to 0.$$

Hence the uniform convergence rate of $|\hat{g}_{\alpha,\lambda}\hat{1}_{\lambda,K} - g_{\alpha,\lambda}1_{\lambda,K}|$ is equal to that of $|\hat{g}_{\alpha,\lambda}1_{\lambda,K} - g_{\alpha,\lambda}1_{\lambda,K}|$. The uniform convergence rate of the latter term follows immediately from Theorem 1L. The wanted result of Corollary 1L(i) follows because the second term in (A.25) becomes zero from some large K on.

(ii) Let $\tilde{\varepsilon}_{\alpha,\lambda} \triangleq \alpha - \tilde{G}_{\alpha,\lambda}$, $\tilde{\delta}_{\alpha,\lambda} \triangleq \tilde{G}_{\alpha,\lambda} - P'_\lambda\beta_\alpha$, and $\tilde{G}_{\alpha,\lambda} \triangleq [g_{\alpha,\lambda_0}(\lambda(X_1)), \ldots, g_{\alpha,\lambda_0}(\lambda(X_n))]'$. First, observe that $\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}1_{n,\lambda}\|\hat{g}_{\alpha,\lambda} - g_{\alpha,\lambda_0}\|_\infty$ is bounded by

$$\begin{aligned}&\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}1_{n,\lambda}\sup_{\bar{\lambda}\in\mathbf{R}}|p^{K\prime}(\bar{\lambda})\hat{Q}_\lambda^{-1}P'_\lambda\tilde{\varepsilon}_{\alpha,\lambda}/n| \\ &\quad + \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}1_{n,\lambda}\sup_{\bar{\lambda}\in\mathbf{R}}|p^{K\prime}(\bar{\lambda})\hat{Q}_\lambda^{-1}P'_\lambda\tilde{\delta}_{\alpha,\lambda}/n| + O(K^{-d}). \qquad \textbf{(A.27)}\end{aligned}$$

The term $O(K^{-d})$ is due to $\|p^{K\prime}\beta_\alpha - g_{\alpha,\lambda_0}\|_\infty = O(K^{-d})$. The second term in (A.27) can be handled in the same manner as in the proof of (C). As for the first term, it is bounded by

$$\begin{aligned}&\sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}1_{n,\lambda}\sup_{\bar{\lambda}\in\mathbf{R}}|p^{K\prime}(\bar{\lambda})\hat{Q}_\lambda^{-1}\{P'_\lambda\tilde{\varepsilon}_{\alpha,\lambda} - P'_{\lambda_0}\varepsilon_{\alpha,\lambda_0}\}/n| \\ &\quad + \sup_{(\alpha,\lambda)\in\mathcal{A}\times\Lambda_n}1_{n,\lambda}\sup_{\bar{\lambda}\in\mathbf{R}}|p^{K\prime}(\bar{\lambda})\hat{Q}_\lambda^{-1}P'_{\lambda_0}\varepsilon_{\alpha,\lambda_0}/n|. \qquad \textbf{(A.28)}\end{aligned}$$

We deal with the first term. This term is bounded by $D_n \times \sup_{\lambda\in\Lambda}\sup_{\bar{\lambda}\in\mathbf{R}}1_{n,\lambda}\|p^{K\prime}(\bar{\lambda})\hat{Q}_\lambda^{-1}\|$, where

$$\begin{aligned}D_n \triangleq &\sup_{(\alpha,\lambda,b)\in\mathcal{A}\times\Lambda_n\times S_K}\left|\frac{1}{n}\sum_{i=1}^n b'\{p^K(\lambda(X_i)) - p^K(\lambda_0(X_i))\}(\alpha(Y_i) - g_{\alpha,\lambda_0}(\lambda(X_i)))\right| \\ &+ \sup_{(\alpha,\lambda,b)\in\mathcal{A}\times\Lambda_n\times S_K}\left|\frac{1}{n}\sum_{i=1}^n b'p^K(\lambda_0(X_i))\{g_{\alpha,\lambda_0}(\lambda(X_i)) - g_{\alpha,\lambda_0}(\lambda_0(X_i))\}\right|. \qquad \textbf{(A.29)}\end{aligned}$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

The first term on the right-hand side is bounded by

$$\zeta_{1,K}\left|\frac{1}{n}\sum_{i=1}^{n}(|\bar{\alpha}(Y_i)|+\sup_{x\in\mathcal{S}_X}\mathbf{E}[|\bar{\alpha}(Y_i)|\,|X_i=x])\right|\sup_{\lambda\in\Lambda_n}\|\lambda-\lambda_0\|_\infty=O_P(n^{-b}\zeta_{1,K})$$

by using Assumption 1(b) and the law of large numbers.

The second term in (A.29) is bounded by $Cn^{-bs}\sup_{b\in S_K}(1/n)\sum_{i=1}^{n}|b'p^K(\lambda_0(X_i))|$ by Assumption 3. Then note that

$$\begin{aligned}&\mathbf{E}^*\left[\sup_{b\in S_K}\left|\frac{1}{n}\sum_{i=1}^{n}|b'p^K(\lambda_0(X_i))|-\mathbf{E}(|b'p^K(\lambda_0(X_i))|)\right|\right]\\&\quad\le\frac{C}{\sqrt{n}}\int_0^{CK^{\sigma/2}}\sqrt{1+\log N(\varepsilon,S_K/K^{\sigma/2},\|\cdot\|)}\,d\varepsilon\\&\quad=O(n^{-1/2}K^{(1/2)+(\sigma/2)}),\end{aligned}$$

using the maximal inequality and following similar steps in the proof of Lemma A1(ii). Therefore,

$$\begin{aligned}\sup_{b\in S_K}\frac{1}{n}\sum_{i=1}^{n}|b'p^K(\lambda_0(X_i))|&\le\sup_{b\in S_K}\mathbf{E}(|b'p^K(\lambda_0(X_i))|)+O_P(n^{-1/2}K^{(1/2)+(\sigma/2)})\\&\le\sup_{b\in S_K}\{b'\mathbf{E}(\{p^K(\lambda_0(X_i))p^K(\lambda_0(X_i))'\}^{M_\sigma})b\}^{1/M_\sigma}\\&\quad+O_P(n^{-1/2}K^{(1/2)+(\sigma/2)}).\end{aligned}$$

The second to the last term is $O(1)$, and hence the last two terms combined are $O_P(1)$ by Assumption 6L(a). This gives the first term in (A.28) as $O_P(\zeta_{0,K}\{n^{-bs}+n^{-b}\zeta_{1,K}\}/c_K)$.

The uniform convergence of the second term in (A.28) can be dealt with by following similar steps in the proof of Theorem 1L(ii). We first bound this term by

$$\frac{\zeta_{0,K}}{c_K}\times\sup_{(\alpha,b)\in\mathcal{A}\times S_K}|b'P'_{\lambda_0}\varepsilon_{\alpha,\lambda_0}/n|$$

and then follow similar steps in the proof of (D). Choose $b_1,\ldots,b_N$ as centers of $2n^{-M_4}$-balls in S_K with $N=Cn^{CKM_4}$ and define $1'_{n,i}\triangleq 1\{F_K(X_i)\tilde{e}(Y_i)<\tau'_n\}$ where $\tau'_n\triangleq\{\sqrt{n}K^{(\sigma p/2)-(1/2)}\}^{1/(p-1)}$. Then,

$$\sup_{(\alpha,b)\in\mathcal{A}\times S_K}|b'P'_{\lambda_0}\varepsilon_{\alpha,\lambda_0}/n|\le\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N}|V^A_{j,n}(\alpha)|+\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N}|V^B_{j,n}(\alpha)|+O_P(n^{-M_4}\zeta_{0,K}),$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

where

$$V_{j,n}^A(\alpha) \triangleq \frac{1}{n}\sum_{i=1}^n b_j' p^K(\lambda_0(X_i))\varepsilon_{\alpha,\lambda_0,i}1'_{n,i} \quad \text{and}$$

$$V_{j,n}^B(\alpha) \triangleq \frac{1}{n}\sum_{i=1}^n b_j' p^K(\lambda_0(X_i))\varepsilon_{\alpha,\lambda_0,i}(1-1'_{n,i}).$$

Note that $\mathbf{E}[\{b_j' p^K(\lambda_0(X_i))\}^2\varepsilon_{\alpha,\lambda_0,i}^2] \le C\mathbf{E}[\{b_j' p^K(\lambda_0(X_i))\}^2] \le C$ by (A.5). Let

$$\mu_n' \triangleq \max_{1\le j\le N}\mathbf{E}^*\left[\sup_{\alpha\in\mathcal{A}}|\sqrt{n}V_{j,n}^A(\alpha)|\right]$$

and apply Theorem 2.14.25 of van der Vaart and Wellner (1996) to obtain

$$\mathbf{P}^*\left\{\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n b_j' p^K(\lambda_0(X_i))\varepsilon_{\alpha,\lambda_0,i}1'_{n,i}\right| > C(\mu_n' + t\sqrt{K\log n})\right\}$$

$$\le N\exp\left(-C\left(\frac{t^2K\log n}{C}\wedge\frac{t\sqrt{K\log n}}{n^{-1/2}\tau_n'}\right)\right) \le Cn^{CKM_4-t^2K}$$

from some sufficiently large K because $n^{-1/2}\sqrt{K\log n}\tau_n' = n^{2-p}K^{(\sigma+1)p-2}(\log n)^{p-1} \to 0$. With sufficiently large t, the last term converges to zero. As in the proof of (D2), we can show that $\mu_n' = O(n^{-1/2})$. Hence we obtain that

$$\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N}|V_{j,n}^A(\alpha)| = O_P(n^{-1/2}\sqrt{K\log n}).$$

We can show that $\sup_{\alpha\in\mathcal{A}}\max_{1\le j\le N}|V_{j,n}^B(\alpha)| = O_P(n^{-1/2}\sqrt{K})$ similarly as in the proof of (D3). Thus the second term in (A.28) is $O_P(n^{-1/2}\zeta_{0,K}\sqrt{K\log n}/c_K)$. However, this latter term is dominated by $O_P(n^{-b}\zeta_{0,K}\zeta_{1,K}/c_K)$ by a condition of this corollary. Combining this with the result for the first term in (A.28), we obtain the wanted result. ■

Proof of Theorem 1U. The proof is almost the same as the proof of Theorem 1L. ■

A.2. Auxiliary Results. We define classes of functions, $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3(b,\lambda)$:

$$\mathcal{H}_1 \triangleq \{\{b'(p^K\circ\lambda)\}^2 : (b,\lambda)\in S_K\times\Lambda\},$$

$$\mathcal{H}_2 \triangleq \{b'(p^K\circ\lambda)(\alpha-(g_{\alpha,\lambda}\circ\lambda)) : (\alpha,\lambda,b)\in\mathcal{A}\times\Lambda\times S_K\}, \quad \text{and}$$

$$\mathcal{H}_3(b,\lambda) \triangleq \{b'(p^K\circ\lambda)(\alpha-(g_{\alpha,\lambda}\circ\lambda)) : \alpha\in\mathcal{A}\}.$$

LEMMA A1. *For any $C>0$, the following holds.*

(i) $\int_0^{CK^\sigma}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{H}_1,\|\cdot\|_2)}\,d\varepsilon = O(K^\sigma\{\varrho_{\Lambda,1}(\zeta_{0,K}\zeta_{1,K}/K^\sigma)+\sqrt{K}\})$.
(ii) $\int_0^{CK^{\sigma/2}}\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{H}_2,\|\cdot\|_2)}\,d\varepsilon = O(K^{(\sigma/2)}\{\varrho_{s,r,q}^K(K^{-(\sigma/2)})+\sqrt{K}\})$.
(iii) $\sup_{(b,\lambda)\in S_K\times\Lambda}\int_0^C\sqrt{1+\log N_{[]}(\varepsilon,\mathcal{H}_3(b,\lambda),\|\cdot\|_2)}d\varepsilon = O(1)$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Proof of Lemma A1. (i) Fix $\varepsilon > 0$ and choose $\{\lambda_1(\cdot), \ldots, \lambda_{N_1}(\cdot)\}$ and $\{\Delta_1(\cdot), \ldots, \Delta_{N_1}(\cdot)\}$ such that for any $\lambda \in \Lambda$, there exists $k \in \{1, \ldots, N_1\}$ with $|\lambda(x) - \lambda_k(x)| \leq \Delta_k(x)$ and $\|\Delta_k\|_q \leq \varepsilon/(\zeta_{0,K}\zeta_{1,K})$. We also choose centers $b_1, \ldots, b_{N_2} \in S_K$ of $2\varepsilon/K^\sigma$-balls that cover S_K. Using the triangular inequality and the mean-value expansion subsequently, we deduce that

$$|\{b'p^K(\lambda(x))\}^2 - \{b_j'p^K(\lambda_k(x))\}^2| \leq 2\zeta_{0,K}\zeta_{1,K}|\lambda(x) - \lambda_k(x)| + |\{b'p^K(\lambda_k(x))\}^2 - \{b_j'p^K(\lambda_k(x))\}^2|.$$

The last term is bounded by $|(b' - b_j')R_K(\lambda_k(x))(b + b_j)|$ or by

$$\sqrt{(b' - b_j')R_K(\lambda_k(x))(b - b_j)}\sqrt{(b' + b_j')R_K(\lambda_k(x))(b + b_j)} \leq \|b - b_j\| \times \|b + b_j\|\sqrt{b_A'R_K(\lambda_k(x))b_A}\sqrt{b_B'R_K(\lambda_k(x))b_B},$$

where $b_A \triangleq (b - b_j)/\|b - b_j\|$ and $b_B \triangleq (b + b_j)/\|b + b_j\|$. However,

$$b_A'R_K(\lambda_k(x))b_A \leq (b_A'R_K(\lambda_k(x))^{M_\sigma}b_A)^{1/M_\sigma}.$$

Hence $|\{b'p^K(\lambda_k(x))\}^2 - \{b_j'p^K(\lambda_k(x))\}^2| \leq 2\|b - b_j\| \times (\mathrm{tr}[R_K(\lambda_k(x))^{M_\sigma}])^{1/M_\sigma}$. Using arguments similar to (A.14), we can bound $\mathbf{E}[\mathrm{tr}(R_K(\lambda_k(x))^{M_\sigma})]$ by CK. Let

$$\Delta_{kj}(x) \triangleq 2\{\zeta_{0,K}\zeta_{1,K}|\Delta_j(x)| + 2\varepsilon K^{-\sigma}(\mathrm{tr}[R_K(\lambda_k(x))^{M_\sigma}])^{1/M_\sigma}\}. \tag{A.30}$$

Then $\|\Delta_{kj}\|_2 \leq C\varepsilon$ by applying Jensen's inequality for the second term noting that $2/M_\sigma < 1$. Hence the set $\{[\Delta_{kj} - b_j'((p^Kp^{K\prime}) \circ \lambda_k)b_j, b_j'((p^Kp^{K\prime}) \circ \lambda_k)b_j + \Delta_{kj}]\}_{k,j=1}^{N_1,N_2}$ forms a set of $C\varepsilon$-brackets that cover the space $\mathcal{H}_1$. Therefore,

$$\log N_{[]}(C\varepsilon, \mathcal{H}_1, \|\cdot\|_2) \leq \log N_{[]}(\varepsilon/(\zeta_{0,K}\zeta_{1,K}), \Lambda, \|\cdot\|_q) + \log N(\varepsilon/K^\sigma, S_K, \|\cdot\|).$$

However, as S_K is a compact set in $\mathbf{R}^K$, $N(\varepsilon, S_K, \|\cdot\|) \leq C\varepsilon^{-CK}$. Thus we deduce that

$$\int_0^{CK^\sigma}\sqrt{1 + \log N_{[]}(C\varepsilon, \mathcal{H}_1, \|\cdot\|_2)}\,d\varepsilon \leq \int_0^{CK^\sigma}\sqrt{1 + \log N_{[]}(C\varepsilon/(\zeta_{0,K}\zeta_{1,K}), \Lambda, \|\cdot\|_2)}\,d\varepsilon + \int_0^{CK^\sigma}\sqrt{1 + \log N(C\varepsilon/K^\sigma, S_K, \|\cdot\|)}\,d\varepsilon.$$

The first term is bounded by $CK^\sigma\varrho_{\Lambda,1}(\zeta_{0,K}\zeta_{1,K}/K^\sigma)$, and the second term is bounded by

$$K^\sigma\int_0^{CK^\sigma/K^\sigma}\sqrt{1 + \log N_{[]}(C\varepsilon, S_K, \|\cdot\|)}\,d\varepsilon \leq CK^\sigma\int_0^2\sqrt{1 + K(-\log\varepsilon)}\,d\varepsilon \leq CK^\sigma\sqrt{K}.$$

The first inequality comes from the fact that $N(\varepsilon, S_K, \|\cdot\|) = 1$ for all $\varepsilon \geq 2$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

(ii) Using similar arguments as for (A.23) in the proof of Theorem 6 in Andrews (1994),

$$\log N_{[]}(C\varepsilon, \mathcal{H}_2, \|\cdot\|_2) \leq \log N_{[]}(\varepsilon, \mathcal{H}_2', \|\cdot\|_2) + \log N_{[]}(\varepsilon, \mathcal{A}', \|\cdot\|_2), \tag{A.31}$$

where $\mathcal{H}_2' \triangleq \{b'(p^K \circ \lambda)\alpha : (\alpha, \lambda, b) \in \mathcal{A} \times \Lambda \times S_K\}$ and

$$\mathcal{A}' \triangleq \{b'(p^K \circ \lambda)(g_{\alpha,\lambda} \circ \lambda) : (\alpha, \lambda, b) \in \mathcal{A} \times \Lambda \times S_K\}. \tag{A.32}$$

We choose similarly as in (i) $\{(\lambda_k, \Delta_k)\}_{k=1}^{N_1}$ and $\{b_j\}_{j=1}^{N_2} \subset S_K$ where $\|\Delta_k\|_2 < \varepsilon/\zeta_{1,K}$ and the set $\{b_j\}_{j=1}^{N_2} \in S_K$ constitutes the centers of $2\varepsilon/K^{\sigma/2}$-balls. We also choose $\{(\alpha_m, \tilde{\Delta}_m)\}_{m=1}^{N_3}$ such that for any $\alpha \in \mathcal{A}$, there exists $m(\alpha) \in \{1, \ldots, N_3\}$ such that $|\alpha(y) - \alpha_{m(\alpha)}(y)| \leq \tilde{\Delta}_{m(\alpha)}(y)$ and $\|\tilde{\Delta}_{m(\alpha)}\|_p < \varepsilon$. Note that $|b'(p^K \circ \lambda)\alpha - b'(p^K \circ \lambda_{k(\lambda)})\alpha| \leq \zeta_{1,K}\bar{\alpha}|\lambda - \lambda_{k(\lambda)}| \leq \zeta_{1,K}\bar{\alpha}\Delta_{k(\lambda)}$ and that

$$\begin{aligned}|b'(p^K \circ \lambda_{k(\lambda)})\alpha - b_{j(b)}'(p^K \circ \lambda_{k(\lambda)})\alpha| &\leq \bar{\alpha}\|b - b_{j(b)}\|(\mathrm{tr}[R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2}])^{1/M_\sigma} \\ &\leq 2\bar{\alpha}\varepsilon(\mathrm{tr}[R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2}])^{1/M_\sigma}/K^{\sigma/2} \quad \text{and}\end{aligned}$$

$$\begin{aligned}&|b_{j(b)}'(p^K \circ \lambda_{k(\lambda)})\alpha - b_{j(b)}'(p^K \circ \lambda_{k(\lambda)})\alpha_{m(\alpha)}| \\ &\qquad \leq |\alpha - \alpha_{m(\alpha)}|(b_{j(b)}' R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2} b_{j(b)})^{1/M_\sigma} \\ &\qquad \leq \tilde{\Delta}_{m(\alpha)}(b_{j(b)}' R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2} b_{j(b)})^{1/M_\sigma}.\end{aligned}$$

Then observe that $\{\mathbf{E}[\bar{\alpha}(Y)^2(\mathrm{tr}[R_K(\lambda_{k(\lambda)}(X))^{M_\sigma/2}])^{2/M_\sigma}]\}^{1/2}$ is equal to

$$\begin{aligned}&\{\mathbf{E}[\mathbf{E}[\bar{\alpha}(Y)^2|X](\mathrm{tr}[R_K(\lambda_{k(\lambda)}(X))^{M_\sigma/2}])^{2/M_\sigma}]\}^{1/2} \\ &\quad \leq C\{\mathbf{E}[(\mathrm{tr}[R_K(\lambda_{k(\lambda)}(X))^{M_\sigma/2}])^{2/M_\sigma}]\}^{1/2} \\ &\quad \leq C\{\mathbf{E}[\mathrm{tr}[R_K(\lambda_{k(\lambda)}(X))^{M_\sigma/2}]]\}^{1/M_\sigma} \leq CK^{\sigma/2}\end{aligned}$$

similarly as in (A.13) and (A.14). Also, note that by the Hölder inequality,

$$\begin{aligned}&\|\tilde{\Delta}_{m(\alpha)}(b_{j(b)}' R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2} b_{j(b)})^{1/M_\sigma}\|_2 \\ &\quad \leq \|\tilde{\Delta}_{m(\alpha)}\|_p \|b_{j(b)}' R_K(\lambda_{k(\lambda)}(\cdot))^{M_\sigma/2} b_{j(b)}\|_{2\sigma p/(p-2)}.\end{aligned} \tag{A.33}$$

However,

$$\begin{aligned}&\mathbf{E}[\{b_{j(b)}' R_K(\lambda_{k(\lambda)}(X))^{M_\sigma/2} b_{j(b)}\}^{2p/(p-2)M_\sigma}] \\ &\quad \leq \mathbf{E}[\{b_{j(b)}' R_K(\lambda_{k(\lambda)}(X))^{M_\sigma} b_{j(b)}\}^{p/(p-2)M_\sigma}] \\ &\quad \leq \{\mathbf{E}[\{b_{j(b)}' R_K(\lambda_{k(\lambda)}(X))^{M_\sigma} b_{j(b)}\}]\}^{p/(p-2)M_\sigma} \leq C,\end{aligned}$$

noting that $\sigma < (p-2)/p$ and applying Jensen's inequality and (A.5). Therefore, the bound in (A.33) is given by $C\varepsilon$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Define

$$\Delta_{k,m,j}(y,x) \triangleq \zeta_{1,K}\bar{\alpha}(y)\Delta_k(x) + 2\bar{\alpha}(y)\varepsilon(\mathrm{tr}[R_K(\lambda_k(x))^{M_\sigma/2}])^{1/M_\sigma}/K^{\sigma/2}$$
$$+ \tilde{\Delta}_m(y)(b_j' R_K(\lambda_k(x))^{M_\sigma/2} b_j)^{1/M_\sigma}.$$

As $\{[b_j'(p^K \circ \lambda_k)\alpha_m - \Delta_{k,m,j}, b_j'(p^K \circ \lambda_k)\alpha_m + \Delta_{k,m,j}]\}_{k,m,j=1}^{N_1,N_2,N_3}$ form $C\varepsilon$-brackets that cover $\mathcal{H}_2'$,

$$N_{[]}(C\varepsilon, \mathcal{H}_2', \|\cdot\|_2) \leq N_{[]}(C\varepsilon/\zeta_{1,K}, \Lambda, \|\cdot\|_2) N_{[]}(C\varepsilon, \mathcal{A}, \|\cdot\|_p) N(C\varepsilon/K^{\sigma/2}, S_K, \|\cdot\|).$$

As for $\mathcal{A}'$, set $\lambda(x) = \lambda(x;\tau)$ and $\lambda_{k(\lambda)}(x) = \lambda(x;\tau_{k(\tau)})$ and observe that

$$|g_{\alpha,\lambda} \circ \lambda - g_{\alpha_{m(\alpha)},\lambda_{k(\lambda)}} \circ \lambda_{k(\lambda)}| \leq |g_{\alpha,\lambda} \circ \lambda - g_{\alpha,\lambda_{k(\lambda)}} \circ \lambda_{k(\lambda)}| \tag{A.34}$$
$$+ |g_{\alpha,\lambda_{k(\lambda)}} \circ \lambda_{k(\lambda)} - g_{\alpha_{m(\alpha)},\lambda_{k(\lambda)}} \circ \lambda_{k(\lambda)}|.$$

Let $g_{\alpha,\tau}(\bar{\lambda}) \triangleq \mathbf{E}[\alpha(Y)|\lambda(X;\tau) = \bar{\lambda}]$ and bound the first term on the right-hand side by

$$\Delta_1(x) \triangleq \sup_{\alpha\in\mathcal{A}} \sup_{\tau' \in \mathcal{T}_0 : \|\tau' - \tau_{k(\tau)}\|_{(1)} < \varepsilon} |g_{\alpha,\tau'}(\lambda(x;\tau')) - g_{\alpha,\tau_{k(\tau)}}(\lambda(x;\tau_{k(\tau)}))|.$$

It turns out that $\|\Delta_1\|_2 \leq C\varepsilon^{qrs/q+2s}$ by Lemma A2(i). The second term in (A.34) is bounded by $\mathbf{E}[\tilde{\Delta}_{m(\alpha)}(Y)|\lambda_{k(\lambda)}(X) = \lambda_{k(\lambda)}(x)]$. Hence similarly as before, we can show that

$$N_{[]}(C\varepsilon, \mathcal{A}', \|\cdot\|_2) \leq N_{[]}(C\varepsilon/\zeta_{1,K}, \Lambda, \|\cdot\|_2) N_{[]}(C\varepsilon, \mathcal{A}, \|\cdot\|_p)$$
$$\times N(\varepsilon^{q+2s/qrs}, \mathcal{T}_0, \|\cdot\|_{(1)}) N(C\varepsilon/K^{\sigma/2}, S_K, \|\cdot\|).$$

However, by Assumption 5L(b), $N_{[]}(C\varepsilon, \Lambda, \|\cdot\|_2) \leq N(C\varepsilon^{1/r}, \mathcal{T}_0, \|\cdot\|_{(1)})$ and hence

$$\int_0^{CK^{\sigma/2}} \sqrt{1 + \log N_{[]}(C\varepsilon/\zeta_{1,K}, \Lambda, \|\cdot\|_2)}\, d\varepsilon$$
$$\leq \zeta_{1,K} \int_0^{CK^{\sigma/2}/\zeta_{1,K}} \sqrt{1 + \log N_{[]}(C\varepsilon, \Lambda, \|\cdot\|_2)}\, d\varepsilon$$
$$\leq \zeta_{1,K} \int_0^{CK^{\sigma/2}/\zeta_{1,K}} \sqrt{1 + \log N_{[]}(C\varepsilon^{1/r}, \mathcal{T}_0, \|\cdot\|_{(1)})}\, d\varepsilon.$$

From (A.31), $\int_0^{CK^{\sigma/2}} \sqrt{1 + \log N_{[]}(C\varepsilon, \mathcal{H}_2, \|\cdot\|_2)} d\varepsilon = O(K^{\sigma/2}\{\varrho_{s,r,q}^K(K^{-(\sigma/2)}) + K^{1/2}\})$, using similar arguments in the proof of (i).

(iii) Fix $(b,\lambda) \in S_K \times \Lambda$ and note that by the mean-value theorem,

$$|b'(p^K \circ \lambda)(\alpha - (g_{\alpha,\lambda} \circ \lambda)) - b'(p^K \circ \lambda)(\alpha_k - (g_{\alpha_k,\lambda} \circ \lambda))|$$
$$\leq (b' R_K(\lambda(\cdot))^{M_\sigma/2} b)^{1/M_\sigma}\{|\alpha - \alpha_k| + |(g_{\alpha,\lambda} \circ \lambda) - (g_{\alpha_k,\lambda} \circ \lambda)|\}.$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

By taking the $\|\cdot\|_p$ norm on both sides and following steps similar to the proof of (i), we conclude that

$$\log N_{[]}(C\varepsilon, \mathcal{H}_3(b,\lambda), \|\cdot\|_p) \leq \log N_{[]}(C\varepsilon, \mathcal{A}, \|\cdot\|_p),$$

which gives the wanted result similarly as before. ■

The following lemma establishes a local uniform $\mathcal{L}_p$-continuity of conditional mean functions in a parameter of the conditioning variable. The lemma uses the same notation as in Section 3.1, but the conditions are formulated to be self-contained, without drawing on assumptions beyond what is assumed in the lemma.

LEMMA A2. *Suppose that for some $p \geq 1$, $\sup_{\alpha\in\mathcal{A}} \sup_{x\in\mathcal{S}_X} \mathbf{E}[|\bar{\alpha}(Y)|^p | X = x] < \infty$, and for some $s \in (0,1]$, there exists $C > 0$ such that for each $\lambda \in \Lambda$ and for all $\delta > 0$, the conditional density $f_\lambda(y,x|\bar{\lambda})$ of (Y,X) given $\lambda(X) = \bar{\lambda}$ with respect to a σ-finite measure w_λ satisfies the following: for all $\bar{\lambda} \in \mathcal{S}_\lambda$,*

$$\sup_{\bar{\lambda}_1\in\mathcal{S}_\lambda:|\bar{\lambda}-\bar{\lambda}_1|<\delta} |f_\lambda(y,x|\bar{\lambda}) - f_\lambda(y,x|\bar{\lambda}_1)| \leq \varphi_\lambda(y,x)\delta^s \tag{A.35}$$

for some $\varphi_\lambda(y,x)$ such that $\sup_{x\in\mathcal{S}_X} \int |\bar{\alpha}(y)|\varphi_\lambda(y,x)w(dy) < C$. Then the following holds.

(i) Suppose that $\Lambda = \{\lambda(\cdot;\tau) : \tau \in \mathcal{T}_0\}$ as in Section 3.1 and there exists $r \in (0,1]$ such that for each $\tau_1 \in \mathcal{T}_0$,

$$\left\{\mathbf{E}^*\left[\sup_{\tau_2\in\mathcal{T}_0:\|\tau_1-\tau_2\|_{(1)}<\delta} |\lambda(X;\tau_1) - \lambda(X;\tau_2)|^q\right]\right\}^{1/q} \leq \delta^r, \quad \text{for all } \delta > 0. \tag{A.36}$$

Then for each $\tau_1 \in \mathcal{T}_0$,

$$\left\{\mathbf{E}^*\left[\sup_{(\alpha,\tau_2)\in\mathcal{A}\times\mathcal{T}_0:\|\tau_1-\tau_2\|_{(1)}<\delta} |g_{\alpha,\tau_1}(\lambda(X;\tau_1)) - g_{\alpha,\tau_2}(\lambda(X;\tau_2))|^p\right]\right\}^{1/p} \leq C\delta^{qrs/(q+ps)}. \tag{A.37}$$

Hence, as for the class $\mathcal{G} \triangleq \{g_{\alpha,\lambda} \circ \lambda : (\alpha,\lambda) \in \mathcal{A}\times\Lambda\}$, there exists $C > 0$ such that for all $\varepsilon \in (0,\infty)$,

$$N_{[]}(C\varepsilon, \mathcal{G}, \|\cdot\|_p) \leq N_{[]}(\varepsilon, \mathcal{A}, \|\cdot\|_p) N(\varepsilon^{(q+ps)/qrs}, \mathcal{T}_0, \|\cdot\|_{(1)}). \tag{A.38}$$

(ii) Suppose that Λ is uniformly bounded. Then for each λ_1 and λ_2 in Λ,

$$\sup_{\alpha\in\mathcal{A}} \|g_{\alpha,\lambda_1} \circ \lambda_1 - g_{\alpha,\lambda_2} \circ \lambda_2\|_\infty \leq C\|\lambda_1 - \lambda_2\|_\infty^s,$$

where C does not depend on the choice of $\lambda_1, \lambda_2 \in \Lambda$. Hence, as for the class $\mathcal{G}$ defined in (i), there exists $C > 0$ such that for all $\varepsilon \in (0,\infty)$,

$$N_{[]}(C\varepsilon, \mathcal{G}, \|\cdot\|_p) \leq N_{[]}(\varepsilon, \mathcal{A}, \|\cdot\|_p) N(\varepsilon^{1/s}, \Lambda, \|\cdot\|_\infty). \tag{A.39}$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Proof of Lemma A2. (i) Fix $\lambda_1(\cdot) \triangleq \lambda(\cdot;\tau_1) \in \Lambda$ for some $\tau_1 \in \mathcal{T}_0$ and set

$$\Delta(x) \triangleq \sup_{\tau_2 \in \mathcal{T}_0 : \|\tau_1 - \tau_2\|_{(1)} < \delta} |\lambda(x;\tau_1) - \lambda(x;\tau_2)|$$

and $\delta_q \triangleq \|\Delta\|_q$ and take $w \triangleq q/(q+ps)$. Hence by (A.36), $\delta_q \leq \delta^r$. We write $\lambda_2(x) \triangleq \lambda(x;\tau_2)$. Define

$$\tilde{g}_{\alpha,\lambda_1}(\bar{\lambda}) \triangleq \frac{\mathbf{E}[\alpha(Y) 1\{|\lambda_1(X) - \bar{\lambda}| \leq \delta_q^w\}]}{\mathbf{P}\{|\lambda_1(X) - \bar{\lambda}| \leq \delta_q^w\}} \quad \text{and}$$

$$\tilde{g}_{\alpha,\lambda_2}(\bar{\lambda}) \triangleq \frac{\mathbf{E}[\alpha(Y) 1\{|\lambda_2(X) - \bar{\lambda}| \leq \delta_q^w\}]}{\mathbf{P}\{|\lambda_2(X) - \bar{\lambda}| \leq \delta_q^w\}}.$$

Then note that $|g_{\alpha,\lambda_1}(\lambda_1(x)) - g_{\alpha,\lambda_2}(\lambda_2(x))|$ is bounded by

$$|\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - g_{\alpha,\lambda_1}(\lambda_1(x))| + |\tilde{g}_{\alpha,\lambda_2}(\lambda_2(x)) - g_{\alpha,\lambda_2}(\lambda_2(x))|$$
$$+ |\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - \tilde{g}_{\alpha,\lambda_2}(\lambda_2(x))|.$$

As for the first two terms on the right-hand side, observe that by the definition of conditional expectation,

$$\frac{\mathbf{E}[\alpha(Y) 1\{\lambda(X) \in [\bar{\lambda} - \delta_q^w, \bar{\lambda} + \delta_q^w]\}]}{\mathbf{P}\{\lambda(X) \in [\bar{\lambda} - \delta_q^w, \bar{\lambda} + \delta_q^w]\}} = \frac{\int_{[\bar{\lambda} - \delta_q^w, \bar{\lambda} + \delta_q^w]} \mathbf{E}[\alpha(Y)|\lambda(X) = \tilde{\lambda}] \, dF_\lambda(\tilde{\lambda})}{\mathbf{P}\{\lambda(X) \in [\bar{\lambda} - \delta_q^w, \bar{\lambda} + \delta_q^w]\}}.$$

Hence we can bound $|\tilde{g}_{\alpha,\lambda_2}(\lambda_2(x)) - g_{\alpha,\lambda_2}(\lambda_2(x))|$ by

$$\left| \frac{\int_{\{\lambda_2(x) \in [\bar{\lambda} - \delta_q^w, \bar{\lambda} + \delta_q^w]\}} \mathbf{E}[\alpha(Y)|\lambda_2(X) = \lambda_2(x)] \, dF_X(x)}{\mathbf{P}\{\lambda_2(X) \in [\lambda_2(x) - \delta_q^w, \lambda_2(x) + \delta_q^w]\}} - \mathbf{E}[\alpha(Y)|\lambda_2(X) = \lambda_2(x)] \right|$$
$$\leq \sup_{v \in [0, \delta_q^w] \cap \operatorname{supp}(\lambda_2(X))} |\mathbf{E}[\alpha(Y)|\lambda_2(X) = \lambda_2(x) + v] - \mathbf{E}[\alpha(Y)|\lambda_2(X) = \lambda_2(x)]|. \tag{A.40}$$

Therefore by (A.35),

$$\sup_{x \in \mathcal{S}_X} \sup_{(\alpha,\lambda_2) \in \mathcal{A} \times \Lambda : \|\tau_1 - \tau_2\|_{(1)} < \delta} |\tilde{g}_{\alpha,\lambda_2}(\lambda_2(x)) - g_{\alpha,\lambda_2}(\lambda_2(x))| \leq C\delta_q^{ws}. \tag{A.41}$$

Similarly, $\sup_{x \in \mathcal{S}_X} \sup_{\alpha \in \mathcal{A}} |\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - g_{\alpha,\lambda_1}(\lambda_1(x))| \leq C\delta_q^{ws}$.

Let us turn to $|\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - \tilde{g}_{\alpha,\lambda_2}(\lambda_2(x))|$. Define $A_j \triangleq 1\{|\lambda_j(X) - \lambda_j(x)| \leq \delta_q^w\}$, $j = 1, 2$.

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

(Case A) Suppose that $\mathbf{E}[A_2] \leq \mathbf{E}[A_1]$. Write $|\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - \tilde{g}_{\alpha,\lambda_2}(\lambda_2(x))|$ as

$$\begin{aligned}\left|\frac{\mathbf{E}[\alpha(Y)A_1]}{\mathbf{E}A_1} - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\right| &= \left|\frac{1}{\mathbf{E}A_1}\left\{\mathbf{E}[\alpha(Y)A_1] - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\mathbf{E}A_1\right\}\right| \\ &= \left|\frac{1}{\mathbf{E}A_1}\mathbf{E}\left[\left\{\alpha(Y) - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\right\}A_1\right]\right| \\ &= \left|\frac{1}{\mathbf{E}A_1}\mathbf{E}\left[\left\{\alpha(Y) - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\right\}(A_1 - A_2)\right]\right|. \end{aligned} \tag{A.42}$$

The second equality follows because

$$\frac{1}{\mathbf{E}A_1}\mathbf{E}\left[\left\{\alpha(Y) - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\right\}A_2\right] = \frac{1}{\mathbf{E}A_1}\left\{\mathbf{E}[\alpha(Y)A_2] - \frac{\mathbf{E}[\alpha(Y)A_2]\mathbf{E}[A_2]}{\mathbf{E}A_2}\right\} = 0.$$

Define $V_1 \triangleq \alpha(Y) - \mathbf{E}[\alpha(Y)A_2]/\mathbf{E}A_2$ and $V_2 \triangleq (\mathbf{E}A_2/\mathbf{E}A_1)\{\alpha(Y) - \mathbf{E}[\alpha(Y)A_2]/\mathbf{E}A_2\}$. Then note that $\mathbf{E}[|V_1||X] < C$ and

$$\mathbf{E}[|V_2||X] = \left\langle\frac{\mathbf{E}A_2}{\mathbf{E}A_1}\right\rangle\mathbf{E}\left[|\alpha(Y) - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}[A_2]}||X\right] < C \tag{A.43}$$

by the condition $\mathbf{E}[|\bar{\alpha}(Y)|^p|X] < \infty$. The last term of (A.42) is bounded by

$$\begin{aligned}&\left|\frac{\mathbf{E}[V_1A_1(A_1 - A_2)]}{\mathbf{E}A_1}\right| + \left|\frac{\mathbf{E}[V_2A_2(A_1 - A_2)]}{\mathbf{E}A_2}\right| \\ &\quad\leq \mathbf{E}[|V_1(1 - A_2)\|\lambda_1(X) = \lambda_1(x)] + \mathbf{E}[|V_2(1 - A_1)\|\lambda_2(X) = \lambda_2(x)] + C\delta_q^{ws},\end{aligned} \tag{A.44}$$

similarly as in (A.40) and (A.41). However, $1 - A_2 = 1\{|\lambda_2(X) - \lambda_2(x)| > \delta_q^w\}$. For the first term,

$$\begin{aligned}\mathbf{E}[|V_1(1 - A_2)\|\lambda_1(X) = \lambda_1(x)] &\leq \mathbf{E}[\mathbf{E}[|V_1\|X](1 - A_2)|\lambda_1(X) = \lambda_1(x)] \\ &\leq C\mathbf{E}[1 - A_2|\lambda_1(X) = \lambda_1(x)] \\ &\leq C\delta_q^{-(wq/p)}\mathbf{E}[|\lambda_2(X) - \lambda_2(x)|^{q/p}|\lambda_1(X) = \lambda_1(x)].\end{aligned} \tag{A.45}$$

As for the last term, $\mathbf{E}[|\lambda_2(X) - \lambda_2(x)|^{q/p}|\lambda_1(X) = \lambda_1(x)]$ is equal to

$$\begin{aligned}&\{\mathbf{E}[|\lambda_2(X) - \lambda_1(X) + \lambda_1(x) - \lambda_2(x)|^{q/p}|\lambda_1(X) = \lambda_1(x)]\} \\ &\quad\leq C\mathbf{E}^*[\Delta(X)^{q/p}|\lambda_1(X) = \lambda_1(x)] + C\Delta(x)^{q/p}.\end{aligned}$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

Hence the first term of (A.44) is bounded by $C\delta_q^{-(wq/p)}\{\mathbf{E}^*[\Delta(X)^{q/p}|\lambda_1(X) = \lambda_1(x)] + \Delta(x)^{q/p}\}$. Likewise, we bound the second term in (A.44) by $C\delta_q^{-(wq/p)}\{\mathbf{E}^*[\Delta(X)^{q/p}|\lambda_2(X) = \lambda_2(x)] + \Delta(x)^{q/p}\}$ using (A.43). Thus

$$
\begin{aligned}
&|\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - \tilde{g}_{\alpha,\lambda_2}(\lambda_2(x))| \\
&\quad \le C\delta_q^{ws} + C\delta_q^{-(wq/p)}\{\mathbf{E}^*[\Delta(X)^{q/p}|\lambda_1(X) = \lambda_1(x)] \\
&\qquad\qquad + \mathbf{E}^*[\Delta(X)^{q/p}|\lambda_2(X) = \lambda_2(x)] + 2\Delta(x)^{q/p}\}. \qquad \textbf{(A.46)}
\end{aligned}
$$

(Case B) Suppose $\mathbf{E}[A_1] \le \mathbf{E}[A_2]$. In this case, we write $|\tilde{g}_{\alpha,\lambda_1}(\lambda_1(x)) - \tilde{g}_{\alpha,\lambda_2}(\lambda_2(x))|$ as

$$
\left|\frac{\mathbf{E}[\alpha(Y)A_1]}{\mathbf{E}A_1} - \frac{\mathbf{E}[\alpha(Y)A_2]}{\mathbf{E}A_2}\right| = \left|\frac{1}{\mathbf{E}A_2}\mathbf{E}\left[\left\{\alpha(Y) - \frac{\mathbf{E}[\alpha(Y)A_1]}{\mathbf{E}A_1}\right\}(A_2 - A_1)\right]\right|,
$$

follow the same steps in Case A with the roles of A_1 and A_2 interchanged, and obtain the same bound in (A.46).

Combined with (A.41), the bound in (A.46) leads to the conclusion that

$$
\begin{aligned}
&|g_{\alpha,\lambda_1}(\lambda_1(x)) - g_{\alpha,\lambda_2}(\lambda_2(x))| \\
&\quad \le C\delta_q^{ws} + C\delta_q^{-(wq/p)}\{\mathbf{E}^*[\Delta(X)^{q/p}|\lambda_1(X) = \lambda_1(x)] \\
&\qquad\qquad + \mathbf{E}^*[\Delta(X)^{q/p}|\lambda_2(X) = \lambda_2(x)] + 2\Delta(x)^{q/p}\}. \qquad \textbf{(A.47)}
\end{aligned}
$$

Therefore, we can bound $\{\mathbf{E}^*[\sup_{(\alpha,\tau_2)\in\mathcal{A}\times\mathcal{T}_0:\|\tau_1-\tau_2\|_{(1)}<\delta}|g_{\alpha,\tau_1}(\lambda(X;\tau_1)) - g_{\alpha,\tau_2}(\lambda(X;\tau_2))|^p]\}^{1/p}$ by

$$
\begin{aligned}
C\delta_q^{ws} + C\delta_q^{-(wq/p)}\{\mathbf{E}^*[\Delta(X)^q]\}^{1/p} &\le C\{\delta_q^{ws} + \delta_q^{-(wq/p)+(q/p)}\} \\
&= C\delta_q^{qs/(q+ps)} \le C\delta^{qrs/(q+ps)},
\end{aligned}
$$

obtaining the wanted result in (A.37).

We choose $\tau_1,\ldots,\tau_{N_1}$ such that for any $\tau \in \mathcal{T}_0$, there exists $j(\lambda) \in \{1,\ldots,N_1\}$ such that $|\tau - \tau_{j(\lambda)}| \le \Delta_{j(\lambda)}$ and $\|\Delta_{j(\lambda)}\|_{(1)} < \varepsilon^{(q+ps)/qrs}$. We also choose $\alpha_1,\ldots,\alpha_{N_2}$ such that for any $\alpha \in \mathcal{A}$, there exists $k(\alpha) \in \{1,\ldots,N_2\}$ such that $|\alpha(y) - \alpha_{k(\alpha)}(y)| \le \tilde{\Delta}_{k(\alpha)}(y)$ and $\|\tilde{\Delta}_{k(\alpha)}\|_p < \varepsilon$. Then note that

$$
\begin{aligned}
\Delta_{k(\alpha),j(\lambda)} &\triangleq \sup_{\alpha':\|\alpha'-\alpha_{k(\alpha)}\|_p<\varepsilon}\sup_{\tau':\|\tau'-\tau_{j(\lambda)}\|_{(1)}<\varepsilon^{(q+ps)/(qrs)}} \\
&\quad \times |g_{\alpha',\tau'}(\lambda(x;\tau')) - g_{\alpha_{k(\alpha)},\tau_{j(\lambda)}}(\lambda(x;\tau_{j(\lambda)}))| \\
&\le \sup_{\alpha':\|\alpha'-\alpha_{k(\alpha)}\|_p<\varepsilon}\sup_{\tau':\|\tau'-\tau_{j(\lambda)}\|_{(1)}<\varepsilon^{(q+ps)/(qrs)}} \\
&\quad \times |g_{\alpha',\tau'}(\lambda(x;\tau')) - g_{\alpha',\tau_{j(\lambda)}}(\lambda(x;\tau_{j(\lambda)}))| \\
&\quad + \sup_{\alpha':\|\alpha'-\alpha_{k(\alpha)}\|_p<\varepsilon}|g_{\alpha',\tau_{j(\lambda)}}(\lambda(x;\tau_{j(\lambda)})) - g_{\alpha_{k(\alpha)},\tau_{j(\lambda)}}(\lambda(x;\tau_{j(\lambda)}))| \\
&\triangleq J_{1,k(\alpha),j(\lambda)}(x) + J_{2,k(\alpha),j(\lambda)}(x), \quad \text{say.}
\end{aligned}
$$

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press

By (A.37), $\|J_{1,k(\alpha),j(\lambda)}\|_p \leq C\varepsilon$. On the other hand, it is easy to see that $\|J_{2,k(\alpha),j(\lambda)}\|_p \leq \|\tilde{\Delta}_{k(\alpha)}\|_p \leq \varepsilon$. Therefore, $\|\Delta_{k(\alpha),j(\lambda)}\|_p \leq C\varepsilon$, leading to the wanted result.

(ii) In the proof of (i), we replace δ_q by $\delta \triangleq 2\|\lambda_1 - \lambda_2\|_\infty$ and w by 1 and follow the steps in (i). Then the bound in (A.44) becomes

$$|\mathbf{E}[|V_1(1-A_2)|\,|\lambda_1(X) = \lambda_1(x)]| + |\mathbf{E}[|V_2(1-A_1)|\,|\lambda_2(X) = \lambda_2(x)]| + C\delta^s.$$

The first two terms disappear because they are bounded by

$$\begin{aligned} &C\mathbf{P}[|\lambda_2(X) - \lambda_2(x)| > \delta|\lambda_1(X) = \lambda_1(x)] + C\mathbf{P}[|\lambda_1(X) - \lambda_1(x)| > \delta|\lambda_2(X) = \lambda_2(x)] \\ &\quad \leq C\mathbf{P}[|\lambda_2(X) - \lambda_1(X) + \lambda_1(x) - \lambda_2(x)| > \delta|\lambda_1(X) = \lambda_1(x)] \\ &\qquad + C\mathbf{P}[|\lambda_1(X) - \lambda_2(X) + \lambda_2(x) - \lambda_1(x)| > \delta|\lambda_2(X) = \lambda_2(x)]. \end{aligned}$$

Both terms are equal to zero by the choice of δ. Following similar steps as in (i), we find that this gives $\|g_{\alpha,\lambda_1} \circ \lambda_1 - g_{\alpha,\lambda_2} \circ \lambda_2\|_\infty \leq C\delta^s$. The statement in (A.39) follows immediately from this result. ■

https://doi.org/10.1017/S0266466608080584 Published online by Cambridge University Press