Statistical Modeling with Spline Functions
Methodology and Theory

Mark H. Hansen
University of California at Los Angeles

Jianhua Z. Huang
University of Pennsylvania

Charles Kooperberg
Fred Hutchinson Cancer Research Center

Charles J. Stone
University of California at Berkeley

Young K. Truong
University of North Carolina at Chapel Hill

January 5, 2006

12
Extended Linear Modeling with Free Knot Splines

As we saw in Chapter 11, many statistical problems of theoretical and practical importance can effectively be treated within the framework of concave extended linear modeling. In that chapter, rates of convergence of maximum likelihood estimates were obtained when the estimating space does not depend on the data. In this chapter, we consider a collection $\mathbb{G}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \Gamma$, of linear estimation spaces having a common dimension that may vary with the sample size. For each fixed $\boldsymbol{\gamma}$, the maximum likelihood estimate is obtained. We will let the data pick which estimation space $\mathbb{G}_{\boldsymbol{\gamma}}$ to use. In particular, $\boldsymbol{\gamma}$ can be thought as the knot positions when the estimation space consists of spline functions, and our interest lies in choosing the knot positions using the data. This chapter contains the theory developed in Stone and Huang (2001a).

Section 12.1 of this chapter contains the basic setup and the main results on rates of convergence. In Section 12.2, we discuss the various properties of spaces of free knot splines and tensor products of such spaces that are needed to verify the conditions in our general results on rates of convergence. In Section 12.3 we verify the conditions in the main results in Section 12.1 in the contexts of density estimation and generalized regression, including ordinary regression as a special case. There, for simplicity, we restrict attention to the saturated model, so that $\eta^* = \eta$. We also restrict attention to spaces $\mathbb{G}_{\boldsymbol{\gamma}}$ that are tensor products of polynomial spline spaces. Section 12.4 contains the proofs of results in Section 12.2.

12.1 Main Results

12.1.1 Statement of Main Results

Consider a concave extended linear model specified by the log-likelihood $l(h, \boldsymbol{W})$ and model space $\mathbb{H}$. Let $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$ be a random sample of size n from the distribution of $\boldsymbol{W}$. When it is well defined, the (normalized) log-likelihood corresponding to this random sample is given by $\ell(h) = n^{-1} \sum_i l(h, \boldsymbol{W}_i)$. Let $\mathbb{G}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \Gamma$, be a collection of finite-dimensional linear subspaces of $\mathbb{H}$. We assume that the functions in each such space $\mathbb{G}_{\boldsymbol{\gamma}}$ are bounded and call $\mathbb{G}_{\boldsymbol{\gamma}}$ an estimation space. For each fixed $\boldsymbol{\gamma} \in \Gamma$, the maximum likelihood estimate is given by $\widehat{\eta}_{\boldsymbol{\gamma}} = \max_{g \in \mathbb{G}_{\boldsymbol{\gamma}}} \ell(g)$. We will let the data pick which estimation space to use. To be specific, we choose $\widehat{\boldsymbol{\gamma}} \in \Gamma$ such that $\ell(\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}}) = \max_{\boldsymbol{\gamma} \in \Gamma} \ell(\widehat{\eta}_{\boldsymbol{\gamma}})$. (Such a $\widehat{\boldsymbol{\gamma}}$ exists under mild conditions; see Lemma 12.1.1 below.) We will study the benefit of allowing the flexibility to pick estimation spaces among a big collection. Specifically we will study the rate of convergence of $\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*$, where, as in the previous chapter, η^* is the best approximation in $\mathbb{H}$ of the function of interest η.

In the above setup, we assume that $\mathbb{G}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \Gamma$, have the same dimension and that the index set Γ is a compact subset of $\mathbb{R}^J$ for some positive integer J. The dimension of $\mathbb{G}_{\boldsymbol{\gamma}}$, Γ and J are allowed to vary with the sample size n. For $\boldsymbol{\gamma} \in \Gamma$, set

$$N_n = \dim(\mathbb{G}_{\boldsymbol{\gamma}}),$$

$$A_{n\boldsymbol{\gamma}} = \sup_{g \in \mathbb{G}_{\boldsymbol{\gamma}}} \frac{\|g\|_\infty}{\|g\|} := \sup_{\substack{g \in \mathbb{G}_{\boldsymbol{\gamma}} \\ \|g\| \neq 0}} \frac{\|g\|_\infty}{\|g\|},$$

and

$$\rho_{n\boldsymbol{\gamma}} = \inf_{g \in \mathbb{G}_{\boldsymbol{\gamma}}} \|g - \eta^*\|_\infty.$$

Fix $n \geq 1$ and suppose that $A_n = \sup_{\boldsymbol{\gamma} \in \Gamma} A_{n\boldsymbol{\gamma}} < \infty$. Then the norms $\|\cdot\|$ and $\|\cdot\|_\infty$ are uniformly equivalent on $\mathbb{G}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \Gamma$, in the sense that $\|g\| \leq \|g\|_\infty \leq A_n \|g\|$ for $\boldsymbol{\gamma} \in \Gamma$ and $g \in \mathbb{G}_{\boldsymbol{\gamma}}$.

It follows from Proposition 11.1.1 that, under regularity conditions,

$$\|\widehat{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2 = O_P\Big(\rho_{n\boldsymbol{\gamma}}^2 + \frac{N_n}{n}\Big)$$

for each fixed $\boldsymbol{\gamma} \in \Gamma$. Let $\boldsymbol{\gamma}^*$ be such that $\rho_{n\boldsymbol{\gamma}^*} = \inf_{\boldsymbol{\gamma} \in \Gamma} \rho_{n\boldsymbol{\gamma}}$. (Such a $\boldsymbol{\gamma}^*$ exists under mild conditions; see Lemma 12.1.1 below.) Then

$$\|\widehat{\eta}_{\boldsymbol{\gamma}^*} - \eta^*\|^2 = O_P\Big(\rho_{n\boldsymbol{\gamma}^*}^2 + \frac{N_n}{n}\Big) = O_P\Big(\inf_{\boldsymbol{\gamma} \in \Gamma} \rho_{n\boldsymbol{\gamma}}^2 + \frac{N_n}{n}\Big).$$

Thus,

$$\inf_{\boldsymbol{\gamma} \in \Gamma} \|\widehat{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2 \leq \|\widehat{\eta}_{\boldsymbol{\gamma}^*} - \eta^*\|^2 = O_P\Big(\inf_{\boldsymbol{\gamma} \in \Gamma} \rho_{n\boldsymbol{\gamma}}^2 + \frac{N_n}{n}\Big).$$

It is natural to expect that, with $\boldsymbol{\gamma}$ estimated by $\widehat{\boldsymbol{\gamma}}$, the squared L_2 norm of the difference between the estimator and the target, i.e., $\|\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2$, will be not much larger than the ideal quantity $\inf_{\boldsymbol{\gamma}\in\Gamma} \|\widehat{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2$. Hence we hope that $\|\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2$ will be not much larger than $\inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}^2 + N_n/n$ in probability. This is confirmed by the following result.

Let $V_n = \bar{O}_P(b_n)$ mean that $\lim_n P(|V_n| \geq cb_n) = 0$ for some $c > 0$, where $b_n > 0$ for $n \geq 1$. Let $V_{n\boldsymbol{\gamma}} = O_P(b_{n\boldsymbol{\gamma}})$ uniformly over $\boldsymbol{\gamma} \in \Gamma$ mean that $\lim_{c\to\infty} \limsup_n P(|V_{n\boldsymbol{\gamma}}| \geq cb_{n\boldsymbol{\gamma}}$ for some $\boldsymbol{\gamma} \in \Gamma) = 0$, where $b_{n\boldsymbol{\gamma}} > 0$ for $n \geq 1$ and $\boldsymbol{\gamma} \in \Gamma$.

As in the previous chapter, it is enlightening to decompose the error into a stochastic part and a systematic part for each fixed $\boldsymbol{\gamma} \in \Gamma$:

$$\widehat{\eta}_{\boldsymbol{\gamma}} - \eta^* = (\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}) + (\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*),$$

where $\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}$ is referred to as the *estimation error* and $\bar{\eta}_{\boldsymbol{\gamma}} - \eta$ as the approximation error.

Proposition 12.1.1. *Suppose Conditions* 12.1.1–12.1.2 *and* 12.1.4–12.1.6 *hold and that* $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}\rho_{n\boldsymbol{\gamma}} = 0$ *and* $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}^2 N_n/n = 0$. *Then, for n sufficiently large, $\bar{\eta}_{\boldsymbol{\gamma}}$ exists uniquely for $\boldsymbol{\gamma} \in \Gamma$ and*

$$\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2 = O(\rho_{n\boldsymbol{\gamma}}^2)$$

uniformly over $\boldsymbol{\gamma} \in \Gamma$. Moreover, except on an event whose probability tends to zero as $n \to \infty$, $\widehat{\eta}_{\boldsymbol{\gamma}}$ exists uniquely for $\boldsymbol{\gamma} \in \Gamma$ and

$$\sup_{\boldsymbol{\gamma}\in\Gamma} \|\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}\|^2 = O_P\Big(\frac{N_n}{n}\Big).$$

Consequently,

$$\|\widehat{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2 = O_P\Big(\rho_{n\boldsymbol{\gamma}}^2 + \frac{N_n}{n}\Big)$$

uniformly over $\boldsymbol{\gamma} \in \Gamma$. In addition,

$$\|\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2 = O_P\Big(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}^2 + (\log n)\frac{N_n}{n}\Big).$$

In the previous theoretical results for fixed knot splines, the squared norms of the approximation error and the estimation error were shown to be bounded above by multiples of $\rho_{n\boldsymbol{\gamma}}^2$ and N_n/n, respectively. Here these results are shown to hold uniformly over the free knot sequences $\boldsymbol{\gamma} \in \Gamma$. Finally, combining the results for the approximation error and the estimation error and incorporating a corresponding result for the maximum likelihood estimation of the knot positions, we get an overall result. In particular, by allowing the knot positions to be selected by the data, we can achieve the best approximation rate among the collection of knot positions with a little inflation (an extra $\log n$ term) of the variability.

The proof of Proposition 12.1.1 is broken up into three theorems (Theorems 12.1.1–12.1.3) that will be given in the following subsections where technical conditions are stated explicitly. The technical conditions will be verified in the contexts of density estimation and generalized regression in Section 12.3 when $\mathbb{G}_{\boldsymbol{\gamma}}$ are spaces of tensor product splines. The $\log n$ term in the final result of this proposition plays an essential role in the proof of that result, but we do not know whether it is essential to the result itself.

12.1.2 Uniformity in Rates of Convergence

If $\boldsymbol{\gamma}$ is predetermined (independent of data) but allowed to increase with sample size, the rate of convergence of $\widehat{\boldsymbol{\gamma}}$ in the context of concave extended linear models is thoroughly treated in Chapter 11. We now show that the rates of convergence results in Theorems 11.2.1 and 11.2.2 hold uniformly in $\boldsymbol{\gamma} \in \Gamma$ if the sufficient conditions in these theorems hold in a uniform sense. Theorems 12.1.1 and 12.1.2 below are in parallel to Theorems 11.2.1 and 11.2.2 and can be proven by similar arguments.

Condition 12.1.1. *The best approximation η^* in $\mathbb{H}$ to η exists and there is a positive constant K_0 such that $\|\eta^*\|_\infty \le K_0$.*

Condition 12.1.2. *For each pair h_1, h_2 of bounded functions in $\mathbb{H}$, $\Lambda(h_1 + \alpha(h_2 - h_1))$ is twice continuously differentiable with respect to α. (i) For any positive constant K, there is a fixed positive number M such that if $h_1, h_2 \in \mathbb{H}$, $\|h_1\|_\infty \le K$, and h_2 is bounded, then*

$$\left|\frac{d}{d\alpha}\Lambda(h_1 + \alpha h_2)\big|_{\alpha=0}\right| \le M\|h_2\|.$$

(ii) For any positive constant K, there are fixed positive numbers M_1 and $M_2 \le M_1$ such that

$$-M_1\|h_2 - h_1\|^2 \le \frac{d^2}{d\alpha^2}\Lambda(h_1 + \alpha(h_2 - h_1)) \le -M_2\|h_2 - h_1\|^2$$

for $h_1, h_2 \in \mathbb{H}$ with $\|h_1\|_\infty \le K$ and $\|h_2\|_\infty \le K$ and $0 \le \alpha \le 1$.

Condition 12.1.1 is the same as Condition 11.2.1, which is restated here for convenience. Condition 12.1.2 strengthens Condition 11.2.2 by putting an adxditional requirement on the first derivative of $\Lambda(\cdot)$. The following result extends Theorem 11.2.1.

Theorem 12.1.1 (Approximation error). *Suppose Conditions* 12.1.1 and 12.1.2 *hold and that $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}\rho_{n\boldsymbol{\gamma}} = 0$. Let K_1 be a positive constant such that $K_1 > K_0$ with K_0 as in Condition* 12.1.1. *Then, for n sufficiently large, $\bar{\eta}_{\boldsymbol{\gamma}}$ exists uniquely and $\|\bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K_1$ for $\boldsymbol{\gamma} \in \Gamma$. Moreover, $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|^2 = O(\rho^2_{n\boldsymbol{\gamma}})$ uniformly over $\boldsymbol{\gamma} \in \Gamma$.*

The following two conditions are strengthened versions of Conditions 11.2.3 and 11.2.4.

Condition 12.1.3. *There is a positive constant* K_0 *such that, for* n *sufficiently large,* $\bar{\eta}_{\boldsymbol{\gamma}}$ *exists uniquely and* $\|\bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K_0$ *for* $\boldsymbol{\gamma} \in \Gamma$.

Condition 12.1.4. *For* $\boldsymbol{\gamma} \in \Gamma$ *and* $g_1, g_2 \in \mathbb{G}_{\boldsymbol{\gamma}}$, $\ell(g_1 + \alpha(g_2 - g_1))$ *is twice continuously differentiable with respect to* $\alpha \in [0,1]$. *(i) The following holds:*

$$\sup_{\boldsymbol{\gamma}\in\Gamma} \sup_{g \in \mathbb{G}_{\boldsymbol{\gamma}}} \frac{\left|\frac{d}{d\alpha}\ell(\bar{\eta}_{\boldsymbol{\gamma}} + \alpha g)|_{\alpha=0}\right|}{\|g\|} = O_P\left(\left(\frac{N_n}{n}\right)^{1/2}\right).$$

(ii) For any positive constant K, *there is a fixed positive number* M *such that*

$$\frac{d^2}{d\alpha^2}\ell(g_1 + \alpha(g_2 - g_1)) \le -M\|g_2 - g_1\|^2, \qquad 0 \le \alpha \le 1,$$

for $\boldsymbol{\gamma} \in \Gamma$ *and* $g_1, g_2 \in \mathbb{G}_{\boldsymbol{\gamma}}$ *with* $\|g_1\|_\infty \le K$ *and* $\|g_2\|_\infty \le K$, *except on an event whose probability tends to zero as* $n \to \infty$.

The following result extends Theorem 11.2.2.

Theorem 12.1.2 (Estimation error). *Suppose Conditions* 12.1.3 *and* 12.1.4 *hold and that* $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A^2_{n\boldsymbol{\gamma}} N_n/n = 0$. *Let* K_1 *be a positive constant such that* $K_1 > K_0$ *with* K_0 *as in Condition* 11.2.3. *Then* $\widehat{\eta}_{\boldsymbol{\gamma}}$ *exists uniquely and* $\|\widehat{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K_1$ *for* $\boldsymbol{\gamma} \in \Gamma$, *except on an event whose probability tends to zero as* $n \to \infty$. *Moreover,* $\sup_{\boldsymbol{\gamma}\in\Gamma} \|\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}\|^2 = O_P(N_n/n)$.

12.1.3 Adaptive Parameter Selection

Condition 12.1.5. *For* $K < \infty$, *the set* $\{(\boldsymbol{\gamma}, g) : \boldsymbol{\gamma} \in \Gamma, g \in \mathbb{G}_{\boldsymbol{\gamma}},\ \text{and}\ \|g\|_\infty \le K\}$ *is compact and* $\ell(\cdot)$ *is continuous on this set.*

When $\mathbb{G}_{\boldsymbol{\gamma}}$ are spaces of tensor product splines in Section 12.2, the first part of Condition 12.1.5 follows from Lemmas 2.1 and 4.1 of Chapter 5 of DeVore and Lorentz (1993). Under the further restriction to density estimation and generalized regression in Section 12.3, the second part of Condition 12.1.5 follows from the corresponding explicit forms of the log-likelihood function.

Lemma 12.1.1. *Suppose Condition* 12.1.5 *holds. Then there is a* $\boldsymbol{\gamma}^* \in \Gamma$ *such that* $\rho_{n\boldsymbol{\gamma}^*} = \inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}$. *Moreover, on the event that* $\widehat{\eta}_{\boldsymbol{\gamma}}$ *exists uniquely and* $\|\widehat{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K_1$ *for* $\boldsymbol{\gamma} \in \Gamma$, *where* K_1 *is a positive constant, there is a* $\widehat{\boldsymbol{\gamma}} \in \Gamma$ *such that* $\ell(\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}}) = \sup_{\boldsymbol{\gamma}\in\Gamma} \ell(\widehat{\eta}_{\boldsymbol{\gamma}})$.

Proof. Given $\boldsymbol{\gamma} \in \Gamma$, choose $g_{\boldsymbol{\gamma}} \in \mathbb{G}_{\boldsymbol{\gamma}}$ such that $\|g_{\boldsymbol{\gamma}} - \eta^*\|_\infty = \rho_{\boldsymbol{\gamma}}$. By Condition 12.1.5, we can choose $\boldsymbol{\gamma}_\nu \in \Gamma$ such that $\boldsymbol{\gamma}_\nu \to \boldsymbol{\gamma}^* \in \Gamma$, $\rho_{\boldsymbol{\gamma}_\nu} \to \inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{\boldsymbol{\gamma}}$, and $\|g_{\boldsymbol{\gamma}_\nu} - g^*\|_\infty \to 0$ as $\nu \to \infty$, where $g^* \in \mathbb{G}_{\boldsymbol{\gamma}^*}$. Then $\|g^* - \eta^*\|_\infty = \inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{\boldsymbol{\gamma}}$, so $\boldsymbol{\gamma}^*$ has its desired property.

It follows from Condition 12.1.5 that, on the indicated event, we can choose $\boldsymbol{\gamma}_\nu \in \Gamma$ such that $\boldsymbol{\gamma}_\nu \to \widehat{\boldsymbol{\gamma}} \in \Gamma$, $\ell(\widehat{\eta}_{\boldsymbol{\gamma}_\nu}) \to \sup_{\boldsymbol{\gamma}\in\Gamma} \ell(\widehat{\eta}_{\boldsymbol{\gamma}})$, and $\widehat{\eta}_{\boldsymbol{\gamma}_\nu} \to g$ as $\nu \to \infty$, where $g \in \mathbb{G}_{\widehat{\boldsymbol{\gamma}}}$. Since $\ell(\cdot)$ is continuous, $\ell(g) = \sup_{\boldsymbol{\gamma}\in\Gamma} \ell(\widehat{\eta}_{\boldsymbol{\gamma}})$, $g = \widehat{\eta}_{\widehat{\boldsymbol{\gamma}}}$ and hence $\widehat{\boldsymbol{\gamma}}$ has its desired property. □

Let $V_{n\boldsymbol{\gamma}} = \bar{O}_P(b_{n\boldsymbol{\gamma}})$ uniformly over $\boldsymbol{\gamma} \in \Gamma$ mean that $\lim_n P(|V_{n\boldsymbol{\gamma}}| \ge cb_{n\boldsymbol{\gamma}}$ for some $\boldsymbol{\gamma} \in \Gamma) = 0$ for some $c > 0$, where $b_{n\boldsymbol{\gamma}} > 0$ for $n \ge 1$ and $\boldsymbol{\gamma} \in \Gamma$.

Condition 12.1.6.

$$\text{(i) } |\ell(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \ell(\eta^*) - [\Lambda(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \Lambda(\eta^*)]| = O_P\left(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}^2 + \frac{N_n}{n}\right)$$

and

$$\text{(ii) } |\ell(\bar{\eta}_{\boldsymbol{\gamma}}) - \ell(\eta^*) - [\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta^*)]| = \bar{O}_P\left((\log^{1/2} n)\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|\left(\frac{N_n}{n}\right)^{1/2} + (\log n)\frac{N_n}{n}\right)$$

uniformly over $\boldsymbol{\gamma} \in \Gamma$.

In Section 12.3, we will verify that Condition 12.1.6 holds under reasonable conditions in the contexts of density estimation and generalized regression. There, we will actually verify a slight strengthening of the second property of Condition 12.1.6:

$$|\ell(\bar{\eta}_{\boldsymbol{\gamma}}) - \ell(\eta^*) - [\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta^*)]| = \bar{O}_P\left((\log^{1/2} n)\left[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|\left(\frac{N_n}{n}\right)^{1/2} + \frac{N_n}{n}\right]\right)$$

uniformly over $\boldsymbol{\gamma} \in \Gamma$.

We have the decomposition

$$\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^* = (\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \bar{\eta}_{\widehat{\boldsymbol{\gamma}}}) + (\bar{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*).$$

Note that Theorem 12.1.2 implies that $\|\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}} - \bar{\eta}_{\widehat{\boldsymbol{\gamma}}}\|^2 = O_P(N_n/n)$, which together with the following theorem yields Proposition 12.1.1.

Theorem 12.1.3 (Parameter selection). *Suppose Conditions* 12.1.1–12.1.6 *hold and that* $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}\rho_{n\boldsymbol{\gamma}} = 0$ *and* $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}^2 N_n/n = 0$. *Then* $\|\bar{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2 = O_P(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}^2) + \bar{O}_P((\log n)N_n/n)$.

Proof. We first show that

$$\ell(\widehat{\eta}_{\boldsymbol{\gamma}}) - \ell(\bar{\eta}_{\boldsymbol{\gamma}}) = O_P\left(\frac{N_n}{n}\right) \qquad \text{uniformly in } \boldsymbol{\gamma} \in \Gamma. \tag{12.1.1}$$

Write

$$f(\alpha) = \ell(\bar{\eta}_{\boldsymbol{\gamma}} + \alpha(\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}})), \qquad \boldsymbol{\gamma} \in \Gamma.$$

By Condition 12.1.4, $f''(\alpha) \le 0$ (except on an event whose probability tends to zero as $n \to \infty$). Thus,

$$0 \le \ell(\widehat{\eta}_{\boldsymbol{\gamma}}) - \ell(\bar{\eta}_{\boldsymbol{\gamma}}) = f(1) - f(0) = f'(0) + \int_0^1 (1-\alpha) f''(\alpha)\, d\alpha \le f'(0).$$

On the other hand, by Condition 12.1.4(i) and Theorem 12.1.2,

$$\begin{aligned} f'(0) &= \frac{d}{d\alpha}\ell(\bar{\eta}_{\boldsymbol{\gamma}} + \alpha(\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}))\Big|_{\alpha=0} \\ &= O_P\Big(\Big(\frac{N_n}{n}\Big)^{1/2}\Big)\|\widehat{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\boldsymbol{\gamma}}\| = O_P\Big(\frac{N_n}{n}\Big) \end{aligned}$$

uniformly in $\boldsymbol{\gamma} \in \Gamma$. The desired result follows.

By Theorem 12.1.1, $\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}$ is bounded. Thus, it follows from Lemma 11.2.1 that, for some positive constant M,

$$M\|\bar{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2 \le \Lambda(\eta^*) - \Lambda(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}).$$

Since $\boldsymbol{\gamma}^* \in \Gamma$ satisfies $\rho_{n\boldsymbol{\gamma}^*} = \inf_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}$, $\|\bar{\eta}_{\boldsymbol{\gamma}^*} - \eta^*\|^2 = O(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho^2_{n\boldsymbol{\gamma}})$ by Theorem 12.1.1. We have the decomposition

$$\begin{aligned} \Lambda(\eta^*) - \Lambda(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}) &= \Lambda(\eta^*) - \Lambda(\bar{\eta}_{\boldsymbol{\gamma}^*}) + \Lambda(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \Lambda(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}) \\ &= I_1 + I_2 - I_3 + I_4, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \Lambda(\eta^*) - \Lambda(\bar{\eta}_{\boldsymbol{\gamma}^*}), \\ I_2 &= \Lambda(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \Lambda(\eta^*) - [\ell(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \ell(\eta^*)], \\ I_3 &= \Lambda(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}) - \Lambda(\eta^*) - [\ell(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}) - \ell(\eta^*)], \\ I_4 &= \ell(\bar{\eta}_{\boldsymbol{\gamma}^*}) - \ell(\bar{\eta}_{\widehat{\boldsymbol{\gamma}}}). \end{aligned}$$

Note that $I_1 = O(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho^2_{n\boldsymbol{\gamma}})$ by Theorem 12.1.1 and Lemma 11.2.1. The terms I_2 and I_3 can be bounded using Condition 12.1.6. Moreover, by using (12.1.1) and $\ell(\widehat{\eta}_{\boldsymbol{\gamma}^*}) \le \ell(\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}})$ [which follows from the definition of $\widehat{\boldsymbol{\gamma}}$], we get that

$$I_4 = \ell(\widehat{\eta}_{\boldsymbol{\gamma}^*}) - \ell(\widehat{\eta}_{\widehat{\boldsymbol{\gamma}}}) + O_P\Big(\frac{N_n}{n}\Big) \le O_P\Big(\frac{N_n}{n}\Big).$$

Hence,

$$\begin{aligned} \|\bar{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|^2 \le \bar{O}_P\bigg((\log^{1/2} n)\|\bar{\eta}_{\widehat{\boldsymbol{\gamma}}} - \eta^*\|\Big(\frac{N_n}{n}\Big)^{1/2} + (\log n)\frac{N_n}{n}\bigg) \\ + O_P\Big(\inf_{\boldsymbol{\gamma}\in\Gamma} \rho^2_{n\boldsymbol{\gamma}} + \frac{N_n}{n}\Big). \end{aligned} \tag{12.1.2}$$

Observe that, for positive numbers B and C, $z^2 \le Bz + C$ implies that $2z^2 \le (B^2 + z^2) + 2C$ and hence that $z^2 \le B^2 + 2C$. Therefore (12.1.2) yields the desired result. □

12.2 Free Knot Splines and Their Tensor Products

In this section we will develop some properties of spaces of free knot splines and tensor products of such spaces, which will be used in Section 12.3 to verify Conditions 12.1.4 and 12.1.6.

For $1 \le l \le L$, let $\mathcal{U}_l = [a_l, b_l]$ be a compact subinterval of $\mathbb{R}$ having positive length $b_l - a_l$ and let $\mathcal{U}$ denote the Cartesian product of $\mathcal{U}_1, \dots, \mathcal{U}_L$. For each l, let m_l be an integer with $m_l \ge 2$ and J_l be a positive integer, and let γ_{lj}, $1 \le j \le J_l$, be such that $a < \gamma_{l1} \le \cdots \le \gamma_{lJ_l} < b$, and $\gamma_{l,j-1} > \gamma_{l,j-m}$ for $2 \le j \le J_l + m_l$, where $\gamma_{lj} = a$ for $1 - m_l \le j \le 0$ and $\gamma_{lj} = b$ for $J_l + 1 \le j \le J_l + m_l$. Let $\mathbb{G}_{l\boldsymbol{\gamma}_l}$ be the space of polynomial splines of order m_l (degree $m_l - 1$) on $\mathcal{U}_l$ with the interior knot sequence $\boldsymbol{\gamma}_l = (\gamma_{l1}, \dots, \gamma_{lJ_l})$, whose dimension $J_l + m_l$ is denoted by N_{nl} to indicate its possible dependence on the sample size n. For $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_L)$, let $\mathbb{G}_{\boldsymbol{\gamma}}$ be the tensor product of $\mathbb{G}_{l\boldsymbol{\gamma}_l}$, $1 \le l \le L$, which has dimension $N_n = \prod_l N_{nl}$.

For $1 \le l \le L$, let $\bar{M}_l \ge 1$ be a fixed positive number, and let Γ_l denote the collection of free knot sequences $\boldsymbol{\gamma}_l = (\boldsymbol{\gamma}_{l1}, \dots, \boldsymbol{\gamma}_{lJ_l})$ on $\mathcal{U}_l$ such that

$$\frac{\gamma_{l,j_2-1} - \gamma_{l,j_2-m_l}}{\gamma_{l,j_1-1} - \gamma_{l,j_1-m_l}} \le \bar{M}_l, \qquad 2 \le j_1, j_2 \le J_l + m_l, \tag{12.2.1}$$

where $\gamma_{l,1-m_l} = \cdots = \gamma_{l0} = a$ and $\gamma_{l,J_l+1} = \cdots = \gamma_{l,J_l+m_l} = b$. Let Γ denote the Cartesian product of Γ_l, $1 \le l \le L$, which can be viewed as a subset of $\mathbb{R}^J$ with $J = \sum_l J_l$. We consider the use of the collection $\mathbb{G}_{\boldsymbol{\gamma}}, \boldsymbol{\gamma} \in \Gamma$, in fitting an extended linear model. Such a collection of free knot splines has some properties we will list below. (The proofs will be given in Section 12.4.) In the technical arguments, we need to approximate Γ by a finite subset of a larger set $\widetilde{\Gamma}$, which is defined in the same way as Γ, but with $\bar{M}_l$ in (12.2.1) replaced by the larger constant $3\bar{M}_l$.

Let ψ denote the uniform distribution on $\mathcal{U}$ and let $\mathrm{vol}(\mathcal{U})$ denote the volume of $\mathcal{U}$. Let $\mathbb{H}$ denote the space of (real-valued) functions on $\mathcal{U}$ that are square-integrable with respect to ψ, and let $\langle \cdot, \cdot \rangle_\psi$ and $\| \cdot \|_\psi$ denote the inner product and norm on $\mathbb{H}$ given by

$$\langle h_1, h_2 \rangle_\psi = \int_{\mathcal{U}} h_1(\boldsymbol{u}) h_2(\boldsymbol{u}) \, \psi(d\boldsymbol{u}) = \frac{1}{\mathrm{vol}(\mathcal{U})} \int_{\mathcal{U}} h_1(\boldsymbol{u}) h_2(\boldsymbol{u}) \, d\boldsymbol{u}$$

and $\|h\|_\psi^2 = \langle h, h \rangle_\psi$.

Let $\boldsymbol{U}$ denote a $\mathcal{U}$-valued random variable that is a transform (function) of $\boldsymbol{W}$ (for example, $\boldsymbol{W} = (\boldsymbol{X}, Y)$ and $\boldsymbol{U} = \boldsymbol{X}$). Partly for simplicity, we consider the theoretical inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ on $\mathbb{H}$ given by $\langle h_1, h_2 \rangle = E[h_1(\boldsymbol{U}) h_2(\boldsymbol{U})]$ and $\|h\|^2 = \langle h, h \rangle = E[h^2(\boldsymbol{U})]$. Define the empirical inner product and empirical norm by $\langle h_1, h_2 \rangle_n = E_n(h_1 h_2) = n^{-1} \sum_i h_1(\boldsymbol{U}_i) h_2(\boldsymbol{U}_i)$ and $\|h\|_n^2 = \langle h, h \rangle_n = n^{-1} \sum_i h^2(\boldsymbol{U}_i)$.

Condition 12.2.1. *The random variable* $\boldsymbol{U}$ *has a density function* $f_{\boldsymbol{U}}$ *such that* $M_1/\mathrm{vol}(\mathcal{U}) \le f_{\boldsymbol{U}} \le M_2/\mathrm{vol}(\mathcal{U})$ *on* $\mathcal{U}$*, where* M_1 *and* M_2 *are fixed positive numbers.*

It follows from Condition 12.2.1 that $M_1 \le 1 \le M_2$ and

$$M_1\|h\|_\psi^2 \le \|h\|^2 \le M_2\|h\|_\psi^2, \qquad h \in \mathbb{H}. \tag{12.2.2}$$

Let $|\cdot|_\infty$ denote the l_∞ norm on any Euclidean space. Let ζ denote the metric on $\mathbb{R}^J$ given by $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) = \max_l 9\bar{M}_l N_{nl}|\boldsymbol{\gamma}_l - \widetilde{\boldsymbol{\gamma}}_l|_\infty/(b_l - a_l)$. The following lemmas will be proved in Section 12.4.

Lemma 12.2.2. *Let* $0 < \epsilon \le 1/2$ *and let* K *be a positive integer. There is a positive constant* M *and there are subsets* Ξ_k*,* $0 \le k \le K$*, of* $\widetilde{\Gamma}$ *such that*

$$\#(\Xi_k) \le \bigl(M\epsilon^{-k}\bigr)^{N_n}, \qquad 1 \le k \le K;$$

every point in Γ *is within* ϵ^K *of some point in* Ξ_K *(in* ζ *distance); and, for* $1 \le k \le K$*, every point in* Ξ_k *is within* ϵ^{k-1} *of some point in* Ξ_{k-1}*.*

Let $0 < \epsilon \le 1/2$ and let Ξ_k, $0 \le k \le K$ be as in Lemma 12.2.2. Given $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, set $\mathbb{B}_{\boldsymbol{\gamma}} = \{g \in \mathbb{G}_{\boldsymbol{\gamma}} : \|g\| \le 1\}$. Let k be a nonnegative integer. If $k = 0$, set $\mathbb{B}_{\boldsymbol{\gamma}k} = \{0\}$; otherwise, let $\mathbb{B}_{\boldsymbol{\gamma}k}$ be a maximal subset of $\mathbb{B}_{\boldsymbol{\gamma}}$ such that any two functions in $\mathbb{B}_{\boldsymbol{\gamma}k}$ are at least ϵ^k apart in the norm $\|\ \|$. Then $\min_{\widetilde{g}\in\mathbb{B}_{\boldsymbol{\gamma}k}} \|g - \widetilde{g}\| \le \epsilon^k$ for $g \in \mathbb{B}_{\boldsymbol{\gamma}}$. Moreover,

$$\#(\mathbb{B}_{\boldsymbol{\gamma}k}) \le \Bigl(\frac{1 + \epsilon^k/2}{\epsilon^k/2}\Bigr)^{N_n} \le (3\epsilon^{-k})^{N_n}.$$

Set $\mathbb{B}_k = \cup_{\boldsymbol{\gamma}\in\Xi_k}\mathbb{B}_{\boldsymbol{\gamma}k}$. Then, by Lemma 12.2.2,

$$\#(\mathbb{B}_k) \le (M'\epsilon^{-2k})^{N_n}, \qquad 1 \le k \le K, \tag{12.2.3}$$

for some constant $M' \ge 1$. Also, set $\mathbb{B} = \{g \in \cup_{\boldsymbol{\gamma}\in\Gamma}\mathbb{G}_{\boldsymbol{\gamma}} : \|g\| \le 1\} = \cup_{\boldsymbol{\gamma}\in\Gamma}\mathbb{B}_{\boldsymbol{\gamma}}$ and $\widetilde{\mathbb{B}} = \{g \in \cup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}}\mathbb{G}_{\boldsymbol{\gamma}} : \|g\| \le 1\} = \cup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}}\mathbb{B}_{\boldsymbol{\gamma}}$.

Lemma 12.2.3. *Suppose, for a given positive integer* n*, that* $\bar{\eta}_{\boldsymbol{\gamma}}$ *exists uniquely and is bounded for* $\boldsymbol{\gamma} \in \widetilde{\Gamma}$*, and that* $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|$ *is a continuous function of* $\boldsymbol{\gamma} \in \widetilde{\Gamma}$*. There is a positive constant* M *such that, for* $0 < \epsilon \le 1$*, there is a subset* $\widetilde{\Gamma}'$ *of* $\widetilde{\Gamma}$ *such that*

$$\#(\widetilde{\Gamma}') \le \exp\bigl(M[\log(2/\epsilon)]N_n\bigr)$$

and every point $\boldsymbol{\gamma}$ *in* Γ *is within* ϵ *(in* ζ *distance) of some point* $\widetilde{\boldsymbol{\gamma}}$ *in* $\widetilde{\Gamma}'$ *such that* $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}} - \eta^*\| \le \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|$*.*

The condition that $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|$ is a continuous function of $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, which is used in the above lemma, follows from the first conclusion of Lemma 12.2.6.

Lemma 12.2.4. *Suppose Condition* 12.2.1 *holds. There is a positive constant* M *such that*

$$\|g\|_\infty \le MN_n^{1/2}\|g\|, \qquad \boldsymbol{\gamma} \in \widetilde{\Gamma} \text{ and } g \in \mathbb{G}_{\boldsymbol{\gamma}}. \tag{12.2.4}$$

Lemma 12.2.5. *There are positive numbers* M_1 *and* M_2 *such that, for* $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ *and* $g \in \mathbb{G}_{\boldsymbol{\gamma}}$*, there is a function* $\widetilde{g} \in \mathbb{G}_{\widetilde{\boldsymbol{\gamma}}}$ *such that* $\|\widetilde{g}\| \le \|g\|$, $\|\widetilde{g} - g\| \le M_1\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})\|g\|$ *and* $\|\widetilde{g} - g\|_\infty \le M_2\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})\|g\|_\infty$*. Suppose Condition* 12.2.1 *holds and that* $\lim_n N_n^2/n = 0$*. Then there is a positive number* M_3 *and an event* Ω_n *such that* $\lim_n P(\Omega_n) = 1$ *and the functions* $\widetilde{g}$ *above can be chosen to satisfy the additional property that* $\|g - \widetilde{g}\|_n \le M_3\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})\|g\|$ *on* Ω_n *for* $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ *and* $g \in \mathbb{G}_{\boldsymbol{\gamma}}$.

Lemma 12.2.6. *Suppose Condition* 12.1.2 *holds. Let* K *be a positive number. There are positive numbers* M_1 *and* M_2 *such that if* $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$, $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \le 1$, $\|\bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K$*, and* $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\| \le K$*, then* $\|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\| \le M_1[\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})]^{1/2}$ *and* $\|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \le M_2N_n^{1/2}[\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})]^{1/2}$*. Suppose in addition Condition* 12.2.1 *holds and that* $\lim_n N_n^2/n = 0$*. Then there is an event* Ω_n *such that* $\lim_n P(\Omega_n) = 1$ *and* $\|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_n \le M_1[\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})]^{1/2}$ *for* $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ *on* Ω_n.

12.3 Verification of Technical Conditions

In this section we verify Conditions 12.1.2, 12.1.4 and 12.1.6 using primitive assumptions in some specific statistical contexts. For simplicity, we focus on two contexts: density estimation in Section 12.3.2 and generalized regression, which includes ordinary regression as a special case, in Section 12.3.3. Again for simplicity, we also restrict attention to the saturated model (that is, there is no structural assumption and $\mathbb{H}$ is the collection of all square integrable functions on $\mathcal{U}$), so that $\eta^* = \eta$. Thus Condition 12.1.1 amounts to the assumption that η is bounded. The case of unsaturated models can be treated similarly at the expense of more complicated notation.

Throughout this section, we take $\mathbb{G}_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, to be tensor product free knot spline spaces as defined in Section 12.2, where $\widetilde{\Gamma}$ is the collection of knot configurations satisfying (12.2.1) with M replaced by the larger constant $\widetilde{M}$. It follows from (12.2.4) that $A_{n\boldsymbol{\gamma}} \le MN_n^{1/2}$ for some constant M, where N_n is the common dimension of $\mathbb{G}_{\boldsymbol{\gamma}}$. Thus the requirements $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}\rho_{n\boldsymbol{\gamma}} = 0$ and $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} A_{n\boldsymbol{\gamma}}^2 N_n/n = 0$, which are used in Proposition 12.1.1, reduce to $\lim_n \sup_{\boldsymbol{\gamma}\in\Gamma} \rho_{n\boldsymbol{\gamma}}N_n^{1/2} = 0$ and $\lim_n N_n^2/n = 0$ respectively.

Condition 12.3.1. $N_n^{-(c-1/2)} \lesssim \log^{-1/2} n$ *and* $N_n^c \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}} \rho_{n\gamma} \lesssim 1$ *for some* $c > 1/2$.

12.3.1 Preliminary Lemmas

Lemma 12.3.7. *Suppose Condition* 12.2.1 *holds and that* $\lim_n N_n^2/n = 0$. *Then*

$$\sup_{\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}} \sup_{f\in\mathbb{G}_{\boldsymbol{\gamma}}} \sup_{g\in\mathbb{G}_{\widetilde{\boldsymbol{\gamma}}}} \frac{|\langle f,g\rangle_n - \langle f,g\rangle|}{\|f\|\|g\|} = o_P(1).$$

Consequently, except on an event whose probability tends to zero as $n\to\infty$,

$$\tfrac{1}{2}\|g\|^2 \le \|g\|_n^2 \le 2\|g\|^2, \qquad \boldsymbol{\gamma}\in\widetilde{\Gamma} \text{ and } g\in\mathbb{G}_{\boldsymbol{\gamma}}.$$

This lemma extends Lemma 11.4.7, which applies to fixed knot splines and other such linear estimation spaces, except that Condition 12.2.1 is not required in Lemma 11.4.7.

Proof of Lemma 12.3.7. It suffices to verify the lemma with $\widetilde{\Gamma}$ replaced by Γ. Let $0<\delta\le 1/4$, let $0<t<\infty$, let $K=K_n$ be a positive integer to be specified later, and let $\mathbb{B}$ and let Ξ_k and $\mathbb{B}_k$, $0\le k\le K$, be as in Lemma 12.2.2 and the following paragraph with $\epsilon=\delta$. We will apply Lemma 11.4.6 with $s=(f,g)$, $V_s=\langle f,g\rangle_n-\langle f,g\rangle=(E_n-E)(fg)$, $\mathbb{S}=\{(f,g): f,g\in\mathbb{B}\}$, $\mathbb{S}_k=\{(f,g): f,g\in\mathbb{B}_k\}$ for $0\le k\le K$, and $\Omega^c=\emptyset$. It follows from (12.2.3) that

$$\#(\mathbb{S}_k)\le (M'\delta^{-2k})^{2N_n}, \qquad 1\le k\le K,$$

and hence that (11.4.5) holds with $C_3=1$ and any $C_4\ge 4\log(M'\delta^{-1})N_n$.

Suppose Condition 12.2.1 holds, let $0<\epsilon=\delta\le 1/4$, let k be a positive integer, let $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}$ with $\zeta(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}})\le\delta^{k-1}$, and let $g\in\mathbb{B}_{\boldsymbol{\gamma}}$. Then, by (12.2.4) and Lemma 12.2.5, there is a function $g'\in\mathbb{B}_{\widetilde{\boldsymbol{\gamma}}}$ such that $\|g-g'\|\le M_1\delta^{k-1}$ and $\|g-g'\|_\infty\le MM_2N_n^{1/2}\delta^{k-1}$. Also, there is a function $\widetilde{g}\in\mathbb{B}_{\widetilde{\boldsymbol{\gamma}},k-1}$ such that $\|g'-\widetilde{g}\|\le\delta^{k-1}$ and hence $\|g'-\widetilde{g}\|_\infty\le MN_n^{1/2}\delta^{k-1}$. Observe that $\|g\|_\infty\le c_1N_n^{1/2}$, $\|\widetilde{g}\|_\infty\le c_1N_n^{1/2}$, $\|g-\widetilde{g}\|\le c_2\delta^{k-1}$, and $\|g-\widetilde{g}\|_\infty\le c_3N_n^{1/2}\delta^{k-1}$, where $c_1=M$, $c_2=M_1+1$, and $c_3=M(M_2+1)$.

Let k be a positive integer, and let $f,\widetilde{f},g,\widetilde{g}$ be functions on $\mathcal{U}$ such that $\|\widetilde{f}\|_\infty\le c_1N_n^{1/2}$, $\|f-\widetilde{f}\|\le c_2\delta^{k-1}$, $\|f-\widetilde{f}\|_\infty\le c_3N_n^{1/2}\delta^{k-1}$, $\|g\|_\infty\le c_1N_n^{1/2}$, $\|g-\widetilde{g}\|\le c_2\delta^{k-1}$, and $\|g-\widetilde{g}\|_\infty\le c_3N_n^{1/2}\delta^{k-1}$. Then

$$\|fg-\widetilde{f}\widetilde{g}\|_\infty\le\|f-\widetilde{f}\|_\infty\|g\|_\infty+\|\widetilde{f}\|_\infty\|g-\widetilde{g}\|_\infty\le 2c_1c_3N_n\delta^{k-1},$$

so $|(E_n-E)(fg-\widetilde{f}\widetilde{g})|\le 4c_1c_3N_n\delta^{k-1}$. Moreover,

$$\begin{aligned}\operatorname{var}(fg-\widetilde{f}\widetilde{g}) &\le 2\operatorname{var}((f-\widetilde{f})g)+2\operatorname{var}(\widetilde{f}(g-\widetilde{g}))\\ &\le 2\|g\|_\infty^2\|f-\widetilde{f}\|^2+2\|\widetilde{f}\|_\infty^2\|g-\widetilde{g}\|^2\\ &\le 4c_1^2c_2^2N_n\delta^{2(k-1)}\end{aligned}$$

Since $0 < 2\delta \le 1$, it now follows from Bernstein's inequality (11.4.2) that, for $t > 0$,

$$P\Big(|(E_n - E)(fg - \widetilde{f}\widetilde{g})| \ge t2^{-(k-1)}\Big) \le 2\exp\left(-\frac{nt^2(2\delta)^{-(k-1)}}{8c_1[c_1c_2^2 + tc_3]N_n}\right). \tag{12.3.1}$$

Let K be such that $4c_1c_3N_n\delta^K \le t$. Given $f, g \in \mathbb{B}$, let $\widetilde{f}, \widetilde{g} \in \mathbb{B}_K$ be such that $\|f - \widetilde{f}\|_\infty \le c_3N_n^{1/2}\delta^K$ and $\|g - \widetilde{g}\|_\infty \le c_3N_n^{1/2}\delta^K$. Then $\|fg - \widetilde{f}\widetilde{g}\|_\infty \le 2c_1c_3N_n\delta^K$, so $|(E_n - E)(fg - \widetilde{f}\widetilde{g})| \le 4c_1c_3N_n\delta^K \le t$. Consequently, (11.4.4) holds with $C_1 = t$ and $C_2 = 0$.

Let $1 \le k \le K$. For $f, g \in \mathbb{B}_k$, let $\widetilde{f}, \widetilde{g} \in \mathbb{B}_{k-1}$ be such that $\|f - \widetilde{f}\| \le c_2\delta^{k-1}$, $\|f - \widetilde{f}\|_\infty \le c_3N_n^{1/2}\delta^{k-1}$, $\|g - \widetilde{g}\| \le c_2\delta^{k-1}$, and $\|g - \widetilde{g}\|_\infty \le c_3N_n^{1/2}\delta^{k-1}$. Since $N_n = o(n^{1/2})$, we now conclude from (12.3.1) that (11.4.6) holds with $C_5 = t$, $C_6 = 2$, and

$$C_4 = \frac{nt^2}{16c_1[c_1c_2^2 + tc_3]N_n} \ge 4\log(M'\delta^{-1})N_n$$

for n sufficiently large. It now follows from Lemma 11.4.6 that, for n sufficiently large,

$$P\Big(\sup_{\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \Gamma} \sup_{f \in \mathbb{B}_{\boldsymbol{\gamma}}} \sup_{g \in \mathbb{B}_{\widetilde{\boldsymbol{\gamma}}}} |\langle f, g\rangle_n - \langle f, g\rangle| \ge 3t\Big) \le \frac{32c_1[c_1c_2^2 + tc_3]N_n}{nt^2},$$

which tends to zero as $n \to \infty$. Since t can be made arbitrarily small, the first conclusion of the lemma is valid, from which the second conclusion follows easily. □

Lemma 12.3.8. *Suppose Condition* 12.2.1 *holds and that* $\lim_n N_n^2/n = 0$, *and let* h_n *be uniformly bounded functions on* $\mathcal{U}$. *Then*

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \sup_{g \in \mathbb{G}_{\boldsymbol{\gamma}}} \frac{|\langle h_n, g\rangle_n - \langle h_n, g\rangle|}{\|g\|} = O_P\left(\left(\frac{N_n}{n}\right)^{1/2}\right).$$

Proof. The proof of this result is a slight simplification of the proof of Lemma 12.3.7. □

The next, obviously valid, lemma is useful in verifying the second property of Condition 12.1.6 in a variety of contexts.

Lemma 12.3.9. *Let* $C_1, \ldots, C_4$ *be fixed positive numbers with* $C_3 > 1$. *Let* $A_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, *be positive numbers that depend on* n, *and let* $V_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, *be random variables that depend on* n. *Suppose that, for* n *sufficiently large,*

$P(|V_{\boldsymbol{\gamma}}| \ge C_1 A_{\boldsymbol{\gamma}}) \le C_2 \exp\big(-2C_3 N_n \log n\big)$ for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$. Let $\widetilde{\Gamma}''$ be a subset of $\widetilde{\Gamma}$ such that

$$\#(\widetilde{\Gamma}'') \le \exp\big(C_3 N_n \log n\big) \qquad \text{for } n \text{ sufficiently large.} \tag{12.3.2}$$

Suppose that, except on an event whose probability tends to zero as $n \to \infty$, for every point $\boldsymbol{\gamma} \in \Gamma$, there is a point $\widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}''$ such that $A_{\widetilde{\boldsymbol{\gamma}}} \le A_{\boldsymbol{\gamma}}$ and $|V_{\boldsymbol{\gamma}} - V_{\widetilde{\boldsymbol{\gamma}}}| \le C_4 A_{\boldsymbol{\gamma}}$. Then $|V_{\boldsymbol{\gamma}}| = \bar{O}_P(A_{\boldsymbol{\gamma}})$ uniformly over $\boldsymbol{\gamma} \in \Gamma$.

12.3.2 Density Estimation

Recall the density estimation setup in Sections 11.1.1 and 11.4.4. In this subsection, we are assuming that $\eta^* = \eta$ and that Assumption 11.4.6 holds or, equivalently, that η is bounded. Thus Condition 12.1.1 holds. We also take $\boldsymbol{U} = \boldsymbol{W} = \boldsymbol{Y}$, so that Condition 12.2.1 holds. In addition, we assume that Condition 12.3.1 holds. We will verify Conditions 12.1.2, 12.1.4 and 12.1.6.

Conditions 12.1.2 and 12.1.4 are strengthened versions of Conditions 11.2.2 and 11.2.4 and their validity follows from arguments similar to those used in Section 11.4.4 to verify Conditions 11.2.2 and 11.2.4, except that Lemma 12.3.8 is used instead of Lemma 11.4.8.

Verification of Condition 12.1.6.

Observe that

$$\ell(\bar{\eta}_{\boldsymbol{\gamma}}) - \ell(\eta) - [\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta)] = (E_n - E)(\bar{\eta}_{\boldsymbol{\gamma}} - \eta). \tag{12.3.3}$$

The first property of Condition 12.1.6 follows from (12.3.3) with $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$, Theorem 12.1.1, and the consequence of Chebyshev's inequality that

$$\begin{aligned}(E_n - E)(\bar{\eta}_{\boldsymbol{\gamma}^*} - \eta) &= O_P\Big(\frac{\|\bar{\eta}_{\boldsymbol{\gamma}^*} - \eta\|}{\sqrt{n}}\Big) \\ &= O_P\Big(\frac{\inf_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}}}{\sqrt{n}}\Big) = O_P\Big(\inf_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}}^2 + \frac{1}{n}\Big).\end{aligned}$$

We claim that

$$\begin{aligned}&|(E_n - E)(\bar{\eta}_{\boldsymbol{\gamma}} - \eta)| \\ &\qquad = \bar{O}_P\bigg((\log^{1/2} n)\Big[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|\Big(\frac{N_n}{n}\Big)^{1/2} + \frac{N_n}{n}\Big]\bigg)\end{aligned} \tag{12.3.4}$$

uniformly over $\boldsymbol{\gamma} \in \Gamma$. The second property of Condition 12.1.6 follows from (12.3.3) and (12.3.4).

Let us now verify (12.3.4). Condition 12.3.1 implies that $N_n^{1/2} \sup_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}} \lesssim \log^{-1/2} n$. Now $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\| \lesssim \sup_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}}$ (uniformly over $\boldsymbol{\gamma} \in \widetilde{\Gamma}$) by Theorem 12.1.1 and $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|_\infty \lesssim N_n^{1/2} \sup_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}} \lesssim \log^{-1/2} n$ by (12.2.4). [Choose

$g^*_{\boldsymbol{\gamma}} \in \mathbb{G}_{\boldsymbol{\gamma}}$ such that $\|g^*_{\boldsymbol{\gamma}} - \eta\|_\infty = \rho_{n\boldsymbol{\gamma}}$]. Let c be a fixed positive number. It follows from Bernstein's inequality (11.4.3) that, for c' a sufficiently large positive number,

$$P\Big(|(E_n - E)(\bar{\eta}_{\boldsymbol{\gamma}} - \eta)| \geq c'(\log^{1/2} n)\big\{\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|(N_n/n)^{1/2} + N_n/n\big\}\Big) \leq 2\exp\big(-2cN_n \log n\big) \qquad (12.3.5)$$

for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$.

Let c be sufficiently large. Then, according to Lemma 12.2.3, there is a subset $\widetilde{\Gamma}''_n$ of $\widetilde{\Gamma}$ such that (12.3.2) holds with $C_3 = c$ and every point $\boldsymbol{\gamma} \in \Gamma$ is within n^{-2} of some point $\widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}''_n$ such that $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}} - \eta\| \leq \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|$. Let $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ be as just described. Then, by Theorem 12.1.1 and Lemma 12.2.6,

$$|(E_n - E)(\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})| \leq 2\|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \lesssim \frac{N_n^{1/2}}{n} \leq \frac{N_n}{n}. \qquad (12.3.6)$$

The desired result (12.3.4) follows from (12.3.5), (12.3.6) and Lemma 12.3.9. This completes the verification of Condition 12.1.6.

12.3.3 Generalized Regression

Recall the generalized regression setup from Sections 11.1 and 11.4.3. Here we are assuming that $\eta^* = \eta$ and that Assumptions 11.4.1–11.4.5 hold. Now η is bounded by Assumption 11.4.3, so Condition 12.1.1 holds. We take $\boldsymbol{W} = (\boldsymbol{X}, Y)$ and $\boldsymbol{U} = \boldsymbol{X}$, so Condition 12.2.1 follows from Assumption 11.4.5. In addition, we assume the following strengthened version of Assumption 11.4.4 holds.

Assumption 12.3.1. There are positive constants M_1 and M_2 such that $E[e^{|Y-\mu(\boldsymbol{X})|/M_1} | \boldsymbol{X} = \boldsymbol{x}] \leq M_2$ for $\boldsymbol{x} \in \mathcal{X}$.

We also assume that Condition 12.3.1 holds. We will verify Conditions 12.1.2, 12.1.4 and 12.1.6.

Let $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ be a random sample of size n from the joint distribution of $\boldsymbol{X}$ and Y. Choose $M'_1 \in (M_1, \infty)$. It follows from Assumption 12.3.1 that $P\big(|Y - \mu(\boldsymbol{X})| \geq M'_1 \log n\big) \leq M_2 n^{-M'_1/M_1}$ and hence that

$$\lim_n P\big(\max_{1 \leq i \leq n} |Y_i - \mu(\boldsymbol{X}_i)| \geq M'_1 \log n\big) = 0. \qquad (12.3.7)$$

Moreover, by the power series expansion of the exponential function, for $m \geq 2$ and $1 \leq i \leq n$,

$$E[|Y_i - \mu(\boldsymbol{X}_i)|^m | \boldsymbol{X}_i] \leq \frac{m!}{2}(2M_1^2 M_2) M_1^{m-2}. \qquad (12.3.8)$$

Thus, by Bernstein's inequality (11.4.2), if h is a bounded function on $\mathcal{X}$, then

$$P(|\langle h, Y-\mu\rangle_n| \geq t|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \leq 2\exp\Big(-\frac{nt^2}{2M_1(2M_1M_2\|h\|_n^2 + t\|h\|_\infty}\Big) \tag{12.3.9}$$

for $t > 0$.

Recall that the log-likelihood based on the random sample and its expected value are given by $\ell(h) = E_n[B(h)Y - C(h)]$ and $\Lambda(h) = E[B(h)Y - C(h)]$.

Verification of Condition 12.1.2.

Observe that

$$\Big|\frac{d}{d\alpha}\Lambda(h_1+\alpha h_2)\big|_{\alpha=0}\Big| = E\big(h_2(\boldsymbol{X})\{B'(h_1(\boldsymbol{X}))\mu(\boldsymbol{X}) - C'(h_1(\boldsymbol{X}))\}, $$

where $\mu(\boldsymbol{x}) = E(Y|\boldsymbol{X}=\boldsymbol{x})$. By Assumptions 11.4.1–11.4.3, $\mu(\cdot)$ is bounded. Since $B'(\cdot)$ and $C'(\cdot)$ are continuous, they are bounded on finite intervals. Condition 12.1.2(i) then follows from the Cauchy–Schwarz inequality. Condition 12.1.2(ii) is the same as Condition 11.2.2 and has been verified in Section 11.4.3.

Verification of Condition 12.1.4.

Lemma 12.3.10. *Suppose* $\lim_n N_n^2/n = 0$*. Then Condition* 12.1.4(ii) *holds.*

Proof. The argument is essentially the same as that of verifying Condition 11.2.4(ii) in Section 11.4.3, except that Lemma 12.3.7 is used instead of Lemma 11.4.7. The requirement that $\lim_n N_n^2/n = 0$ ensures the applicability of Lemma 12.3.7. □

Lemma 12.3.11. *Suppose* $\lim_n N_n^2/n = 0$ *and* $\sup_{\boldsymbol{\gamma}} \rho_{n\boldsymbol{\gamma}} = O(N_n^{-c})$ *for some* $c > 1/2$*. Then Condition* 12.1.4(i) *holds.*

Proof. In this proof, set $\bar{\rho}_n = \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}} \rho_{n\boldsymbol{\gamma}}$. By Theorem 12.1.1 applied to $\widetilde{\Gamma}$ there is a positive constant K_1 such that, for n sufficiently large, $\bar{\eta}_{\boldsymbol{\gamma}}$ exists uniquely and $\|\bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \leq K_1$ for $\boldsymbol{\gamma}\in\widetilde{\Gamma}$. Let $\boldsymbol{\gamma}\in\widetilde{\Gamma}$ and $g \in \mathbb{G}_{\boldsymbol{\gamma}}$. Then

$$\frac{d}{d\alpha}\ell(\bar{\eta}_{\boldsymbol{\gamma}}+\alpha g)\Big|_{\alpha=0} = E_n[gD(\bar{\eta}_{\boldsymbol{\gamma}})] + E_n[gB'(\bar{\eta}_{\boldsymbol{\gamma}})(Y-\mu)]$$

where $D(\bar{\eta}_{\boldsymbol{\gamma}}) = B'(\bar{\eta}_{\boldsymbol{\gamma}})\mu - C'(\bar{\eta}_{\boldsymbol{\gamma}})$ and $E[gD(\bar{\eta}_{\boldsymbol{\gamma}})] = 0$.

Let $0 < \delta \leq 1/4$. Since $N_n^{1/2}\bar{\rho}_n \lesssim N_n^{-(c-1/2)}$ for some $c > 1/2$ by Condition 12.3.1, there is an $\epsilon \in (0, \delta^2)$ and there is a fixed positive number c_1 such that, for n sufficiently large,

$$N_n^{1/2}\bar{\rho}_n \leq c_1(N_n^{1/2})^{-(\log 1/\delta)/(\log \delta/\epsilon^{1/2})}$$

and hence

$$\min(c_1^{-1}N_n^{1/2}\bar\rho_n, N_n^{1/2}\epsilon^{(k-1)/2}) \le \delta^{k-1} \tag{12.3.10}$$

for $k \ge 1$. (If $N_n^{1/2}\epsilon^{(k-1)/2} \ge \delta^{k-1}$, then $N_n^{1/2}\bar\rho_n \le c_1\delta^{k-1}$.)

Let Ω_n, $\lim_n P(\Omega_n) = 1$, be an event that depends only on $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, and is such that the statements in Lemma 12.2.5 and Lemma 12.2.6 hold.

Let k be a positive integer, and let $\boldsymbol\gamma, \widetilde{\boldsymbol\gamma} \in \widetilde\Gamma$ be such that $\zeta(\boldsymbol\gamma, \widetilde{\boldsymbol\gamma}) \le \epsilon^{k-1}$. Then, by Lemma 12.2.6, $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\| \le c_2\epsilon^{(k-1)/2}$, $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\|_n \le c_2\epsilon^{(k-1)/2}$ on Ω_n, and $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\|_\infty \le c_3 N_n^{1/2}\epsilon^{(k-1)/2}$ (for some fixed positive constants c_2, c_3). Let $\mathbb{B}_{\widetilde{\boldsymbol\gamma},k}$ be as in Section 12.2 and let $g \in \mathbb{B}_{\boldsymbol\gamma}$. Then, by Lemmas 12.2.4 and 12.2.5 (see the proof of Lemma 12.3.7), there is a $\widetilde g \in \mathbb{B}_{\widetilde{\boldsymbol\gamma},k-1}$ such that $\|g - \widetilde g\| \le c_4\epsilon^{k-1}$, $\|g-\widetilde g\|_n \le c_5\epsilon^{k-1}$ on Ω_n, and $\|g-\widetilde g\|_\infty \le c_6 N_n^{1/2}\epsilon^{k-1}$. Now

$$gB'(\bar\eta_{\boldsymbol\gamma}) - \widetilde g B'(\bar\eta_{\widetilde{\boldsymbol\gamma}}) = (g - \widetilde g)B'(\bar\eta_{\boldsymbol\gamma}) + \widetilde g[B'(\bar\eta_{\boldsymbol\gamma}) - B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})]. \tag{12.3.11}$$

Observe that, $\|(g-\widetilde g)B'(\bar\eta_{\boldsymbol\gamma})\|_n \le c_7\epsilon^{k-1}$ on Ω_n and $\|(g-\widetilde g)B'(\bar\eta_{\boldsymbol\gamma})\|_\infty \le c_7 N_n^{1/2}\epsilon^{k-1}$. Observe also that, $\|\widetilde g[B'(\bar\eta_{\boldsymbol\gamma}) - B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})]\|_n \le c_8 N_n^{1/2}\epsilon^{(k-1)/2}$ on Ω_n and $\|\widetilde g[B'(\bar\eta_{\boldsymbol\gamma}) - B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})]\|_\infty \le c_8 N_n\epsilon^{(k-1)/2}$. Consequently, $\|gB'(\bar\eta_{\boldsymbol\gamma}) - \widetilde g B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_n \le c_9 N_n^{1/2}\epsilon^{(k-1)/2}$ on Ω_n and $\|gB'(\bar\eta_{\boldsymbol\gamma}) - \widetilde g B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_\infty \le c_9 N_n\epsilon^{(k-1)/2}$. By the same argument, c_9 can be chosen so that, in addition, $\|gD(\bar\eta_{\boldsymbol\gamma}) - \widetilde g D(\bar\eta_{\widetilde{\boldsymbol\gamma}})\| \le c_9 N_n^{1/2}\epsilon^{(k-1)/2}$ and $\|gD(\bar\eta_{\boldsymbol\gamma}) - \widetilde g D(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_\infty \le c_9 N_n \epsilon^{(k-1)/2}$.

Alternatively, by Theorem 12.1.1 and Lemma 12.3.7,

$$\sup_{\boldsymbol\gamma\in\widetilde\Gamma} \frac{\|\bar\eta_{\boldsymbol\gamma} - \eta\|}{\rho_{n\boldsymbol\gamma}} = O(1) \quad \text{and} \quad \sup_{\boldsymbol\gamma\in\widetilde\Gamma} \frac{\|\bar\eta_{\boldsymbol\gamma} - \eta\|_n}{\rho_{n\boldsymbol\gamma}} = O(1)[1 + o_P(1)].$$

(Choose $g^* \in \mathbb{G}_{\boldsymbol\gamma}$ such that $\|g^* - \eta\|_\infty = \rho_{n\boldsymbol\gamma}$.) Consequently, for n sufficiently large, $\|\bar\eta_{\boldsymbol\gamma} - \eta\| \le c_{10}\bar\rho_n$ and $\|\bar\eta_{\boldsymbol\gamma} - \eta\|_n \le c_{10}\bar\rho_n$ on Ω_n for $\boldsymbol\gamma \in \widetilde\Gamma$ (provided that Ω_n is suitably chosen).

Given $\boldsymbol\gamma, \widetilde{\boldsymbol\gamma} \in \widetilde\Gamma$, we have that $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\| \le 2c_{10}\bar\rho_n$ and $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\|_n \le 2c_{10}\bar\rho_n$ on Ω_n. Choose $\eta'_{\boldsymbol\gamma} \in \mathbb{G}_{\boldsymbol\gamma}$ and $\eta'_{\widetilde{\boldsymbol\gamma}} \in \mathbb{G}_{\widetilde{\boldsymbol\gamma}}$ such that $\|\eta'_{\boldsymbol\gamma} - \eta\|_\infty \le \bar\rho_n$ and $\|\eta'_{\widetilde{\boldsymbol\gamma}} - \eta\|_\infty \le \bar\rho_n$. It follows from the triangle inequality and (12.2.4) that $\|\eta'_{\boldsymbol\gamma} - \bar\eta_{\boldsymbol\gamma}\|_\infty \le M(c_{10}+1)N_n^{1/2}\bar\rho_n$. Thus $\|\bar\eta_{\boldsymbol\gamma} - \eta\|_\infty \le [M(c_{10}+1)N_n^{1/2}+1]\bar\rho_n$. Similarly, $\|\bar\eta_{\widetilde{\boldsymbol\gamma}} - \eta\|_\infty \le [M(c_{10}+1)N_n^{1/2}+1]\bar\rho_n$. Hence $\|\bar\eta_{\boldsymbol\gamma} - \bar\eta_{\widetilde{\boldsymbol\gamma}}\|_\infty \le 2[M(c_{10}+1)N_n^{1/2}+1]\bar\rho_n$.

Let $\zeta(\boldsymbol\gamma, \widetilde{\boldsymbol\gamma}) \le \epsilon^{k-1}$ and let $g \in \mathbb{B}_{\boldsymbol\gamma}$ and $\widetilde g \in \mathbb{B}_{\widetilde{\boldsymbol\gamma},k-1}$ be as above. Then [recall (12.2.4), (12.3.10), and (12.3.11)], $\|gB'(\bar\eta_{\boldsymbol\gamma}) - \widetilde g B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_n \le c_{11}\delta^{k-1}$ on Ω_n, $\|gB'(\bar\eta_{\boldsymbol\gamma}) - \widetilde g B'(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_\infty \le c_{11}N_n^{1/2}\delta^{k-1}$, $\|gD(\bar\eta_{\boldsymbol\gamma}) - \widetilde g D(\bar\eta_{\widetilde{\boldsymbol\gamma}})\| \le c_{11}\delta^{k-1}$, and $\|gD(\bar\eta_{\boldsymbol\gamma}) - \widetilde g D(\bar\eta_{\widetilde{\boldsymbol\gamma}})\|_\infty \le c_{11}N_n^{1/2}\delta^{k-1}$.

Let $K = K_n$ be a positive integer satisfying the two inequalities specified in the next paragraph, and let Ξ_k, $\mathbb{B}_{\boldsymbol\gamma k}$ for $\boldsymbol\gamma \in \widetilde\Gamma$, and $\mathbb{B}_k$, $0 \le k \le K$, be as in Lemma 12.2.2 and the following paragraph with the current value of ϵ. We will apply Lemma 11.4.6 with $s = (\boldsymbol{\gamma}, g)$, $V_s = E_n\{g[D(\bar{\eta}_{\boldsymbol{\gamma}}) + B'(\bar{\eta}_{\boldsymbol{\gamma}})(Y-\mu)]\}$, $\mathbb{S} = \{(\boldsymbol{\gamma}, g) : \boldsymbol{\gamma} \in \mathbb{B}$ and $g \in \mathbb{G}_{\boldsymbol{\gamma}}\}$, and $\mathbb{S}_k = \{(\boldsymbol{\gamma}, g) : \boldsymbol{\gamma} \in \Xi_k$ and $g \in \mathbb{B}_{\boldsymbol{\gamma} k}\}$. Now $\#(\mathbb{S}_k) \le (M'\epsilon^{-2k})^{N_n}$ for $1 \le k \le K$ by (12.2.3), so (11.4.5) holds with $C_3 = 1$ and any $C_4 \ge 2\log(M'\epsilon^{-1})N_n$.

Let Ω_{n0} denote the event that $\max_{1\le i\le n} |Y_i - \mu(\boldsymbol{X}_i)| \le M_1' \log n$ with M_1' as in (12.3.7). Then $\lim_n P(\Omega_{n0}) = 1$. Choose $\boldsymbol{\gamma} \in \Gamma$ and $g \in \mathbb{B}_{\boldsymbol{\gamma}}$. Let $\widetilde{\boldsymbol{\gamma}} \in \Xi_K$ be such that $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \le \epsilon^K$. Then there is a $\widetilde{g} \in \mathbb{B}_{\widetilde{\boldsymbol{\gamma}} K}$ such that $\|gD(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_\infty \le c_{11} N_n^{1/2}\delta^K$ and $\|gB'(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_\infty \le c_{11}N_n^{1/2}\delta^K$. Thus $\|gD(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_\infty \le (N_n/n)^{1/2}$ provided that K satisfies the inequality $c_{11}\delta^K \le n^{-1/2}$ and $|E_n\{[gB'(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})](Y-\mu)\}| \le (N_n/n)^{1/2}$ on Ω_{n0} provided that K satisfies the inequality $M_1' c_{11}\delta^K \le 1/(n^{1/2}\log n)$. Let K satisfy both inequalities. Then (11.4.4) holds with $C_1 = 2(N_n/n)^{1/2}$, $C_2 = 0$, and $\Omega = \Omega_{n0}$.

Let $1 \le k \le K$. Given $\boldsymbol{\gamma} \in \Xi_k$ and $g \in \mathbb{B}_{\boldsymbol{\gamma} k}$, choose $\widetilde{\boldsymbol{\gamma}} \in \Xi_{k-1}$ and $\widetilde{g} \in \mathbb{B}_{\widetilde{\boldsymbol{\gamma}}, k-1}$ such that $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \le \epsilon^{k-1}$, $\|gB'(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_n \le c_{11}\delta^{k-1}$ on Ω_n, $\|gB'(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_\infty \le c_{11}N_n^{1/2}\delta^{k-1}$, $\|gD(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\| \le c_{11}\delta^{k-1}$, and $\|gD(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})\|_\infty \le c_{11}N_n^{1/2}\delta^{k-1}$. Write $s = (\boldsymbol{\gamma}, g)$ and $V_s = V_{1s} + V_{2s}$, where $V_{1s} = E_n[gD(\bar{\eta}_{\boldsymbol{\gamma}})]$ and $V_{2s} = E_n[gB'(\bar{\eta}_{\boldsymbol{\gamma}})(Y-\mu)]$. Similarly, write $\widetilde{s} = (\widetilde{\boldsymbol{\gamma}}, \widetilde{g})$ and $V_{\widetilde{s}} = V_{1\widetilde{s}} + V_{2\widetilde{s}}$, where $V_{1\widetilde{s}} = E_n[\widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})]$ and $V_{2\widetilde{s}} = E_n[\widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})(Y-\mu)]$. Observe that $V_{1s} - V_{1\widetilde{s}} = (E_n - E)[gD(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}D(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})]$. Since $0 < 2\delta \le 1$, it follows from Bernstein's inequality (11.4.2) that, for $C > 0$,

$$P\big(|V_{1s} - V_{1\widetilde{s}}| \ge C2^{-(k-1)}(N_n/n)^{1/2}\big) \le 2\exp\Big(-\frac{C^2(2\delta)^{-(k-1)}N_n}{2c_{11}(c_{11} + Cn^{-1/2}N_n)}\Big).$$

Similarly, $V_{2s} - V_{2\widetilde{s}} = E_n\{[gB'(\bar{\eta}_{\boldsymbol{\gamma}}) - \widetilde{g}B'(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})](Y-\mu)\}$, so it follows from (12.3.9) that

$$P\big(|V_{2s} - V_{2\widetilde{s}}| \ge C2^{-(k-1)}(N_n/n)^{1/2}\big|\Omega_n\big) \le 2\exp\Big(-\frac{C^2(2\delta)^{-(k-1)}N_n}{2c_{11}(c_{11} + Cn^{-1/2}N_n)}\Big).$$

provided that c_{11} is sufficiently large. Hence

$$P\big(|V_s - V_{\widetilde{s}}| \ge 2C2^{-(k-1)}(N_n/n)^{1/2}; \Omega_n\big) \le 4\exp\Big(-\frac{C^2(2\delta)^{-(k-1)}N_n}{2c_{11}(c_{11} + Cn^{-1/2}N_n)}\Big),$$

so (11.4.6) holds with

$$C_4 = \frac{C^2 N_n}{4c_{11}(c_{11} + Cn^{-1/2}N_n)} \ge 2\log(M'\epsilon^{-1})N_n$$

for C sufficiently large, $C_5 = 2C(N_n/n)^{1/2}$, $C_6 = 4$, and $\Omega = \Omega_n$. Consequently, by Lemma 11.4.6,

$$P\Big(\sup_{\gamma\in\Gamma}\sup_{g\in\mathbb{B}_\gamma}\Big|\frac{d}{d\alpha}\ell(\bar\eta_\gamma+\alpha g)\Big|_{\alpha=0}\Big| \geq 2(1+2C)(N_n/n)^{1/2}\Big) \leq \frac{16c_{11}(c_{11}+Cn^{-1/2}N_n)}{C^2N_n} + P\big((\Omega_n\cap\Omega_{n0})^c\big),$$

which can be made arbitrarily close to zero by making n and C sufficiently large. □

Verification of Condition 12.1.6.

It follows from (12.3.8) and Bernstein's inequality (11.4.3) (with $H = M_1A$) that if h is a bounded function on $\mathcal{X}$ and $A \geq \|h\|_\infty$, then

$$P\Big(|E_n\{h(Y-\mu)\}| \geq tM_1^{-1}A^{-1}\big[(2M_1^2M_2)^{1/2}\|h\|_n(N_n/n)^{1/2} + N_n/n\big]\Big|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n\Big) \leq 2\exp\Big(-\frac{tM_1^{-2}A^{-2}N_n}{2}\Big) \tag{12.3.12}$$

for $t \geq 1$.

Observe that

$$\begin{aligned}\ell(\bar\eta_\gamma)-\ell(\eta)-[\Lambda(\bar\eta_\gamma)-\Lambda(\eta)] &= (E_n-E)\{[B(\bar\eta_\gamma)-B(\eta)]\mu-[C(\bar\eta_\gamma)-C(\eta)]\}\\ &\quad + E_n\{[B(\bar\eta_\gamma)-B(\eta)](Y-\mu)\}.\end{aligned} \tag{12.3.13}$$

Lemma 12.3.12. *Suppose Condition* 12.3.1 *holds. Then*

$$(E_n-E)\{[B(\bar\eta_\gamma)-B(\eta)]\mu-[C(\bar\eta_\gamma)-C(\eta)]\} = \bar O_P\bigg((\log^{1/2}n)\Big[\|\bar\eta_\gamma-\eta\|\Big(\frac{N_n}{n}\Big)^{1/2}+\frac{N_n}{n}\Big]\bigg)$$

uniformly over $\gamma\in\Gamma$.

Proof. The proof of this result is similar to that of Condition 12.1.6(ii) in the density estimation context. □

Lemma 12.3.13. *Suppose Condition* 12.3.1 *holds. Then*

$$|E_n\{[B(\bar\eta_\gamma)-B(\eta)](Y-\mu)\}| = \bar O_P\bigg((\log^{1/2}n)\Big[\|\bar\eta_\gamma-\eta\|\Big(\frac{N_n}{n}\Big)^{1/2}+\frac{N_n}{n}\Big]\bigg)$$

uniformly over $\gamma\in\Gamma$.

Proof. Note that $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\| \lesssim \sup_{\boldsymbol{\gamma}\in\widetilde{\Gamma}} \rho_{n\boldsymbol{\gamma}}$ and $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|_\infty \lesssim \log^{-1/2} n$ uniformly over $\boldsymbol{\gamma} \in \widetilde{\Gamma}$. (See the arguments in Section 12.3.2.) Set $h_{\boldsymbol{\gamma}} = B(\bar{\eta}_{\boldsymbol{\gamma}}) - B(\eta)$ for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$. Then $\|h_{\boldsymbol{\gamma}}\| \lesssim \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|$, $\|h^2_{\boldsymbol{\gamma}}\| \lesssim (\log^{-1/2} n)\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|$ and $\|h^2_{\boldsymbol{\gamma}}\|_\infty \lesssim \log^{-1} n$ uniformly over $\boldsymbol{\gamma} \in \widetilde{\Gamma}$. Let c_1 be a fixed positive number. It now follows from Bernstein's inequality (11.4.2) [note that $\|h_{\boldsymbol{\gamma}}\|_n^2 = E_n(h^2_{\boldsymbol{\gamma}})$] that, for c_2 a sufficiently large positive number,

$$P\biggl(\|h_{\boldsymbol{\gamma}}\|_n^2 - \|h_{\boldsymbol{\gamma}}\|^2 \ge c_2^2\Bigl[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|\Bigl(\frac{N_n}{n}\Bigr)^{1/2} + \frac{N_n}{n}\Bigr]\biggr) \le 2\exp\bigl(-2c_1 N_n \log n\bigr)$$

for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$ and hence that, for c_2 a sufficiently large positive number,

$$P(\Omega^c_{n\boldsymbol{\gamma}}) \le 2\exp\bigl(-2c_1 N_n \log n\bigr), \qquad \boldsymbol{\gamma} \in \widetilde{\Gamma}.$$

where $\Omega_{n\boldsymbol{\gamma}}$ denotes the event that $\|h_{\boldsymbol{\gamma}}\|_n \le c_2[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\| + (N_n/n)^{1/2}]$. It follows from (12.3.12) that, for a sufficiently large positive number c_3,

$$P\Bigl(|E_n\{h_{\boldsymbol{\gamma}}(Y-\mu)\}| \ge c_3(\log^{1/2} n)\bigl[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|(N_n/n)^{1/2} + N_n/n\bigr]\Big|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\Bigr) \le 2\exp\bigl(-2c_1 N_n \log n\bigr)$$

on $\Omega_{n\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$ and hence that

$$P\Bigl(|E_n\{h_{\boldsymbol{\gamma}}(Y-\mu)\}| \ge c_3(\log^{1/2} n)\bigl[\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|(N_n/n)^{1/2} + N_n/n\bigr]\Bigr) \le 4\exp\bigl(-2c_1 N_n \log n\bigr) \tag{12.3.14}$$

for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$.

Let c_1 be sufficiently large. Then, according to Lemma 12.2.3, there is a subset $\widetilde{\Gamma}''_n$ of $\widetilde{\Gamma}$ such that (12.3.2) holds with $C_3 = c_1$ and every point $\boldsymbol{\gamma} \in \Gamma$ is within n^{-3} of some point $\widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}''_n$ such that $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}} - \eta\| \le \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta\|$. Let $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ be as just described. By Lemma 12.2.6,

$$\begin{aligned} |E_n\{[B(\bar{\eta}_{\boldsymbol{\gamma}}) - B(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})](Y-\mu)\}| &\lesssim \|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \max_{1\le i\le n} |Y_i - \mu(\boldsymbol{X}_i)| \\ &\lesssim N_n^{1/2} n^{-3/2} \log n \\ &\lesssim N_n/n \end{aligned} \tag{12.3.15}$$

provided that $|Y_i - \mu(\boldsymbol{X}_i)| \le M'_1 \log n$ for $1 \le i \le n$.

The desired result follows from (12.3.2), (12.3.7), (12.3.14), (12.3.15), and Lemma 12.3.9. □

Lemma 12.3.14. *Suppose Condition* 12.3.1 *holds. Then Condition* 12.1.6 *holds.*

Proof. Now $E\big(E_n\{[B(\bar\eta_{\boldsymbol{\gamma}^*})-B(\eta)](Y-\mu)\}\big|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\big)=0$ and

$$\operatorname{var}\big(E_n\{[B(\bar\eta_{\boldsymbol{\gamma}^*})-B(\eta)](Y-\mu)\}\big|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\big)=O\Big(\frac{\|\bar\eta_{\boldsymbol{\gamma}^*}-\eta\|_n^2}{n}\Big),$$

so

$$E\big[\big(E_n\{[B(\bar\eta_{\boldsymbol{\gamma}^*})-B(\eta)](Y-\mu)\}\big)^2\big]=O_P\Big(\frac{\|\bar\eta_{\boldsymbol{\gamma}^*}-\eta\|^2}{n}\Big).$$

Since $\|\bar\eta_{\boldsymbol{\gamma}^*}-\eta\|=\inf_{\boldsymbol{\gamma}\in\widetilde\Gamma}\rho_{n\boldsymbol{\gamma}}$, it follows from Chebyshev's inequality that

$$E_n\{[B(\bar\eta_{\boldsymbol{\gamma}^*})-B(\eta)](Y-\mu)\}=O_P\Big(\frac{\inf_{\boldsymbol{\gamma}\in\widetilde\Gamma}\rho_{n\boldsymbol{\gamma}}}{\sqrt{n}}\Big).$$

Similarly,

$$(E_n-E)\{[B(\bar\eta_{\boldsymbol{\gamma}^*})-B(\eta)]\mu-[C(\bar\eta_{\boldsymbol{\gamma}^*})-C(\eta)]\}=O_P\Big(\frac{\inf_{\boldsymbol{\gamma}\in\widetilde\Gamma}\rho_{n\boldsymbol{\gamma}}}{\sqrt{n}}\Big).$$

The first property of Condition 12.1.6 now follows from (12.3.13) with $\boldsymbol{\gamma}=\boldsymbol{\gamma}^*$. The second property follows from (12.3.13) and Lemmas 12.3.12 and 12.3.13. □

Ordinary regression.

The framework of generalized regression, as considered above, includes ordinary regression as a special case. Specifically, let $B(\eta)=2\eta$ for $\eta\in\mathbb{R}$ and $\Psi(dy)=\pi^{-1/2}e^{-y^2}dy$ for $y\in\mathbb{R}$. Then $S=\mathbb{R}$. Also, $C(\eta)=\eta^2$ and $A(\eta)=\eta$ for $\eta\in\mathbb{R}$, so the regression function μ equals the response function η. Suppose that Y has finite second moment. The pseudo-log-likelihood and its expectation are given, respectively, by $l(h;\boldsymbol{X},Y)=2h(\boldsymbol{X})Y-h^2(\boldsymbol{X})=-[Y-h(\boldsymbol{X})]^2+Y^2$ and $\Lambda(h)=-E\{[Y-h(\boldsymbol{X})]^2\}+E(Y^2)$. Assumption 11.4.3 is that the regression function is bounded. Let h_1 and h_2 be bounded functions on $\mathcal{X}$. Then

$$\frac{d}{d\alpha}\Lambda(h_1+\alpha h_2)\Big|_{\alpha=0}=2E\{h_2(\boldsymbol{X})[\mu(\boldsymbol{X})-h_1(\boldsymbol{X})]\}$$

and

$$\frac{d^2}{d\alpha^2}\Lambda(h_1+\alpha(h_2-h_1))=-2\|h_2-h_1\|^2,$$

so Condition 12.1.2 follows from the boundedness of the regression function and of the density function of $\boldsymbol{X}$. Also,

$$\frac{d}{d\alpha}\ell(\bar\mu_{\boldsymbol{\gamma}}+\alpha g)\Big|_{\alpha=0}=2E_n\{g[Y-\bar\mu_{\boldsymbol{\gamma}}(\boldsymbol{X})]\}$$

and

$$\frac{d^2}{d\alpha^2}\ell(g_1+\alpha(g_2-g_1)) = -2\|g_2-g_1\|_n^2.$$

Thus Condition 12.1.4(ii) follows from Lemma 12.3.7, while Condition 12.1.4(i) requires Lemma 12.3.11 for its verification.

12.4 Proofs of Lemmas in Section 12.2

In this section we verify Lemmas 12.2.2–12.2.6.

Consider a free knot sequence $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)$ such that $a < \gamma_1 \le \cdots \le \gamma_J < b$ and

$$\frac{\gamma_{j_2-1}-\gamma_{j_2-m}}{\gamma_{j_1-1}-\gamma_{j_1-m}} \le \bar{M}, \qquad 2 \le j_1, j_2 \le J+m, \tag{12.4.1}$$

where $\gamma_{1-m} = \cdots = \gamma_0 = a$ and $\gamma_{J+1} = \cdots = \gamma_{J+m} = b$.

Observe that

$$\begin{aligned}\sum_{j=1}^{J+m}(\gamma_{j-1}-\gamma_{j-m}) &= \sum_{j=0}^{J+m-1}\gamma_j - \sum_{j=1-m}^{J}\gamma_j \\ &= \sum_{j=J+1}^{J+m-1}\gamma_j - \sum_{j=1-m}^{-1}\gamma_j \\ &= (m-1)(b-a).\end{aligned}$$

Thus it follows from (12.4.1) that

$$\gamma_{j-1}-\gamma_{j-m} \ge \frac{(m-1)(b-a)}{\bar{M}(J+m)}, \qquad 2 \le j \le J+m. \tag{12.4.2}$$

The requirement (12.4.1) is stronger than the bound on the global mesh ratio of $\boldsymbol{\gamma}$ that was considered by de Boor (1976). To see this, note let $\boldsymbol{\gamma} \in \Gamma$ and note that $\gamma_1 - \gamma_{1-m} = \gamma_1 - \gamma_{2-m}$, $\gamma_{J+m} - \gamma_J = \gamma_{J+m-1} - \gamma_J$, and

$$\frac{\gamma_{j_2}-\gamma_{j_2-m}}{\gamma_{j_1}-\gamma_{j_1-m}} \le \frac{\gamma_{j_2-1}-\gamma_{j_2-m}+\gamma_{j_2}-\gamma_{j_2+1-m}}{\frac{\gamma_{j_1-1}-\gamma_{j_1-m}}{2}+\frac{\gamma_{j_1}-\gamma_{j_1+1-m}}{2}}$$

for $1 \le j_1, j_2 \le J+m$ (the numerator is increased and the denominator is decreased), so it follows from (12.4.1) that

$$\frac{\gamma_{j_2}-\gamma_{j_2-m}}{\gamma_{j_1}-\gamma_{j_1-m}} \le 2\bar{M}, \qquad 1 \le j_1, j_2 \le J+m. \tag{12.4.3}$$

Example 12.4.1. Suppose that $J = m-1$ and hence that $J+m = 2m-1$, and suppose also that $\gamma_j = (a+b)/2$ for $1 \le j \le m-1$. Then

$\gamma_{j-1} - \gamma_{j-m} = (b-a)/2$ for $2 \le j \le 2m-1$, so 1 is the smallest value of m that satisfies (12.4.1). Also, $\gamma_j - \gamma_{j-m} = (b-a)/2$ for $1 \le j \le m-1$ and for $m+1 \le j \le 2m-1$, while $\gamma_m - \gamma_0 = b-a$, so 1 is also the smallest value of $\bar{M}$ that satisfies (12.4.3).

Observe that $\sum_{j=1}^{J+m}(\gamma_j - \gamma_{j-m}) = m(b-a)$. Thus it follows from (12.4.3) that

$$\gamma_j - \gamma_{j-m} \ge \frac{m(b-a)}{2\bar{M}(J+m)}, \qquad 1 \le j \le J+m, \tag{12.4.4}$$

and

$$\gamma_j - \gamma_{j-m} \le \frac{2\bar{M}m(b-a)}{J+m}, \qquad 1 \le j \le J+m, \tag{12.4.5}$$

Proof of Lemma 12.2.2. We first verify this result when $L = 1$, $J = J_1 \ge 1$, $\gamma_j = \gamma_{1j}$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1$, $\mathcal{U} = \mathcal{U}_1 = [a,b] = [a_1,b_1]$, $m = m_1 \ge 2$, and $N_n = N_{n1} = J+m$. Here the metric ζ is given by $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) = 9\bar{M}N_n|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}|_\infty/(b-a)$. Let $0 \le \epsilon_1 \le 2$, let $\boldsymbol{\gamma} \in \Gamma$, and let $\widetilde{\boldsymbol{\gamma}}$ be a free knot sequence such that $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \le \epsilon_1$ and hence

$$2|\boldsymbol{\gamma} - \widetilde{\boldsymbol{\gamma}}|_\infty \le \frac{\epsilon_1(b-a)}{4\bar{M}N_n}.$$

Thus, by (12.4.2), $\widetilde{\boldsymbol{\gamma}}$ satisfies (12.4.1) with $\bar{M}$ replaced by

$$\bar{M}\frac{m-1+\epsilon_1/4}{m-1-\epsilon_1/4} \le 3\bar{M},$$

so $\widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$. Let $\widetilde{\Gamma}_{\epsilon_1}$ denote the collection of all such free knot sequences $\widetilde{\boldsymbol{\gamma}}$ as $\boldsymbol{\gamma}$ ranges over Γ. Then $\widetilde{\Gamma}_{\epsilon_1} \subset \widetilde{\Gamma}$ and $\widetilde{\Gamma}_0 = \Gamma$.

Given a positive integer Λ, let $\phi(u;\Lambda)$ denote the function on $[a,b]$ defined by

$$\phi(u;\Lambda) = a + \frac{b-a}{\Lambda}\left[\Lambda\frac{u-a}{b-a} + \frac{1}{2}\right], \qquad a \le u \le b,$$

where $[\cdot]$ denotes the greatest integer function. Observe that $\phi(u;\Lambda)$ is nondecreasing in u, $\phi(a;\Lambda) = a$, $\phi(b;\Lambda) = b$, $\phi(u;\Lambda) \in \{a + i(b-a)/\Lambda : i = 0, \dots, \Lambda\}$, and

$$u - \frac{b-a}{2\Lambda} < \phi(u;\Lambda) \le u + \frac{b-a}{2\Lambda}, \qquad a \le u \le b.$$

Given the free knot sequence $\boldsymbol{\gamma}$, consider the transformed sequence $\phi(\boldsymbol{\gamma};\Lambda) = (\phi(\gamma_j;\Lambda))$. Let $0 < \epsilon \le 1$. Observe that

$$|\phi(\boldsymbol{\gamma};\Lambda) - \boldsymbol{\gamma}|_\infty \le \frac{b-a}{2\Lambda}$$

and hence that if

$$\Lambda \ge 4\epsilon^{-1}\bar{M}N_n, \tag{12.4.6}$$

then $\zeta(\boldsymbol{\gamma}, \phi(\boldsymbol{\gamma}, \Lambda)) \le \epsilon$. Let Λ be the smallest integer satisfying (12.4.6). Then $\Lambda - 1 \le 4\epsilon^{-1}\bar{M}N_n$. [Observe also that if (12.4.6) holds, $u_1, u_2 \in [a, b]$, and

$$u_2 - u_1 \ge \frac{(b-a)\epsilon}{4\bar{M}N_n},$$

then

$$\frac{\Lambda(u_2 - u_1)}{b-a} \ge 1$$

and hence $\phi(u_2; \Lambda) > \phi(u_1; \Lambda)$.]

Suppose that (12.4.6) holds and let $0 \le \epsilon_0 \le 1$. Set $\widetilde{\Gamma}'_{\epsilon_0,\epsilon} = \{\phi(\boldsymbol{\gamma}; \Lambda) : \boldsymbol{\gamma} \in \widetilde{\Gamma}_{\epsilon_0}\} \subset \widetilde{\Gamma}_{\epsilon_0+\epsilon}$. Then every point in $\widetilde{\Gamma}_{\epsilon_0}$ is within ϵ of some point in $\widetilde{\Gamma}'_{\epsilon_0,\epsilon}$. Observe that

$$\#(\widetilde{\Gamma}'_{\epsilon_0,\epsilon}) \le \binom{(m-1)(\Lambda-1)}{J}.$$

(Note that the multiplicity of each free knot is at most $m-1$.)

Let I be an integer with $I \ge J$. Then

$$1 = \sum_{y=0}^{I} \binom{I}{y} \Big(\frac{J}{I}\Big)^y \Big(1 - \frac{J}{I}\Big)^{I-y} \ge \binom{I}{J} \Big(\frac{J}{I}\Big)^J \Big(1 - \frac{J}{I}\Big)^{I-J},$$

so

$$\begin{aligned}\binom{I}{J} &\le \Big(\frac{I}{J}\Big)^J \Big(1 - \frac{J}{I}\Big)^{J-I} \\ &= \Big(\frac{I}{J}\Big)^J \bigg(\Big(1 - \frac{J}{I}\Big)^{-\left(\frac{I}{J}-1\right)}\bigg)^J \\ &\le \Big(\frac{I}{J}\Big)^J e^J.\end{aligned}$$

(Observe that $(d/dx)[x + (1-x)\log(1-x)] > 0$ for $0 < x < 1$, so $x + (1-x)\log(1-x) > 0$ for $0 < x < 1$ and hence $(1-x)^{-(1/x-1)} < e$ for $0 < x < 1$.) Consequently,

$$\#(\widetilde{\Gamma}'_{\epsilon_0,\epsilon}) \le \Big[4e\epsilon^{-1}\bar{M}(m-1)\Big(1 + \frac{m}{J}\Big)\Big]^J \le (4e\epsilon^{-1}m^2\bar{M})^{N_n}.$$

Consider now the general case $L \ge 1$. Here $\zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) = \max_l \zeta_l(\boldsymbol{\gamma}_l, \widetilde{\boldsymbol{\gamma}}_l)$ and

$$\Lambda_l \ge 4\epsilon^{-1}\bar{M}_l N_{nl}, \qquad 1 \le l \le L. \tag{12.4.7}$$

Let $\widetilde{\Gamma}$ be the Cartesian product of $\widetilde{\Gamma}_l$, $1 \le l \le L$, and let $\widetilde{\Gamma}_{\epsilon_1}$ denote the Cartesian product of $\widetilde{\Gamma}_{l\epsilon_1}$, $1 \le l \le L$. Then $\widetilde{\Gamma}_{\epsilon_1} \subset \Gamma$ and $\widetilde{\Gamma}_0 = \Gamma$. Let $\widetilde{\Gamma}'_{\epsilon_0,\epsilon} \subset \widetilde{\Gamma}_{\epsilon_0+\epsilon}$ denote the Cartesian product of $\widetilde{\Gamma}'_{l\epsilon_0,\epsilon}$, $1 \le l \le L$. Then every point in $\widetilde{\Gamma}_{\epsilon_0}$ is within ϵ of some point in $\widetilde{\Gamma}'_{\epsilon_0,\epsilon}$. Now $N_n = \prod_l N_{nl} \ge \sum_l N_{nl}$, so

$$\#(\widetilde{\Gamma}_{\epsilon_0,\epsilon}) \le (4e\epsilon^{-1} \max_l \bar{M}_l m_l^2)^{N_n}.$$

Let $0 < \epsilon \le 1/2$, let K be a positive integer, and set $\Xi_K = \widetilde{\Gamma}_{0,\epsilon^K} \subset \widetilde{\Gamma}$ and $\Xi_k = \widetilde{\Gamma}_{\epsilon^K+\cdots+\epsilon^{k+1},\epsilon^k} \subset \widetilde{\Gamma}$ for $0 \le k \le K-1$. Then

$$\#(\Xi_k) \le (4e\epsilon^{-k} \max_l \bar{M}_l m_l^2)^{N_n}, \qquad 1 \le k \le K.$$

Moreover, every point in $\Gamma = \widetilde{\Gamma}_0$ is within ϵ^K of some point in $\widetilde{\Gamma}_{0,\epsilon^K} = \Xi_K$; and, for $1 \le k \le K$, every point in $\Xi_k \subset \widetilde{\Gamma}_{\epsilon^K+\cdots+\epsilon^k}$ is within ϵ^{k-1} of some point in $\widetilde{\Gamma}_{\epsilon^K+\cdots+\epsilon^k,\epsilon^{k-1}} = \Xi_{k-1}$. □

Proof of Lemma 12.2.3. For each point $\boldsymbol{\gamma}' \in \widetilde{\Gamma}'_{0,\epsilon/2}$ (which is defined as in the proof of Lemma 12.2.2), there is a point $\widetilde{\boldsymbol{\gamma}}$ in the compact set $\{\boldsymbol{\gamma} \in \widetilde{\Gamma} : \zeta(\boldsymbol{\gamma}',\boldsymbol{\gamma}) \le \epsilon/2\}$ that minimizes the function $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|$ over this set. Let $\widetilde{\Gamma}_{0,\epsilon}$ denote the collection of all such points $\widetilde{\boldsymbol{\gamma}}$. Then $\widetilde{\Gamma}_{0,\epsilon} \subseteq \widetilde{\Gamma}_\epsilon$ and $\#(\widetilde{\Gamma}_{0,\epsilon}) \le \#(\widetilde{\Gamma}'_{0,\epsilon/2}) \le (8e\epsilon^{-1} \max_l \bar{M}_l m_l^2)^{N_n}$. Given $\boldsymbol{\gamma} \in \Gamma$, choose $\boldsymbol{\gamma}' \in \widetilde{\Gamma}'_{0,\epsilon/2}$ such that $\zeta(\boldsymbol{\gamma},\boldsymbol{\gamma}') \le \epsilon/2$ and let $\widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}_{0,\epsilon}$ be as defined above. Then $\zeta(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}) \le \epsilon$ and $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}} - \eta^*\| \le \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta^*\|$. □

Suppose that $L = 1$. Let $B_{\boldsymbol{\gamma}j}$ be the normalized B-spline corresponding to the knot sequence $\gamma_{j-m},\dots,\gamma_j$. According to Theorem 4.2 of DeVore and Lorentz (1993, Chapter 5), there is a positive constant $D_m \le 1$ such that

$$\frac{D_m^2}{m(b-a)} \sum_j b_j^2(\gamma_j - \gamma_{j-m}) \le \left\| \sum_j b_j B_{\boldsymbol{\gamma}j} \right\|_\psi^2 \le \frac{1}{m(b-a)} \sum_j b_j^2(\gamma_j - \gamma_{j-m}) \tag{12.4.8}$$

and

$$D_m \max_j |b_j| \le \left\| \sum_j b_j B_{\boldsymbol{\gamma}j} \right\|_\infty \le \max_j |b_j|. \tag{12.4.9}$$

It follows from (12.4.4), (12.4.5) and (12.4.8) that

$$\frac{D_m^2}{2\bar{M}(J+m)} \sum_j b_j^2 \le \left\| \sum_j b_j B_{\boldsymbol{\gamma}j} \right\|_\psi^2 \le \frac{2\bar{M}}{J+m} \sum_j b_j^2, \qquad \boldsymbol{\gamma} \in \Gamma. \tag{12.4.10}$$

For general L, set $m = \prod_l m_l$, $D = \prod_l D_{m_l}$, and $\bar{M} = \prod_l \bar{M}_l$,and note that $N_n = \prod_l (J_l + m_l)$. Also, let $\mathcal{J}$ denote the Cartesian product of the sets $\{1,\dots,J_{l+m_l}\}, 1 \le l \le L$ and, for $j = (j_1,\dots,j_L) \in \mathcal{J}$, consider the tensor product B-spline $B_{\boldsymbol{\gamma}j}(\boldsymbol{u}) = B_{\boldsymbol{\gamma}_1 j_1}(u_1)\dots B_{\boldsymbol{\gamma}_L j_L}(u_L)$. The *support* supp$(h)$ of a function h on a set $\mathcal{U}$ is defined by supp$(h) = \{\boldsymbol{u} \in \mathcal{U} : h(\boldsymbol{u}) \ne 0\}$.

Lemma 12.4.15. *Let $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ and $j \in \mathcal{J}$. Then*

$$\frac{D^2}{6^L \bar{M} N_n} \sum_j b_j^2 \le \Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} \Big\|_\psi^2 \le \frac{6^L \bar{M}}{N_n} \sum_j b_j^2; \tag{12.4.11}$$

$$D \max_j |b_j| \le \Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} \Big\|_\infty \le \max_j |b_j|; \tag{12.4.12}$$

$$\psi\big(\operatorname{supp}(B_{\boldsymbol{\gamma} j})\big) \le \frac{6^L \bar{M} m}{N_n}; \tag{12.4.13}$$

$$\#\{j \in \mathcal{J} : B_{\boldsymbol{\gamma} j}(\boldsymbol{u}) \ne 0\} \le m \qquad \text{for } \boldsymbol{u} \in \mathcal{U}; \tag{12.4.14}$$

$$\#\{k \in \mathcal{J} : B_{\boldsymbol{\gamma} j} B_{\widetilde{\boldsymbol{\gamma}} k} \text{ is not identically zero on } \mathcal{U}\} \le 38^L \bar{M}^2 m; \tag{12.4.15}$$

$$\|B_{\boldsymbol{\gamma} j} - B_{\widetilde{\boldsymbol{\gamma}} j}\|_\infty \le L \zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}); \tag{12.4.16}$$

$$\|B_{\boldsymbol{\gamma} j} - B_{\widetilde{\boldsymbol{\gamma}} j}\|_\psi^2 \le \frac{L^2 6^L 2 \bar{M} m}{N_n} \zeta^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}); \tag{12.4.17}$$

$$\Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} - \sum_j b_j B_{\widetilde{\boldsymbol{\gamma}} j} \Big\|_\psi^2 \le \frac{8 L^2 6^{2L} 38^L \bar{M}^4 m^2}{D^2} \zeta^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} \Big\|_\psi^2; \tag{12.4.18}$$

and

$$\Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} - \sum_j b_j B_{\widetilde{\boldsymbol{\gamma}} j} \Big\|_\infty \le \frac{2mL}{D} \zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \Big\| \sum_j b_j B_{\boldsymbol{\gamma} j} \Big\|_\infty. \tag{12.4.19}$$

Proof. Equation (12.4.11) follows from (12.4.8), with $\bar{M}$ replaced by $3\bar{M}$, and induction; (12.4.12) follows from (12.4.9) and induction; since $\psi(\operatorname{supp}(B_{\boldsymbol{\gamma} j})) = \prod_l [(\gamma_{l,j} - \gamma_{l,j-m_l})/(b_l - a_l)]$, (12.4.13) follows from (12.4.5) with $\bar{M}_l$ replaced by $3\bar{M}_l$.

To verify (12.4.14), let $u_l \in \mathcal{U}_l$ and suppose first that u_l is not a knot. Then $\gamma_{l,j_0} < u_l < \gamma_{l,j_0+1}$ for some j_0. If $B_{\boldsymbol{\gamma}_l j}(u_l) > 0$, then $\gamma_{l,j-m_l} < u_l < \gamma_{l,j}$ and hence $j_0 + 1 \le j \le j_0 + m_l$. Suppose, instead, that $u = \gamma_{l,j_0}$. If $B_{\boldsymbol{\gamma}_l j}(u_l) > 0$, then $\gamma_{l,j-m_l} < \gamma_{l,j_0} < \gamma_{l,j}$, so $j_0 + 1 \le j \le j_0 + m_l - 1$. In either case,

$$\#\{j \in \mathcal{J} : B_{\boldsymbol{\gamma} j} \ne 0\} = \prod_l \#\{j \in \mathcal{J}_l : B_{\boldsymbol{\gamma}_l j}(u_l) \ne 0\} \le \prod_l m_l = m.$$

To verify (12.4.15), given $j \in \mathcal{J}_l$, let k_1 (k_2) be the smallest (largest) value of k in $\mathcal{J}_l$ such that $B_{\boldsymbol{\gamma}_l j} B_{\widetilde{\boldsymbol{\gamma}}_l k}$ is not identically zero. Then $\widetilde{\gamma}_{l,k_1} > \gamma_{l,j-m_l}$ and $\widetilde{\gamma}_{l,k_2-m_l} < \gamma_{l,j}$. It follows from (12.4.5) (with $\bar{M}_l$ replaced by $3\bar{M}_l$) that

$$\gamma_{l,k_2-m_l} < \gamma_{l,j} \le \gamma_{l,j-m_l} + \frac{6\bar{M}_l m_l (b-a)}{J_l + m_l}.$$

Let I be the smallest integer such that $I \ge 6^2 \bar{M}_l^2$. It follows from (12.4.4) that

$$\widetilde{\gamma}_{l,k_1+Im_l} \ge \widetilde{\gamma}_{l,k_1} + \frac{Im_l(b-a)}{2\bar{M}_l(J_l+m_l)} > \gamma_{l,j-m_l} + \frac{Im_l(b-a)}{6\bar{M}_l(J_l+m_l)} \ge \gamma_{l,k_2-m_l}$$

and hence that $k_2 < k_1 + (I+1)m_l$. Consequently,

$$\begin{aligned}\#\{k \in \mathcal{J}_l : B_{\boldsymbol{\gamma}_l j} B_{\widetilde{\boldsymbol{\gamma}}_l k} \text{ is not identically zero on } \mathcal{U}_l\} \\ \le (I+1)m_l \le (6^2\bar{M}_l^2 + 2)m_l \le 38\bar{M}_l^2 m_l,\end{aligned}$$

which yields the desired result.

To verify (12.4.16), we first observe that, as a consequence of Definition 4.12 and Theorems 2.51, 2.55, and 4.27 of Schumaker (1981), the partial derivative of $B_{\gamma_l j}$ with respect to the knot $\gamma_{l,k}$ for $j - m_l \le k \le j$ is bounded in absolute value by

$$\max\left\{\frac{1}{\gamma_{l,j-1} - \gamma_{l,j-m_l}}, \frac{1}{\gamma_{l,j} - \gamma_{l,j+1-m_l}}\right\}.$$

Thus, by (12.4.2),

$$\begin{aligned}\|B_{\boldsymbol{\gamma}_l j} - B_{\widetilde{\boldsymbol{\gamma}}_l j}\|_\infty &\le \frac{3\bar{M}_l(m_l+1)N_l}{(m_l-1)(b_l-a_l)}|\boldsymbol{\gamma}_l - \widetilde{\boldsymbol{\gamma}}_l|_\infty \\ &\le \frac{m_l+1}{3(m_l-1)}\zeta_l(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \le \zeta_l(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}).\end{aligned}$$

The desired result now follows from the observation that normalized B-splines lie between 0 and 1.

Equation (12.4.17) follows from (12.4.13) and (12.4.16).

Set

$$A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}j} = \{k \in \mathcal{J} : \langle B_{\boldsymbol{\gamma}j} - B_{\widetilde{\boldsymbol{\gamma}}j}, B_{\boldsymbol{\gamma}k} - B_{\widetilde{\boldsymbol{\gamma}}k}\rangle_\psi \ne 0\}, \quad \boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma} \text{ and } j \in \mathcal{J}.$$

Then $\#(A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}j}) \le 38^L 4\bar{M}^2 m$ by (12.4.15). Consequently, by (12.4.11) and (12.4.17),

$$\begin{aligned}
&\Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} - \sum_j b_j B_{\widetilde{\boldsymbol{\gamma}}j} \Big\|_\psi^2 \\
&\qquad = \sum_j \sum_{k \in A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}j}} b_j b_k \langle B_{\boldsymbol{\gamma}j} - B_{\widetilde{\boldsymbol{\gamma}}j}, B_{\boldsymbol{\gamma}k} - B_{\widetilde{\boldsymbol{\gamma}}k} \rangle_\psi \\
&\qquad \le \sum_j \sum_{k \in A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}j}} \Big(\frac{b_j^2 + b_k^2}{2} \Big) \Big(\frac{\|B_{\boldsymbol{\gamma}j} - B_{\widetilde{\boldsymbol{\gamma}}j}\|_\psi^2 + \|B_{\boldsymbol{\gamma}k} - B_{\widetilde{\boldsymbol{\gamma}}k}\|_\psi^2}{2} \Big) \\
&\qquad \le \frac{8L^2 6^L 38^L \bar{M}^3 m^2}{N_n} \zeta^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \sum_j b_j^2 \\
&\qquad \le \frac{8L^2 6^{2L} 38^L \bar{M}^4 m^2}{D^2} \zeta^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} \Big\|_\psi^2,
\end{aligned}$$

so (12.4.18) holds.

It follows from (12.4.14) that, for $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \Gamma$ and $\boldsymbol{u} \in \mathcal{U}$, there are at most $2m$ values of $j \in \mathcal{J}$ such that $B_{\boldsymbol{\gamma}j}(\boldsymbol{u}) - B_{\widetilde{\boldsymbol{\gamma}}j}(\boldsymbol{u}) \ne 0$. Thus, by (12.4.12) and (12.4.16),

$$\Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} - \sum_j b_j B_{\widetilde{\boldsymbol{\gamma}}j} \Big\|_\infty \le \frac{2mL}{D} \zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} \Big\|_\infty,$$

so (12.4.19) holds. □

Proof of Lemma 12.2.4. It follows from (12.4.11) and (12.4.12) that

$$\Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} \Big\|_\infty^2 \le \max_j b_j^2 \le \frac{6^L \bar{M} N_n}{D^2} \Big\| \sum_j b_j B_{\boldsymbol{\gamma}j} \Big\|_\psi^2.$$

The desired result now follows from (12.2.2). □

Recall that $\boldsymbol{U}$ is defined as a transform of $\boldsymbol{W}$. Let $\boldsymbol{U}_1, \dots, \boldsymbol{U}_n$ be the corresponding transforms of $\boldsymbol{W}_1, \dots, \boldsymbol{W}_n$, respectively. Recall the definition of empirical inner product and empirical norm in Section 12.2. Observe that $E_n(h) = \langle 1, h \rangle_n$.

Lemma 12.4.16. *Suppose Condition* 12.2.1 *holds and that* $N_n = o(n^{1-\epsilon})$ *for some* $\epsilon > 0$. *Then there is a constant* M *and there is an event* Ω_n *such that* $\lim_n P(\Omega_n) = 1$ *and*

$$\Big\| \sum_j \beta_j B_{\boldsymbol{\gamma}j} - \sum_j \beta_j B_{\widetilde{\boldsymbol{\gamma}}j} \Big\|_n^2 \le M \zeta^2(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}) \Big\| \sum_j \beta_j B_{\boldsymbol{\gamma}j} \Big\|^2 \qquad \text{on } \Omega_n$$

for $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ *and* $\beta_j \in \mathbb{R}$ *for* $j \in \mathcal{J}$.

Proof. It follows from (12.4.16) that

$$\begin{aligned}\|B_{\boldsymbol{\gamma}j}-&B_{\widetilde{\boldsymbol{\gamma}}j}\|_n^2\\ &\le \|B_{\boldsymbol{\gamma}j}-B_{\widetilde{\boldsymbol{\gamma}}j}\|_\infty^2\frac{1}{n}\#\big(\{i:\boldsymbol{U}_i\in\operatorname{supp}(B_{\boldsymbol{\gamma}j})\cup\operatorname{supp}(B_{\widetilde{\boldsymbol{\gamma}}j})\}\big)\\ &\le L^2\zeta^2(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}})\frac{1}{n}\#\big(\{i:\boldsymbol{U}_i\in\operatorname{supp}(B_{\boldsymbol{\gamma}j})\cup\operatorname{supp}(B_{\widetilde{\boldsymbol{\gamma}}j})\}\big)\end{aligned}$$

for $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}$ and $j\in\mathcal{J}$. It follows from Condition 12.2.1, (12.4.13), and the assumption on N_n by a straightforward application of Bernstein's inequality (11.4.2) [or by Theorem 12.2 of Breiman, Friedman, Olshen and Stone (1984)] that

$$\sup_{\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}}\max_{j\in\mathcal{J}}\frac{1}{n}\#\big(\{i:\boldsymbol{U}_i\in\operatorname{supp}(B_{\boldsymbol{\gamma}j})\cup\operatorname{supp}(B_{\widetilde{\boldsymbol{\gamma}}j})\}\big)\le\frac{6^L2\bar{M}M_2m}{N_n}[1+o_P(1)].$$

Let Ω_n denote the event that

$$\sup_{\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}}\max_{j\in\mathcal{J}}\frac{1}{n}\#\big(\{i:\boldsymbol{U}_i\in\operatorname{supp}(B_{\boldsymbol{\gamma}j})\cup\operatorname{supp}(B_{\widetilde{\boldsymbol{\gamma}}j})\}\big)\le\frac{6^L4\bar{M}M_2m}{N_n}.$$

Then $\lim_n P(\Omega_n)=1$ and

$$\|B_{\boldsymbol{\gamma}j}-B_{\widetilde{\boldsymbol{\gamma}}j}\|_n^2\le\frac{4L^26^L\bar{M}M_2m}{N_n}\zeta^2(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}})\text{ on }\Omega_n \tag{12.4.20}$$

for $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}$ and $j\in\mathcal{J}$.

Set $A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}jn}=\{k\in\mathcal{J}:\langle B_{\boldsymbol{\gamma}j}-B_{\widetilde{\boldsymbol{\gamma}}j},B_{\boldsymbol{\gamma}k}-B_{\widetilde{\boldsymbol{\gamma}}k}\rangle_n\ne0\}$ for $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}$ and $j\in\mathcal{J}$. Then $\#(A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}jn})\le 38^L4\bar{M}^2m$ by (12.4.15). Consequently, by (12.4.11), (12.4.20), and Condition 12.2.1,

$$\begin{aligned}\Big\|\sum_j\beta_jB_{\boldsymbol{\gamma}j}-\sum_j\beta_jB_{\widetilde{\boldsymbol{\gamma}}j}\Big\|_n^2&\\ &=\sum_k\sum_{k\in A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}jn}}\beta_j\beta_k\langle B_{\boldsymbol{\gamma}j}-B_{\widetilde{\boldsymbol{\gamma}}j},B_{\boldsymbol{\gamma}k}-B_{\widetilde{\boldsymbol{\gamma}}k}\rangle_n\\ &\le\sum_j\sum_{k\in A_{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}jn}}\Big(\frac{\beta_j^2+\beta_k^2}{2}\Big)\Big(\frac{\|B_{\boldsymbol{\gamma}j}-B_{\widetilde{\boldsymbol{\gamma}}j}\|_n^2+\|B_{\boldsymbol{\gamma}k}-B_{\widetilde{\boldsymbol{\gamma}}k}\|_n^2}{2}\Big)\\ &\le\frac{16L^26^L38^L\bar{M}^3M_2m^2}{N_n}\zeta^2(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}})\sum_j\beta_j^2\\ &\le\frac{16L^26^{2L}38^L\bar{M}^4M_2m^2}{D^2M_1}\zeta^2(\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}})\Big\|\sum_j\beta_jB_{\boldsymbol{\gamma}j}\Big\|^2\end{aligned}$$

on Ω_n for $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}}\in\widetilde{\Gamma}$ and $\beta_j\in\mathbb{R}$ for $j\in\mathcal{J}$, as desired. □

Proof of Lemma 12.2.5. Set $\epsilon = \zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})$. Write $g = \sum_j \beta_j B_{\boldsymbol{\gamma}j}$ and set $g' = \sum_j \beta_j B_{\widetilde{\boldsymbol{\gamma}}j}$. It follows from (12.2.2) and (12.4.18) that $\|g - g'\| \le c_1\epsilon\|g\|$ for some constant c_1, and it follows from (12.4.19) that $\|g - g'\|_\infty \le c_2\epsilon\|g\|_\infty$ for some constant c_2.

If $\|g'\| \le \|g\|$, then $\widetilde{g} = g'$ has the properties specified in the first result of the lemma. Suppose, instead, that $\|g'\| > \|g\|$ and set $\lambda = \|g\|/\|g'\|$. Then $(1 + c_1\epsilon)^{-1} \le \lambda < 1$, $\|\lambda g'\| = \|g\|$, and $\|g - \lambda g'\| \le \|g - g'\| \le c_1\epsilon\|g\|$. (Note that $\langle g, g'\rangle \le \|g\|\|g'\|$ by the Cauchy–Schwarz inequality.) Moreover,

$$\begin{aligned}\|g' - \lambda g'\|_\infty &= (\|g'\| - \|g\|)\frac{\|g'\|_\infty}{\|g'\|} \\ &\le \frac{\|g' - g\|(\|g\|_\infty + \|g' - g\|_\infty)}{\|g\|} \\ &\le c_1\epsilon(1 + c_2)\|g\|_\infty,\end{aligned}$$

so $\|g - \lambda g'\|_\infty \le (c_1 + c_2 + c_1c_2)\epsilon\|g\|_\infty$ and hence $\widetilde{g} = \lambda g'$ has the properties specified in the first result.

Let Ω_{n1} be the event Ω_n in Lemma 12.4.16, let Ω_{n2} be the event that $\|g\|_n^2 \le 2\|g\|^2$ for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$, and set $\Omega_n = \Omega_{n1} \cup \Omega_{n2}$. It follows from Lemmas 12.4.16 and 12.3.7 that $\lim_n P(\Omega_n) = 1$. Let ϵ, g', and λ be as in the proof of the first result of the lemma. Then for some constant c_3, $\|g - g'\|_n \le c_3\epsilon\|g\|$ on Ω_n. If $\|g'\| \le \|g\|$, then $\widetilde{g} = g'$ satisfies the desired additional property. Otherwise,

$$\begin{aligned}\|g - \lambda g'\|_n &\le \|g - g'\|_n + (1 - \lambda)\|g'\|_n \\ &\le c_3\epsilon\|g\| + 2\Big(\frac{1}{\lambda} - 1\Big)\|g\| \\ &\le (c_3 + 2c_1)\epsilon\|g\|\end{aligned}$$

on Ω_n, so $\widetilde{g} = \lambda g'$ satisfies the desired additional property. □

Proof of Lemma 12.2.6. Let $K_1 > K$. Choose $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}} \in \widetilde{\Gamma}$ such that $\|\bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \le K$ and $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \le K$, and set $\epsilon = \zeta(\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}})$. By Lemma 12.2.5, there is a fixed positive number c_1 (not depending on $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}$) and there are functions $\eta'_{\boldsymbol{\gamma}} \in \mathbb{G}_{\boldsymbol{\gamma}}$ and $\eta'_{\widetilde{\boldsymbol{\gamma}}} \in \mathbb{G}_{\widetilde{\boldsymbol{\gamma}}}$ such that $\|\eta'_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \le c_1\epsilon$ and $\|\eta'_{\widetilde{\boldsymbol{\gamma}}} - \bar{\eta}_{\boldsymbol{\gamma}}\|_\infty \le c_1\epsilon$. Without loss of generality, we can assume that $\epsilon \le 1$ and that ϵ is sufficiently small that $\|\eta'_{\boldsymbol{\gamma}}\|_\infty \le K_1$ and $\|\eta'_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \le K_1$. Then, by Condition 12.1.2, there is a fixed positive number c_2 such that $\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta'_{\widetilde{\boldsymbol{\gamma}}}) \le c_2\epsilon$ and $\Lambda(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}) - \Lambda(\eta'_{\boldsymbol{\gamma}}) \le c_2\epsilon$. Since $\Lambda(\eta'_{\widetilde{\boldsymbol{\gamma}}}) \le \Lambda(\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}})$, we conclude that $\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta'_{\boldsymbol{\gamma}}) \le 2c_2\epsilon$. On the other hand, by Condition 12.1.2, $\Lambda(\bar{\eta}_{\boldsymbol{\gamma}}) - \Lambda(\eta'_{\boldsymbol{\gamma}}) \ge c_3\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\|^2$ for some constant c_3, so $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\| \le (2c_2c_3^{-1}\epsilon)^{1/2}$ and hence

$$\begin{aligned}\|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\| &\le \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\| + \|\eta'_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\| \\ &\le (2c_2c_3^{-1}\epsilon)^{1/2} + c_1\epsilon \\ &\le [(2c_2c_3^{-1})^{1/2} + c_1]\epsilon^{1/2}.\end{aligned}$$

Moreover, by (12.2.2), (12.4.11), and (12.4.12), $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\|_\infty \le c_4 N_n^{1/2} \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\|$. So,

$$\begin{aligned} \|\bar{\eta}_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty &\le \|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\|_\infty + \|\eta'_{\boldsymbol{\gamma}} - \bar{\eta}_{\widetilde{\boldsymbol{\gamma}}}\|_\infty \\ &\le \left(\frac{2c_2c_4^2}{c_3}\right)^{1/2} N_n^{1/2}\epsilon^{1/2} + c_1\epsilon. \end{aligned}$$

By Lemma 12.3.7, there is an event Ω_n such that $\lim_n P(\Omega_n) = 1$ and $\|g\|_n \le 2\|g\|$ on Ω_n for $\boldsymbol{\gamma} \in \widetilde{\Gamma}$ and $g \in \mathbb{G}_{\boldsymbol{\gamma}}$. Thus, by the first paragraph of this proof, $\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\|_n \le 2\|\bar{\eta}_{\boldsymbol{\gamma}} - \eta'_{\boldsymbol{\gamma}}\| \le 2(2c_2c_3^{-1}\epsilon)^{1/2}$ and hence $\|\bar{\eta}_{\widetilde{\boldsymbol{\gamma}}} - \bar{\eta}_{\boldsymbol{\gamma}}\|_n \le [2(2c_2c_3^{-1})^{1/2} + c_1)]\epsilon^{1/2}$ on Ω_n for $\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}$ as in the first paragraph. □