Convergence Rates for Smoothing Spline Estimators in Varying Coefficient Models

P.P.B. Eggermont†, R.L. Eubank‡*, V.N. LaRiccia‡

†University of Delaware and ‡Arizona State University

May 18, 2005

Abstract

We consider the estimation of a multiple regression model in which the coefficients change slowly in “time”, with “time” being an additional covariate. Under reasonable smoothness conditions, we prove the usual expected mean square error bounds for the smoothing spline estimators of the coefficient functions.

1. Introduction

Since their introduction by Hastie and Tibshirani (1993), varying coefficient models have become an increasingly popular option for dimension reduction in nonparametric regression with multiple predictors. An important special case of the general varying coefficient formulation is the time-varying coefficient model that utilizes only one effect-modifying covariate (“time”). These later models have applications in various contexts such as functional regression analysis, see, e.g., Hoover, Chiang, Rice and Wu (2001) or Eubank et al. (2004), and the analysis of longitudinal data, e.g., Wu and

*Research supported by NSF

Keywords and phrases: Time-varying coefficient model, spline smoothing, convergence rates, reproducing kernel Hilbert space.

AMS 2000 Subject Classification: 62G08, 62G20.
Chiang (2000). In this paper, we study the efficacy of smoothing spline estimation in time-varying coefficient models.

To be precise, we consider the data

\[(Y_{in}, X_{in}, t_{in}), \quad i = 1, 2, \ldots, n,\]

where the \(Y_{in}\) are the responses, the \(X_{in} \in \mathbb{R}^{1 \times p}\) are the predictors (here, \(p\) is some fixed positive integer), and the time points \(t_{in}\) are the additional covariates. It is assumed that locally (for nearby \(t_{in}\)), the usual linear model

\[Y_{in} = X_{in} \beta + \varepsilon_{in} \quad i = 1, 2, \ldots, n,\]

provides a good fit for some fixed \(\beta \in \mathbb{R}^{p}\), but not globally in that \(\beta\) varies with time. Then, it is sensible to let the predictors change with time as well, so that the new and improved model is

\[Y(t_{in}) = X(t_{in}) \beta(t_{in}) + \varepsilon(t_{in}), \quad i = 1, 2, \ldots, n,\]

where \(\beta(t)\) is a smooth vector-valued function of time, \(X(t)\) is a suitable stochastic process modeling a random design, and \(\varepsilon(t)\) is white noise independent of the \(X(t)\)-process with (unknown) variance \(\sigma^2\).

Thus, we have a family of linear models (1.3), with one observation per model. If the model is changing smoothly, then we can pretend that observations for nearby models are observations for the “current” model and we can perform the usual multiple regression estimation. Note that the preceding implies that the \(X(t_{in})\) should vary anything but smoothly with time. As a matter of fact, we want the nearby \(X(t_{in})\) to be as far apart as possible. The technical assumption is that \(X(t_{1,n}), X(t_{2,n}), \ldots, X(t_{n,n})\) are mutually independent for all time points \(t_{in}\). (If there are replicate time points, there is a slight notational hitch we shall ignore.)

In this paper, we prove convergence rates for the smoothing spline estimator of \(\beta(t)\) in the model (1.3). This estimator was proposed (but not further studied) by Hoover, Rice, Wu and Chiang (2004), and is defined as the solution to

\[
\text{minimize } \frac{1}{n} \sum_{i=1}^{n} \left| X(t_{in}) b(t_{in}) - Y(t_{in}) \right|^2 + \sum_{j=1}^{p} h_{j}^{2m} \left\| b_{j}^{(m)} \right\|^2 ,
\]

where the minimization is over all smooth functions \(b_{j}(t), \ j = 1, 2, \ldots, p\). Here, \(\left\| b_{j}^{(m)} \right\|\) denotes the \(L^2\)-norm of the \(m\)-th derivative of \(b_{j}\). The existence and uniqueness of the estimator follows, e.g., from Eubank et al.
(2004), provided that the matrix $H \in \mathbb{R}^{n \times mp}$ with $i$-th row defined as the Kronecker product

$$X(t_{in}) \otimes (1, t_{in}, t_{in}^2, \ldots, t_{in}^{m-1}),$$

has full column rank. As an aside, the same authors show that the resulting estimator can be computed efficiently using the Kalman filter (with associated Bayesian confidence intervals) in $O(n)$ operations. Also, note that the usual spline smoothing problem is subsumed in (1.4) by taking $p = 1$ and $X(t) = 1$ for all $t$. Taking some of the $b_j(t)$ constant over time would cover partially linear models, see, e.g., Green, Jennison and Seheult (1985). Finally, note that in (1.4), each coefficient of $b(t)$ has its own smoothing parameter. It is a question of practical importance whether these $h_j$ can be chosen (near) optimally by data-driven methods, but we shall not address this issue here.

A number of authors have studied the large sample properties of kernel estimators for $\beta(t)$ in varying coefficient models. Convergence rates for Nadaraya-Watson kernel estimators in longitudinal versions of (1.2) with time varying covariates have been derived in Wu, Chiang and Hoover (1998), Hoover, Rice, Wu and Yang (1998), Wu, Yu and Chiang (2000) and Wu and Chiang (2000). Similar results for local polynomial based estimators are provided in Fan and Zhang (1999, 2000), and Cai, Fan and Li (2000).

For the univariate case, when in addition, there are replicated responses at each time $t_{in}$, Chiang, Rice and Wu (2001) employ a smoothing spline estimator differing from (1.4). In this situation, they collect the observed $Y$ responses at each $t_{in}$ and obtain pilot estimators of the $\beta(t_{in})$ via simple linear regression. Thus, an unbiased estimator of $\beta(t_{in})$ is provided by the minimizer $b = \hat{\beta}(t_{in})$ of

$$\minimize_{b \in \mathbb{R}^p} \sum_{t=t_{in}} |X(t) b - Y(t)|^2$$

Finally, they estimate the coefficient function by smoothing the $\hat{\beta}(t_{in})$, rather than the $Y(t_{in})$ as in (1.4), using a cubic ($m = 2$) smoothing spline. Chiang, Rice and Wu (2001) obtain point-wise converge rates of order $n^{-8/9}$ when $\beta(t)$ has four Lipschitz-continuous derivatives and satisfies the natural boundary conditions $\beta^{(2)}(0) = \beta^{(2)}(1) = \beta^{(3)}(0) = \beta^{(3)}(1) = 0$.

In contrast to Chiang, Rice and Wu (2001), we do not assume replicate responses and, instead, deal with the estimator from (1.4) while explicitly
allowing \( t \) and \( X \) to be related. Under suitable minimal conditions, we derive the \( \mathcal{O}(n^{-2m/(2m+1)}) \) bound for the mean squared error of the spline estimator (1.4). When the above mentioned natural boundary conditions hold for each coefficient of \( \beta(t) \), then the rate improves to \( \mathcal{O}(n^{-4m/(4m+1)}) \).

The paper is laid out as follows. In the next section, we formulate the assumptions and state the theorem on the convergence rates. The remaining sections develop the proof of theorem. Most notably, in \( \S \) 3, we introduce the Sobolev spaces \( W^{m,2}(0,1) \) with suitable inner products depending on the (scalar) smoothing parameter \( \lambda \) and the associated reproducing kernels. The reproducing kernels are used to show appropriate bounds on random sums of the kind

\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) f(t_{in})
\]

where \( \varepsilon(t) \) is a white noise process and \( f \) is a random function in \( W^{m,2}(0,1) \).

In \( \S \) 4, we prove the main theorem on convergence rates, formulating various lemmas which are then proved in later sections.

### 2. The Problem, Assumptions and Main Result

In this section, we give a precise description of the estimation problem and state the main result on the convergence rates of the estimator.

We consider the data

\[
(2.1) \quad (Y(t_{in}), X(t_{in}), t_{in}) , \quad i = 1, 2, \ldots, n ,
\]

following the model

\[
(2.2) \quad Y(t) = X(t) \beta(t) + \varepsilon(t) , \quad 0 \leq t \leq 1 .
\]

Here, \( \beta(t) \) is a deterministic vector-valued function with values in \( \mathbb{R}^p \) for some fixed integer \( p \), and \( X(t) \) is a random function with values in \( \mathbb{R}^{1 \times p} \).

The noise process \( \varepsilon(t) \) is independent of \( X(t) \) and satisfies

\[
(2.3) \quad \mathbb{E}[\varepsilon(t)] = 0 , \quad \mathbb{E}[|\varepsilon(t)|^2] = \sigma^2 , \quad \mathbb{E}[\varepsilon(t) \varepsilon(s)] = 0 \text{ for } t \neq s .
\]

The “design” process \( X(t) \) is assumed to be independent and bounded, i.e.,

\[
(2.4) \quad X(t_{1,n}), X(t_{2,n}), \ldots, X(t_{n,n}) \text{ are mutually independent ,}
\]
and there exists a constant $C$ such that for all $t$,

\[(2.5) \quad \|X(t)\|_{\mathbb{R}^p} \leq C \text{ almost surely.}\]

Moreover, it is required to be “full” in that $\mathbb{E}[X(t)^TX(t)]$ should be positive definite, uniformly in $t$, i.e., for some positive constant $\rho$,

\[(2.6) \quad \mathbb{E}[X(t)^TX(t)] - \rho I \text{ is semi-positive-definite for all } t.\]

Finally, the time points $t_{1,n}, t_{2,n}, \ldots, t_{n,n}$ are assumed to be deterministic and super-quasi-uniformly distributed on $[0, 1]$. (More or less equally spaced time points will do. For the precise statement, see Definition 1 in § 4.) The smoothness requirement on $\beta$ is interpreted as

\[(2.7) \quad \|\beta\|_2^2 + \|\beta_{j}^{(m)}\|_2^2 \leq C_1, \quad j = 1, 2, \ldots, p,\]

for a (known) integer $m \geq 1$ and an (unknown) constant $C_1$. Here, $\| \cdot \|$ denotes the $L^2$ norm,

\[(2.8) \quad \|f\|_2^2 = \langle f, f \rangle \text{ where } \langle f, g \rangle = \int_0^1 f(t)g(t)\,dt .\]

Thus, the function space of interest is, for $m \geq 1$,

\[(2.9) \quad W^{m,2}(0,1) = \left\{ f \in C[0,1] : f^{(m)}(t) \text{ absolutely continuous} \quad f^{(m)} \in L^2(0,1) \right\} .\]

We then concisely state the smoothness assumption (2.7) on $\beta$ as

\[(2.10) \quad \beta \in (W^{m,2}(0,1))^p .\]

The estimator of $\beta$ under consideration here is the solution $b_{nh}$ of the spline smoothing problem

\[(2.11) \quad \text{minimize } S_n(b) + J_h(b) \text{ subject to } b \in (W^{m,2}(0,1))^p ,\]

where

\[(2.12) \quad S_n(b) = \frac{1}{n} \sum_{i=1}^n \left| X(t_{in})b(t_{in}) - Y(t_{in}) \right|^2 ,\]

\[(2.13) \quad J_h(b) = \sum_{j=1}^p h_j^{2m} \|b_j^{(m)}\|^2 .\]
Here, the $h_j$ are the smoothing parameters, and as typical, have to be chosen appropriately. The problem (2.11) always has solutions, while the condition (1.5) guarantees uniqueness. Also, the usual considerations show that the coefficients of the solution(s) are natural splines of degree $2m - 1$. For more on splines smoothing, see, e.g., WAHBA (1990) and EUBANK (1999).

We have the following result on the mean squared error of $b^{nh}$.

**Theorem 1. (Convergence Rates)** (a) Under the model (2.1)–(2.7) and super-quasi-uniformly distributed, deterministic time points,

$$
E\left[\sum_{j=1}^{p} \| b_j^{nh} - \beta_j \|^2 \right] = O\left( \sum_{j=1}^{m} h_j^{2m} + (nh_j)^{-1} \right),
$$

for deterministic $h_j$ with $h_j \to 0$ and $nh_j^2/\log n \to \infty$.

(b) If in addition, $\beta \in (W^{2m,2}(0,1))^p$ and the components of $\beta$ satisfy the natural boundary conditions

$$
\beta_j^{(k)}(0) = \beta_j^{(k)}(1) = 0, \quad k = m, m + 1, \ldots, 2m - 1,
$$

for all $j$, then the bias terms $h_j^{2m}$ in (a) improve to $h_j^{4m}$, i.e.,

$$
E\left[\sum_{j=1}^{p} \| b_j^{nh} - \beta_j \|^2 \right] = O\left( \sum_{j=1}^{m} h_j^{4m} + (nh_j)^{-1} \right).
$$

It follows that if $h_j \asymp n^{-1/(2m+1)}$ for all $j$, then we get the usual rate $n^{-2m/(2m+1)}$ for the mean integrated squared error, and the rate $n^{-4m/(4m+1)}$ for $h_j \asymp n^{-1/(4m+1)}$ when the natural boundary conditions apply.

**3. The Setting**

In this section, we lay the groundwork for the proof of the main theorem. The observation that suggests everything that follows is that the coefficient functions $\beta_j(t)$ live in a reproducing kernel Hilbert space (not a surprise in the spline world) with inner products depending on a smoothing parameter. The associated reproducing kernels turn out to be the right tool for studying the random sums that pop-up in various places. This is the approach for ordinary spline smoothing, taken in EGGERMONT and LARICCIA (in preparation), which carries over nicely to the present problem.
In §2, we already made the case for $W^{m,2}(0,1)$ being the appropriate space of functions. The smoothness conditions on the components $\beta_j$ of $\beta$ may then be expressed as

$$\|\beta_j\|_{W^{m,2}(0,1)}^2 \leq C,$$

where for all $f$,

$$(3.1) \quad \| f \|_{W^{m,2}(0,1)} = \{ \| f \|^2 + \| f^{(m)} \|^2 \}^{1/2}.$$

However, in view of the spline smoothing problem, the norms $\| \cdot \|_{m,\lambda}$ and inner products $\langle \cdot, \cdot \rangle_{m,\lambda}$ suggest themselves,

$$(3.2) \quad \| f \|_{m,\lambda} = \{ \langle f, f \rangle_{m,\lambda} \}^{1/2},$$

$$\langle f, g \rangle_{m,\lambda} = \langle f, g \rangle + \lambda^{2m} \langle f^{(m)}, g^{(m)} \rangle.$$ 

It is well-known that $W^{m,2}(0,1)$ is a reproducing kernel Hilbert space but that the reproducing kernel depends on the choice of the inner product. For the inner product (3.2), we denote the reproducing kernel as $R_{m\lambda}(t, \cdot)$. Thus, $R_{m\lambda}(t, \cdot) \in W^{m,2}(0,1)$ for all $t$ and satisfies

$$(3.3) \quad f(t) = \langle R_{m\lambda}(t, \cdot), f \rangle_{m,\lambda} \quad \text{for all} \quad t \in [0, 1], \quad f \in W^{m,2}(0,1).$$

It is not too hard to show that uniformly in $t$,

$$(3.4) \quad \| R_{m\lambda}(t, \cdot) \|_{m,\lambda} = O(\lambda^{-1/2}), \quad \lambda \to 0,$$

e.g., by proving that $W^{m,2}(0,1)$ is a reproducing kernel Hilbert space in the form of the inequalities

$$(3.5) \quad |f(t)| \leq c \lambda^{-1/2} \| f \|_{1,\lambda} \leq c_m \lambda^{-1/2} \| f \|_{m,\lambda},$$

for suitable constants $c$ and $c_m$, independent of $t$. See, e.g., ADAMS and FOURNIER (2003).

The reproducing kernels $R_{m\lambda}$ are useful for determining the behavior of sums

$$\frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) f(t_{in})$$

with random $\varepsilon(t)$ as in (2.3) and random $f$, i.e., $f$ a function of the $\varepsilon(t_{in})$. 
Theorem 2. (Bounds on Random Sums) Let \( \theta_n = (\theta_{1,n}, \theta_{2,n}, \ldots, \theta_{n,n})^T \) be a vector of random variables satisfying
\[
E[\theta_n] = 0 \quad , \quad E[\theta_n \theta_n^T] = \sigma_\theta^2 I_{n \times n},
\]
with \( \sigma_\theta < \infty \). Then, there exists a constant \( c \) such that for all (random) \( f \in W^{m,2}(0,1) \), and all deterministic \( \lambda, 0 < \lambda < 1 \),
\[
E\left[ \frac{1}{n} \sum_{i=1}^n \theta_i \, f(t_i) \right] \leq c \, (n\lambda)^{-1/2} \left( E\left[ \| f \|_{m,\lambda}^2 \right] \right)^{1/2}.
\]
Proof. With (3.3), we have for all \( f \in W^{m,2}(0,1) \), random or not,
\[
\frac{1}{n} \sum_{i=1}^n \theta_i \, f(t_i) = \langle \frac{1}{n} \sum_{i=1}^n \theta_i \mathcal{R}_{m,\lambda}(t_i, \cdot) , f \rangle_{m,\lambda} 
\]
\[
\leq \| \frac{1}{n} \sum_{i=1}^n \theta_i \mathcal{R}_{m,\lambda}(t_i, \cdot) \|_{m,\lambda} \| f \|_{m,\lambda}.
\]
(3.6)

Now, employing (3.4), one shows that
\[
E\left[ \left\| \frac{1}{n} \sum_{i=1}^n \theta_i \mathcal{R}_{m,\lambda}(t_i, \cdot) \right\|_{m,\lambda}^2 \right] = O\left( (n\lambda)^{-1} \right),
\]
and then Cauchy-Schwarz applied to (3.6) gives the required result. \( \square \)

This approach to random sums should be contrasted with the metric entropy approach. See, e.g., VAN DE GEER (2000) and references therein.

4. The Proof of Theorem 1

The starting point in the proof is the following inequality. It is useful to introduce the error
\[
\delta(t) = b_{nh}(t) - \beta(t), \quad 0 \leq t \leq 1.
\]

Lemma 1. The error \( \delta(t) \) satisfies
\[
\frac{2}{n} \sum_{i=1}^n \left| X(t_i) \delta(t_i) \right|^2 + J_h(\delta) \leq \frac{2}{n} \sum_{i=1}^n \epsilon(t_i) X(t_i) \delta(t_i) + J_h(\beta) - J_h(b_{nh}).
\]
The proof of this lemma does not use much beyond the fact that (2.11) is a quadratic minimization problem, see, e.g., van de Geer (2000). The hard work is to show that the inequality implies that
\[ \Delta^2 \leq c (n\mathcal{H})^{-1/2} \Delta + c \mathcal{H}^{2m}, \]
where \( \Delta^2 = \mathbb{E} \left[ \sum_{j=1}^{p} \| \delta_j \|_{mh}^2 \right] \)
and \( \mathcal{H} \) is the maximum of the \( h_j \). This would imply part (a) of the theorem.

We first state bounds on the various parts appearing in the inequality of Lemma 1, and then prove the theorem. The bounds themselves are proved in the following sections.

**Lemma 2.** Let \( m \geq 1 \). For the model (2.1)–(2.7), there exists a constant \( c \) such that for random \( v \in \left( W^{m,2}(0,1) \right)^p \),
\[
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X(t_{in}) v(t_{in}) \right] \leq c \sum_{j=1}^{p} (nh_j)^{-1/2} \left( \mathbb{E} \left[ \| v_j \|_{m,h_j}^2 \right] \right)^{1/2},
\]
provided \( nh_j \to \infty, h_j \to 0 \) for all \( j \).

Scanning our progress so far, we now have the \( L^2 \) integral of \( \delta \) on the right of the inequality (4.2), but a sum on the left. We need a quadrature result for “regular” distributions of the time points \( t_{in} \).

**Definition 1.** The family of time points \( \{ t_{in} : i = 1, 2, \ldots, n \} \) is quasi-uniform if there exists a constant \( c \) such that for all \( f \) with integrable derivative,
\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(t_{in}) - \int_0^1 f(t) \, dt \right| \leq c n^{-1} \| f' \|_{L^1(0,1)}.
\]
The family of time points \( \{ t_{in} : i = 1, 2, \ldots, n \} \) is super-quasi-uniform if there exists positive constants \( c \) and \( c_1 \) such that for all \( f \) with integrable derivative,
\[
\frac{1}{n} \sum_{i=1}^{n} f(t_{in}) \geq c \int_0^1 f(t) \, dt - c_1 n^{-1} \| f' \|_{L^1(0,1)}.
\]

It is a standard exercise from numerical analysis to show that both of the uniform time point designs
\[
(4.3) \quad t_{in} = \frac{i-1}{n-1} \quad \text{and} \quad t_{in} = \frac{i-1/2}{n}, \quad i = 1, 2, \ldots, n,
\]
are quasi-uniform in the sense of Definition 1. See, e.g., Kincaid and Cheney (1990) or §7. A family of time points is super-quasi-uniform if it "contains" a quasi-uniform family consisting asymptotically of at least \( r n \) points for some fixed \( r > 0 \).

The above two lemmas get combined into the following.

**Lemma 3.** Suppose the time points \( \{ t_{in} : i = 1, 2, \cdots, n \} \) are super-quasi-uniform. Then, there exists positive constants \( c \) and \( c_1 \) such that for all random \( v \in (W^{m,2}(0,1))^p \), and all \( h_j \) satisfying \( nh_j^2/\log n \to \infty, h_j \to 0 \),

\[
\mathbb{E}\left[ \frac{1}{n} \sum_{i=1}^{n} \left| X(t_{in}) v(t_{in}) \right|^2 \right] \geq \sum_{j=1}^{p} \left\{ c \mathbb{E}\left[ \| v_j \|_2 \right] - c_1 (nH^2/\log n)^{-1/2} \mathbb{E}\left[ \| v_j \|_{m,h_j}^2 \right] \right\},
\]

with \( H^{-1} = \sum_{j=1}^{p} h_j^{-1} \).

**Proof of Theorem 1.** Consider the equality (4.2). Applying Lemma 3 to the left hand side gives the lower bound, for a positive constant \( c \),

\[
( c - c_1 (nH^2/\log n)^{-1/2}) \sum_{j=1}^{n} \eta_j^2 ,
\]

where, with \( \delta_j = b_{nh}^j - \beta \),

\[
\eta_j^2 = \mathbb{E}[\| \delta_j \|_{m,h_j}^2] .
\]

Of course, under the conditions of Theorem 1 on the \( h_j \), we may ignore the \( (nH^2/\log n)^{-1/2} \) term in the above lower bound.

Then, applying Lemma 2 to the right hand side of (4.2) gives

\[
(4.6) \quad c \sum_{j=1}^{p} \eta_j^2 \leq c_1 \sum_{j=1}^{p} (nh_j)^{-1/2} \eta_j + J_h(\beta) - \mathbb{E}[J_h(b_{nh})] .
\]

Since we may obviously drop the term \( -\mathbb{E}[J_h(b_{nh})] \) in (4.6), and since the assumption \( \beta \in (W^{m,2}(0,1))^p \) implies that

\[
J_h(\beta) \leq c \sum_{j=1}^{p} h_j^{2m} ,
\]
then part (a) of Theorem 1 follows.

To prove part (b), we take a closer look at the term $J_h(\beta) - \mathbb{E}[J_h(b^{nh})]$ on the right hand side of the inequality (4.6). Consider the identity, valid for all $f, g \in W^{m,2}(0,1)$,

$$
\| f^{(m)} \|^2 - \| g^{(m)} \|^2 = 2 \left\langle f^{(m)}, f^{(m)} - g^{(m)} \right\rangle - \| f^{(m)} - g^{(m)} \|^2.
$$

Obviously, we need not worry about the term $-\| f^{(m)} - g^{(m)} \|^2$. Now, suppose that $f \in W^{2m,2}(0,1)$ and that $f$ satisfies the natural boundary conditions

$$
f^{(k)}(0) = f^{(k)}(1) = 0, \quad k = m, m+1, \ldots, 2m - 1.
$$

Then, integration by parts $m$ times yields

$$
\left\langle f^{(m)}, f^{(m)} - g^{(m)} \right\rangle = (-1)^m \left\langle f^{(2m)}, f - g \right\rangle,
$$

which may be bounded by $\| f^{(2m)} \| \| f - g \|$.

Applying this to each component of $J_h(\beta) - J_h(b^{nh})$ gives

$$
J_h(\beta) - J_h(b^{nh}) \leq \sum_{j=1}^{p} 2 h_{j}^{2m} \| \beta^{(2m)}_j \| \| \delta_j \|,
$$

so that after taking expectations (and Cauchy-Schwarz),

$$
J_h(\beta) - \mathbb{E}[J_h(b^{nh})] \leq \sum_{j=1}^{p} 2 h_{j}^{2m} \| \beta^{(2m)}_j \| \eta_j,
$$

with $\eta_j$ as in (4.6). Substituting the above bound into (4.5) then gives

$$
c \sum_{j=1}^{p} \eta_j^2 \leq c_2 \sum_{j=1}^{p} \left\{ (nh_{j})^{-1/2} + h_{j}^{2m} \right\} \eta_j,
$$

and part (b) follows.

This completes the proof of the theorem. In the remaining sections, we prove the lemmas.

5. Quadrature

In this section, we first prove that the uniform time points (4.3) are indeed quasi-uniform in the sense of Definition 1. We also prove Lemma 3.
Lemma 4. For the uniform time points \( t_{in} = \frac{i-1}{n-1} \), \( i = 1, 2, \ldots, n \), and for every \( f \) with integrable derivative,

\[
\left| \frac{1}{n} \sum_{i=1}^{n} f(t_{in}) - \int_{0}^{1} f(t) \, dt \right| \leq \frac{1}{n-1} \int_{0}^{1} |f'(t)| \, dt .
\]

For the time points \( t_{in} = \frac{i-1/2}{n} \), \( i = 1, 2, \ldots, n \), the inequality holds with the factor \( \frac{1}{n-1} \) replaced by \( \frac{1}{n} \).

Proof. We only consider the first part. The first step is the following amusing identity,

\[
(5.1) \quad \frac{1}{n} \sum_{i=1}^{n} c_{in} = \frac{1}{n-1} \sum_{i=1}^{n-1} \left\{ a_{in} c_{in} + b_{in} c_{i+1,n} \right\} ,
\]

for all \( c_{in} \), \( i = 1, 2, \ldots, n \), where \( a_{in} = (n-i)/n \), \( b_{in} = i/n \). Then, with the intervals \( \omega_{in} = (t_{in}, t_{i+1,n}) \),

\[
\frac{1}{n-1} \left\{ a_{in} f(t_{in}) + b_{in} f(t_{i+1,n}) \right\} - \int_{\omega_{in}} f(t) \, dt = \sum_{i=1}^{n-1} \left\{ a_{in} \int_{\omega_{in}} \{ f(t_{in}) - f(t) \} \, dt + b_{in} \int_{\omega_{in}} \{ f(t) - f(t_{i+1,n}) \} \, dt \right\} .
\]

Now, for \( t \in \omega_{in} \),

\[
|f(t) - f(t_{in})| = \left| \int_{t_{in}}^{t} f'(s) \, ds \right| \leq \int_{\omega_{in}} |f'(s)| \, ds ,
\]

so

\[
\int_{\omega_{in}} |f(t) - f(t_{in})| \, dt \leq \frac{1}{n-1} \int_{\omega_{in}} |f'(t)| \, dt .
\]

The same bound applies to \( \int_{\omega_{in}} |f(t) - f(t_{i+1,n})| \, dt \). Then, adding these bounds gives

\[
\left| \frac{1}{n-1} \left\{ a_{in} f(t_{in}) + b_{in} f(t_{i+1,n}) \right\} - \int_{\omega_{in}} f(t) \, dt \right| \leq \frac{1}{n-1} \int_{\omega_{in}} |f'(t)| \, dt ,
\]

and then adding these over \( i = 1, 2, \ldots, n - 1 \), gives the required result. \( \square \)
Lemma 5. Let the time points $t_{1,n}, t_{2,n}, \ldots, t_{n,n}$ be super-quasi-uniform and let $m \geq 1$. Then, there exists positive constants $c$ and $c_1$ such that for all functions $f \in W^{m,2}(0,1)$, and $n\lambda \to \infty$, $\lambda \to 0$,

$$\frac{1}{n} \sum_{i=1}^{n} |f(t_{in})|^2 \geq c \int_{0}^{1} |f(t)|^2 dt - c_1 (n\lambda)^{-1} \| f \|_{m,\lambda}^2.$$  

Proof. By the super-quasi-uniformity of the time points, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} |f(t_{in})|^2 \geq c \int_{0}^{1} |f(t)|^2 dt - c n^{-1} \| \{ f \} \|_{L^1(0,1)}.$$  

Since

$$\| \{ f \} \|_{L^1(0,1)} = 2 \| f \|_{L^1(0,1)} \leq 2 \lambda^{-1} \| f \| \{ \lambda f \},$$

the last inequality by (3.5), the lemma follows. \hfill \Box

We are now ready for Lemma 3.

Proof of Lemma 3. First, write

$$|X(t_{in})v(t_{in})|^2 = v(t_{in})^T X(t_{in})^T X(t_{in}) v(t_{in}).$$

Then,

$$\frac{1}{n} \sum_{i=1}^{n} |X(t_{in}) v(t_{in})|^2 = \frac{1}{n} \sum_{i=1}^{n} v(t_{in})^T \mathbb{E}[X(t_{in})^T X(t_{in})] v(t_{in}) + \text{rem},$$

where the remainder “rem” is given by

$$\text{rem} = \frac{1}{n} \sum_{i=1}^{n} v(t_{in})^T \Theta(t_{in}) v(t_{in}),$$

with $\Theta(t) = X(t)^T X(t) - \mathbb{E}[X(t)^T X(t)]$.

Now, by assumption (2.6) and Lemma 5,

$$\frac{1}{n} \sum_{i=1}^{n} v(t_{in})^T \mathbb{E}[X(t_{in})^T X(t_{in})] v(t_{in}) \geq \rho \frac{1}{n} \sum_{i=1}^{n} v(t_{in})^T v(t_{in}) \geq \sum_{j=1}^{p} \left\{ c \| v_j \|^2 - c_1 (nh_j)^{-1} \| v_j \|_{m,h_j}^2 \right\}.$$
So, this term is better than advertised.

For the remainder “rem”, we write $\text{rem} = \sum_{j, \ell=1}^{p} S_{j\ell}$, with

$$S_{j\ell} = \frac{1}{n} \sum_{i=1}^{n} v_{j}(t_{in}) v_{\ell}(t_{in}) [\Theta(t_{in})]_{j,\ell}.$$  

Note that by (2.4) and (2.5), the random variables $[\Theta(t_{in})]_{j,\ell}, i = 1, 2, \ldots, n,$ are independent and bounded (uniformly in the time points). Now, by our reproducing kernel Hilbert spaces trick, as in §3, for all $\lambda$ (actually, for $\lambda = (h_{j} h_{\ell})^{1/2}$),

$$|S_{j\ell}| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} [\Theta(t_{in})]_{j,\ell} R_{1,\lambda}(t_{in}, \cdot) \right\|_{1,\lambda} \left\| v_{j} v_{\ell} \right\|_{1,\lambda}.$$  

The McDiarmid-Devroye exponential inequality, see Lemma 7 below, yields the almost sure bound

$$\left(5.2\right) \left\| \frac{1}{n} \sum_{i=1}^{n} [\Theta(t_{in})]_{j,\ell} R_{1,\lambda}(t_{in}, \cdot) \right\|_{1,\lambda} = O\left(\left(\frac{n\lambda}{\log n}\right)^{-1/2}\right).$$  

Next, below in Lemma 6, we show for $\lambda = (h_{j} h_{\ell})^{1/2}$ that

$$\lambda^{-1/2} \left\| v_{j} v_{\ell} \right\|_{1,\lambda} \leq c \left( h_{j}^{-1} + h_{\ell}^{-1} \right) \left( \left\| v_{j} \right\|_{m,h_{j}}^{2} + \left\| v_{\ell} \right\|_{m,h_{\ell}}^{2} \right).$$  

This gives a bound on $\mathbb{E}[ | S_{j\ell} | ]$, and hence

$$\mathbb{E}[ \text{rem} ] \leq c \left( \frac{n}{\log n} \right)^{-1/2} H^{-1} \sum_{j=1}^{p} \left\| v_{j} \right\|_{m,h_{j}}^{2},$$  

with $H^{-1} = h_{1}^{-1} + h_{2}^{-1} + \cdots + h_{p}^{-1}$. The lemma follows.  

**Lemma 6.** There exists a constant $c_{m}$ such that for all $f, g \in W^{1,2}(0, 1)$ and all positive $\lambda, \kappa, \nu$, with $\nu = (\lambda \kappa)^{1/2}$,

$$\nu^{-1/2} \left\| f g \right\|_{1,\nu} \leq c \left( \lambda^{-1} + \kappa^{-1} \right) \left( \left\| f \right\|_{1,\lambda}^{2} + \left\| g \right\|_{1,\kappa}^{2} \right).$$  

**Proof.** First, we have the inequality

$$\nu^{-1/2} \left\| f g \right\|_{1,\nu} \leq (\lambda \kappa)^{-1/4} \left\| f \right\| + (\lambda \kappa)^{1/4} \left\| (f g)' \right\|.$$  

$$\lambda^{-1/2} \left\| v_{j} v_{\ell} \right\|_{1,\lambda} \leq c \left( h_{j}^{-1} + h_{\ell}^{-1} \right) \left( \left\| v_{j} \right\|_{m,h_{j}}^{2} + \left\| v_{\ell} \right\|_{m,h_{\ell}}^{2} \right).$$
The first term on the right is easy: By the Arithmetic-Geometric Mean inequality, for all \( r > 0 \), we have \( 2 f g \leq r f^2 + r^{-1} g^2 \) (pointwise), so that for \( r = (\kappa/\lambda)^{1/2} \),

\[
2 \left( \lambda \kappa \right)^{-1/4} \| f g \| \leq \lambda^{-1/2} \| f \|^2 + \kappa^{-1/2} \| g \|^2 .
\]

(5.3)

For the remaining term, obviously

\[
\| (f g)' \| \leq \| f g' \| + \| f' g \| .
\]

Now,

\[
\| f g' \| \leq \| f \|_{L^\infty(0,1)} \| g' \| \leq c \lambda^{-1/2} \| f \|_{1,\lambda} \| g' \| ,
\]

the last inequality by (3.5). Using the Arithmetic-Geometric Mean inequality once (no, twice) more, then gives

\[
\nu^{1/2} \| f g' \| \leq c \lambda^{-1/4} \kappa^{1/4} \| f \|_{1,\lambda} \| g' \|
\leq c \lambda^{-1/2} \kappa^{-1/2} \| f \|_{1,\lambda}^2 + c \kappa \| g \|^2
\leq c (\lambda^{-1} + \kappa^{-1}) \| f \|_{1,\lambda}^2 + c \kappa^{-1} \| g \|_{1,\kappa}^2 .
\]

Adding the corresponding inequality for \( \nu^{1/2} \| f' g \| \) proves the lemma. \( \square \)

The example \( \lambda = \kappa = \nu \) and \( f(x) = g(x) = \max(1-x/\lambda, 0) \) on the interval \((0,1)\) shows that Lemma 6 cannot easily be improved. We leave the details to the interested reader.

6. Random Sums

**Proof of Lemma 2.** First we write

\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X(t_{in}) v(t_{in}) = \frac{1}{n} \sum_{j=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X_j(t_{in}) v_j(t_{in}) \right\} .
\]

Then, as in the proof of Theorem 2, we write, with \( \lambda = h_j \),

\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X_j(t_{in}) v_j(t_{in}) = \left( \frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X_j(t_{in}) \mathcal{R}_{m\lambda}(t_{in}, \cdot) , v_j \right)_{m,\lambda} \leq \| v_j \|_{m\lambda} S ,
\]

(6.1)
where
\[ S = \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon(t_{in}) X_j(t_{in}) \mathcal{R}_{m\lambda}(t_{in}, \cdot) \right\|_{m,h} \].

Now, \( E[S^2] \) is equal to
\[ n^{-2} \sum_{i,k=1}^{n} E[\varepsilon(t_{in}) \varepsilon(t_{kn}) X_j(t_{in}) X_j(t_{kn})] \langle \mathcal{R}_{m\lambda}(t_{in}, \cdot), \mathcal{R}_{m\lambda}(t_{kn}, \cdot) \rangle_{m,\lambda} \].

Further, from (2.3),
\[ E[\varepsilon(t_{in}) \varepsilon(t_{kn}) X_j(t_{in}) X_j(t_{kn})] = 0 \]
for \( i \neq k \), and with (2.5)
\[ E[|\varepsilon(t_{in})|^2 | X_j(t_{in})|^2] \leq c E[|\varepsilon(t_{in})|^2] \leq c \sigma^2 . \]

Finally, with (3.4),
\[ E[S^2] \leq c (nh)^{-1} , \]
and the lemma follows with Cauchy-Schwarz.

The left hand side of the inequality (4.2) also contains a random sum. This was handled in (5.2) by referring to the following lemma.

**Lemma 7.** Let \( \theta_{1,n}, \theta_{2,n}, \ldots, \theta_{n,n} \) be independent random variables with mean zero and
\[ | \theta_{in} | \leq \gamma , \ i = 1, 2, \ldots, n . \]
Then, for all \( \lambda \) with \( 0 < \lambda < 1 \),
\[ \left\| \frac{1}{n} \sum_{i=1}^{n} \theta_{in} \mathcal{R}_{m\lambda}(t_{in}, \cdot) \right\|_{m,\lambda} = \mathcal{O}\left( (n\lambda/\log n)^{-1/2} \right) \text{ almost surely} . \]

**Proof.** Let \( \theta_n = (\theta_{1,n}, \theta_{2,n}, \ldots, \theta_{n,n})^T \) and define
\[ \varphi(\theta_{1,n}, \theta_{2,n}, \ldots, \theta_{n,n}) \equiv \varphi(\theta_n) = \left\| \frac{1}{n} \sum_{i=1}^{n} \theta_{in} \mathcal{R}_{m\lambda}(t_{in}, \cdot) \right\|_{m,\lambda} . \]
Then, for all \( |a| \leq \gamma , |b| \leq \gamma , |\zeta_i| \leq \gamma \), the maximal variation of \( \varphi(\theta_n) \) over its first argument,
\[ s_{1,n} = \sup_{|a| \leq 1, |b| \leq 1} \left| \varphi(a, \zeta_2, \ldots, \zeta_n) - \varphi(b, \zeta_2, \ldots, \zeta_n) \right| , \]
satisfies
\[ s_{1,n} \leq \sup_{a,b} \frac{1}{n} \left\| (a - b) \mathcal{R}_{m\lambda}(t_{in}, \cdot) \right\|_{m,\lambda} \leq 2 \gamma n^{-1} \left\| \mathcal{R}_{m\lambda}(t_{in}, \cdots) \right\|_{m,\lambda} \leq c n^{-1} \lambda^{-1/2}, \]
for the appropriate constant \( c \), not depending on \( n \). The same bound, with the same \( c \), applies to \( s_{in} \) when the total variation of \( \varphi \) over the \( i \)-th argument is considered, for all \( i \).

The McDiarmid–Devroye exponential inequality, see, e.g., Devroye (1991) or Devroye, Györfi and Lugosi (1996), now implies that
\begin{equation}
\mathbb{P}\left[ \left| \varphi(\theta_n) - \mathbb{E}[\varphi(\theta_n)] \right| \geq u \right] \leq 2 \exp\left( -\frac{1}{2} \frac{u^2}{s_n^2} \right),
\end{equation}
where
\[ s_n^2 = \sum_{i=1}^{n} s_{in}^2 \leq c (n\lambda)^{-1}. \]
Now, take
\[ u = u_n = 2 s_n (\log n)^{1/2}, \]
to conclude that
\begin{equation}
\mathbb{P}\left[ \left| \varphi(\theta_n) - \mathbb{E}[\varphi(\theta_n)] \right| \geq u_n \right] \leq 2 n^{-2},
\end{equation}
so that with Borel-Cantelli,
\[ \left| \varphi(\theta_n) - \mathbb{E}[\varphi(\theta_n)] \right| = \mathcal{O}(u_n) = \mathcal{O}\left( (n\lambda/\log n)^{-1/2} \right), \quad \text{almost surely.} \]
Since \( \mathbb{E}[\varphi(\theta_n)] \leq \left\{ \mathbb{E}\left[ |\varphi(\theta_n)|^2 \right] \right\}^{1/2} \leq c (n\lambda)^{-1/2} \), this proves the lemma.

\section*{References}


Splines in Varying Coefficient Models


**Food and Resource Economics**  
**University of Delaware**  
**Newark, DE 19717-1303**

**Department of Mathematics and Statistics**  
**Arizona State University**  
**Tempe, AZ 85287-1804.**