The Annals of Statistics
1993, Vol. 21, No. 1, 124–145

NONPARAMETRIC ESTIMATION IN THE COX MODEL[1]

BY FINBARR O'SULLIVAN

University of Washington

Nonparametric estimation of the relative risk in a generalized Cox model with multivariate time dependent covariates is considered. Estimation is based on a penalized partial likelihood. Using techniques from Andersen and Gill, and Cox and O'Sullivan, upper bounds on rate of convergence in a variety of norms are obtained. These upper bounds match the optimal rates available for linear nonparametric regression and density estimation. The results are uniform in the smoothing parameter, which is an important step for the analysis of data dependent rules for the selection of the smoothing parameter.

1. Introduction. The Cox (1972) proportional hazards model is widely used in modern survival analysis. In this model the hazard rate or intensity of failure for the survival time of an individual with covariate vector x, which may depend on time t, is expressed as

$$\lambda(t; x(t)) = \lambda_0(t)\exp\{\theta_0(x(t))\}, \qquad t \geq 0. \tag{1.1}$$

θ_0 is the relative risk function and λ_0 is the underlying baseline hazard. Both θ_0 and λ_0 are unknown, but statistical inference is typically restricted to the relative risk, see Andersen and Gill (1982) for example. This paper develops some asymptotic properties for a nonparametric penalized partial likelihood estimator of the relative risk, proposed in O'Sullivan (1988). These results complement the work of Zucker and Karr (1990), who have considered an alternative approach to nonparametric estimation in the Cox model.

The method of analysis here makes use of some rudimentary multivariate counting process techniques for the Cox model developed by Andersen and Gill (1982) and Gill (1984). We begin with a brief summary of the counting process framework for the case of iid observations. We suppose we have n independent subjects which are continuously monitored over a finite time period, say the interval $[0,1]$. For each subject, there is a process $(N_i(t), x_i(t), y_i(t))$ for $0 \leq t \leq 1$. Here $N_i(t)$ is a counting process recording events, such as death, hospital visits and so on, to occur up to time t, $x_i(t)$ is a d-dimensional covariate process, and $y_i(t)$ is either 1 or 0 depending on whether or not the subject is under observation immediately prior to time t. In survival analysis, $y_i(t) = 0$ for $t > c$ corresponds to right censoring at time c. Formally, as in Gill

---

Received June 1986; revised March 1992.

[1]Partially supported by NSF Grant MCS-84-03239 and by the National Institutes of Health Grant AI-29168.

AMS 1991 *subject classifications*. Primary 62G05; secondary 62P10, 41A35, 41A25, 47A53, 45L10, 45M05.

Key words and phrases. Cox model, martingales, penalized partial likelihood, rates of convergence, relative risk.

(1984), the generalized Cox model specifies that $N^{(n)} = (N_1, N_2, \dots, N_n)$ is a multivariate counting process with a random intensity process $\lambda^{(n)} = (\lambda_1, \lambda_2, \dots, \lambda_n)$ for which

$$\lambda_i(t) = y_i(t) \cdot \exp\{\theta_0(x_i(t))\} \cdot \lambda_0(t). \tag{1.2}$$

The underlying baseline hazard λ_0 and the relative risk θ_0: $\mathbb{R}^d \to \mathbb{R}$ are fixed unknown quantities. As discussed in Gill (1984), a family of right continuous nondecreasing sub σ-algebras $\{F_t^{(n)}: t \in [0,1]\}$ are defined on the n-dimensional sample space, with $F_{t-}^{(n)}$ representing the history of the n-dimensional process up to time t. All processes are adapted to this family of σ-algebras. $y_i(\cdot)$ is a predictable process taking values in $\{0,1\}$. The d-dimensional covariate process $x_i(\cdot)$ is predictable and takes values in a fixed bounded subset of $\mathbb{R}^d$. Specification of $\lambda^{(n)}$ as an intensity process means that the process

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(\tau)\, d\tau, \qquad i = 1, 2, \dots, n \text{ and } t \in [0,1], \tag{1.3}$$

is a local martingale with mean zero, $EM_i(t) = 0$. The predictable covariation of $M^{(n)} = (M_1, M_2, \dots, M_n)$ is given by

$$\langle M_i, M_i\rangle(t) = \int_0^t \lambda_i(\tau)\, d\tau, \qquad \langle M_i, M_j\rangle = 0, i \neq j. \tag{1.4}$$

1.1. *Definition of the penalized partial likelihood and some assumptions.* Inferences for the relative risk θ_0 will be based on the penalized partial likelihood functional

$$l_{n\mu}(\theta) = l_n(\theta) + \mu J(\theta), \qquad \mu > 0. \tag{1.5}$$

Here $l_n(\theta)$ is the analogue of the negative logarithm of the partial likelihood used by Andersen and Gill (1982):

$$l_n(\theta) = \int_0^1 \log\left[\frac{1}{n}\sum_{i=1}^n y_i(\tau) e^{\theta(x_i(\tau))}\right] d\overline{N}(\tau) - \frac{1}{n}\sum_{i=1}^n \int_0^1 \theta(x_i(\tau))\, dN_i(\tau) \tag{1.6}$$

with $\overline{N}(t) = (1/n)\sum_{i=1}^n N_i(t)$. J is a penalty functional designed to incorporate prior notions about the smoothness of the relative risk. Examples of commonly used penalty functionals are given in Cox (1984), Cox and O'Sullivan (1990) and Wahba (1990). The penalized likelihood estimator of the relative risk $\theta_{n\mu}$ is defined as the minimizer of $l_{n\mu}$ over a class of functions Θ. Θ is the nominal parameter space. This method of estimation can be thought of in the light of Tikhonov's method of regularization [Tikhonov and Arsenin (1977) and Cox and O'Sullivan (1990)]. Issues related to the numerical computation of the penalized partial likelihood estimate along with a proposed data dependent procedure for selecting the smoothing parameter μ are discussed in O'Sullivan (1988).

1.1.1. *Some assumptions.* Assumptions on the measurement model and the nature of the parameter space follow. The assumptions in this paper combine the standard kinds of conditions set out in the analysis of the Cox model and in the analysis of nonparametric regression estimators, see Section 4 of Andersen and Gill (1982) and Assumptions 1–4 of Cox (1984).

ASSUMPTION A (Measurement model).

(i) (N_i, y_i, x_i), $i = 1, 2, \ldots, n$, are iid replicates of a fixed triple of random processes $(N, y, x))$ observed on $[0, 1]$.

(ii) Let $\Lambda_0(t) = \int_0^t \lambda_0(\tau)\, d\tau$. It is assumed that λ_0 is bounded away from zero and infinity.

(iii) For each $t \in [0, 1]$, the random variable $x(t)$ has density $h(\cdot \mid t)$. Let

$$p(x,t) = P[y(t) = 1 \mid x(t) = x],$$

and let $q(x,t) = p(x,t)h(x \mid t)$. We suppose that there are strictly positive constants, k_1 and k_2 (independent of x and t), such that

$$k_1 < q(x,t) < k_2 \quad \text{and} \quad \left|\frac{\partial}{\partial t} q(x,t)\right| < k_2.$$

(iv) $\{x(t),\ t \in [0, 1]\} \subseteq \underline{X} \subseteq \mathbb{R}^d$, where $\underline{X}$ is a bounded open simply connected set with C^∞ boundary [see Definition 3.2.1.2 of Triebel (1978)].

Assumption A(i) is used by Andersen and Gill, for example. Assumption A(ii) is stronger than the more typical condition that $0 < \Lambda(1) < \infty$ used by Andersen and Gill. The additional strength here is used to obtain a certain uniformity in the main result. Assumptions A(iii) and A(iv) are more technical. Assumption A(iii) is used in proving results concerning derivatives of a continuous version of the partial likelihood and (iv) is used is to obtain growth behavior on eigenvalues which in turn determine the ultimate rates of convergence for the approach. For the assumptions on the parameter space, let $W_2^k(\underline{X}, \mathbb{R})$ denote the Sobolev space of real valued L_2 functions defined on $\underline{X}$ whose kth order partial derivatives are square integrable; see Adams (1975). Sobolev spaces may also be defined for noninteger k [Adams (1975)] and throughout this paper the order k can assume any positive real value. $W_{02}^k(\underline{X}, \mathbb{R})$ is the subspace of $W_2^k(\underline{X}, \mathbb{R})$ which consists of functions which integrate to zero, $\int_{\underline{X}} \theta(x)\, dx = 0$, for all $\theta \in W_{02}^k(\underline{X}, \mathbb{R})$.

ASSUMPTION B (Parameter space).

(i) Θ is a Hilbert space of functions θ: $\underline{X} \to \mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. The elements of Θ are constrained to integrate to zero.

(ii) For some $m > 3d/2$, $\Theta = W_{02}^m(\underline{X}; \mathbb{R})$ (meaning the spaces are equal as sets and they have equivalent norms). The true function parameter θ_0 is in W_{02}^{mp} for some $3d/2m < p \le 1$.

(iii) The penalty functional $J(\theta) = (1/2)\langle \theta, W\theta \rangle$, where W is a bounded linear operator on Θ, which is self-adjoint and nonnegative definite and there are positive constants k_1, k_2 such that for all $\theta \in \Theta$,

$$k_1 \|\theta\|^2 \le \langle \theta, W\theta \rangle + \|\theta\|_{L_2}^2 \le k_2 \|\theta\|^2.$$

The Hilbert space structure is convenient for the spectral analysis of linear operators which arise. Note that the true parameter is not assumed to be an element of Θ. It is only necessary that the true parameter be approximable by elements of Θ. The condition that m and mp be greater than $3d/2$ covers the standard cubic smoothing spline situation in which $m = 2$, $p = 1$ and $d = 1$. The assumption on the penalty functional holds for the standard Laplacian penalty functionals used to define multivariate smoothing splines [Cox (1984)].

1.2. *Asymptotic convergence result.* The estimation error is written as the sum of a systematic and stochastic error

$$\theta_{n\mu} - \theta_0 = \theta_\mu - \theta_0 + \theta_{n\mu} - \theta_\mu, \tag{1.7}$$

where θ_μ is the minimizer of limiting penalized partial likelihood with $l_n(\theta)$ replaced by $l(\theta) = \lim_n l_n(\theta)$; see (2.4) in the next section. The existence of θ_μ for μ sufficiently small is discussed in Theorem 3.1. For $0 \le b \le 1$, let $\|\cdot\|_b$ be the Sobolev W_2^{bm}-norm. The object of the paper is to establish the following asymptotic convergence result.

THEOREM 1. *Suppose Assumptions A and B hold.*

(i) *There is some μ_0 such that for all $\mu \in [0, \mu_0]$ θ_μ is uniquely defined. There is a positive constant C_p such that for $0 \le b \le (p - d/2m)/2$*

$$\|\theta_\mu - \theta_0\|_b^2 \le C_p \mu^{(p-b)} \|\theta_0\|_p^2 \quad \text{as } \mu \to 0.$$

(ii) *Let μ_n be a sequence satisfying $\mu_0 > \mu_n = O(n^{-\delta})$ with $\delta > 0$ and for some $\alpha \in (d/2m, (p - d/2m)/2)$, $n^{-1}\mu_n^{-2(\alpha + d/2m)} \to 0$. Let $0 \le b \le \alpha$, and let μ_{*n} be any deterministic sequence in $[\mu_n, \mu_0]$. For a sufficiently large positive constant M the event: $\theta_{n\mu}$ exists uniquely and satisfies*

$$\|\theta_{n\mu_{*n}} - \theta_{\mu_{*n}}\|_b^2 \le M n^{-1} \mu_{*n}^{-(b + d/2m)}$$

and

$$\sup_{\mu \in [\mu_n, \mu_0]} \left\{ \frac{\|\theta_{n\mu} - \theta_\mu\|_b^2}{n^{-1}\mu^{-(b+d/2m)}} \right\} \le M \log(\mu_n^{-1})$$

occurs with probability approaching unity as $n \to \infty$.

PROOF. The results follow from Theorems 3.1 and 3.3 of Section 3. □

1.3. *Discussion and outline of the paper.* The first part of the theorem gives the order of the systematic error and the second part gives the order of the stochastic error. From the theorem it follows that if μ_{*n} is $O(n^{-2m/(2mp+d)})$, then $\|\theta_{n\mu_{*n}} - \theta_0\|_b^2$ is bounded by $O_p(n^{-2m(p-b)/(2mp+d)})$. In particular, if $\theta_0 \in W_2^m$ (i.e., $p = 1$), the integrated square error $\|\theta_{n\mu_{*n}} - \theta_0\|_0^2$ is bounded by $O_p(n^{-2m/(2m+d)})$. Thus in the analogue of the standard one-dimensional cubic smoothing spline setup ($m = 2$, $d = 1$ and

$p = 1$) we get the familiar rate of $O_p(n^{-4/5})$. Uniformity in the smoothing parameter is noteworthy. We expect that this will be most useful in the analysis of data dependent rules for choosing the smoothing parameter. Theorem 1 yields upper bounds on the rates of convergence. These upper bounds correspond to the optimal rates of convergence obtained by Stone (1982) for nonparametric regression and density estimation in Hölder spaces.

The paper represents an application of techniques developed in Cox and O'Sullivan (1990), hereafter abbreviated CO. The approach is based on Taylor series expansions of the penalized partial likelihood in function space. The theoretical framework is elaborated in Section 2, and some bounds on derivatives of the partial likelihood are noted. The main results are proved in Section 3. The uniformity result requires a strengthening of a theorem in CO. This result is proved in the Appendix along with a more technical lemma used in obtaining bounds on derivatives.

2. Theoretical framework. The analysis considers solutions to the variational equation obtained from the penalized partial likelihood and, by analogy with Cramér's approach to asymptotics for parametric maximum likelihood estimators, the main results follow from investigation of one-term Taylor series expansions for the variational equation. Quantities related to the second and third order derivative operators play an important role. Several assumptions must be verified in order to apply the theory developed in CO. We note that Assumptions A.1 and A.2 of CO follow directly from our Assumption B. [The operator U is given in (2.13). The compactness is easily established.] This section elaborates the theoretical framework and shows that Assumptions A.3 and A.4 of CO also hold. We begin by developing a representation for the estimator and derivatives of interest. The discussion of derivatives uses Sobolev spaces $W_2^{m\alpha}(X; \mathbb{R})$ with $d/2m < \alpha < p$ (see Theorem 1). The condition that $\alpha > d/2m$ guarantees by Sobolev's imbedding theorem [Theorem 5.4 of Adams (1975)] that pointwise evaluation is a continuous linear functional in $W_2^{m\alpha}(X; \mathbb{R})$. We let $S(R, \alpha)$ be the ball of radius R in $W_2^{\alpha m}(\underline{X}; \mathbb{R})$ and let $S_{\theta_0}(R, \alpha) = \{\theta_0\} \oplus S(R, \alpha)$. Since $\alpha < p$, $\theta_0 \in W_{02}^{m\alpha}(X; \mathbb{R})$. Let $N_{\theta_0} = S_{\theta_0}(R, \alpha)$ for some $R > 0$. N_{θ_0} is a neighborhood of θ_0 in $W_{02}^{m\alpha}(X; \mathbb{R})$. We use M and M_R to denote generic finite positive constants in proofs of lemmas and theorems. Successive appearances of such constants will typically correspond to larger values.

2.1. *Derivatives and a representation for the estimate.* Following Andersen and Gill, let $s_n(\theta, t) = (1/n)\sum_{i=1}^n y_i(t) e^{\theta(x_i(t))}$ so the negative logarithm of the partial likelihood is

$$l_n(\theta) = \int_0^1 \log[s_n(\theta, t)]\, d\overline{N}(t) - \frac{1}{n}\sum_{i=1}^n \int_0^1 \theta(x_i(t))\, dN_i(t). \tag{2.1}$$

Derivatives of $l_n(\theta)$ will involve derivatives of $s_n(\theta, t)$. Computing formally, the first, second and third order Frechet derivatives, of $s_n(\theta, t)$ are given by

$$
\begin{aligned}
Ds_n(\theta,t)\phi &= \frac{1}{n}\sum_{i=1}^{n}\phi(x_i(t))y_i(t)e^{\theta(x_i(t))},\\
D^2s_n(\theta,t)\phi\psi &= \frac{1}{n}\sum_{i=1}^{n}\phi(x_i(t))\psi(x_i(t))y_i(t)e^{\theta(x_i(t))},\qquad (2.2)\\
D^3s_n(\theta,t)\phi\psi\xi &= \frac{1}{n}\sum_{i=1}^{n}\phi(x_i(t))\psi(x_i(t))\xi(x_i(t))y_i(t)e^{\theta(x_i(t))},
\end{aligned}
$$

where ϕ, ψ, ξ are functions. Since evaluation is a continuous linear functional in $W_2^{m\alpha}(X; \mathbb{R})$ the above quantities are the Frechet derivatives of $s_n(\theta, t)$ in $W_2^{\alpha m}(X; \mathbb{R})$. We note that the Frechet derivative is the function space generalization of the total derivative used in standard multivariate calculus [Rall (1969)]. The continuity of the derivatives follows since $e^{\theta(x)}$ is continuous in θ for $\theta \in S(R, \alpha)$ because $\alpha > d/2m$.

Fixing $\theta \in W_2^{\alpha m}$, by the strong law of large numbers $s_n(\theta, t) \to s(\theta, t)$ almost surely, where $s(\theta, t) = E[y(t)e^{\theta(x(t))}] = \int e^{\theta(x)}q(x, t)\,dx$. In addition under Assumption A and B, $\sup_t|s_n(\theta, t) - s(\theta, t)|$ converges to zero in probability. This follows as in the justification of Condition B in Theorem 4.1 of Andersen and Gill. The Frechet derivatives of $s(\theta, t)$ are given by

$$
\begin{aligned}
Ds(\theta,t)\phi &= \int \phi(x)e^{\theta(x)}q(x,t)\,dx,\\
D^2s(\theta,t)\phi\psi &= \int \phi(x)\psi(x)e^{\theta(x)}q(x,t)\,dx,\qquad (2.3)\\
D^3s(\theta,t)\phi\psi\xi &= \int \phi(x)\psi(x)\xi(x)e^{\theta(x)}q(x,t)\,dx.
\end{aligned}
$$

These derivatives are clearly continuous. The following lemma will be useful.

LEMMA 2.1. *For any $R > 0$ and $\alpha > d/2m$ there are constants $0 < m_R \le M_R < \infty$ such that for all $\theta, \theta_* \in S(R, \alpha)$ and $t \in [0, 1]$ the following hold:*

(a) $m_R < s_n(\theta, t)/s_n(\theta_*, t) < M_R$.

(b) $m_R < s(\theta, t) < M_R$ *and* $s_n(\theta, t) < M_R$.

(c) *If* $\phi, \psi \in \Theta$, *then* $\{Ds(\theta, t)\phi\}^2 \le M_R\|\phi\|_{L_2}^2$ *and* $\{D^2s(\theta, t)\phi\psi\}^2 \le M_R\|\phi\|_{L_2}^2\|\psi\|_{L_2}^2$.

PROOF. (a) follows from the definition of s_n, Assumption A(iv) and the uniform boundedness of θ and θ_* [they are both elements of $S(R, \alpha)$ and $\alpha > d/2m$]. (b) follows from Assumptions A(iii) and A(iv) and the boundedness of $\theta(x)$ for $x \in X$ and $y_i(t)$. Applications of Hölder's inequality gives (c). □

Using the uniform convergence of $s_n(\theta, t)$ to $s(\theta, t)$ mentioned above, it may be shown that $l_n(\theta)$ converges in probability to $l(\theta)$, where

$$\begin{aligned} l(\theta) &= \int \log[s(\theta,t)]s(\theta_0,t)\lambda_0(t)\,dt \\ &\quad - \int\int \theta(x)e^{\theta_0(x)}q(x,t)\,dx\,\lambda_0(t)\,dt. \end{aligned} \tag{2.4}$$

By the chain rule the first three derivatives of l and l_n are well defined and continuous on $W_2^{m\alpha}(X; \mathbb{R})$. The first and second derivatives of $l(\theta)$ are given by

$$\begin{aligned} Dl(\theta)\phi &= \int \frac{Ds(\theta,t)\phi}{s(\theta,t)} s(\theta_0,t)\lambda_0(t)\,dt \\ &\quad - \int\int \phi(x)e^{\theta_0(x)}q(x,t)\,dx\,\lambda_0(t)\,dt \\ &= \int \frac{Ds(\theta,t)\phi}{s(\theta,t)} s(\theta_0,t)\lambda_0(t)\,dt - \int Ds(\theta_0,t)\phi\lambda_0(t)\,dt, \end{aligned} \tag{2.5}$$

$$D^2 l(\theta)\phi\psi = \int \left[\frac{D^2 s(\theta,t)\phi\psi}{s(\theta,t)} - \frac{Ds(\theta,t)\phi}{s(\theta,t)} \frac{Ds(\theta,t)\psi}{s(\theta,t)} \right] s(\theta_0,t)\lambda_0(t)\,dt. \tag{2.6}$$

It is clear that θ_0 satisfies $Dl(\theta_0) \equiv 0$. Some algebra gives

$$D^2 l(\theta)\phi\phi = \int\int \left[\phi(x) - \bar{\phi}_t\right]^2 p_x(t)\,dx\,s(\theta_0,t)\lambda_0(t)\,dt, \tag{2.7}$$

where $p_x(t) = e^{\theta(x)}q(x,t)/s(\theta,t)$ and $\bar{\phi}_t = \int\phi(x)p_x(t)\,dx$. From this we have that $l(\theta)$ is convex. Using (2.7), Lemma 2.1(b) and Assumptions A(ii) and A(iii) there are positive constants c_1 and c_2 such that for $\theta_* \in S_{\theta_0}(R, \alpha)$

$$\begin{aligned} c_1 \int\int \left[\phi(x) - \bar{\phi}_t\right]^2 dx\,\lambda_0(t)\,dt &\le D^2 l(\theta)\phi\phi \\ &\le c_2 \int\int \left[\phi(x) - \bar{\phi}_t\right]^2 dx\,\lambda_0(t)\,dt. \end{aligned} \tag{2.8}$$

Letting m_x be the Lebesgue measure of $\underline{X}$ and restricting to ϕ's which integrate to zero,

$$\int\int \left[\phi(x) - \bar{\phi}_t\right]^2 dx\,\lambda_0(t)\,dt = \int \phi^2(x)\,dx\,\Lambda_0(1) + m_x \int \left[\bar{\phi}_t\right]^2 \lambda_0(t)\,dt. \tag{2.9}$$

Applying the Cauchy–Schwarz inequality to $[\bar{\phi}_t]^2$ and using Assumption A(ii), we have, for some $0 < c_1 \le c_2 < \infty$,

$$c_1 \int [\phi(x)]^2\,dx \le D^2 l(\theta)\phi\phi \le c_2 \int [\phi(x)]^2\,dx \tag{2.10}$$

for all $\theta \in S_{\theta_0}(R, \alpha)$ and all $\phi \in L_2(X; \mathbb{R})$ which integrate to zero. Thus $l(\theta)$ is strictly convex in N_{θ_0} and since $Dl(\theta_0) = 0$, θ_0 is the unique minimizer of $l(\theta)$ in N_{θ_0}. Since $l(\theta)$ is globally convex, θ_0 is in fact the global minimizer of $l(\theta)$

over $W_{02}^{m\alpha}(X; \mathbb{R})$. This result along with existence of continuous derivatives up to order three guarantees that Assumption A.3 of CO holds.

The first and second order Frechet derivatives of l_n are given by

$$(2.11)\quad \begin{aligned} Dl_n(\theta)\phi &= \int_0^1 \frac{Ds_n(\theta,t)\phi}{s_n(\theta,t)}\, d\bar{N}(t) - \frac{1}{n}\sum_{i=1}^n \int_0^1 \phi(x_i(t))\, dN_i(t), \\ D^2 l_n(\theta)\phi\psi &= \int_0^1 \left\{ \frac{D^2 s_n(\theta,t)\phi\psi}{s_n(\theta,t)} - \frac{Ds_n(\theta,t)\phi}{s_n(\theta,t)} \frac{Ds_n(\theta,t)\psi}{s_n(\theta,t)} \right\} d\bar{N}(t). \end{aligned}$$

Again, with $p_i(t) = [(1/n)y_i(t)e^{\theta(x_i(t))}]/s_n(\theta,t)$ for $i = 1,2,\dots,n$, and $\bar{\phi}_t = \sum_{i=1}^n \phi(x_i(t))p_i(t)$,

$$(2.12)\quad \frac{D^2 s_n(\theta,t)\phi\phi}{s_n(\theta,t)} - \frac{Ds_n(\theta,t)\phi}{s_n(\theta,t)} \frac{Ds_n(\theta,t)\phi}{s_n(\theta,t)} = \sum_{i=1}^n p_i(t)\left[\phi(x_i(t)) - \bar{\phi}_t\right]^2,$$

and it follows that l_n and $l_{n\mu}$ are convex. A straightforward argument, along the lines given in the appendix of O'Sullivan, Yandell and Raynor (1986), shows that the penalized partial likelihood estimator must lie in the subspace $\Theta_n = N(W) + \operatorname{Span}\{\xi(x_i(t_{ij}))\}$, where $\xi(x)$ is the Riesz representer of evaluation at x and $N(W)$ is the null space of the linear operator W. Span stands for the span of the given set, where t_{ij} ranges over the event times of the counting process $N_i(t)$ for $t \in [0,1]$ and $i = 1,2,\dots,n$. If $N(W)$ is finite, then Θ_n is a finite dimensional space, although its dimension will in general be larger than n. When W corresponds to the usual Laplacian penalty functional used to generate thin plate smoothing splines [Wahba (1990)] the penalized partial likelihood estimator can be represented as a generalized Laplacian smoothing spline [O'Sullivan (1988)]. Following the argument in O'Sullivan, Yandell and Raynor (1986), a sufficient condition for the existence of a unique minimizer of the penalized partial likelihood in (2.7) is that there exist a unique minimizer of the negative logarithm of the partial likelihood over $N(W)$. These results are summarized in the following theorem.

THEOREM 2.2. *Under Assumptions A and B, θ_0 is the unique root of $Dl(\theta)$ in $W_{02}^{\alpha m}$ for $\alpha > d/2m$. If the dimension of the null space of W is finite, the minimizer of the $l_{n\mu}$ must lie in the finite dimensional subspace Θ_n. A sufficient condition for the existence of a unique minimizer of $l_{n\mu}$ is that there exists a unique minimizer of l_n over $N(W)$.*

2.2. *Spectral decomposition and convergence norms.* Let U be an operator defined on Θ by

$$(2.13)\quad \langle \psi, U\phi \rangle = \int \phi(x)\psi(x)\, dx$$

for $\psi, \phi \in \Theta$. From Assumption B(iii), $\|\theta\|_1^2 = \langle \theta, W\theta \rangle + \langle \theta, U\theta \rangle$ is an equivalent norm on Θ. It follows from Section 3.3 of Weinberger (1974) and the construction in Section 2 of Cox (1988) that there is a sequence of eigenfunctions $\{\phi_\nu: \nu = 1, 2, \dots\}$ and corresponding eigenvalues $\{\gamma_\nu: \nu = 1, 2, \dots\}$ satisfying

$$\begin{aligned} \langle \phi_\mu, U\phi_\nu \rangle &= \delta_{\nu\mu}, \\ \langle \phi_\mu, W\phi_\nu \rangle &= \gamma_\nu \delta_{\nu\mu}, \end{aligned} \tag{2.14}$$

for all pairs ν, μ of positive integers, where $\delta_{\nu\mu}$ is Kronecker's delta. If Θ were equivalent to $W_2^m(\underline{X}; \mathbb{R})$, Assumption A(iv) and standard results concerning elliptic differential operators [see, e.g., Cox (1984)] would give that $\gamma_\nu = O(\nu^{2m/d})$ as $\nu \to \infty$. The elements of Θ are constrained to integrate to zero but using the interlacing property [Corollary 1 of Theorem 9.1 on page 63 of Weinberger (1974)], this single linear constraint cannot affect the rate of growth of the eigenvalues. Thus $\gamma_\nu = O(\nu^{2m/d})$.

For $b \geq 0$ let

$$\|\theta\|_b = \left\{ \sum_{\nu=1}^{\infty} (1 + \gamma_\nu^b) \langle \theta, U\phi_\nu \rangle^2 \right\}^{1/2} \tag{2.15}$$

and let Θ_b denote the normed linear space obtained by completing $\{\theta \in \Theta: \|\theta\|_b < \infty\}$ in $\|\cdot\|_b$ norm. Θ_b is a Hilbert space with inner product

$$\langle \theta, \zeta \rangle_b = \sum_{\nu=1}^{\infty} (1 + \gamma_\nu^b) \langle \theta, U\phi_\nu \rangle \langle \zeta, U\phi_\nu \rangle. \tag{2.16}$$

It is easily shown using standard interpolation theory that Θ_b is equivalent to $W_{02}^{bm}(\underline{X}; \mathbb{R})$ for $b \in [0, 1]$. We denote $W_{02}^{bm}(\underline{X}; \mathbb{R})$ by Θ_b from here on. The operators U and W extend to linear operators on Θ_b for $b \in [0, 1]$ (Lemma 2.1 of CO).

Let $U(\theta)$ be defined by

$$\langle \psi, U(\theta)\phi \rangle = D^2 l(\theta)\phi\psi. \tag{2.17}$$

By definition of the derivative $U(\theta)$ is a bounded linear operator on Θ_α since $\alpha > d/2m$. From (2.10) we have the next lemma which implies Assumption A.4 of CO.

LEMMA 2.3. *There are constants* $0 < c_1 \leq c_2 < \infty$, *such that for all* $\theta_* \in S(R, \alpha)$,

$$c_1 \langle \theta, U\theta \rangle \leq \langle \theta, U(\theta_*)\theta \rangle \leq c_2 \langle \theta, U\theta \rangle$$

for all $\theta \in \Theta$.

Replacing U by $U(\theta_*)$ in (2.14) we obtain for each $\theta_* \in N_{\theta_0}$ sequences of eigenvalues $\{\gamma_{*\nu};\ \nu = 1, 2, \dots\}$ and corresponding eigenfunctions $\{\phi_{*\nu};\ \nu = 1, 2, \dots\}$. This leads to a norm

$$\|\theta\|_{*b} = \left\{ \sum_{\nu=1}^{\infty} [1 + \gamma_{*\nu}^b] \langle \theta, U(\theta_*)\phi_{*\nu} \rangle^2 \right\}^{1/2} \tag{2.18}$$

and corresponding Hilbert space Θ_{*b}. These spaces are uniformly equivalent for $\theta_* \in N_{\theta_0}$. $U(\theta_*)$ extends to a bounded linear operator on Θ_{*b} for $b \in [0,1]$ and the linear operator

$$G_\mu(\theta_*) = U(\theta_*) + \mu W$$

is bounded and invertible on Θ_{*b}. From Lemma 2.1 of CO we have the following lemma.

LEMMA 2.4. *For $R > 0$ and $\theta_* \in S(R, \alpha)$, $b > 0$ and $\nu = 1, 2, \ldots$ we have the following*:

(i) $\|\phi_\nu\|_b^2 = 1 + \gamma_\nu^b$ *and* $\|\phi_{*\nu}\|_{*b}^2 = 1 + \gamma_{*\nu}^b$.

(ii) $[U + \mu W]^{-1} U \phi_\nu = (1 + \mu\gamma_\nu)^{-1}\phi_\nu$ *and* $[U(\theta_*) + \mu W]^{-1} U(\theta_*)\phi_{*\nu} = (1 + \mu\gamma_{*\nu})^{-1}\phi_{*\nu}$.

(iii) *For $b \geq 0$ and $c \geq 0$ with $b + c < 2 - d/2m$, uniformly in $\theta_* \in N_{\theta_0}$*,

$$\sum (1 + \gamma_{*\nu}^b)(1 + \gamma_{*\nu}^c)(1 + \mu\gamma_{*\nu})^{-2} \approx \mu^{-(b+c+d/2m)} \quad \textit{as } \mu \to 0,$$

meaning that the supremum, over $\theta_* \in N_{\theta_0}$, of the ratio of the quantity on the left to that on the right remains bounded away from 0 and ∞ as $\mu \to 0$.

PROOF. The first two results follow directly from the definitions. The last result comes from the fact that $\gamma_{*\nu} = O(\nu^{2m/d})$ uniformly for $\theta_* \in S(R, \alpha)$ and then approximation of sums by integrals as on page 479 of Craven and Wahba (1979). □

Finally note that if $\zeta \in \Theta_\alpha$, then for $\theta_* \in N_{\theta_0}$,

$$\left\|G_\mu(\theta_*)^{-1}\zeta\right\|_{*b}^2 = \sum_\nu (1 + \gamma_{*\nu}^b)(1 + \mu\gamma_{*\nu})^{-2}\langle \zeta, U(\theta_*)\phi_{*\nu}\rangle^2. \tag{2.19}$$

2.3. *Bounds on derivatives.* It will be useful to have bounds on various quantities related to the first, second and third order derivatives of the partial likelihood. For $0 \leq b \leq \alpha$, $\mu > 0$, $\theta_1, \theta_2 \in N_{\theta_0}$ and u, v unit elements in Θ_α (so $\|u\|_\alpha = \|v\|_\alpha = 1$) let

$$\begin{aligned} K_{2n}(\mu, b) &= \sup_{\theta_1,\theta_2} \sup_u \left\|G_\mu(\theta_1)^{-1}\left[D^2 l_n(\theta_2)u - D^2 l(\theta_2)u\right]\right\|_b, \\ K_3(\mu, b) &= \sup_{\theta_1,\theta_2} \sup_{u,v} \left\|G_\mu(\theta_1)^{-1}\left[D^3 l(\theta_2)uv\right]\right\|_b, \qquad (2.20) \\ K_{3n}(\mu, b) &= \sup_{\theta_1,\theta_2} \sup_{u,v} \left\|G_\mu(\theta_1)^{-1}\left[D^3 l_n(\theta_2)uv\right]\right\|_b. \end{aligned}$$

The next lemma provides bounds on the behavior of these quantities.

LEMMA 2.5. *There are a constant* $0 < M < \infty$ *and a random variable* $A_n = O_p(1)$ *such that for* $b \le 2 - \alpha - d/2m$ *we have the following*:

(i) $K_3(\mu, b)^2 \le M\mu^{-(b+d/2m)}$.
(ii) $K_{2n}(\mu, b)^2 \le A_n M n^{-1}\mu^{-(b+d/2m)}\mu^{-\alpha}$.
(iii) $K_{3n}(\mu, b)^2 \le A_n M\mu^{-(b+d/2m)}\{1 + n^{-1}\mu^{-\alpha}\}$.

Both M *and* A_n *are independent of* μ *and* b.

PROOF. The results follow from Lemma A.1 in the Appendix because $N_{\theta_0} \subset S(R, \alpha)$ for some R. We consider part (iii), which illustrates the technicalities. Let $\theta_* = \theta_1$ and $\theta = \theta_2$:

$$\begin{aligned}
&\left\| G_\mu(\theta_*)^{-1} D^3 l_n(\theta) uv \right\|_b \\
&\quad \le M \left\| G_\mu(\theta_*)^{-1} D^3 l_n(\theta) uv \right\|_{*b} \qquad (2.21)\\
&\quad = M \sum_{\nu=1}^{\infty} [1 + \gamma_{*\nu}]^b [1 + \mu\gamma_{*\nu}]^{-2} \{D^3 l_n(\theta_*) uv\phi_{*\nu}\}^2;
\end{aligned}$$

but, from Lemma A.1(iii),

$$\begin{aligned}
&\le A_n \cdot M\|u\|_\alpha^2 \cdot \|v\|_\alpha^2 \cdot \sum_{\nu=1}^{\infty} [1 + \gamma_{*\nu}]^b [1 + \mu\gamma_{*\nu}]^{-2} \\
&\quad \times \left\{ \|\phi_{*\nu}\|_0^2 + n^{-1}\|\phi_{*\nu}\|_\alpha^2 \right\},
\end{aligned}$$

and from Lemma 2.4(i) and (iii) (the condition that $b < 2 - d/2m - \alpha$ is used here)

$$\le A_n \cdot M \cdot \|u\|_\alpha^2 \cdot \|v\|_\alpha^2 \cdot \mu^{-(b+d/2m)}\left\{1 + n^{-1}\mu^{-\alpha}\right\}. \qquad (2.22)$$

This proves part (iii). The arguments for parts (i) and (ii) are very similar, just replace the application of Lemma A.1(iii) by Lemma A.1(i) and (ii), respectively. □

3. Error analysis. Second order Taylor series expansions of the penalized partial likelihood (and its limiting version) yield linear approximations to the systematic (bias) and stochastic (variance) components of the estimation error. The linear approximations are defined by the following:

(i) Continuous linearization:

$$\bar{\theta}_\mu - \theta_0 = -G_\mu(\theta_0)^{-1} Dl_\mu(\theta_0). \qquad (3.1)$$

(ii) Discrete linearization:

$$\bar{\theta}_{n\mu} - \theta_\mu = -G_\mu(\theta_\mu)^{-1}\left[Dl_n(\theta_\mu) - Dl(\theta_\mu)\right]. \qquad (3.2)$$

We will show that $\theta_\mu - \theta_0$ and $\theta_{n\mu} - \theta_\mu$ may be used to approximate $\bar{\theta}_\mu - \theta_0$ and $\bar{\theta}_{n\mu} - \theta_\mu$, respectively. Here θ_μ is defined as the minimizer of the limiting penalized partial likelihood $l_\mu(\theta)$; see (1.7). Recall $\theta_0 \in \Theta_p = W_{02}^{mp}(X; \mathbb{R})$. For the systematic error we have the following result.

THEOREM 3.1 (Systematic error bound). *There is some $\mu_0 > 0$ such that for any $\mu \in [0, \mu_0]$, θ_μ exists and is uniquely defined. Also for $\alpha < (p - d/2m)/2$ there is a constant C_p such that for $0 \le b \le \alpha$ as $\mu \to 0$,*

$$\|\theta_\mu - \theta_0\|_b^2 \le C_p \mu^{(p-b)} \|\theta_0\|_p^2. \tag{3.3}$$

PROOF. We apply Theorem 3.1 of CO which requires $d(\mu, \alpha) \to 0$ and $r(\mu, \alpha) \to 0$ as $\mu \to 0$, where $d(\mu, b) = \|\bar{\theta}_\mu - \theta_0\|_b$ and $r(\mu, b) = K_3(\mu, b) d(\mu, \alpha)$. Since $Dl_\mu(\theta_0) = Dl(\theta_0) + \mu W \theta_0 = \mu W \theta_0$ we have the representation

$$\bar{\theta}_\mu - \theta_0 = \left[U(\theta_0) + \mu W\right]^{-1} U(\theta_0)\theta_0 - \theta_0. \tag{3.4}$$

Thus using Theorem 2.3(c) of Cox (1988) $d(\mu, b)^2 \le C_p \mu^{p-b} \cdot \|\theta_0\|_p^2$. Combining this with Lemma 2.5(ii) gives $r(\mu, b)^2 \le C_p \|\theta_0\|_p^2 \cdot \mu^{-(b+d/2m)} \cdot \mu^{p-\alpha}$. Since $\alpha < p$, $d(\mu, \alpha) \to 0$, and since $2\alpha < p - d/2m$, $r(\mu, \alpha) \to 0$ as $\mu \to 0$. Thus Assumption A.5 of CO holds and from Theorem 3.1 of CO, the conclusion of the theorem follows. □

Before considering the stochastic component of the error we analyze $\|\bar{\theta}_{n\mu} - \theta_\mu\|_b$ for $b \le \alpha$.

LEMMA 3.2. *Let μ_n be any sequence tending to zero such that for some α satisfying $d/2m < \alpha < (p - d/2m)/2$, $n^{-1}\mu_n^{-2(\alpha+d/2m)} \to 0$. Let $0 \le b \le \alpha$ and let μ_{*n} be a deterministic sequence in $[\mu_n, \mu_0]$ (μ_0 as in Theorem 3.1). We have*

$$\|\theta_{n\mu*n} - \theta_{\mu*n}\|_b^2 = O_p\left(n^{-1}\mu_{*n}^{-(b+d/2m)}\right)$$

and

$$\sup_{\mu \in [\mu_n, \mu_0]} \left\{ \frac{\|\theta_{n\mu} - \theta_\mu\|_b^2}{n^{-1}\mu^{-(b+d/2m)}} \right\} = O_p\left(\log(\mu_n^{-1})\right).$$

PROOF. By first order Taylor series expansion about θ_0 [see Rall (1969)]

$$\begin{aligned} \bar{\theta}_{n\mu} - \theta_\mu &= G_\mu(\theta_\mu)^{-1}\left[Dl_n(\theta_0) - Dl(\theta_0)\right] \\ &\quad + \int_0^1 G_\mu(\theta_\mu)^{-1}\left[D^2 l_n(\theta^{(s)}) - Dl^2(\theta^{(s)})\right](\theta_\mu - \theta_0) s\, ds, \end{aligned} \tag{3.5}$$

where $\theta^{(s)} = \theta_0 + s(\theta_\mu - \theta_0)$. Thus

$$\begin{aligned} \|\bar{\theta}_{n\mu} - \theta_\mu\|_b^2 &\le \left\| G_\mu(\theta_\mu)^{-1}\{Dl_n(\theta_0) - Dl(\theta_0)\} \right\|_b^2 \\ &\quad + \tfrac{1}{2} K_{2n}(\mu, b)^2 \cdot \|\theta_\mu - \theta_0\|_\alpha^2. \end{aligned} \tag{3.6}$$

Note, since $\alpha < (p - d/2m)/2$ and $p \le 1$, $\alpha < 2 - \alpha - d/2m$. Thus using Theorem 3.1 and Lemma 2.5(iii) the second term is bounded above by $A_n M n^{-1} \mu^{-(b+d/2m)} \mu^{-\alpha} \mu^{p-\alpha}$, where $A_n = O_p(1)$. Since $\alpha < (p - d/2m)/2$

and $b \le \alpha$, this is less than $A_n M n^{-1} \mu^{-(b+d/2m)}$. To prove the theorem we need to develop bounds for the first term. Let $R_\mu = G_\mu(\theta_0)^{-1}[U(\theta_\mu) - U(\theta_0)]$ and $x_n(\mu) = G_\mu(\theta_0)^{-1}[Dl_n(\theta_0) - Dl(\theta_0)]$ so

$$G_\mu(\theta_\mu)^{-1}[Dl_n(\theta_0) - Dl(\theta_0)] = [I + R_\mu]^{-1} x_n(\mu). \tag{3.7}$$

R_μ is clearly a bounded linear operator on Θ_α. Let

$$|R_\mu|_{\alpha,b} = \sup_{\|\theta\|_\alpha = 1} \|R_\mu \theta\|_b. \tag{3.8}$$

By Taylor series expansion, $R_\mu = \int_0^1 G_\mu(\theta_0)^{-1} D^3 l(\theta^{(s)})(\theta_\mu - \theta_0) s\, ds$, where $\theta^{(s)} = \theta_0 + s(\theta_\mu - \theta_0)$. Thus from the definition of K_3,

$$|R_\mu|_{\alpha,b} \le \tfrac{1}{2} K_3(\mu, b) \|\theta_\mu - \theta_0\|_\alpha \le M \cdot \mu^{-(b+d/2m)/2} \cdot \mu^{(p-\alpha)/2}, \tag{3.9}$$

where M is a positive constant. Since $\alpha + b < 2\alpha < p - d/2m$ we can choose μ_0 such that for all $\mu < \mu_0$, $|R_\mu|_{\alpha,\alpha} < 1/2$. Since

$$\|R_\mu^k \theta\|_b \le |R_\mu|_{\alpha,b} \cdot |R_\mu|_{\alpha,\alpha}^{k-1} \|\theta\|_\alpha$$

we have by expansion in a Neumann series $[I + R_\mu]^{-1} = \sum_{k=0}^\infty (-1)^k R_\mu^k$ (which is valid for elements in Θ_α)

$$\begin{aligned} \left\| [I + R_\mu]^{-1} x_n(\mu) \right\|_b &\le \sum_{k=0}^{\infty} \left\| R_\mu^k x_n(\mu) \right\|_b \\ &\le \| x_n(\mu) \|_b + |R_\mu|_{\alpha,b} \cdot \| x_n(\mu) \|_\alpha \cdot \sum_{k=0}^{\infty} 2^{-k} \\ &\le \| x_n(\mu) \|_b + 2 |R_\mu|_{\alpha,b} \cdot \| x_n(\mu) \|_\alpha. \end{aligned} \tag{3.10}$$

Thus

$$\begin{aligned} \frac{\|\bar{\theta}_{n\mu} - \theta_\mu\|_b^2}{n^{-1}\mu^{-(b+d/2m)}} &\le \frac{\| x_n(\mu) \|_b^2}{n^{-1}\mu^{-(b+d/2m)}} \\ &\quad + 2M\mu^{p-2\alpha-d/2m} \frac{\| x_n(\mu) \|_\alpha^2}{n^{-1}\mu^{-(\alpha+d/2m)}} + A_n M \\ &\le M \left[\frac{\| x_n(\mu) \|_b^2}{n^{-1}\mu^{-(b+d/2m)}} + \frac{\| x_n(\mu) \|_\alpha^2}{n^{-1}\mu^{-(\alpha+d/2m)}} + A_n \right]. \end{aligned} \tag{3.11}$$

To prove the results we will now show that for $b \le \alpha$,

$$\begin{aligned} &\frac{\| x_n(\mu) \|_b^2}{n^{-1}\mu^{-(b+d/2m)}} \le M O_p(1) \quad \text{and} \\ &\sup_{\mu \in [\mu_n, \mu_0]} \frac{\| x_n(\mu) \|_b^2}{n^{-1}\mu^{-(b+d/2m)}} \le M O_p\left(\log(\mu_n^{-1}) + 1\right). \end{aligned} \tag{3.12}$$

From the mapping principle in Weinberger (1974), Assumption A.4 and Proposition 2.1 of CO we have, for $\theta_* \in N_{\theta_0}$ and $b \in [0,1]$, $\|\theta\|_b \le M\|\theta\|_{*b}$ for all $\theta \in \Theta_b$. Also for all ν greater than some ν_0, $0 < m \le (\gamma_{*\nu}/\nu^{2m/d}) \le M < \infty$. Thus

$$\begin{aligned} \left\|G_\mu(\theta_0)^{-1}\{Dl_n(\theta_0) - Dl(\theta_0)\}\right\|_b^2 &\le M\left\|G_\mu(\theta_0)^{-1}\{Dl_n(\theta_0) - Dl(\theta_0)\}\right\|_{0b}^2 \\ &= M\sum_{\nu=1}^{\infty}[1+\gamma_{0\nu}]^b[1+\mu\gamma_{0\nu}]^{-2}B_\nu^{(n)} \\ &\le M\sum_{\nu=1}^{\infty}[1+\bar{\gamma}_\nu]^b[1+\mu\bar{\gamma}_\nu]^{-2}B_\nu^{(n)}, \end{aligned} \tag{3.13}$$

where $\bar{\gamma}_\nu = \nu^{2m/d}$ and $B_\nu^{(n)} = \{Dl_n(\theta_0)\phi_{0\nu} - Dl(\theta_0)\phi_{0\nu}\}^2$. But

$$\begin{aligned} Dl_n(\theta_0)\phi_{0\nu} - Dl(\theta_0)\phi_{0\nu} &= \int\left[\frac{Ds_n(\theta_0,t)\phi_{0\nu}}{s_n(\theta_0,t)} - \frac{Ds(\theta_0,t)\phi_{0\nu}}{s(\theta_0,t)}\right]d\bar{N}(t) \\ &\quad + \int\frac{Ds(\theta_0,t)\phi_{0\nu}}{s(\theta_0,t)}\left[d\bar{N}(t) - s(\theta_0,t)\lambda_0(t)\,dt\right] \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\int\phi_{0\nu}(x_i(t))\,dM_i(t) \\ &\quad - \int\left[Ds_n(\theta_0,t)\phi_{0\nu} - Ds(\theta_0,t)\phi_{0\nu}\right]\lambda_0(t)\,dt. \end{aligned} \tag{3.14}$$

The expected values of the squares of the last three terms are easily computed and shown to be bounded by $Mn^{-1}\|\phi_{0\nu}\|_{00}^2$. For the first term,

$$\begin{aligned} &\left\{\int\left[\frac{Ds_n(\theta_0,t)\phi_{0\nu}}{s_n(\theta_0,t)} - \frac{Ds(\theta_0,t)\phi_{0\nu}}{s(\theta_0,t)}\right]d\bar{N}(t)\right\}^2 \\ &\quad \le \left\{\int\left[\frac{Ds_n(\theta_0,t)\phi_{0\nu} - Ds(\theta_0,t)\phi_{0\nu}}{s_n(\theta_0,t)}\right]d\bar{N}(t)\right\}^2 \\ &\quad\quad + \left\{\int\left[\frac{Ds(\theta_0,t)\phi_{0\nu}}{s(\theta_0,t)}\,\frac{s_n(\theta_0,t) - s(\theta_0,t)}{s_n(\theta_0,t)}\right]d\bar{N}(t)\right\}^2. \end{aligned}$$

Applying arguments used in Lemma A.1(i) in the Appendix gives bounds of $O(n^{-1}\|\phi_{0\nu}\|_0^2)$ for the expectations of both terms on the right-hand side of the latter expression. Thus

$$EB_\nu^{(n)} \le Mn^{-1}\|\phi_{0\nu}\|_{00}^2, \tag{3.15}$$

where M does not depend on n or ν. The first part of (3.12) now follows directly by Markov's inequality because

$$
\begin{aligned}
E\|x_n(\mu)\|_b^2 &\le \sum_{\nu=1}^{\infty}[1+\bar{\gamma}_\nu]^b[1+\mu\bar{\gamma}_\nu]^{-2}EB_\nu^{(n)} \\
&\le Mn^{-1}\sum_{\nu=1}^{\infty}[1+\bar{\gamma}_\nu]^b[1+\mu\bar{\gamma}_\nu]^{-2} \le Mn^{-1}\mu^{-(b+d/2m)}
\end{aligned} \tag{3.16}
$$

as $\mu \to 0$. The last inequality comes from Lemma 2.4(iii). For the second part of (3.12) consider

$$
h_n(\mu) = \frac{\sum_{\nu=1}^{\infty}[1+\bar{\gamma}_\nu]^b[1+\mu\bar{\gamma}_\nu]^{-2}B_\nu^{(n)}}{n^{-1}\mu^{-(b+d/2m)}}. \tag{3.17}
$$

Clearly

$$
\sup_{\mu\in[\mu_n,\mu_0]}|h_n(\mu)| \le |h_n(\mu_n)| + \int_{\mu_n}^{\mu_0}|h_n'(\mu)|\,d\mu. \tag{3.18}
$$

The condition on α and the previous argument implies $|h_n(\mu_n)| \le M\,O_p(1)$. For the second term by direct computation $|h_n'(\mu)| \le \mu^{-1}|h_n(\mu)|$ so

$$
E|h_n'(\mu)| \le c\mu^{-1}\left\{\frac{\sum_{\nu=1}^{\infty}[1+\bar{\gamma}_\nu]^b[1+\mu\bar{\gamma}_\nu]^{-2}}{\mu^{-(b+d/2m)}}\right\}. \tag{3.19}
$$

The term in brackets is bounded (uniformly in $\mu \in [0,\mu_0]$) by Lemma 2.4(iii). Thus there is a constant M such that

$$
E\int_{\mu_n}^{\mu_0}|h_n'(\mu)|\,d\mu \le M\int_{\mu_n}^{\mu_0}\mu^{-1}\,d\mu \le M\log(\mu_n^{-1}). \tag{3.20}
$$

From (3.18) $\sup_{\mu\in[\mu_n,\mu_0]}|h_n(\mu)| \le M\,O_p(1+\log(\mu_n^{-1}))$. □

The final result follows by applying Theorem A.2 of the Appendix.

THEOREM 3.3 (Stochastic error bound). *Let $\mu_n = O(n^{-\delta})$ for some $\delta > 0$ and suppose for some α satisfying $d/2m < \alpha < (p - d/2m)/2$,*

$$
n^{-1}\mu_n^{-2(\alpha+d/2m)} \to 0.
$$

Let $0 \le b \le \alpha$ and μ_{*n} be a deterministic sequence in $[\mu_n,\mu_0]$. There is $\mu_0 > 0$ and some constant $0 < M < \infty$ independent of μ and n such that the event: $\theta_{n\mu}$ exists for all $\mu \in [\mu_n,\mu_0]$ satisfying

$$
\|\theta_{n\mu_{*n}} - \theta_{\mu_{*n}}\|_b^2 \le Mn^{-1}\mu_{*n}^{-(b+d/2m)}
$$

and

$$
\sup_{\mu\in[\mu_n,\mu_0]}\left\{\frac{\|\theta_{n\mu}-\theta_\mu\|_b^2}{n^{-1}\mu^{-(b+d/2m)}}\right\} \le M\log(\mu_n^{-1}),
$$

occurs with probability approaching unity as $n \to \infty$.

PROOF. Let $d_n(\mu, b) = \|\bar{\theta}_{n\mu} - \theta_\mu\|_b$ and $r_n(\mu, b) = K_{2n}(\mu, b) + K_{3n}(\mu, b)d_n(\mu, \alpha)$. By definition of μ_n, $n^{-1}\mu_n^{-2(\alpha+d/2m)} = O(n^{-\varepsilon})$ for some $\varepsilon > 0$. Thus $n^{-1}\mu_n^{-2(\alpha+d/2m)}\log(\mu_n^{-1}) \to 0$. From Theorem 3.1, by choice of μ_0, $\|\theta_\mu - \theta_0\|_\alpha < R/2$ for all $\mu \in [0, \mu_0]$. Thus with $\tau = R/2$, $S_{\theta_\mu}(\tau, \alpha) \subset S(R, \alpha) = N_{\theta_0}$, for all $\mu \in [0, \mu_0]$. From Lemma 3.2, μ_0 can be chosen so that

$$\sup_{\mu \in [\mu_n, \mu_0]} d_n(\mu, \alpha)^2 \le O_p\big(n^{-1}\log(\mu_n^{-1})\mu_n^{-(\alpha+d/2m)}\big). \tag{3.21}$$

This tends to zero in probability. Using Lemma 2.5,

$$\begin{aligned}\sup_{\mu \in [\mu_n, \mu_0]} & r_n(\mu, \alpha)^2 \\ &\le O_p\big(n^{-1}\mu_n^{-(2\alpha+d/2m)} + \log(\mu_n^{-1})\mu_n^{-2(\alpha+d/2m)}\big) \to_p 0.\end{aligned} \tag{3.22}$$

Thus Assumption C in the Appendix is satisfied and so by Theorem A.2 the event that $\theta_{n\mu}$ uniquely exists for all $\mu \in [\mu_n, \mu_0]$ with

$$\|\theta_{n\mu} - \bar{\theta}_{n\mu}\|_b \le r_n(\mu, b)d_n(\mu, \alpha)$$

occurs with probability approaching unity as $n \to \infty$. On this event,

$$\begin{aligned}\|\theta_{n\mu} - \theta_\mu\|_b &\le \|\bar{\theta}_{n\mu} - \theta_\mu\|_b + \|\theta_{n\mu} - \bar{\theta}_{n\mu}\|_b \\ &\le d_n(\mu, b) + r_n(\mu, b)d_n(\mu, \alpha).\end{aligned} \tag{3.23}$$

Using Lemma 3.2,

$$\begin{aligned}\sup_{\mu \in [\mu_n, \mu_0]} & \left\{ \frac{r_n(\mu, b)^2 d_n(\mu, \alpha)^2}{n^{-1}\mu^{-(b+d/2m)}} \right\} \\ &\le O_p\big(n^{-1}\log(\mu_n^{-1})\{\mu_n^{-(2\alpha+d/2m)} + \mu_n^{-2(\alpha+d/2m)}\}\big) \to_p 0.\end{aligned} \tag{3.24}$$

Thus the convergence rate of $\|\theta_\mu - \theta_\mu\|_b$ is determined by the convergence of $d_n(\mu, b)$. From here the results follow by Lemma 3.2. □

APPENDIX

The first result in this Appendix concerns bounds on the derivatives of the partial likelihood and its limiting form (see Section 2). Assumptions A and B are in force throughout. Spectral decomposition of U and W (see Section 2) in $W_2^m(X; \mathbb{R})$ leads to a representation for the W_2^{bm}-norm:

$$\|u\|_b^2 = \sum_\nu [1 + \gamma_\nu]^b u_\nu^2, \tag{A.1}$$

where $u_\nu = \int \phi_\nu(x)u(x)\,dx$ and γ_ν $(\equiv \nu^{d/2m})$ and ϕ_ν are the eigenvalues and eigenfunctions arising from the spectral decomposition.

LEMMA A.1 (Bounds on derivatives). *Let $\alpha > d/2m$ and $R > 0$ be given. There is a constant $0 < M_R < \infty$ and a random variable $A_n = O_p(1)$ such that for all $\theta_* \in S(R, \alpha)$ we have the following*:

(i) $\{D^3 l(\theta_*)uvw\}^2 \le M_R \|u\|_\alpha^2 \|v\|_\alpha^2 \|w\|_0^2$.
(ii) $\{D^2 l_n(\theta_*)uv - D^2 l(\theta_*)uv\}^2 \le A_n M_R n^{-1} \|u\|_\alpha^2 \|v\|_\alpha^2$.
(iii) $\{D^3 l_n(\theta_*)uvw\}^2 \le A_n M_R \|u\|_\alpha^2 \|v\|_\alpha^2 \{\|w\|_0^2 + n^{-1}\|w\|_\alpha^2\}$.

PROOF. The constant M_R is used generically in the proof, successive appearances will typically involve larger values. The technicalities are illustrated by parts (i) and (ii). For part (i), computing directly

$$
\begin{aligned}
D^3 l(\theta_*)uvw = \int \Bigg[& \frac{D^3 s(\theta_*, t)uvw}{s(\theta_*, t)} \\
& - \frac{D^2 s(\theta_*, t)vw}{s(\theta_*, t)} \frac{Ds(\theta_*, t)u}{s(\theta_*, t)} \\
& - \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)} \frac{Ds(\theta_*, t)w}{s(\theta_*, t)} \\
& - \frac{D^2 s(\theta_*, t)uw}{s(\theta_*, t)} \frac{Ds(\theta_*, t)v}{s(\theta_*, t)} \\
& + \frac{Ds(\theta_*, t)u}{s(\theta_*, t)} \frac{Ds(\theta_*, t)v}{s(\theta_*, t)} \frac{Ds(\theta_*, t)w}{s(\theta_*, t)} \Bigg] s(\theta_0, t)\lambda_0(t)\, dt.
\end{aligned} \tag{A.2}
$$

Each term in this expression is analyzed separately. Fortunately the analysis is very similar for the different terms. For the first term, direct application of the Cauchy–Schwarz inequality and Lemma 2.2 gives

$$
\begin{aligned}
& \left\{ \int \frac{D^3 s(\theta_*, t)uvw}{s(\theta_*, t)} s(\theta_0, t)\lambda_0(t)\, dt \right\}^2 \\
& \le M_R \int \left\{ \int u(x)v(x)w(x)e^{\theta_*(x)}\, dx \right\}^2 \lambda_0(t)\, dt \\
& \le M_R \sup_{x \in X} |u(x)|^2 \sup_{x \in X} |v(x)|^2 \|w \mid L_2\|^2 \\
& \le M_R \|u\|_\alpha^2 \|v\|_\alpha^2 \|w\|_0^2.
\end{aligned} \tag{A.3}
$$

The last inequality follows from Sobolev's imbedding theorem because $\alpha > d/2m$. Similar analysis of the other terms leads to the bound in (i).

For part (ii),

$$
\begin{aligned}
&D^2 l_n(\theta_*)uv - D^2 l(\theta_*)uv \\
&\quad = \int \left[\frac{D^2 s_n(\theta_*, t)uv}{s_n(\theta_*, t)} - \frac{D s_n(\theta_*, t)u}{s_n(\theta_*, t)} \frac{D s_n(\theta_*, t)v}{s_n(\theta_*, t)} \right] d\overline{N}(t) \\
&\qquad - \int \left[\frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)} - \frac{D s(\theta_*, t)u}{s(\theta_*, t)} \frac{D s(\theta_*, t)v}{s(\theta_*, t)} \right] s(\theta_0, t)\lambda_0(t)\, dt.
\end{aligned} \tag{A.4}
$$

We analyze a representative term,

$$
\begin{aligned}
&\int \frac{D^2 s_n(\theta_*, t)uv}{s_n(\theta_*, t)}\, d\overline{N}(t) - \int \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)} s(\theta_0, t)\lambda_0(t)\, dt \\
&\quad = \int \frac{D^2 s_n(\theta_*, t)uv}{s_n(\theta_*, t)} - \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)}\, d\overline{N}(t) \\
&\qquad + \int \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)} \left[d\overline{N}(t) - s(\theta_0, t)\lambda_0(t)\, dt \right].
\end{aligned} \tag{A.5}
$$

Here write

$$
\begin{aligned}
&\int \frac{D^2 s_n(\theta_*, t)uv}{s_n(\theta_*, t)} - \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)}\, d\overline{N}(t) \\
&\quad = \int \frac{D^2 s_n(\theta_*, t)uv - D^2 s(\theta_*, t)uv}{s_n(\theta_*, t)}\, d\overline{N}(t) \\
&\qquad + \int \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)} \frac{[s_n(\theta_*, t) - s(\theta_*, t)]}{s_n(\theta_*, t)}\, d\overline{N}(t).
\end{aligned} \tag{A.6}
$$

Let $\psi(x) = e^{\theta_*(x)}u(x)v(x)$. Now $\|\psi\|_\alpha = \|e^{\theta_*}uv\|_\alpha \le \|e^{\theta_*}\|_\alpha \|u\|_\alpha \|v\|_\alpha$ for $\alpha > d/2m$ (see Appendix of CO) and by series representation of the exponential $\|e^{\theta_*}\|_\alpha \le M_R e^{\|\theta_*\|_\alpha}$. Thus $\psi \in W_2^{m\alpha}$. Let $\psi(x) = \sum_\nu \psi_\nu \phi_\nu(x)$, where $\psi_\nu = \int \psi(x)\phi_\nu(x)\, dx$. Also $r_n(t)\phi_\nu = (1/n)\sum_{i=1}^n y_i(t)\phi_\nu(x_i(t))$ and $r(t)\phi_\nu = \int \phi_\nu(x) q(x,t)\, dx$. Substituting for ψ and applying the Cauchy–Schwarz inequality to the summation over ν gives

$$
\begin{aligned}
&\left\{ \int \left[\frac{D^2 s_n(\theta_*, t)uv - D^2 s(\theta_*, t)uv}{s_n(\theta_*, t)} \right] d\overline{N}(t) \right\}^2 \\
&\quad = \left\{ \sum_\nu [1 + \gamma_\nu]^{\alpha/2} \psi_\nu \cdot [1 + \gamma_\nu]^{-\alpha} \left\{ \int \left[\frac{r_n(t)\phi_\nu - r(t)\phi_\nu}{s_n(\theta_*, t)} \right] d\overline{N}(t) \right\}^2 \right\} \\
&\quad \le \|\psi\|_\alpha^2 \sum_\nu [1 + \gamma_\nu]^{-\alpha} \left\{ \int \left[\frac{r_n(t)\phi_\nu - r(t)\phi_\nu}{s_n(\theta_*, t)} \right] d\overline{N}(t) \right\}^2.
\end{aligned}
$$

Above we have used the fact that $\Sigma_\nu[1+\gamma_\nu]^\alpha\psi_\nu^2 \le M_R\Sigma_\nu[1+\gamma_\nu^\alpha]\psi_\nu^2 = M_R\|\psi\|_\alpha^2$. Using the Cauchy–Schwarz inequality and Lemma 2.1(i),

$$\left\{\int\left[\frac{r_n(t)\phi_\nu - r(t)\phi_\nu}{s_n(\theta_*,t)}\right]d\bar{N}(t)\right\}^2$$

$$\text{(A.7)}\qquad \le \left\{\int\frac{[r_n(t)\phi_\nu - r(t)\phi_\nu]^2}{s_n(\theta_*,t)}\,d\bar{N}(t)\right\}\cdot\left\{\int\frac{1}{s_n(\theta_*,t)}\,d\bar{N}(t)\right\}$$

$$\le M_R\cdot\left\{\int\frac{[r_n(t)\phi_\nu - r(t)\phi_\nu]^2}{s_n(\theta_0,t)}\,d\bar{N}(t)\right\}\cdot\left\{\int\frac{1}{s_n(\theta_0,t)}\,d\bar{N}(t)\right\}.$$

Now $E\{\int[1/s_n(\theta_0,t)]\,d\bar{N}(t)\} = E\{\int(1/s_n(\theta_0,t))s_n(\theta_0,t)\lambda_0(t)\,dt\} = \Lambda_0(1) < \infty$. Also

$$E\left\{\int\frac{[r_n(t)\phi_\nu - r(t)\phi_\nu]^2}{s_n(\theta_0,t)}\,d\bar{N}(t)\right\}$$

$$\text{(A.8)}\qquad = \int E[r_n(t)\phi_\nu - r(t)\phi_\nu]^2\lambda_0(t)\,dt$$

$$\le M_R n^{-1}\int \operatorname{Var}\phi_\nu(x(t))\,dt \le M_R n^{-1}\|\phi_\nu\|_0^2.$$

Hence by Markov's inequality

$$\text{(A.9)}\qquad \left\{\int\frac{D^2s_nn(\theta_*,t)uv - D^2s(\theta_*,t)uv}{s_n(\theta_*,t)}\,d\bar{N}(t)\right\}^2 \le M_R A_n\|u\|_\alpha^2\|v\|_\alpha^2 n^{-1},$$

where A_n (independent of u, v and θ_*) is $O_p(1)$. Also

$$E\left\{\int\frac{D^2s(\theta_*,t)uv}{s(\theta_*,t)}\,\frac{[s_n(\theta_*,t) - s(\theta_*,t)]}{s_n(\theta_*,t)}\,d\bar{N}(t)\right\}^2$$

$$\text{(A.10)}\qquad \le M_R\left\{\sup_t D^2s(\theta_*,t)uv\right\}^2\cdot\int\frac{[s_n(\theta_*,t) - s(\theta_*,t)]^2}{s_n(\theta_0,t)}\,d\bar{N}(t)$$

$$\cdot\int\frac{1}{s_n(\theta_0,t)}\,d\bar{N}(t).$$

Here Assumption A(iii) and the Cauchy–Schwarz inequality gives

$$\sup_t D^2s(\theta_*,t)uv \le M_R\|u\|_0\|v\|_0,$$

so by repeating the previous argument leading to (A.9) we obtain an upper bound of $\|u\|_\alpha^2\|v\|_\alpha^2 n^{-1}M_R A_n$ for (A.10).

To deal with the remaining term in (A.5) let $g(t) = [D^2 s(\theta_*, t)uv/s(\theta_*, t)]$. From Assumption A(iii),

$$\|g\|_{L_2}^2 + \left\|\frac{dg}{dt}\right\|_{L_2}^2 \le M_R \|u\|_{L_2}^2 \|v\|_{L_2}^2. \tag{A.11}$$

Therefore $g \in W_2^1[0,1]$. Thus g has the representation $g(t) = \Sigma_\nu g_\nu b_\nu(t)$, where b_ν are $L_2[0,1]$-orthonormal functions and $\|g\|_{W_2^1}^2 = \Sigma_\nu [1+\nu]^2 g_\nu^2$. Substituting the series expansion for g and applying the Cauchy–Schwarz inequality to the sum over ν [as in the argument before (A.7)] gives

$$\begin{aligned}&\left\{\int \frac{D^2 s(\theta_*, t)uv}{s(\theta_*, t)}\left[d\overline{N}(t) - s(\theta_0, t)\lambda_0(t)\,dt\right]\right\}^2\\ &\quad\le \left\{\sum_\nu [1+v]^2 g_\nu^2\right\}\\ &\qquad\cdot \sum_\nu [1+v]^{-2}\left\{\int b_\nu(t)\left[d\overline{N}(t) - s(\theta_0, t)\lambda_0(t)\,dt\right]\right\}^2.\end{aligned} \tag{A.12}$$

But

$$\begin{aligned}&E\left\{\int b_\nu(t)\left[d\overline{N}(t) - s(\theta_0, t)\lambda_0(t)dt\right]\right\}^2\\ &\quad\le E\left\{\int b_\nu(t)\,d\overline{M}(t)\right\}^2 + E\left\{\int b_\nu(t)[s_n(\theta_0, t) - s(\theta_0, t)]\lambda_0(t)\,dt\right\}^2\\ &\quad\le Mn^{-1}\|b_\nu\|_{L_2}^2 = Mn^{-1}.\end{aligned} \tag{A.13}$$

Combining results gives the required bound for (A.5). Similar arguments are applied for the other part of (A.4) and this gives the bound in part (ii). The proof of part (iii) uses techniques already encountered in the analysis of parts (i) and (ii). □

Uniform linearization result. A generalization of a linearization result in CO for penalized likelihood estimators is now proved. Consider a penalized likelihood functional $l_{n\mu}$ defined on a real Hilbert space Θ by

$$l_{n\mu}(\theta) = l_n(\theta) + \frac{\mu}{2}\langle\theta, W\theta\rangle \tag{A.14}$$

for $\mu > 0$. The limiting version of $l_n(\theta)$ is denoted $l(\theta)$. We assume that l_n, l, W and Θ satisfy Assumptions A.1–A.4 of CO. From Assumption A.3 of CO there is a bounded linear compact operator U on Θ such that $D^2 l(\theta_0)\phi\phi$ is equivalent to $\langle\phi, U\phi\rangle$ for $\phi \in \Theta$. Considering the spectral decomposition of W relative to U (as in Section 2 or Section 2 of CO) leads to a set of spaces Θ_b with norms $\|\cdot\|_b$ for $b \in [0,1]$. The true parameter is in $N_{\theta_0} \subset \Theta_\alpha$ for some α. Let $Z_{n\mu}(\theta) = Dl_n(\theta) + \mu W\theta$, where D is the Frechet derivative operator with respect to θ. Define $U(\theta)$ by $\langle \phi, U(\theta)\phi\rangle = D^2 l(\theta)\phi\phi$ for $\theta \in N_{\theta_0}$ and for $\theta_\mu \in N_{\theta_0}$ let

$$\text{(A.15)}\qquad \begin{aligned} \bar{\theta}_{n\mu} - \theta_\mu &= \left[U(\theta_\mu) + \mu W\right]^{-1} Z_{n\mu}(\theta_\mu) \\ &= \left[U(\theta_\mu) + \mu W\right]^{-1}\left\{Dl_n(\theta_\mu) - Dl(\theta_\mu)\right\}. \end{aligned}$$

The existence of these quantities is justified in Section 2 of CO. Let

$$d_n(\mu, b) = \|\bar{\theta}_{n\mu} - \theta_\mu\|_b \quad \text{and} \quad r_n(\mu, b) = K_{2n}(\mu, b) + K_{3n}(\mu, b) d_n(\mu, \alpha),$$

where K_{2n} and K_{3n} are norms of the second and third order derivative operators defined as in Section 2 of this paper, see also Section 2 of CO. We make the following assumption.

ASSUMPTION C. For some $\tau > 0$, $S_{\theta_\mu}(\alpha, \tau) \subset N_{\theta_0}$ for $\mu \in [\mu_n, \mu_0]$, and

$$\text{(A.16)}\qquad \sup_{\mu \in [\mu_n, \mu_0]} d_n(\mu, \alpha) \to_p 0 \quad \text{and} \quad \sup_{\mu \in [\mu_n, \mu_0]} r_n(\mu, \alpha) \to_p 0.$$

THEOREM A.2 (Uniform existence of $\theta_{n\mu}$ and linearization). *Let α and $b \in [0, \alpha]$ be given and suppose μ_0 and μ_n satisfy Assumption* C. *With probability tending to unity as $n \to \infty$, for all $\mu \in [\mu_n, \mu_0]$ we have the following*:

(i) *There is a unique root $\theta_{n\mu}$ of $Z_{n\mu}(\theta) = 0$ in $S_{\theta_\mu}(2d_n(\mu, \alpha), \alpha)$, and*
(ii) $\|\theta_{n\mu} - \bar{\theta}_{n\mu}\|_b \le 2 \cdot r_n(\mu, b) \cdot d_n(\mu, \alpha)$.

PROOF. For $\tau > 0$, let E_τ^n be the event that

$$\text{(A.17)}\qquad \sup_{\mu \in [\mu_n, \mu_0]} d_n(\mu, \alpha) < m/2 \quad \text{and} \quad \sup_{\mu \in [\mu_n, \mu_0]} r_n(\mu, \alpha) < \tau/2.$$

The probability of E_τ^n can be made arbitrarily large for all $n > n_0$. Let $t_{n\mu} = 2d_n(\mu,\alpha)$ and choose $m < 1$ so that Assumption C holds. Now we restrict to the event E_τ^n. Here $S_{\theta_\mu}(t_{n\mu}, \alpha) \subset N_{\theta_0}$ for all $n > n_0$ and $\sup_{\mu \in [\mu_n, \mu_0]} r_n(\mu, \alpha) < 1/2$. Let $F_{n\mu}(\phi) = \phi - [U(\theta_\mu) + \mu W]^{-1} Z_{n\mu}(\theta_\mu + \phi)$ for $\phi \in \Theta_\alpha$. Repeating the computations in Theorem 3.2 of CO gives

$$\text{(A.18)}\qquad \left\| F_{n\mu}(\phi) \right\|_\alpha \le \left[r_n(\mu, \alpha) + \tfrac{1}{2}\right] t_{n\mu} < \left[\tfrac{1}{2} + \tfrac{1}{2}\right] t_{n\mu} = t_{n\mu}$$

and

$$\text{(A.19)}\qquad \left\| F_{n\mu}(\phi_1) - F_{n\mu}(\phi_2) \right\|_b \le \left[\tfrac{1}{2}\right] \|\phi_1 - \phi_2\|_b,$$

which hold for all $\mu \in [\mu_n, \mu_0]$. Thus $F_{n\mu}$ is a contraction on $S_{\theta_\mu}(t_{n\mu}, \alpha)$ for all $\mu \in [\mu_n, \mu_0]$. From here the argument of Theorem 3.2 in CO is repeated to obtain parts (i) and (ii) of the theorem. □

Acknowledgments. I am very grateful to Professors John Crowley and Grace Wahba for stimulating discussions related to this paper.

REFERENCES

ADAMS, R. (1975). *Sobolev Spaces*. Academic, New York.

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120.

COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

COX, D. D. (1988). Approximation of method of regularization estimators. *Ann. Statist.* **16** 694–712.

COX, D. D. and O'SULLIVAN, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18** 1676–1695.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31** 377–403.

GILL, R. D. (1984). Understanding Cox's regression model: A martingale approach *J. Amer. Statist. Assoc.* **79** 441–447.

O'SULLIVAN, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Statist. Comp.* **9** 531–542.

O'SULLIVAN, F., YANDELL, B and RAYNOR W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–104.

RALL, L. B. (1969). *Computational Solution of Nonlinear Operator Equations*. Wiley, New York.

STONE, C. J. (1982). Optimal global convergence rates for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

TIKHONOV, A. and ARENIN, V. (1977). *Solutions of Ill-Posed Problems*. Wiley, New York.

TRIEBEL, H. (1978). *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam.

WAHBA, G. (1990). *Spline Models in Statistics*. SIAM, Philadelphia.

WEINBERGER, H. F. (1974). *Variational Methods for Eigenvalue Approximation*. SIAM, Philadelphia.

ZUCKER, D. M. and KARR, A. F. (1990). Nonparametric survival analysis with time dependent covariates: A penalized partial likelihood approach. *Ann. Statist.* **18** 329–353.

DEPARTMENTS OF BIOSTATISTICS
AND STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195