The Annals of Statistics
1994, Vol. 22, No. 1, 211–237

DATA-DRIVEN EFFICIENT ESTIMATORS FOR A PARTIALLY LINEAR MODEL

BY HUNG CHEN[1] AND JYH-JEN HORNG SHIAU[2]

State University of New York, Stony Brook and National Chiao-Tung University

Chen and Shiau showed that a two-stage spline smoothing method and the partial regression method lead to efficient estimators for the parametric component of a partially linear model when the smoothing parameter is a deterministic sequence tending to zero at an appropriate rate. This paper is concerned with the large-sample behavior of these estimators when the smoothing parameter is chosen by the generalized cross validation (GCV) method or Mallows' C_L. Under mild conditions, the estimated parametric component is asymptotically normal with the usual parametric rate of convergence for both spline estimation methods. As a by-product, it is shown that the "optimal rate" for the smoothing parameter, with respect to expected average squared error, is the same for the two estimation methods as it is for ordinary smoothing splines.

1. Introduction. In this paper, we study the asymptotic behavior of the two efficient estimators for the parametric component of a partially linear model discussed in Chen and Shiau (1991) when the smoothing parameter is chosen either by the generalized cross validation (GCV) method proposed by Craven and Wahba (1979) or by the Mallows C_L criterion [Mallows (1973)]. As in Chen and Shiau (1991), we consider a semiparametric regression model

$$y_{in} = \mathbf{x}_{in}^T\boldsymbol{\beta} + g(t_{in}) + e_{in}, \qquad i = 1, \dots, n, \tag{1}$$

where both the $\mathbf{x}_{in} = (x_{i1n}, \dots, x_{idn})^T$ (a d-vector) and $t_{in} \in [0, 1]$ are observed design variables, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ is a vector of unknown regression coefficients, g is a smooth function to be estimated and the $\{e_{in}\}$ are independent and identically distributed errors when mean zero and variance σ^2.

Several estimation methods for model (1) have been proposed in the literature. See Chen and Shiau (1991) and the references cited therein. Chen and Shiau (1991) discussed the asymptotic behavior of the following three estimators.

(i) The *partial spline estimator* [proposed by Engle, Granger, Rice and Weiss (1986), Wahba (1984, 1986) and Shiau, Wahba and Johnson (1986), among

Received June 1991; revised July 1992.

[1] Supported by NSF Grant DMS-89-01556.

[2] Part of this work was done while J. Shiau was at the Engineering Research Center, AT&T Bell Laboratories.

AMS 1991 *subject classifications.* Primary 62G05, 62G99, 62J99.

Key words and phrases. Partial splines, semiparametric regression, smoothing splines, rate of convergence, partial regression, generalized cross validation, Mallows' C_L, efficient estimators.

others] is the solution to the following variational problem:

$$\min_{\beta \in R^d,\, g \in W_2^m} \frac{1}{n} \sum_{i=1}^{n} [y_{in} - \mathbf{x}_{in}^T \beta - g(t_{in})]^2 + \lambda \int_0^1 [g^{(m)}(t)]^2 \, dt, \tag{2}$$

where W_2^m is the Sobolev space $\{f | f$ has $m-1$ absolutely continuous derivatives and $f^{(m)} \in L_2[0,1]\}$ and λ is the smoothing parameter controlling the tradeoff between fidelity to data and roughness of the solution. It is known that the partial spline estimators for β and $\mathbf{g} = (g(t_{1n}), \dots, g(t_{nn}))^T$ are

$$\widehat{\beta}_\lambda = (X^T(I - S_\lambda)X)^{-1} X^T (I - S_\lambda)\mathbf{y} \quad \text{and} \quad \widehat{\mathbf{g}}_\lambda = S_\lambda(\mathbf{y} - X\widehat{\beta}_\lambda), \tag{3}$$

where $X = (x_{irn})$ is the $n \times d$ design matrix for the parametric component of (1), $\mathbf{y} = (y_{1n}, \dots, y_{nn})^T$ and S_λ is the smoother matrix for ordinary spline smoothing [i.e., when $\beta = \mathbf{0}$ in (2)].

(ii) The *partial regression estimator* was proposed independently by Denby (1986) and Speckman (1988). Motivated by the partial regression scheme in linear regression, the partial regression estimator is obtained by first smoothing X and $\mathbf{y}$, respectively, by the smoother matrix S_λ and then regressing the residuals of $\mathbf{y}$ on the residuals of X. Specifically, we have the partial regression estimator defined by

$$\widehat{\beta}_{1\lambda} = (X^T(I - S_\lambda)^2 X)^{-1} X^T (I - S_\lambda)^2 \mathbf{y} \quad \text{and} \quad \widehat{\mathbf{g}}_{1\lambda} = S_\lambda(\mathbf{y} - X\widehat{\beta}_{1\lambda}). \tag{4}$$

(iii) The *two-stage spline smoothing estimator* was recently proposed by Chen and Shiau (1991). For simplicity, we shall discuss a simplified version of the estimator when the same smoothing parameter is used in both stages of smoothing, namely,

$$\begin{aligned} \widehat{\beta}_{0\lambda} &= (X^T(I - S_\lambda)^3 X)^{-1} X^T (I - S_\lambda)^2 \mathbf{y}, \\ \widehat{\mathbf{g}}_{0\lambda} &= S_\lambda(\mathbf{y} - X\widehat{\beta}_{0\lambda}) - (I - S_\lambda) S_\lambda X \widehat{\beta}_{0\lambda}. \end{aligned} \tag{5}$$

The basic idea behind this estimator is to modify the partial spline method so that roughness of the parametric component is penalized as well as that of the nonparametric component. Thus we first smooth X to obtain the residuals $(I - S_\lambda)X$ for the purpose of extracting the smooth part from the parametric component, and then we apply the partial spline technique to smooth $\mathbf{y}$ over $(I - S_\lambda)X$. This two-stage smoothing gives (5).

In general, the smoother matrix S_λ in (3), (4) and (5) can be replaced by any commonly used smoother matrix. Of course, estimators obtained by different smoothers may behave differently. See Chen and Shiau (1991) for some remarks. In this paper, we only study the case that S_λ is the smoothing spline smoother.

To use these three methods to estimate β and g in practice, it is necessary to specify a value of the smoothing parameter λ. In the context of nonparametric regression, it is well known that the choice of λ is very crucial to the solution. A popular data-driven method of choosing λ is the generalized cross validation (GCV) method (to be described in Section 2). Numerically, the GCV method has been proven to be a good method. Speckman (1981) and Li (1986) gave some nice theoretical results on the GCV method. However, the use of the GCV method for determining the value of λ in (3), (4) or (5) has not yet been thoroughly examined. To our knowledge the only relevant reference is Speckman (1988), who gave a weak GCV theorem as in Craven and Wahba (1979) for the partial regression estimator (4) in the context of kernel smoothing.

There have been some studies on the asymptotic behavior of the preceding three estimators, when λ is a deterministic quantity depending on n, in the setting that $x_{irn} = h_r(t_{in}) + z_{irn}$, where the h_r's are smooth functions and $\{(z_{i1n}, \dots, z_{idn})\}_{1 \le i \le n}$ are independent and identically distributed error vectors with zero mean and positive definite covariance matrix. For the partial spline estimator with spline smoothing, Rice (1986) pointed out that $\widehat{\beta}_\lambda - \beta$ can achieve the usual parametric rate of convergence as in parametric regression, namely, $O(n^{-1/2})$, only at the expense of undersmoothing the nonparametric component g. Thus Rice (1986) concluded that the use of the GCV method for choosing λ is questionable in this case.

On the other hand, Speckman (1988), for the partial regression estimator with kernel smoothing, and Chen and Shiau (1991), for the two-stage spline smoothing estimator as well as the partial regression estimator with spline smoothing, showed that the negative result reported in Rice (1986) disappears. More specifically, by choosing an appropriate rate for λ, the convergence rate of $\widehat{\beta}_{0\lambda} - \beta$ or $\widehat{\beta}_{1\lambda} - \beta$ reaches the parametric rate $O(n^{-1/2})$ while $\widehat{\mathbf{g}}_{0\lambda}$ or $\widehat{\mathbf{g}}_{1\lambda}$ can still estimate $\mathbf{g} = (g(t_{1n}), \dots, g(t_{nn}))^T$ with the same optimal convergence rate as that of the ordinary nonparametric regression estimator, which is achievable by the GCV estimator of λ. Basically, Chen and Shiau (1991) demonstrated that the goal of obtaining an estimate for the regression surface $g(\cdot)$ with an "optimal" nonparametric convergence rate does not conflict with the goal of obtaining an estimate for the parametric component β with the parametric convergence rate. Since "optimal" estimates of the regression surface can be obtained by the method of GCV for the nonparametric regression context, we expect that the parametric convergence rate can be achieved for some estimators of β, such as (4) and (5), for the semiparametric model (1). The following conjecture is hence reasonably made by Speckman (1988) for kernel smoothing and Chen and Shiau (1991) for spline smoothing.

CONJECTURE. *The GCV method can be used to choose the value of* λ *in* (4) *or* (5) *such that* $\widehat{\beta}_{1\lambda}$ *or* $\widehat{\beta}_{0\lambda}$ *can still estimate* β *with* $n^{-1/2}$ *rate.*

The main objective of this paper is to prove this conjecture when S_λ is the smoother matrix for ordinary spline smoothing. We also prove that the same result holds if λ is chosen by the criterion of Mallows' C_L. We remark that although the problem of determining smoothing parameters for nonparametric regression based on data only is studied extensively in the literature [see Li (1986) and references therein], those results are not applicable in general to the problem posed in this article. A further remark on this is given in Section 2.

The main results are summarized in Theorems 1 and 2 (Section 2), in which the asymptotic distributions of $\widehat{\beta}_{0\lambda} - \beta$ and $\widehat{\beta}_{1\lambda} - \beta$ are derived when the smoothing parameter is determined by either the (restricted) GCV method or (restricted) Mallows' C_L. Descriptions of these two methods are given in Section 2. Most of the proofs are given in the remaining sections.

As a by-product of proving Theorems 1 and 2, it is shown in Propositions 1(b) and 3(b) that the "optimal rate" for the smoothing parameter, with respect to expected average squared error, is the same for the two estimation methods as it is for ordinary smoothing splines.

As suggested by a referee, we also have looked into the situation studied by Heckman (1986). When $h_r \equiv$ constant, Heckman (1986) established asymptotic normality for the partial spline estimator of β and showed that its bias is asymptotically negligible. According to the preceding discussion, it is expected that the GCV method can be used to choose the value of λ in (3) such that $\widehat{\beta}_\lambda$ can still estimate β with $n^{-1/2}$ rate under the setting of Heckman (1986). This conjecture is also confirmed for a more general case where the h_r's are polynomial of degree less than m, and the result is presented as Theorem 3 in Section 2.

2. Data-driven methods and main results. In this section we describe the (restricted) GCV method and (restricted) Mallows' C_L for determining the value of λ in (4) and (5) and present the main results of this paper. We first introduce some notation. Write

$$X\widehat{\beta}_{0\lambda} + \widehat{\mathbf{g}}_{0\lambda} = [S_\lambda + (I - S_\lambda)^2 X(X^T(I - S_\lambda)^3 X)^{-1} X^T (I - S_\lambda)^2]\mathbf{y} = A_{0\lambda}\mathbf{y},$$
$$X\widehat{\beta}_{1\lambda} + \widehat{\mathbf{g}}_{1\lambda} = [S_\lambda + (I - S_\lambda) X(X^T(I - S_\lambda)^2 X)^{-1} X^T (I - S_\lambda)^2]\mathbf{y} = A_{1\lambda}\mathbf{y},$$

where $A_{0\lambda}$ and $A_{1\lambda}$ are so-called hat matrices or influence matrices. Let $\widehat{\lambda}_{0G}$ be the minimizer of the generalized cross validation function (GCV function)

$$V_0(\lambda) = \frac{n^{-1}\left\|(I - A_{0\lambda})\mathbf{y}\right\|^2}{[n^{-1}\operatorname{tr}(I - A_{0\lambda})]^2}$$

over $\lambda \in [\lambda_1, \lambda_2]$ where $\lambda_1 = n^{-\delta_1}\log^m n$ with $\delta_1 = 2m/5$, and $\lambda_2 = n^{-\delta_2}$ for any δ_2 satisfying $0 < \delta_2 < 2m/(4m+1)$. Also $\|(I - A_{0\lambda})\mathbf{y}\|^2 = \mathbf{y}^T(I - A_{0\lambda})^T(I - A_{0\lambda})\mathbf{y}$, the residual sum of squares. Similarly, let $\widehat{\lambda}_{0C}$ denote the minimizer of Mallows' C_L

$$C_{0L}(\lambda) = n^{-1}\left\|(I - A_{0\lambda})\mathbf{y}\right\|^2 + 2n^{-1}\sigma^2 \operatorname{tr} A_{0\lambda}$$

over $\lambda \in [\lambda_1, \lambda_2]$, where σ^2 is assumed known. For the partial regression method, $\widehat{\lambda}_{1G}$, $V_1(\lambda)$, $\widehat{\lambda}_{1C}$ and $C_{1L}(\lambda)$ are defined accordingly for $A_{1\lambda}\mathbf{y}$.

It is known that there exists a common orthonormal basis for all S_λ (with λ being the running index), for example, a Demmler–Reinsch basis [Demmler and Reinsch (1975)]. In other words, all S_λ can be diagonalized simultaneously by this basis. Further details of this basis are given in Section 3. Unfortunately, it is not clear whether there exists such a common orthonormal basis for all $A_{0\lambda}$ or $A_{1\lambda}$ in general. Although both the GCV method and Mallows' C_L have been studied in the context of nonparametric regression when S_λ is the smoother matrix for smoothing splines, these results are not applicable to our problem since the arguments used to prove these results depend strongly on the existence of a common orthonormal basis for all S_λ.

Throughout the rest of the paper, we assume that $\{\mathbf{x}_{in}\}$ is a random sample from $\mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_d)^T$, $x_r = h_r(t) + z_r$, for $1 \le r \le d$, $t \in [0, 1]$ and the h_r's are smooth functions. Set $g_0 = \Sigma_{r=1}^d \beta_r h_r + g$. We also assume that the following conditions hold.

(A1) $Ez_r = 0$, $\mathrm{Var}((z_1, \dots, z_d)) = \Sigma = (\sigma_{rs})$ and $Ez_r^4 < \infty$, for $1 \le r \le d$, where Σ is a $d \times d$ positive definite matrix.

(A2) $\int_0^1 (g_0^{(m)}(t))^2 \, dt = \gamma > 0$ and $m \ge 2$.

(A3) The points t_{in} are generated by $(2i-1)/2n = \int_0^{t_{in}} p(t)\, dt$ for some density function $p(t)$ on [0, 1].

(A4) The errors $e_{1n}, \dots, e_{nn}$ are i.i.d. having a distribution independent of n and t, and $Ee_{in}^4 < \infty$, for $i = 1, 2, \dots, n$.

(A5) $g, h_r, g_0 \in \mathcal{F} = \{f\colon f \in W_2^{2m}[0,1], f^{(k)}(0) = f^{(k)}(1) = 0,\ m \le k \le 2m-1\}$ for $1 \le r \le d$.

Under (A3), we can find the magnitude of $\mathrm{tr}\, S_\lambda^l$ for $l = 1, 2, \dots$ over $[\lambda_1, \lambda_2]$ based on Lemma 5.1 of Speckman (1981). This result is summarized in Lemma 2(c). Under (A5), functions in $\mathcal{F}$ are the so-called very smooth functions defined in Wahba (1977). When A2 also holds, it follows from Speckman [(1981), (3.2) and Lemma 3.1] that an exact bound can be obtained for $\mathbf{g}_0^T(I - S_\lambda)^2 \mathbf{g}_0$, where $\mathbf{g}_0 = (g_0(t_{1n}), \dots, g_0(t_{nn}))^T$. This bound is given in Lemma 2(b) in Section 3.

We now discuss the assumption (A5), which states that g_0 and h_r must satisfy boundary conditions on some high derivatives. (A5) is considered because it and (A2) give an explicit asymptotic expression for the expectation of the averaged squared error loss. Then this expression can be used to determine the asymptotic behavior of λ determined by either the GCV method or Mallows' C_L. Using the bias reduction approach developed by Eubank and Speckman (1991), g_0 and h_r can be modified (by construction) to satisfy the boundary conditions specified in (A5). It is then conjectured that a result similar to that of this paper without (A5) will still hold as long as an explicit asymptotic expression for the expectation of the averaged squared error loss exists after the boundary adjustment. However, no proof is available now.

The asymptotic distribution of $\widehat{\beta}_{0\hat{\lambda}}$ and $\widehat{\beta}_{1\hat{\lambda}}$ are summarized in Theorems 1 and 2, respectively, when the value of λ is determined by either the GCV method or Mallows' C_L.

THEOREM 1. *Under* (A1)–(A5), $\sqrt{n}(\widehat{\beta}_{0\hat{\lambda}} - \beta)$ *converges in distribution to* $N(0, \sigma^2\Sigma^{-1})$ *for* $\widehat{\lambda} = \widehat{\lambda}_{0G}$ *or* $\widehat{\lambda}_{0C}$.

THEOREM 2. *Under* (A1)–(A5), $\sqrt{n}(\widehat{\beta}_{1\hat{\lambda}} - \beta)$ *converges in distribution to* $N(0, \sigma^2\Sigma^{-1})$ *for* $\widehat{\lambda} = \widehat{\lambda}_{1G}$ *or* $\widehat{\lambda}_{1C}$.

Now we describe the (restricted) GCV method and (restricted) Mallows' C_L for determining the value of λ in (3) under the assumption that h_r's are polynomial of degree less than m [i.e., $h_r^{(m)}(t) \equiv 0$]. The results are summarized in Theorem 3. First write

$$X\widehat{\beta}_\lambda + \widehat{\mathbf{g}}_\lambda = [S_\lambda + (I - S_\lambda)X(X^T(I - S_\lambda)X)^{-1}X^T(I - S_\lambda)]\mathbf{y} = A_\lambda \mathbf{y}.$$

Let $\widehat{\lambda}_G$ and $\widehat{\lambda}_C$ be the minimizer of the corresponding GCV function and Mallows' C_L, respectively, over $\lambda \in [\lambda_1, \lambda_2]$.

THEOREM 3. *Under* (A1)–(A5) *and* $h_r^{(m)}(t) \equiv 0$, *for* $1 \le r \le d$, $\sqrt{n}(\widehat{\beta}_{\widehat{\lambda}} - \beta)$ *converges in distribution to* $N(0, \sigma^2\Sigma^{-1})$, *for* $\widehat{\lambda} = \widehat{\lambda}_G$ *or* $\widehat{\lambda}_C$.

Let $L_{0n}(\lambda)$ denote the averaged squared error loss over design points, that is, $n^{-1}\|A_{0\lambda}\mathbf{y} - X\beta - \mathbf{g}\|^2$, and λ_{0R} denote the value of λ that minimizes the risk $R_{0n}(\lambda) = EL_{0n}(\lambda)$ over $[\lambda_1, \lambda_2]$. Note that here the expectation is taken with respect to e only, that is, conditioned on $(\mathbf{x}, t)$. We will prove Theorem 1 via the following three steps. Since the GCV method or Mallows' C_L attempts to provide a data-based estimate of λ_{0R}, we first try to locate λ_{0R}. Let $\Lambda_0 = [\lambda_1, n^{-\delta_3}]$, where $\delta_1 > 2m/(4m+1) > \delta_3 > \delta_2 > \frac{1}{4}$. Note that Λ_0 is contained in $[\lambda_1, \lambda_2]$. We show in Proposition 1 that $\lambda_{0R} \in \Lambda_0$. Next, we show in Proposition 2 that the choice of λ based on either the GCV method or Mallows' C_L does fall in Λ_0 in probability. Finally, we show that $\sqrt{n}(\widehat{\beta}_{0\widehat{\lambda}} - \beta)$ is asymptotically normal.

Set $\mathbf{h}_r = (h_r(t_{1n}), \ldots, h_r(t_{nn}))^T$, for $1 \le r \le d$, and

$$c_l = \pi \left[\int_0^1 p^{1/2m}(v)\, dv\right]^{-1} \int_0^\infty (1 + v^{2m})^{-l}\, dv.$$

The proofs of the following two propositions are given in Section 4.

PROPOSITION 1. *Under* (A1)–(A5) *and* $\lambda \in [\lambda_1, \lambda_2]$, *when n tends to infinity, we have* (a) $R_{0n}(\lambda) \approx \lambda^2 + n^{-1}\lambda^{-1/2m}$ *and* (b) $\lambda_{0R} \approx n^{-2m/(4m+1)}$.

Here the symbol $a(n) \approx b(n)$ means that $a(n)/b(n)$ is bounded away from zero and infinity. Note that $\lambda_{0R} \in \Lambda_0$ is an immediate result of (b).

PROPOSITION 2. *Under* (A1)–(A5) *and* $\lambda \in [\lambda_1, \lambda_2]$, $\lim_n P(\widehat{\lambda} \in \Lambda_0) = 1$, *for* $\widehat{\lambda} = \widehat{\lambda}_{0G}$ *or* $\widehat{\lambda}_{0C}$.

To prove Theorem 1, we use the following technical lemma to pave the way. Set $A_0(\lambda) = n^{-1}X^T(I - S_\lambda)^3X$, $A_1(\lambda) = n^{-1}X^T(I - S_\lambda)^2X$, $Z = (z_{irn})_{n \times d}$ and $H = (h_r(t_{in}))_{n \times d}$.

LEMMA 1. *Assume that* (A1)–(A5) *hold and that* $g, h_r \in \mathcal{F}$, *for* $1 \le r \le d$. *Then the following hold uniformly over all* $\lambda \in [\lambda_1, \lambda_2]$:

(a) $A_0(\lambda) = \Sigma(I + o_p(1))$;
(b) $A_1(\lambda) = \Sigma(I + o_p(1))$;

and the following hold uniformly over all $\lambda \in \Lambda_0$:

(c) $n^{-1/2}X^T(I - S_\lambda)^2 S_\lambda X = o_p(1)$;
(d) $n^{-1/2}X^T(I - S_\lambda)^2\mathbf{g} = o_p(1)$;
(e) $n^{-1/2}H^T(I - S_\lambda)^2\mathbf{e} = o_p(1)$;
(f) $n^{-1/2}Z^T S_\lambda^l \mathbf{e} = o_p(1)$, *for* $l = 1, 2$;
(g) $n^{-1/2}Z^T(I - S_\lambda^l)\mathbf{g} = o_p(1)$.

The proof of Lemma 1 is given at the end of Section 3. Note that the notation $o_p(1)$ used in this paper denotes either the usual convention or a $d \times d$ (or $d \times 1$) matrix such that the magnitude of each element is $o_p(1)$.

Now the proof of Theorem 1 becomes fairly simple.

PROOF OF THEOREM 1. Rewrite

$$A_0(\widehat{\lambda})n^{1/2}(\widehat{\beta}_{0\widehat{\lambda}} - \beta) = n^{-1/2}Z^T\mathbf{e} + \mathrm{Rem}(\widehat{\lambda}),$$

where

$$\mathrm{Rem}(\lambda) = n^{-1/2}\{X^T(I - S_\lambda)^2(S_\lambda X\beta + \mathbf{g}) + H^T(I - S_\lambda)^2\mathbf{e} + Z^T[(I - S_\lambda)^2 - I]\mathbf{e}\}.$$

It follows from Lemma 1(c)–(f) that $\sup_{\lambda\in\Lambda_0}|\mathrm{Rem}(\lambda)| = o_p(1)$. Although any realization of $\widehat{\lambda}$ is in $[\lambda_1, \lambda_2]$, which is a wider interval than Λ_0, by noting that, for any $c > 0$,

$$\begin{aligned} P\left(\left|\mathrm{Rem}(\widehat{\lambda})\right| > c\right) &\le P(\widehat{\lambda} \notin \Lambda_0) + P\left(|\mathrm{Rem}(\widehat{\lambda})| > c \text{ and } \widehat{\lambda} \in \Lambda_0\right) \\ &\le P(\widehat{\lambda} \notin \Lambda_0) + P\left(\sup_{\lambda\in\Lambda_0}|\mathrm{Rem}(\lambda)| > c\right), \end{aligned}$$

we can conclude that $\mathrm{Rem}(\widehat{\lambda}) = o_p(1)$ by Proposition 2.

By Lemma 1(a), $\sup_{\lambda\in[\lambda_1,\lambda_2]}|A_0(\lambda) - \Sigma| = o_p(1)$. Since

$$|A_0(\widehat{\lambda}) - \Sigma| \le \sup_{\lambda\in[\lambda_1,\lambda_2]}|A_0(\lambda) - \Sigma| = o_p(1),$$

we have $A_0(\widehat{\lambda}) \to \Sigma$ in probability. It is shown in Chen and Shiau (1991) that $n^{-1/2}Z^T\mathbf{e} \to N(\mathbf{0}, \sigma^2\Sigma)$ in distribution. We then conclude $\sqrt{n}(\widehat{\beta}_{0\widehat{\lambda}} - \beta) \to N(\mathbf{0}, \sigma^2\Sigma^{-1})$ by the above argument and Slutsky's theorem. □

We now turn to the partial regression estimator (4). Observe that

$$\begin{aligned} A_1(\lambda)n^{1/2}(\widehat{\beta}_{1\lambda} - \beta) = {}& n^{-1/2}Z^T\mathbf{e} + n^{-1/2}X^T(I - S_\lambda)^2\mathbf{g} + n^{-1/2}H^T(I - S_\lambda)^2\mathbf{e} \\ &+ n^{-1/2}Z^T[(I - S_\lambda)^2 - I]\mathbf{e}. \end{aligned}$$

Similarly, the proof of Theorem 2 can be performed via the following two propositions and Lemma 1(b)–(f).

Let the loss function $L_{1n}(\lambda) = n^{-1}\|A_{1\lambda}\mathbf{y} - X\beta - \mathbf{g}\|^2$, and let λ_{1R} denote the value of the smoothing parameter that minimizes the risk $R_{1n}(\lambda) = EL_{1n}(\lambda)$ over $\lambda \in [\lambda_1, \lambda_2]$.

PROPOSITION 3. *Under* (A1)–(A5) *and* $\lambda \in [\lambda_1, \lambda_2]$, *when* n *tends to infinity, we have* (a) $R_{1n}(\lambda) \approx \lambda^2 + n^{-1}\lambda^{-1/2m}$ *and* (b) $\lambda_{1R} \approx n^{-2m/(4m+1)}$.

PROPOSITION 4. *Under* (A1)–(A5) *and* $\lambda \in [\lambda_1, \lambda_2]$, $\lim_n P(\widehat{\lambda} \in \Lambda_0) = 1$ *for* $\widehat{\lambda} = \widehat{\lambda}_{1G}$ *or* $\widehat{\lambda}_{1C}$.

We now turn to the partial spline estimator (3) when $h_r^{(m)}(t) \equiv 0$.

PROOF OF THEOREM 3. Set $A_2(\lambda) = n^{-1}X^T(I - S_\lambda)X$. Rewrite

$$A_2(\widehat{\lambda})n^{1/2}(\widehat{\beta}_{\widehat{\lambda}} - \beta) = n^{-1/2}Z^T\mathbf{e} + \mathrm{Rem}(\widehat{\lambda}),$$

where

$$\mathrm{Rem}(\lambda) = n^{-1/2}\left\{Z^T(I - S_\lambda)\mathbf{g} + H^T(I - S_\lambda)(\mathbf{e} + \mathbf{g}) - Z^T S_\lambda \mathbf{e}\right\}.$$

Note that $H^T(I - S_\lambda)(\mathbf{e} + \mathbf{g}) \equiv 0$ because the h_r's are polynomials of degree less than m and S_λ is the smoother matrix for ordinary spline smoothing. Using the same proof to show Lemma 1(a), we have $\sup_{\lambda \in [\lambda_1, \lambda_2]} |A_2(\lambda) - \Sigma| = o_p(1)$. It follows from Lemma 1(f) and (g) that $\sup_{\lambda \in \Lambda_0} |\mathrm{Rem}(\lambda)| = o_p(1)$. We then conclude $\sqrt{n}(\widehat{\beta}_{\widehat{\lambda}} - \beta) \to N(\mathbf{0}, \sigma^2\Sigma^{-1})$ by the above discussion and the argument used in proving Theorem 1. □

3. Technical lemmas. In this section we state two more technical lemmas and summarize some properties of smoothing splines that are needed in the sequel. Lemma 1 is proved as an immediate result of these lemmas.

It is well known that smoothing splines are in the space of natural polynomial splines of order $2m$ on $[0, 1]$ with knot set $\{t_{in}\}_{i=1}^n$. According to Demmler and Reinsch (1975), a basis for natural splines is $\{\phi_{jn}(t)\}_{1 \le j \le n}$ with the following biorthogonality property:

$$\frac{1}{n}\sum_{i=1}^{n} \phi_{jn}(t_{in})\phi_{kn}(t_{in}) = \delta_{jk}, \qquad \int_0^1 \phi_{jn}^{(m)}(t)\phi_{kn}^{(m)}(t)\,dt = \lambda_{kn}\delta_{jk}.$$

Here $\{\lambda_{kn}\}$ is a nondecreasing sequence of nonnegative numbers, and the eigenvalues of S_λ are $(1 + \lambda\lambda_{kn})^{-1}$ for $1 \le k \le n$. Hence, S_λ is a nonnegative definite matrix and has the eigenvalue decomposition $\Gamma^T D_\lambda \Gamma$, where D_λ is a diagonal $n \times n$ matrix with k-th diagonal value $(1 + \lambda_{kn}\lambda)^{-1}$ and Γ is an orthogonal $n \times n$ matrix with the ij-th element $n^{-1/2}\phi_{in}(t_{jn})$. Therefore, $(I - S_\lambda)^l S_\lambda^k = S_\lambda^l (I - S_\lambda)^k$ for any positive integers l and k.

Let

$$B_{1p}^2 = n^{-1}\mathbf{g}^T(I - S_\lambda)^2\mathbf{g}, \qquad B_{2rp}^2 = n^{-1}\mathbf{h}_r^T(I - S_\lambda)^2\mathbf{h}_r$$
$$\text{and} \quad B_{3p}^2 = n^{-1}\mathbf{g}_0^T(I - S_\lambda)^2\mathbf{g}_0.$$

Note that B_{1p}^2 is the averaged squared bias of the ordinary smoothing spline estimate of $\mathbf{g}$. A similar interpretation is applicable to B_{2rp}^2 and B_{3p}^2.

The following lemma is due to Speckman [(1981), Lemma 3.1, (3.2) and Theorem 2.4].

LEMMA 2. *Suppose that* (A3) *holds. When* $\lambda \in [\lambda_1, \lambda_2]$ *and* $m \geq 2$, (a) $B_{3p}^2 = O(\lambda^2)$ *if* $g_0 \in \mathcal{F}$, (b) $B_{3p}^2 = \gamma\lambda^2(1 + o(1))$ *if* $g_0 \in \mathcal{F}$ *and* (A2) *holds, and* (c) $\operatorname{tr} S_\lambda^l = \sum_k (1 + \lambda_{kn}\lambda)^{-l} = c_l\lambda^{-1/2m}(1 + o(1))$ *for positive integer* l.

Thus Lemma 2(a) also implies that $B_{1p}^2 = O(\lambda^2)$ and $B_{2rp}^2 = O(\lambda^2)$, if $g, h_r \in \mathcal{F}$. Lemma 3 summarizes the convergence rates for some terms to be used later in the proofs of Lemma 1 and Propositions 1–4. Let $x_{irn} = h_r(t_{in}) + z_{irn}$ and $\mathbf{z}_r = (z_{1rn}, \dots, z_{nrn})^T$.

LEMMA 3. *Assume that* (A1)–(A4) *hold and that* $h_r, f, f_1, f_2 \in \mathcal{F}$, *for* $1 \leq r \leq d$. *Let* a, a_0 *and* a_1 *be constants satisfying* $1 < a < 1/a_0 < 5$ *and* $a < a_1$. *Then, for any finite positive integer* l, *the following statements hold uniformly over all* $\lambda \in [\lambda_1, \lambda_2]$ *and* $1 \leq r, s \leq d$:

(a) $\mathbf{z}_r^T S_\lambda^l \mathbf{z}_s = c_l\sigma_{rs}\lambda^{-1/2m} + o_p\left(\lambda^{-1/4ma_0}\right)$;

(b) $\mathbf{e}^T S_\lambda^l \mathbf{e} = \sigma^2 c_l \lambda^{-1/2m}(1 + o_p(1))$;

(c) $n^{-1/2}\mathbf{f}^T(I - S_\lambda)^l\mathbf{e} = O_p(\lambda^{1/a_1}) = o_p(1)$, *where* $\mathbf{f} = (f(t_{1n}), \dots, f(t_{nn}))^T$;

(d) $n^{-1/2}\mathbf{f}^T(I - S_\lambda)^l\mathbf{z}_r = O_p(\lambda^{1/a_1}) = o_p(1)$;

(e) $n^{-1}\mathbf{f}_1^T(I - S_\lambda)^l\mathbf{f}_2 = O(\lambda^2)$, *where* $\mathbf{f}_i = (f_i(t_{1n}), \dots, f_i(t_{nn}))^T$, $i = 1, 2$ *and* $l \geq 2$;

(f) $n^{-1}\mathbf{x}_r^T(I - S_\lambda)^2\mathbf{z}_s = \sigma_{rs} + o_p(1)$;

(g) $n^{-1}\mathbf{x}_r^T(I - S_\lambda)^l\mathbf{x}_s = \sigma_{rs} + o_p(1)$, *for* $l \geq 2$;

(h) $n^{-1/2}\mathbf{x}_r^T(I - S_\lambda)^2 S_\lambda \mathbf{x}_s = o_p(1) + O(n^{1/2}\lambda^2)$;

(i) $\mathbf{z}_r^T S_\lambda^l \mathbf{e} = o_p\left(\lambda^{-1/4ma_0}\right)$;

(j) $n^{-1}\mathbf{z}_r^T(I - S_\lambda)^l\mathbf{e} = O_p(n^{-1/2})$;

(k) $$\mathbf{x}_r^T(I - S_\lambda)^3\mathbf{x}_s - \mathbf{x}_r^T(I - S_\lambda)^2\mathbf{z}_s = -(c_1 - 2c_2 + c_3)\sigma_{rs}\lambda^{-1/2m} + o_p\left(\lambda^{-1/4ma_0}\right) + O_p(n^{1/2}\lambda^{1/a_1}) + \mathbf{h}_r^T(I - S_\lambda)^3\mathbf{h}_s;$$

(l) $$\mathbf{x}_r^T(I - S_\lambda)^3\mathbf{x}_s - \mathbf{x}_i^T(I - S_\lambda)^4\mathbf{x}_s = (c_1 - 3c_2 + 3c_3 - c_4)\sigma_{rs}\lambda^{-1/2m} + o_p(\lambda^{-1/4ma_0}) + O_p(n^{1/2}\lambda^{1/a_1}) + \mathbf{h}_r^T(I - S_\lambda)^3 S_\lambda \mathbf{h}_s.$$

Lemma 3 immediately gives the results of Lemma 1 as shown below. The nontrivial proof of Lemma 3 is deferred to Section 6.

PROOF OF LEMMA 1. First note that $\lambda^{-1/4ma_0} = o(n^{-1/2})$, for all $\lambda \in [\lambda_1, \lambda_2]$ since $1/a_0 < 5$. It is easy to see that (a) and (b) hold by Lemma 3(g). Note that $\lambda^2 = o(n^{-1/2})$, for all $\lambda \in \Lambda_0$. Then it is easy to see that (c) holds by Lemma 3(h); (d) holds by Lemma 3(d) and 3(e); (e) holds by Lemma 3(c); (f) holds by Lemma 3(i); (g) holds by Lemma 3(d). □

4. Proof for two-stage spline smoothing estimate. We prove Propositions 1 and 2 for the two-stage spline smoothing estimates in this section. The following technical lemma summarizes the convergence rates for some terms to be used in the proofs. The proof of the lemma is deferred to Section 7.

LEMMA 4. *Assume that* (A1)–(A4) *hold and that* $g, h_r \in \mathcal{F}$, *for* $1 \le r \le d$. *We further assume that the constants* a, a_0 *and* a_1 *specified in Lemma* 3 *satisfy the further constraint that* $4m/(4m-1) > a_1 > a$ *and* $a_0 > \frac{1}{2}$. *Then the following statements hold uniformly over all* $\lambda \in [\lambda_1, \lambda_2]$:

(a) $n^{-1}\operatorname{tr}A_{0\lambda} = c_1 n^{-1}\lambda^{-1/2m}(1 + o_p(1))$;
(b) $n^{-1}\operatorname{tr}A_{0\lambda}^2 = c_2 n^{-1}\lambda^{-1/2m}(1 + o_p(1))$;
(c) $n^{-1}\mathbf{g}_0^T(I - A_{0\lambda})^2\mathbf{g}_0 = \gamma\lambda^2(1 + o_p(1))$;
(d) $n^{-1}\beta^T Z^T(I - A_{0\lambda})^2 Z\beta$

$$= \left[n^{-1}\left(\sum_r \beta_r \mathbf{h}_r\right)^T (I - S_\lambda)^4 \left(\sum_r \beta_r \mathbf{h}_r\right) + (c_2 - 2c_3 + c_4)\beta^T\Sigma\beta n^{-1}\lambda^{-1/2m}\right](1 + o_p(1));$$

(e) $|n^{-1}\mathbf{g}_0^T(I - A_{0\lambda})^2\mathbf{e}| = o_p(R_{0n}(\lambda))$;
(f) $|n^{-1}(Z\beta)^T(I - A_{0\lambda})^2\mathbf{e}| = o_p(R_{0n}(\lambda))$;
(g) $|n^{-1}|\mathbf{e}^T(2A_{0\lambda} - A_{0\lambda}^2)\mathbf{e} - \sigma^2(2\operatorname{tr}A_{0\lambda} - \operatorname{tr}A_{0\lambda}^2)| = o_p(R_{0n}(\lambda))$.

PROOF OF PROPOSITION 1. Write $A_{0\lambda}\mathbf{y} - X\beta - \mathbf{g} = (A_{0\lambda} - I)(X\beta + \mathbf{g}) + A_{0\lambda}\mathbf{e}$. Hence

$$R_{0n}(\lambda) = n^{-1}(X\beta + \mathbf{g})^T(I - A_{0\lambda})^2(X\beta + \mathbf{g}) + n^{-1}\sigma^2\operatorname{tr}A_{0\lambda}^2.$$

Note that $X\beta + \mathbf{g} = Z\beta + \mathbf{g}_0$. Then

$$R_{0n}(\lambda) = \Bigg\{[c_2\sigma^2 + (c_2 - 2c_3 + c_4)\beta^T\Sigma\beta]n^{-1}\lambda^{-1/2m} + \gamma\lambda^2 + n^{-1}\left(\sum_r \beta_r\mathbf{h}_r\right)^T (I - S_\lambda)^4 \left(\sum_r \beta_r\mathbf{h}_r\right)\Bigg\}(1 + o_p(1)), \tag{6}$$

by Lemma 4(b)–(d).

Note that $n^{-1}(\Sigma_r \beta_r \mathbf{h}_r)^T (I - S_\lambda)^4 (\Sigma_r \beta_r \mathbf{h}_r) \geq 0$ and its order is $O(\lambda^2)$, by Lemma 3(e) and that the eigenvalues of S_λ are between 0 and 1. Also, $(c_2 - 2c_3 + c_4)\beta^T \Sigma \beta \geq 0$, by the fact that $c_2 - 2c_3 + c_4 > 0$ and Σ is positive definite. Hence, Proposition 1(a) holds by (6), and Proposition 1(b) follows easily from Proposition 1(a). □

PROOF OF PROPOSITION 2. Recall that $C_{0L}(\lambda) = n^{-1}\|(I - A_{0\lambda})\mathbf{y}\|^2 + 2n^{-1}\sigma^2 \operatorname{tr} A_{0\lambda}$, which can be written as

$$
\begin{aligned}
C_{0L}(\lambda) &= n^{-1}\mathbf{e}^T\mathbf{e} + R_{0n}(\lambda) + 2n^{-1}(Z\beta + \mathbf{g}_0)^T (I - A_{0\lambda})^2 \mathbf{e} \\
&\quad + n^{-1}\{\sigma^2(2 \operatorname{tr} A_{0\lambda}) - \operatorname{tr} A_{0\lambda}^2) - \mathbf{e}^T(2A_{0\lambda} - A_{0\lambda}^2)\mathbf{e}\} \\
&= n^{-1}\mathbf{e}^T\mathbf{e} + R_{0n}(\lambda) + o_p(R_{0n}(\lambda)),
\end{aligned} \tag{7}
$$

by Lemma 4(e)–(g).

Recall that the GCV function $V_0(\lambda) = n^{-1}\|(I - A_{0\lambda})\mathbf{y}\|^2 [n^{-1}\operatorname{tr}(I - A_{0\lambda})]^{-2}$. Write $A_{0\lambda} = S_\lambda + B_{0\lambda}$, where $B_{0\lambda} = n^{-1}(I - S_\lambda)^2 X A_0^{-1}(\lambda) X^T (I - S_\lambda)^2$. It follows from Lemmas 1(a) and 3(g) that

$$
\operatorname{tr} B_{0\lambda} = \operatorname{tr}\left(A_0^{-1}(\lambda) n^{-1} X^T (I - S_\lambda)^4 X\right) = \operatorname{tr}(I_{d\times d} + o_p(1)) = O_p(1).
$$

Also Lemma 2(c) gives that $\operatorname{tr} S_\lambda = O(\lambda^{-1/2m})$. We then have

$$
\left[n^{-1}\operatorname{tr}(I - A_{0\lambda})\right]^{-2} = 1 + 2n^{-1}\operatorname{tr} A_{0\lambda} + o(n^{-1}\operatorname{tr} A_{0\lambda}).
$$

Observe that

$$
\begin{aligned}
n^{-1}\|(I - A_{0\lambda})\mathbf{y})\|^2 &= R_{0n}(\lambda) + 2n^{-1}(Z\beta + \mathbf{g}_0)^T (I - A_{0\lambda})^2 \mathbf{e} + n^{-1}\mathbf{e}^T\mathbf{e} \\
&\quad - n^{-1}\left[\mathbf{e}^T(2A_{0\lambda} - A_{0\lambda}^2)\mathbf{e} - \sigma^2(2 \operatorname{tr} A_{0\lambda} - \operatorname{tr} A_{0\lambda}^2)\right] \\
&\quad - 2\sigma^2 n^{-1}\operatorname{tr} A_{0\lambda}.
\end{aligned}
$$

The fourth term on the right-hand side is equal to $o_p(R_{0n}(\lambda))$, by Lemma 4(g). The second term is also of the order $o_p(R_{0n}(\lambda))$, by Lemma 4(e) and (f). We thus get

$$
\begin{aligned}
V_0(\lambda) &= \left[n^{-1}\mathbf{e}^T\mathbf{e} + R_{0n}(\lambda) + o_p(R_{0n}(\lambda)) - 2\sigma^2 n^{-1}\operatorname{tr} A_{0\lambda}\right] \\
&\quad \times \left[1 + 2n^{-1}\operatorname{tr} A_{0\lambda} + o(n^{-1}\operatorname{tr} A_{0\lambda})\right] \\
&= n^{-1}\mathbf{e}^T\mathbf{e} + R_{0n}(\lambda) + o_p(R_{0n}(\lambda)) + 2(n^{-1}\operatorname{tr} A_{0\lambda})(n^{-1}\mathbf{e}^T\mathbf{e} - \sigma^2) \\
&= n^{-1}\mathbf{e}^T\mathbf{e} + R_{0n}(\lambda) + o_p(R_{0n}(\lambda)),
\end{aligned} \tag{8}
$$

by Lemma 4(a), Proposition 1(a) and the law of large numbers. From (7) and (8), we have

$$
C_{0L}(\lambda) - C_{0L}(\lambda_{0R}) = R_{0n}(\lambda) - R_{0n}(\lambda_{0R}) + o_p(R_{0n}(\lambda))
$$

and

$$
V_0(\lambda) - V_0(\lambda_{0R}) = R_{0n}(\lambda) - R_{0n}(\lambda_{0R}) + o_p(R_{0n}(\lambda)),
$$

respectively. When $R_{0n}(\lambda)/R_{0n}(\lambda_{0R}) \to \infty$, it follows easily that $C_{0L}(\lambda) > C_{0L}(\lambda_{0R})$ and $V_0(\lambda) > V_0(\lambda_{0R})$ in probability. Since $\widehat{\lambda}$ is the minimizer of $C_{0L}(\lambda)$ or $V_0(\lambda)$, this implies that $R_{0n}(\widehat{\lambda})/R_{0n}(\lambda_{0R}) \to 1$ in probability. Let $\{\delta_n\}$ be any sequence that tends to infinity. Note that $R_{0n}(\lambda_{0R}\delta_n)/R_{0n}(\lambda_{0R}) \to \infty$ and $R_{0n}(\lambda_{0R}/\delta_n)/R_{0n}(\lambda_{0R}) \to \infty$ by Proposition 1(a). Hence, $R_{0n}(\lambda)/R_{0n}(\lambda_{0R}) \to \infty$ for any $\lambda \geq \lambda_{0R}\delta_n$ or $\lambda \leq \lambda_{0R}/\delta_n$. Since $R_{0n}(\widehat{\lambda})/R_{0n}(\lambda_{0R})$ cannot go to infinity, we have that

$$\lim_n P\big(\lambda_{0R}/\delta_n \leq \widehat{\lambda} \leq \lambda_{0R}\delta_n\big) = 1.$$

Since $\{\delta_n\}$ is any sequence that tends to infinity, $\widehat{\lambda}$ cannot be too far away from λ_{0R} in probability. Thus $\lim_n P(\widehat{\lambda} \in \Lambda_0) = 1$. $\square$

5. Proof for the partial regression estimate. First, we state a technical lemma that summarizes the convergence rates for some terms to be used in the proofs of Propositions 3 and 4. We defer the proof of this lemma to Section 7.

LEMMA 5. *Assume that* (A1)–(A4) *hold and that* $g, h_r \in \mathcal{F}$, *for* $1 \leq r \leq d$. *We further assume that the constants* a, a_0 *and* a_1 *specified in Lemma* 3 *satisfy the further constraint that* $4m/(4m-1) > a_1 > a$ *and* $a_0 > \frac{1}{2}$. *Then the following statements hold uniformly over all* $\lambda \in [\lambda_1, \lambda_2]$:

(a) $n^{-1}\mathbf{g}_0^T(I - A_{1\lambda})^T(I - A_{1\lambda})\mathbf{g}_0 = \gamma\lambda^2(1 + o_p(1))$;

(b)
$$\begin{aligned} &n^{-1}\beta^T Z^T (I - A_{1\lambda})^T (I - A_{1\lambda})Z\beta \\ &\quad = n^{-1}\left[\left(\sum_r \beta_r \mathbf{h}_r\right)^T (I - S_\lambda)^2 \left(\sum_r \beta_r \mathbf{h}_r\right) + o_p\left(\lambda^{-1/2m}\right)\right](1 + o_p(1)); \end{aligned}$$

(c) $|n^{-1}(X\beta + \mathbf{g})^T(I - A_{1\lambda})^T(I - A_{1\lambda})\mathbf{e}| = o_p(R_{1n}(\lambda))$.

PROOF OF PROPOSITION 3. Simple algebra leads to

$$R_{1n}(\lambda) = n^{-1}(X\beta + \mathbf{g})^T(I - A_{1\lambda})^T(I - A_{1\lambda})(X\beta + \mathbf{g}) + n^{-1}\sigma^2 \operatorname{tr} A_{1\lambda}^T A_{1\lambda}.$$

Set $A_{1\lambda} = S_\lambda + B_{1\lambda}$, where $B_{1\lambda} = n^{-1}(I - S_\lambda)XA_1^{-1}(\lambda)X^T(I - S_\lambda)^2$. By Lemmas 1(b), 3(g) and 2(c), we have

$$\begin{aligned} \operatorname{tr} B_{1\lambda}^T B_{1\lambda} &= \operatorname{tr} A_1^{-1}(\lambda)\big[n^{-1}X^T(I - S_\lambda)^2 X\big]A_1^{-1}(\lambda)\big[n^{-1}X^T(I - S_\lambda)^4 X\big] \\ &= O_p(1) \end{aligned} \tag{9}$$

and

$$\begin{aligned} &\operatorname{tr} S_\lambda^T B_{1\lambda} = \operatorname{tr} A_1^{-1}(\lambda)\Big\{n^{-1}X^T\big[(I - S_\lambda)^3 - (I - S_\lambda)^4\big]X\Big\} = o_p(1), \\ &\operatorname{tr} S_\lambda^2 = c_2\lambda^{-1/2m}(1 + o(1)). \end{aligned} \tag{10}$$

Hence,

$$n^{-1} \operatorname{tr} A_{1\lambda}^T A_{1\lambda} = c_2 n^{-1} \lambda^{-1/2m}(1 + o_p(1)). \tag{11}$$

Then since $X\beta + \mathbf{g} = Z\beta + \mathbf{g}_0$, we have

$$\begin{aligned} R_{1n}(\lambda) = \Bigg[& c_2\sigma^2 n^{-1}\lambda^{-1/2m} + \gamma\lambda^2 \\ & + n^{-1}\Big(\sum_r \beta_r \mathbf{h}_r\Big)^T (I - S_\lambda)^2 \Big(\sum_r \beta_r \mathbf{h}_r\Big)\Bigg](1 + o_p(1)), \end{aligned} \tag{12}$$

by Lemmas 5(a), 5(b) and (11).

Note that $n^{-1}(\sum_r \beta_r \mathbf{h}_r)^T \ (I - S_\lambda)^2 (\sum_r \beta_r \mathbf{h}_r) \geq 0$ and its order is $O(\lambda^2)$ by Lemma 3(e). Hence, (a) holds; (b) follows easily from (a). □

PROOF OF PROPOSITION 4. We first observe that $\operatorname{tr} B_{1\lambda} = O_p(1)$ by Lemmas 1(b) and 3(g). Then, by Lemma 5(c), it remains to show that

$$n^{-1}|\sigma^2 \operatorname{tr} A_{1\lambda}^T A_{1\lambda} - \mathbf{e}^T(A_{1\lambda} + A_{1\lambda}^T - A_{1\lambda}^T A_{1\lambda})\mathbf{e}| = o(R_{1n}(\lambda)), \tag{13}$$

$$n^{-1}|\sigma^2(2 \operatorname{tr} A_{1\lambda} - \operatorname{tr} A_{1\lambda}^T A_{1\lambda}) - \mathbf{e}^T(A_{1\lambda} + A_{1\lambda}^T - A_{1\lambda}^T A_{1\lambda})\mathbf{e}| = o(R_{1n}(\lambda)) \tag{14}$$

hold uniformly over all $\lambda \in [\lambda_1, \lambda_2]$, so that

$$\begin{aligned} C_{1L}(\lambda) &= n^{-1}\mathbf{e}^T\mathbf{e} + R_{1n}(\lambda) + o_p(R_{1n}(\lambda)), \\ V_{1L}(\lambda) &= n^{-1}\mathbf{e}^T\mathbf{e} + R_{1n}(\lambda) + o_p(R_{1n}(\lambda)). \end{aligned}$$

Then, by applying the same argument employed in Proposition 2, we have Proposition 4.

It follows from Lemmas 1(b), 3(c), 3(j) and 3(g) that

$$\begin{aligned} n^{-1}\mathbf{e}^T B_{1\lambda}\mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)(Z + H)]A_1^{-1}(\lambda)[n^{-1}(Z + H)^T(I - S_\lambda)^2\mathbf{e}] \\ &= o_p(R_{1n}(\lambda)), \end{aligned} \tag{15}$$

$$\begin{aligned} n^{-1}\mathbf{e}^T B_{1\lambda} S_\lambda \mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)^2(Z + H)]A_1^{-1}(\lambda) \\ &\quad \times \{n^{-1}(Z + H)^T[(I - S_\lambda) - (I - S_\lambda)^2]\mathbf{e}\} \\ &= o(R_{1n}(\lambda)), \end{aligned} \tag{16}$$

$$\begin{aligned} n^{-1}\mathbf{e}^T B_{1\lambda}^T B_{1\lambda}\mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)^2(Z + H)]A_1^{-1}(\lambda) \\ &\quad \times [n^{-1}X^T(I - S_\lambda)^2 X]A_1^{-1}(\lambda) \\ &\quad \times [n^{-1}(Z + H)^T(I - S_\lambda)^2\mathbf{e}] \\ &= o_p(R_{1n}(\lambda)). \end{aligned} \tag{17}$$

It follows from Lemmas 3(b) and 2(c) that

$$n^{-1}[\mathbf{e}^T(2S_\lambda - S_\lambda^2)\mathbf{e} - \sigma^2 \operatorname{tr}\,(2S_\lambda - S_\lambda^2)] = o_p(R_{1n}(\lambda)). \tag{18}$$

We conclude (13) and (14) by (9), (10) and (15)–(18). □

6. Proof of Lemma 3. We begin with a technical lemma which is an extension of Lemma 4.4 in Speckman (1985) to the case when the random variables are not independent. Therefore, the Gaussian assumption in Speckman (1985) or Li (1986) is removed.

LEMMA 6. *Let $W_1, \dots, W_n$ be random variables with zero mean and finite variance. Suppose that there exist nonnegative numbers $\{u_k\}$ such that*

$$E\left[\sum_{k=\mu}^{\nu} W_k\right]^2 \le \sum_{k=\mu}^{\nu} u_k, \quad \textit{for all } \mu < \nu.$$

Then, for any $c > 0$,

$$P\left\{\sup_{0\le c_1\le\cdots\le c_n\le c_0}\left|\sum_{k=1}^{n} c_k W_k\right| \ge c\right\} \le c^{-2}c_0^2(\log_2 4n)^2\sum_{k=1}^{n} u_k.$$

PROOF. By the argument used in Lemma 4.4 of Speckman (1985), we have

$$\sup_{0\le c_1\le\cdots\le c_n\le c_0}\left|\sum_{k=1}^{n} c_k W_k\right| = c_0 \max_{1\le i\le n}\left|\sum_{k=1}^{i} W_k\right|.$$

Then, by the first two theorems stated in Serfling [(1970), page 1228],

$$E\left[\max_{1\le i\le n}\left|\sum_{k=1}^{i} W_k\right|\right]^2 \le (\log_2 4n)^2 \sum_{k=1}^{n} u_k.$$

Hence, this lemma holds by Chebyshev's inequality. □

REMARK 1. When $EW_kW_l = 0$, for $k \ne l$, Lemma 6 holds with $u_k = \mathrm{Var}(W_k)$.

REMARK 2. Lemma 6 also holds when $0 \le c_n \le \cdots \le c_1 \le c_0$.

Define

$$\xi_{krn} = n^{-1/2}\sum_{i=1}^{n} z_{irn}\phi_{kn}(t_{in}), \qquad h_{krn} = n^{-1/2}\sum_{i=1}^{n} h_r(t_{in})\phi_{kn}h_r(t_{in}),$$

$$c_{kn} = n^{-1/2}\sum_{i=1}^{n} g(t_{in})\phi_{kn}(t_{in}), \qquad \varepsilon_{kn} = n^{-1/2}\sum_{i=1}^{n} e_{in}\phi_{kn}(t_{in}),$$

for $1 \le k \le n$ and $1 \le r \le d$. Lemma 6 will be applied to $\{\xi_{krn}\xi_{ksn}\}_{1\le k\le n}$ and $\{\xi_{krn}\varepsilon_{kn}\}_{1\le k\le n}$, for $1 \le r, s \le d$, later on in the proof of Lemma 3. Thus we need to show that these two sequences of random variables satisfy the assumption of Lemma 6.

LEMMA 7. *For any finite positive integer l and $1 \le r, s \le d$, both*

$$\left\{\left(\xi_{krn}\xi_{ksn} - \sigma_{rs}\right)\left(1 + \lambda_{kn}\lambda\right)^{-l}\right\}_{1\le k\le n} \quad \text{and} \quad \left\{\xi_{krn}\varepsilon_{kn}\left(1 + \lambda_{kn}\lambda\right)^{-l}\right\}_{1\le k\le n}$$

satisfy the assumption of Lemma 6 *with* $u_k = c^*(1 + \lambda_{kn}\lambda)^{-2l}$, *for some constant* c^*.

PROOF. Recall that $S_\lambda = \Gamma^T D_\lambda \Gamma$. Set $D_{\mu\nu} = (d_{ik})_{n\times n}$, where $d_{ik} = 1$, if $\mu \le i = k \le \nu$, and $d_{ik} = 0$, otherwise. In other words, $D_{\mu\nu}$ is an $n \times n$ diagonal matrix with the diagonal entry equal to 1 from the μ-th row to the ν-th row and zero otherwise. Then

$$\begin{aligned} &\sum_{k=\mu}^{\nu} \frac{\xi_{krn}\varepsilon_{kn}}{(1 + \lambda_{kn}\lambda)^l} = \mathbf{z}_r^T(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l \mathbf{e}, \\ &\sum_{k=\mu}^{\nu} \frac{\xi_{krn}\xi_{ksn} - \sigma_{rs}}{(1 + \lambda_{kn}\lambda)^l} = \mathbf{z}_r^T(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l \mathbf{z}_s - \sigma_{rs}\ \mathrm{tr}(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l. \end{aligned} \tag{19}$$

By (A1), (A4) and a conditioning argument, we have

$$\begin{aligned} E\left[\sum_{k=\mu}^{\nu} \frac{\xi_{krn}\varepsilon_{kn}}{(1 + \lambda_{kn}\lambda)^l}\right]^2 &= \sigma^2 E\mathbf{z}_r^T(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^{2l}\mathbf{z}_r, \\ &= \sigma^2\sigma_{rr}\ \mathrm{tr}(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^{2l} \\ &= \sigma^2\sigma_{rr}\left[\sum_{k=\mu}^{\nu}(1 + \lambda_{kn}\lambda)^{-2l}\right]. \end{aligned}$$

Letting $u_k = \sigma^2\sigma_{rr}(1 + \lambda_{kn}\lambda)^{-2l}$, we have shown that the assumption of Lemma 6 holds for $\{\xi_{krn}\varepsilon_{kn}(1 + \lambda_{kn}\lambda)^{-l}\}_{1\le k\le n}$.

Next, by (19) we have

$$E\left[\sum_{k=\mu}^{\nu} \frac{\xi_{krn}\xi_{ksn} - \sigma_{rs}}{(1 + \lambda_{kn}\lambda)^l}\right]^2 = \mathrm{Var}\left(\mathbf{z}_r^T(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l\mathbf{z}_s\right),$$

since $E(\mathbf{z}_r^T(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l\mathbf{z}_s) = \sigma_{rs}\ \mathrm{tr}(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l$. We first show that, for any symmetric matrix $A = (a_{ij})_{n\times n}$,

$$\mathrm{Var}\left(\mathbf{z}_r^T A\mathbf{z}_s\right) \le c^0\ \mathrm{tr}\, A^2 \tag{20}$$

for $1 \le r \le s \le d$, where c^0 is a constant depending on $Ez_r^2 z_s^2$ and Σ only. For notational simplicity, we only demonstrate the case of $r = 1$ and $s = 2$. First, we note that $E\mathbf{z}_1^T A\mathbf{z}_2 = \sigma_{12}\ \mathrm{tr}\, A$ and

$$(\mathbf{z}_1^T A\mathbf{z}_2)^2 = \sum_i\sum_j\sum_k\sum_l a_{ij}a_{kl}z_{i1n}z_{j2n}z_{k1n}z_{l2n}.$$

Since $\{(z_{i1n}, z_{i2n})\}_{1\le i\le n}$ are mutually independent with mean (0, 0), we have

$$Ez_{i1n}z_{j2n}z_{k1n}z_{l2n} = \begin{cases} Ez_1^2z_2^2, & i=j=k=l,\\ \sigma_{12}^2, & i=j,\, k=l,\, i\ne k,\\ \sigma_{11}\sigma_{22}, & i=k,\, j=l,\, i\ne j,\\ 0, & \text{otherwise.}\end{cases}$$

Hence

$$\operatorname{Var}(\mathbf{z}_1^T A\mathbf{z}_2) = (Ez_1^2z_2^2 - \sigma_{12}^2)\sum_i a_{ii}^2 + \sigma_{11}\sigma_{22}\sum_{i\ne j} a_{ij}^2 \le c\sum_{i,j} a_{ij}^2,$$

where $c = \max(Ez_1^2z_2^2 - \sigma_{12}^2, \sigma_{11}\sigma_{22})$. Since A is symmetric and $\sum_{i,j}a_{ij}^2 = \operatorname{tr} A^2$, (20) holds.

Let $A = (\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^l$. By (19) and (20), we have

$$E\left[\sum_{k=\mu}^{\nu} \frac{\xi_{krn}\xi_{ksn} - \sigma_{rs}}{(1+\lambda_{kn}\lambda)^l}\right]^2 = \operatorname{Var}(\mathbf{z}_r^T A\mathbf{z}_s) \le c^0 \operatorname{tr}(\Gamma^T D_{\mu\nu} D_\lambda D_{\mu\nu}\Gamma)^{2l}$$

$$= c^0\left[\sum_{k=\nu}^{h}(1+\lambda_{kn}\lambda)^{-2l}\right].$$

Thus $\{(\xi_{krn}\xi_{ksn} - \sigma_{rs})(1+\lambda_{kn}\lambda)^{-l}\}_{1\le k\le n}$ satisfies the assumption of Lemma 6 by identifying $u_k = c^0(1+\lambda_{kn}\lambda)^{-2l}$. □

PROOF OF PART (a). First, we show the case of $l = 1$, that is, to show that

$$\mathbf{z}_r^T S_\lambda \mathbf{z}_s = \sigma_{rs} \operatorname{tr} S_\lambda + o_p(\lambda^{-1/4ma_0}) \tag{21}$$

holds uniformly for all $\lambda \in [\lambda_1, \lambda_2]$ and its proof argument will be used throughout the proof of Lemma 3. Since $\delta_2 < \delta_1$, there exists $a > 1$ such that $a\delta_2 < \delta_1$. Define the index set $\Lambda = \{\delta\colon \delta = a^i\delta_2$, for some positive integer i and $\delta < \delta_1\}$. Then Λ is a finite partition of $[\delta_1, \delta_2]$. Correspondingly, $\{n^{-\delta}, \delta \in \Lambda\}$ is a finite partition of $[\lambda_1, \lambda_2]$. For any $\tau = n^{-a\delta}$ with $a\delta \in \Lambda$, $E\mathbf{z}_r^T S_\tau \mathbf{z}_s = \sigma_{rs} \operatorname{tr} S_\tau$ and $\operatorname{Var}(\mathbf{z}_r^T S_\tau \mathbf{z}_s) \le c^0 \operatorname{tr} S_\tau^2 = O(\tau^{-1/2m})$ by (20) and Lemma 2(c). Thus by the Chebyshev inequality, we have

$$\mathbf{z}_r^T S_\tau \mathbf{z}_s - \sigma_{rs} \operatorname{tr} S_\tau = O_p(\tau^{-1/4m}). \tag{22}$$

Write

$$\begin{aligned}(\mathbf{z}_r^T S_\lambda \mathbf{z}_s - \sigma_{rs} \operatorname{tr} S_\lambda) &- (\mathbf{z}_r^T S_\tau \mathbf{z}_s - \sigma_{rs} \operatorname{tr} S_\tau)\\ &= \sum_{k=1}^{n}\left(\frac{1}{1+\lambda_{kn}\lambda} - \frac{1}{1+\lambda_{kn}\tau}\right)(\xi_{krn}\xi_{ksn} - \sigma_{rs})\\ &= \frac{\tau-\lambda}{\lambda}\sum_{k=1}^{n}\left(1 - \frac{1}{1+\lambda_{kn}\lambda}\right)\frac{\xi_{krn}\xi_{ksn} - \sigma_{rs}}{1+\lambda_{kn}\tau}.\end{aligned} \tag{23}$$

Note that $(1+\lambda_{kn}\lambda)^{-1}$ are nonincreasing in k and bounded above by 1, and that $\{(\xi_{krn}\xi_{ksn}-\sigma_{rs})(1+\lambda_{kn}\tau)^{-1}\}_{1\le k\le n}$ satisfy the assumption of Lemma 6 with $u_k = c^*(1+\lambda_{kn}\tau)^{-2}$. Then, for any $c > 0$ and $\delta \in \Lambda$, we have

$$P\left\{\sup_{n^{-a\delta}<\lambda\le n^{-\delta}}\left|\sum_{k=1}^{n}\frac{1}{1+\lambda_{kn}\lambda}\frac{\xi_{krn}\xi_{ksn}-\sigma_{rs}}{1+\lambda_{kn}\tau}\right|\ge c\right\}$$
$$\le P\left\{\sup_{0\le 1/(1+\lambda_{nn}\lambda)\le\cdots\le 1/(1+\lambda_{1n}\lambda)\le 1}\left|\sum_{k=1}^{n}\frac{1}{1+\lambda_{kn}\lambda}\frac{\xi_{krn}\xi_{ksn}-\sigma_{rs}}{1+\lambda_{kn}\tau}\right|\ge c\right\}$$
$$\le c^{-2}(\log_2 4n)^2\sum_{k=1}^{n}u_k,$$

by applying Lemma 6 to (23). Since

$$\sum_k u_k = c^*\left(1+\lambda_{kn}\tau\right)^{-2} = c^*c_2\tau^{-1/2m}\left(1+o(1)\right)$$

by Lemma 2(c), these arguments lead to

$$(24)\quad \left(\mathbf{z}_r^T S_\lambda \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\lambda\right) - \left(\mathbf{z}_r^T S_\tau \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\tau\right) = O_p\left(\frac{\lambda-\tau}{\lambda}\tau^{-1/4m}\log n\right)$$

uniformly for all $\lambda \in [n^{-a\delta}, n^{-\delta}]$. Then, by (22), for $n^{-a\delta} < \lambda \le n^{-\delta}$,

$$\left(\mathbf{z}_r^T S_\lambda \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\lambda\right) = O_p\left(\tau^{-1/4m}\right) + O_p\left(\tau^{-1/4m}\log n\right) = o_p\left(\lambda^{-1/4ma_0}\right),$$

where a_0 is any fixed constant satisfying $1/a_0 > a > 1$. Since the cardinality of Λ is finite, (21) holds.

Now it remains to study the case when $l \ge 2$. Note that

$$E\mathbf{z}_r^T S_\tau^l \mathbf{z}_s = \sigma_{rs}\operatorname{tr} S_\tau^l = c_l\sigma_{rs}\tau^{-1/2m}\left(1+o(1)\right) \quad\text{and}$$
$$\operatorname{Var}\left(\mathbf{z}_r^T S_\tau^l \mathbf{z}_s\right) \le c^0 \operatorname{tr} S_\tau^{2l} \approx \tau^{-1/2m}$$

by (20) and Lemma 2(c). Hence $\mathbf{z}_r^T S_\tau^l \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\tau^l = O_p(\tau^{-1/4m})$ by the Chebyshev inequality. Some algebra shows that

$$\left(\mathbf{z}_r^T S_\lambda^l \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\lambda^l\right) - \left(\mathbf{z}_r^T S_\tau^l \mathbf{z}_s - \sigma_{rs}\operatorname{tr} S_\tau^l\right)$$
$$= \sum_k\left[\frac{1}{(1+\lambda_{kn}\lambda)^l} - \frac{1}{(1+\lambda_{kn}\tau)^l}\right]\left(\xi_{krn}\xi_{ksn}-\sigma_{rs}\right)$$
$$= \frac{\tau-\lambda}{\lambda}\sum_{k=1}^{n}\sum_{i=0}^{l-1}\left[\frac{1}{(1+\lambda_{kn}\lambda)^i} - \frac{1}{(1+\lambda_{kn}\lambda)^{i+1}}\right]\frac{\xi_{krn}\xi_{ksn}-\sigma_{rs}}{(1+\lambda_{kn}\tau)^{l-i}}.$$

Note that $(1+\lambda_{kn}\lambda)^{-i}$ are nonincreasing in k and bounded above by 1. Hence, (a) holds by applying Lemma 6 to each term on the right-hand side of the above expression and by the argument used in showing (21). □

PROOF OF PART (b). (b) follows from (a) by identifying $\mathbf{z}_r$ and $\mathbf{z}_s$ in (a) with $\mathbf{e}$ in (b). □

PROOF OF PART (c). For any finite positive integer l, observe that $En^{-1/2}\mathbf{f}^T(I - S_\tau)^l\mathbf{e} = 0$ and

$$\operatorname{Var}[n^{-1/2}\mathbf{f}^T(I - S_\tau)^l\mathbf{e}] = n^{-1}\sigma^2\mathbf{f}^T(I - S_\tau)^{2l}\mathbf{f} \le n^{-1}\sigma^2\mathbf{f}^T(I - S_\tau)^2\mathbf{f} = O(\tau^2)$$

since the eigenvalues of S_τ are between 0 and 1. Hence, for any given $\tau \in [\lambda_1, \lambda_2]$,

$$n^{-1/2}\mathbf{f}^T(I - S_\tau)^l\mathbf{e} = O_p(\tau). \tag{25}$$

For $\tau = n^{-\delta}$ with $\delta \in \Lambda$, write

$$\begin{aligned} &\mathbf{f}^T[(I - S_\lambda)^l - (I - S_\tau)^l]\mathbf{e} \\ &\quad = \frac{\lambda - \tau}{\tau}\sum_{k=1}^{n}\left[\sum_{i=1}^{l}\left(1 - \frac{1}{1 + \lambda_{kn}\lambda}\right)^{l-i}\left(1 - \frac{1}{1 + \lambda_{kn}\tau}\right)^{i}\right] \\ &\qquad \times \frac{1}{1 + \lambda_{kn}\lambda}\frac{\lambda_{kn}\tau}{1 + \lambda_{kn}\tau}f_{kn}\varepsilon_{kn}, \end{aligned} \tag{26}$$

where $f_{kn} = n^{-1/2}\sum_{i=1}^{n} f(t_{in})\phi_{kn}(t_{in})$. Note that $\{[\lambda_{kn}\tau/(1 + \lambda_{kn}\tau)]f_{kn}\varepsilon_{kn}\}$ does not depend on λ, $E(f_{kn}\varepsilon_{kn})(f_{ln}\varepsilon_{ln}) = 0$, for $k \ne l$, that $\{(1+\lambda_{kn}\lambda)^{-i}(1+\lambda_{kn}\tau)^{-j}\}$, for $1 \le i, j \le l$, are nonincreasing in k, and that

$$\begin{aligned} \operatorname{Var}\left(\frac{\lambda_{kn}\tau}{1 + \lambda_{kn}\tau}f_{kn}\varepsilon_{kn}\right) &= n^{-1}\sigma^2\sum_{k=1}^{n}\left(\frac{\lambda_{kn}\tau}{1 + \lambda_{kn}\tau}\right)^2 f_{kn}^2 \\ &= n^{-1}\sigma^2\mathbf{f}^T(I - S_\tau)^2\mathbf{f} = O(\tau^2). \end{aligned}$$

It follows from Remark 1 following Lemma 6 that we can apply Lemma 6 to each term on the right-hand side of (26). Thus we conclude that

$$n^{-1/2}[\mathbf{f}^T(I - S_\lambda)^l\mathbf{e} - \mathbf{f}^T(I - S_\tau)^l\mathbf{e}] = O_p((\tau - \lambda)\log n) \tag{27}$$

holds uniformly for all $\lambda \in [n^{-a\delta}, n^{-\delta}]$. By (25) and (27), for any $a_1 > a$,

$$n^{-1/2}\mathbf{f}^T(I - S_\lambda)^l\mathbf{e} = O_p(\lambda^{1/a_1})$$

holds uniformly for all $\lambda \in [\lambda_1, \lambda_2]$ and finite positive integer l. Hence, (c) holds. □

PROOF OF PART (d). (d) can be shown similarly by identifying $\mathbf{e}$ in (c) with $\mathbf{z}_r$ in (d). □

PROOF OF PART (e). Note that

$$\left|n^{-1}\mathbf{f}_1^T(I - S_\lambda)^l\mathbf{f}_2\right| \le (n^{-1}\mathbf{f}_1^T(I - S_\lambda)^l\mathbf{f}_1)^{1/2}(n^{-1}\mathbf{f}_2^T(I - S_\lambda)^l\mathbf{f}_2)^{1/2},$$

by the Cauchy–Schwarz inequality. Since that the eigenvalues of S_λ are between 0 and 1, (e) holds by Lemma 2(a). □

PROOF OF PART (f). Write $X^T(I-S_\lambda)^2 Z = Z^TZ+H^T(I-S_\lambda)^2Z+Z^T(S_\lambda^2-2S_\lambda)Z$. Then, by (A1) (in Section 2) and the law of large numbers, $n^{-1}Z^TZ = \Sigma+o_p(1)$. Hence, (f) follows from (a) and (d). □

PROOF OF PART (g). Note that

$$n^{-1}\mathbf{x}_r^T(I-S_\lambda)^l\mathbf{x}_s = n^{-1}\mathbf{z}_r^T(I-S_\lambda)^l\mathbf{z}_s + n^{-1}\mathbf{h}_r^T(I-S_\lambda)^l\mathbf{z}_s + n^{-1}\mathbf{h}_s^T(I-S_\lambda)^l\mathbf{z}_r + n^{-1}\mathbf{h}_r^T(I-S_\lambda)^l\mathbf{h}_s.$$

Recall that $n^{-1}Z^TZ = \Sigma + o_p(1)$. Hence, it follows easily from (a), (d) and (e) that (g) holds. □

PROOF OF PART (h). Write

$$\mathbf{x}_r^T(I-S_\lambda)^2S_\lambda\mathbf{x}_s = \mathbf{z}_r^T(S_\lambda - 2S_\lambda^2 + S_\lambda^3)\mathbf{z}_s + \mathbf{h}_r^T(I-S_\lambda)^2S_\lambda\mathbf{h}_s + \mathbf{h}_r^T[(I-S_\lambda)^2-(I-S_\lambda)^3]\mathbf{z}_s + \mathbf{h}_s^T[(I-S_\lambda)^2-(I-S_\lambda)^3]\mathbf{z}_r.$$

Since $|\mathbf{h}_r^T(I-S_\lambda)^2S_\lambda\mathbf{h}_s| \le nB_{2rp}B_{2sp} = O(n\lambda^2)$, (h) holds by (a) and (d). □

PROOF OF PART (i). Observe that $E\mathbf{z}_r^TS_\tau^l\mathbf{e} = 0$ and

$$\operatorname{Var}\left(\mathbf{z}_r^TS_\tau^l\mathbf{e}\right) = \sigma^2 \operatorname{Var}\left(\mathbf{z}_r^TS_\tau^{2l}\mathbf{z}_r\right) \approx \tau^{-1/2m},$$

by (20) and Lemma 2(c). Hence $\mathbf{z}_r^TS_\tau^l\mathbf{e} = O_p(\tau^{-1/4m})$, for any given sequence $\tau = n^{-a\delta}$ with $a\delta \in \Lambda$. Write

$$\text{(28)}\quad \mathbf{z}_r^T(S_\lambda^l - S_\tau^l)\mathbf{e} = \frac{\tau-\lambda}{\lambda}\sum_{k=1}^{n}\sum_{\nu=0}^{l-1}\left[\frac{1}{(1+\lambda_{kn}\lambda)^\nu} - \frac{1}{(1+\lambda_{kn}\lambda)^{\nu+1}}\right]\frac{\xi_{krn}\varepsilon_{kn}}{(1+\lambda_{kn}\tau)^{l-\nu}}.$$

Note that $(1+\lambda_{kn}\lambda)^{-i}$ are nonincreasing in k and bounded above by 1. By Lemma 7, $\{\xi_{krn}\varepsilon_{kn}(1+\lambda_{kn}\tau)^{-j}\}$ satisfies the assumption of Lemma 6. By applying Lemma 6 to each term on the right-hand side of (28), we conclude that

$$\text{(29)}\qquad \mathbf{z}_r^TS_\lambda^l\mathbf{e} = O_p\left((1-\lambda^{-1}\tau)\tau^{-1/4m}\log n\right) = O_p\left(\lambda^{-1/4ma_0}\right)$$

holds uniformly over $\lambda \in [n^{-a\delta}, n^{-\delta}]$. Hence, (i) holds. □

PROOF OF PART (j). Write

$$n^{-1/2}\mathbf{z}_r^T(I-S_\lambda)^2\mathbf{e} = n^{-1/2}\mathbf{z}_r^T\mathbf{e} + n^{-1/2}\mathbf{z}_r^T(-2S_\lambda + S_\lambda^2)\mathbf{e}.$$

Then by the central limit theorem and (i), (j) holds. □

PROOF OF PARTS (k) AND (l). It follows from (a) and (c) that

$$\begin{aligned}
&\mathbf{x}_r^T(I-S_\lambda)^3\mathbf{x}_s - \mathbf{x}_r^T(I-S_\lambda)^2\mathbf{z}_s \\
&\quad = -\mathbf{z}_r^T(I-S_\lambda)^2 S_\lambda \mathbf{z}_s - \mathbf{h}_r^T(I-S_\lambda)^2\mathbf{z}_s \\
&\qquad + \mathbf{h}_r^T(I-S_\lambda)^3\mathbf{z}_s + \mathbf{h}_s^T(I-S_\lambda)^3\mathbf{z}_r + \mathbf{h}_r^T(I-S_\lambda)^3\mathbf{h}_s \\
&\quad = -(c_1 - 2c_2 + c_3)\sigma_{rs}\lambda^{-1/2m} + o_p(\lambda^{-1/4ma_0}) + O_p(n^{1/2}\lambda^{1/a_1}) \\
&\qquad + \mathbf{h}_r^T(I-S_\lambda)^3\mathbf{h}_s,
\end{aligned}$$

$$\begin{aligned}
&\mathbf{x}_r^T(I-S_\lambda)^3\mathbf{x}_s - \mathbf{x}_r^T(I-S_\lambda)^4\mathbf{x}_s \\
&\quad = \mathbf{z}_r^T(I-S_\lambda)^3 S_\lambda \mathbf{z}_s + \mathbf{h}_s^T[(I-S_\lambda)^3 - (I-S_\lambda)^4]\mathbf{z}_r \\
&\qquad + h_r^T[(I-S_\lambda)^3 - (I-S_\lambda)^4]\mathbf{z}_s + \mathbf{h}_r^T(I-S_\lambda)^3 S_\lambda \mathbf{h}_s \\
&\quad = (c_1 - 3c_2 + 3c_3 - c_4)\sigma_{rs}\lambda^{-1/2m} + o_p(\lambda^{-1/4ma_0}) + O_p(n^{1/2}\lambda^{1/a_1}) \\
&\qquad + \mathbf{h}_r^T(I-S_\lambda)^3 S_\lambda \mathbf{h}_s.
\end{aligned}$$

Hence, we conclude (k) and (l). □

7. Proofs of Lemmas 4 and 5.

PROOF OF LEMMA 4. Recall $R_{0n}(\lambda) \approx \lambda^2 + n^{-1}\lambda^{-1/2m}$. From now on, we require that the three constants a, a_0 and a_1 in Lemma 3 satisfy $4m/(4m-1) > a_1 > a$ and $a_0 > \frac{1}{2}$ so that, for $\lambda \in [\lambda_1, \lambda_2]$,

$$n^{-1}\lambda^{-1/4ma_0} = o(\lambda^2 + n^{-1}\lambda^{-1/2m}) = o(R_{0n}(\lambda)) \tag{30}$$

and

$$n^{-1/2}\lambda^{1/a_1} = o(\lambda^2 + n^{-1}\lambda^{-1/2m}) = o(R_{0n}(\lambda)). \tag{31}$$

Equations (30) and (31) can be verified by simple algebra. Recall that $A_{0\lambda} = S_\lambda + B_{0\lambda}$, where $B_{0\lambda} = n^{-1}(I - S_\lambda)^2 X A_0^{-1}(\lambda) X^T(I - S_\lambda)^2$ and $A_{0\lambda} = n^{-1}X^T(I - S_\lambda)^3 X$.

[(a) and (b)] By Lemma 3(g), we have

$$\operatorname{tr} B_{0\lambda} = \operatorname{tr}\{A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^4 X]\} = O_p(1), \tag{32}$$

$$\operatorname{tr} B_{0\lambda}^2 = \operatorname{tr}\{A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^4 X]\}^2 = O_p(1), \tag{33}$$

$$\operatorname{tr} S_\lambda B_{0\lambda} = \operatorname{tr} A_0^{-1}(\lambda)\{n^{-1}X^T[(I-S_\lambda)^5 - (I-S_\lambda)^4]X\} = o_p(1). \tag{34}$$

This, together with Lemma 2(c), proves (a) and (b).

(c) It follows from Lemma 2(b) that

$$n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2\mathbf{g}_0 = \gamma\lambda^2(1+o(1)). \tag{35}$$

Also, by Lemma (a) and Lemma 3(d) and (e), we have

$$\begin{aligned} n^{-1}\mathbf{g}_0^T B_{0\lambda}^2 \mathbf{g}_0 &= [n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2 H + n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2 Z] \\ &\quad \times \{A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^4 X]A_0^{-1}(\lambda)\} \\ &\quad \times [n^{-1}H^T(I-S_\lambda)^2\mathbf{g}_0 + n^{-1}Z^T(I-S_\lambda)^2\mathbf{g}_0] \\ &= [O(\lambda^2)+o_p(n^{-1/2})][\Sigma^{-1}+o_p(1)][O(\lambda^2)+o_p(n^{-1/2})] = o(\lambda^2). \end{aligned}$$

This, together with the Cauchy–Schwarz inequality and (35), leads to (c).

(d) Write

$$\begin{aligned} n^{-1}Z^T(I-A_{0\lambda})^2 Z = n^{-1}\Big\{ & Z^T(S_\lambda^2 - 2S_\lambda)Z + Z^T[(I-B_{0\lambda}) - (B_{0\lambda}-B_{0\lambda}^2)]Z \\ & + Z^T S_\lambda B_{0\lambda} Z + Z^T B_{0\lambda} S_\lambda Z\Big\}. \end{aligned}$$

By Lemma 3(a), we have the first term

$$n^{-1}(Z\beta)^T(S_\lambda^2 - 2S_\lambda)(Z\beta) = (c_2 - 2c_1)\beta^T\Sigma\beta n^{-1}\lambda^{-1/2m}(1+o_p(1)). \tag{36}$$

It follows from Lemma 3(a), (d) and (f), Lemma 1(a), (30) and (31) that the third term

$$\begin{aligned} n^{-1}\beta^T Z^T S_\lambda B_{0\lambda} Z\beta &= \beta^T[n^{-1}Z^T S_\lambda(I-S_\lambda)^2 X]A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2 Z]\beta \\ &= (c_1 - 2c_2 + c_3)\beta^T\Sigma\beta n^{-1}\lambda^{-1/2m}(1+o_p(1)). \end{aligned} \tag{37}$$

The fourth term has the same rate.

Observe that

$$\begin{aligned} n^{-1}Z^T(I-B_{0\lambda})Z &= [n^{-1}Z^T Z A_0^{-1}(\lambda)][A_0(\lambda) - n^{-1}X^T(I-S_\lambda)^2 Z] \\ &\quad + \Big\{[n^{-1}Z^T(2S_\lambda - S_\lambda^2)Z] - [n^{-1}Z^T(I-S_\lambda)^2 H]\Big\} \\ &\quad \times A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2 Z], \\ n^{-1}Z^T(B_{0\lambda}-B_{0\lambda}^2)Z &= [n^{-1}Z^T(I-S_\lambda)^2 X]A_0^{-1}(\lambda)[A_0(\lambda) - n^{-1}X^T(I-S_\lambda)^4 X] \\ &\quad \times A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2 Z]. \end{aligned}$$

Then by Lemma 3(d) and (f), we have

$$n^{-1}H^T(I-S_\lambda)^2 Z = o_p(\lambda^2 + n^{-1}\lambda^{-1/2m}) \quad \text{and} \quad n^{-1}X^T(I-S_\lambda)^2 Z = \Sigma + o_p(1).$$

Also, by (A1) (in Section 2) and the law of large numbers, we see that $n^{-1}Z^TZ = \Sigma + o_p(1)$. Hence, it follows from Lemma 1(a) that

$$n^{-1}Z^T Z A_0^{-1}(\lambda) = I + o_p(1). \tag{38}$$

Hence, by Lemma 1(a), (36) and Lemma 3(k) and (l), we conclude that

$$\begin{aligned} &n^{-1}(Z\beta)^T\big[(I-B_{0\lambda})-(B_{0\lambda}-B_{0\lambda}^2)\big](Z\beta)\\ &\quad= n^{-1}\big[(H\beta)^T(I-S_\lambda)^4(H\beta)+O_p(\lambda^{-1/4ma_0})+O_p(n^{-1/2}\lambda^{1/a_1})\\ &\qquad+(4c_2-4c_3+c_4)\beta^T\Sigma\beta\lambda^{-1/2m}\big](1+o_p(1)). \end{aligned}$$

This, together with (36) and (37), proves (d).

(e) Note that $n^{-1}\mathbf{g}_o^T(I-S_\lambda)^l\mathbf{h}_r = O(\lambda^2)$, for $l \geq 2$, and

$$n^{-1}X^T(I-S_\lambda)^2\mathbf{e} = O_p(n^{-1/2}\lambda^{1/a_1})+O_p(n^{-1/2}) = O_p(n^{-1/2}) \tag{39}$$

by Lemma 3(c) and (j). Then by (31), (39), Lemma 1(a) and Lemma 3 (d), (e) and (g),

$$\begin{aligned} &n^{-1}\mathbf{g}_0^TB_{0\lambda}^2\mathbf{e}\\ &\quad= [n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2X]A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^4X]A_0^{-1}(\lambda)\\ &\qquad\times[n^{-1}X^T(I-S_\lambda)^2\mathbf{e}]\\ &\quad\approx(n^{-1/2}\lambda^{1/a_1}+\lambda^2)\big[n^{-1/2}\lambda^{1/a_1}+O_p(n^{-1/2})\big]\\ &\quad= o_p(R_{0n}(\lambda)). \end{aligned}$$

Similarly, we have

$$\begin{aligned} n^{-1}\mathbf{g}_0^T(I-S_\lambda)B_{0\lambda}\mathbf{e} &= [n^{-1}\mathbf{g}_0^T(I-S_\lambda)^3X]A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2\mathbf{e}]\\ &= o_p(R_{0n}(\lambda)). \end{aligned}$$

By Lemma 3(c) and (31), $n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2\mathbf{e} = o_p(R_{0n}(\lambda))$. Putting these results together, we have (e).

(f) Write $(I-A_{0\lambda})^2 = S_\lambda^2-(I-B_{0\lambda})S_\lambda-S_\lambda(I-B_{0\lambda})+(I-B_{0\lambda})-(B_{0\lambda}-B_{0\lambda}^2)$. By the central limit theorem, $n^{-1}Z^T\mathbf{e} = O_p(n^{-1/2})$. Then it follows from Lemma 1(a) and Lemma 3 (c), (k), (f) and (i) that

$$\begin{aligned} n^{-1}Z^T(I-B_{0\lambda})\mathbf{e} &= [A_0(\lambda)-n^{-1}Z^T(I-S_\lambda)^2X]A_0^{-1}(\lambda)(n^{-1}Z^T\mathbf{e})\\ &\quad-[n^{-1}Z^T(I-S_\lambda)^2X]A_0^{-1}(\lambda)\\ &\quad\times[n^{-1}H^T(I-S_\lambda)^2\mathbf{e}-n^{-1}Z^T(2S_\lambda-S_\lambda^2)\mathbf{e}]\\ &= o(R_{0n}(\lambda)). \end{aligned}$$

By Lemma 3(l), we have $A_0(\lambda)-n^{-1}X^T(I-S_\lambda)^4X = O_p(R_{0n}(\lambda))$. It then follows by (39) and Lemma 3(f) that

$$\begin{aligned} n^{-1}Z^T(B_{0\lambda}-B_{0\lambda}^2)\mathbf{e} &= [n^{-1}Z^T(I-S_\lambda)^2X]A_0^{-1}(\lambda)\\ &\quad\times[A_0(\lambda)-n^{-1}X^T(I-S_\lambda)^4X]A_0^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2\mathbf{e}]\\ &= o(R_{0n}(\lambda)). \end{aligned}$$

Next,

$$
\begin{aligned}
n^{-1}Z^T(I - B_{0\lambda})S_\lambda \mathbf{e} &= n^{-1}Z^T S_\lambda \mathbf{e} - [n^{-1}Z^T(I - S_\lambda)^2 X]A_0^{-1}(\lambda) \\
&\qquad \times \{n^{-1}X^T[(I - S_\lambda)^2 - (I - S_\lambda)^3]\mathbf{e}\} \\
&= o(R_{0n}(\lambda)),
\end{aligned} \tag{40}
$$

since $n^{-1}Z^T S_\lambda \mathbf{e} = o_p(R_{0n}(\lambda))$, by Lemma 3(i) and the fact that

$$
\begin{aligned}
n^{-1}X^T[(I - S_\lambda)^2 - (I - S_\lambda)^3]\mathbf{e} &= n^{-1}H^T[(I - S_\lambda)^2 - (I - S_\lambda)^3]\mathbf{e} \\
&\qquad + n^{-1}Z^T(S_\lambda - 2S_\lambda^2 + S_\lambda^3)\mathbf{e} \\
&= o(R_{0n}(\lambda)),
\end{aligned} \tag{41}
$$

by Lemma 3(c) and (i). Similarly, $n^{-1}Z^T S_\lambda(I - B_{0\lambda})\mathbf{e} = o(R_{0n}(\lambda))$. Finally,

$$n^{-1}Z^T S_\lambda^l \mathbf{e} = o_p(R_{0n}(\lambda))$$

by Lemma 3(i). Combining all the terms, we have (f).

(g) Write $A_{0\lambda}^2 - 2A_{0\lambda} = S_\lambda^2 - 2S_\lambda + (B_{0\lambda}S_\lambda + S_\lambda B_{0\lambda}) + B_{0\lambda}^2 - 2B_{0\lambda}$. First, we note that

$$
n^{-1}(\mathbf{e}^T S_\lambda^l \mathbf{e} - \sigma^2 \operatorname{tr} S_\lambda^l) = o_p(n^{-1}\lambda^{-1/4ma_0}) = o_p(R_{0n}(\lambda)), \tag{42}
$$

by the proofs of Lemma 3(a) and (b) and (30). Next, by Lemma 1(a), (39) and (41) and Lemma 3(g),

$$
\begin{aligned}
n^{-1}\mathbf{e}^T B_{0\lambda}\mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)^2 X]A_0^{-1}(\lambda)[n^{-1}X^T(I - S_\lambda)^2\mathbf{e}] \\
&\approx [n^{-1/2}\lambda^{1/a_1} + O_p(n^{-1/2})]^2 = o_p(R_{0n}(\lambda)),
\end{aligned} \tag{43}
$$

$$
\begin{aligned}
n^{-1}\mathbf{e}^T B_{0\lambda} S_\lambda \mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)^2 X]A_0^{-1}(\lambda)[n^{-1}X^T(I - S_\lambda)^2 S_\lambda \mathbf{e} \\
&= o_p(R_{0n}(\lambda)),
\end{aligned} \tag{44}
$$

$$
\begin{aligned}
n^{-1}\mathbf{e}^T B_{0\lambda}^2 \mathbf{e} &= [n^{-1}\mathbf{e}^T(I - S_\lambda)^2 X]A_0^{-1}(\lambda)[n^{-1}X^T(I - S_\lambda)^4 X] \\
&\qquad \times A_0^{-1}(\lambda)[n^{-1}X^T(I - S_\lambda)^2\mathbf{e}] = o_p(R_{0n}(\lambda)).
\end{aligned} \tag{45}
$$

Part (g) holds by (32)–(34), (43)–(45) and (42). □

PROOF OF LEMMA 5. Recall that $A_{1\lambda} = S_\lambda + B_{1\lambda}$, where

$$B_{1\lambda} = n^{-1}(I - S_\lambda)XA_1^{-1}(\lambda)X^T(I - S_\lambda)^2 \quad \text{and} \quad A_1(\lambda) = n^{-1}X^T(I - S_\lambda)^2 X.$$

(a) By Lemma 3(d), (e) and (g),

$$
\begin{aligned}
n^{-1}\mathbf{g}_0^T B_{1\lambda}^T B_{1\lambda}\mathbf{g}_0 &= [n^{-1}\mathbf{g}_0^T(I - S_\lambda)^2 X]A_1^{-1}(\lambda)[n^{-1}X^T(I - S_\lambda)^2 X]A_1^{-1}(\lambda) \\
&\qquad \times [n^{-1}X^T(I - S_\lambda)^2\mathbf{g}_0] \\
&= O_p((\lambda^2 + n^{-1/2}\lambda^{1/a_1})^2).
\end{aligned}
$$

It can be easily verified that $O_p(n^{-1}\lambda^{2/a_1}) = o_p(\lambda^2)$. This, together with the Cauchy–Schwarz inequality and (35), proves (a).

(b) Write

$$\beta^T Z^T (I - A_{1\lambda})^T (I - A_{1\lambda}) Z\beta = \beta^T Z^T (S_\lambda^2 - 2S_\lambda) Z\beta + 2\beta^T Z^T S_\lambda^T B_{1\lambda} Z\beta + \beta^T Z^T \left[(I - B_{1\lambda}) - B_{1\lambda}^T (I - B_{1\lambda})\right] Z\beta.$$

By (36), Lemma 3(a) and (f) and Lemma 1 (b), we have the second term,

$$\begin{aligned} n^{-1}\beta^T Z^T S_\lambda^T B_{1\lambda} Z\beta &= \beta^T [n^{-1} Z^T S_\lambda (I - S_\lambda) X] A_1^{-1}(\lambda) [n^{-1} X^T (I - S_\lambda)^2 Z]\beta \\ &= (c_1 - c_2)\beta^T \Sigma \beta n^{-1} \lambda^{-1/2m} + o_p(n^{-1}\lambda^{-1/4ma_0}) \\ &\quad + O_p(n^{-1/2}\lambda^{1/a_1}). \end{aligned} \tag{46}$$

Observe that

$$\begin{aligned} n^{-1} Z^T (I - B_{1\lambda}) Z &= [n^{-1} Z^T Z A_1^{-1}(\lambda)][A_1(\lambda) - n^{-1} X^T (I - S_\lambda)^2 Z] \\ &\quad + (n^{-1} Z^T S_\lambda Z) A_1^{-1}(\lambda) [n^{-1} X^T (I - S_\lambda)^2 Z] \\ &\quad - [n^{-1} Z^T (I - S_\lambda) H] A_1^{-1}(\lambda) [n^{-1} X^T (I - S_\lambda)^2 Z], \end{aligned}$$

$$\begin{aligned} n^{-1} Z^T B_{1\lambda}^T (I - B_{1\lambda}) Z &= [n^{-1} Z^T (I - S_\lambda) X] A_1^{-1}(\lambda) \\ &\quad \times \{[n^{-1} X^T (I - S_\lambda) Z - n^{-1} X^T (I - S_\lambda)^2 X] \\ &\quad + [n^{-1} X^T (I - S_\lambda)^2 X] A_1^{-1}(\lambda) \\ &\quad \times [A_1(\lambda) - n^{-1} X^T (I - S_\lambda)^2 Z]\}. \end{aligned}$$

It follows from Lemma 3(a) and (d) that

$$\begin{aligned} \mathbf{x}_r^T (I - S_\lambda)^2 \mathbf{x}_s - \mathbf{x}_r^T (I - S_\lambda)^2 \mathbf{z}_s &= \mathbf{x}_r^T (I - S_\lambda)^2 \mathbf{h}_s \\ &= \mathbf{h}_r^T (I - S_\lambda)^2 \mathbf{h}_s + O_p(n^{1/2}\lambda^{1/a_1}) \\ &= \mathbf{h}_r^T (I - S_\lambda)^2 \mathbf{h}_s + o_p(n\lambda^2 + \lambda^{-1/2m}), \end{aligned}$$

$$\begin{aligned} \mathbf{x}_r^T (I - S_\lambda) \mathbf{z}_s - \mathbf{x}_r^T (I - S_\lambda)^2 \mathbf{x}_s &= (c_1 - c_2)\sigma_{rs}\lambda^{-1/2m} - \mathbf{h}_r^T (I - S_\lambda)^2 \mathbf{h}_s \\ &\quad + o_p(\lambda^{-1/4ma_0}) + O_p(n^{1/2}\lambda^{1/a_1}) \\ &= (c_1 - c_2)\sigma_{rs}\lambda^{-1/2m} - \mathbf{h}_r^T (I - S_\lambda)^2 \mathbf{h}_s \\ &\quad + o_p(n\lambda^2 + \lambda^{-1/2m}). \end{aligned}$$

Note that $|n^{-1}\mathbf{h}_r^T (I - S_\lambda)^2 \mathbf{h}_s| = O(\lambda^2)$ by Lemma 3(e). Hence

$$\begin{aligned} &\beta^T [A_1(\lambda) - n^{-1} X^T (I - S_\lambda)^2 Z]\beta \\ &\quad = n^{-1}[o_p(n\lambda^2 + \lambda^{-1/2m}) + (H\beta)^T (I - S_\lambda)^2 H\beta](1 + o_p(1)), \end{aligned} \tag{47}$$

$$\begin{aligned} &\beta^T [n^{-1} X^T (I - S_\lambda) Z - n^{-1} X^T (I - S_\lambda)^2 X]\beta \\ &\quad = n^{-1}[(c_1 - c_2)\beta^T \Sigma \beta \lambda^{-1/2m} - (H\beta)^T (I - S_\lambda)^2 H\beta](1 + o_p(1)). \end{aligned} \tag{48}$$

By Lemma 1(b), (38), (47), (48) and Lemma 3(a), (d), (f) and (g), we conclude that

$$(49)\quad \begin{aligned} &n^{-1}(Z\beta)^T[(I-B_{1\lambda})-B_{1\lambda}^T(I-B_{1\lambda})](Z\beta)\\ &\quad = n^{-1}[(H\beta)^T(I-S_\lambda)^2(H\beta)+c_2\beta^T\Sigma\beta\lambda^{-1/2m}](1+o_p(1)). \end{aligned}$$

Part (b) holds by (36), (46) and (49).

(c) Write $(I-A_{1\lambda})^T(I-A_{1\lambda}) = (I-S_\lambda)^2 + B_{1\lambda}^T B_{1\lambda} - (I-S_\lambda)B_{1\lambda} - B_{1\lambda}^T(I-S_\lambda)$. By Lemma 1(b), (48) and Lemma 3(c)–(e), (g) and (j), we have

$$\begin{aligned} &n^{-1}\mathbf{g}_0^T B_{1\lambda}^T B_{1\lambda}\mathbf{e}\\ &\quad = [n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2X]A_1^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2X]A_1^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2\mathbf{e}]\\ &\quad \approx (n^{-1/2}\lambda^{1/a_1}+\lambda^2)[n^{-1/2}\lambda^{1/a_1}+O_p(n^{-1/2})]\\ &\quad = o_p(R_{1n}(\lambda)), \end{aligned}$$

$$\begin{aligned} &n^{-1}\mathbf{g}_0^T(I-S_\lambda)B_{1\lambda}\mathbf{e}\\ &\quad = n^{-1}\mathbf{g}_0^T B_{1\lambda}^T(I-S_\lambda)\mathbf{e}\\ &\quad = [n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2X]A_1^{-1}(\lambda)[n^{-1}X^T(I-S_\lambda)^2\mathbf{e}]\\ &\quad = o_p(R_{1n}(\lambda)). \end{aligned}$$

By Lemma 3(c), $n^{-1}\mathbf{g}_0^T(I-S_\lambda)^2\mathbf{e} = o_p(R_{1n}(\lambda))$. Hence,

$$(50)\quad n^{-1}\mathbf{g}_0^T(I-A_{1\lambda})^T(I-A_{1\lambda})\mathbf{e} = o(R_{1n}(\lambda)).$$

Write

$$\begin{aligned} (I-A_{1\lambda})^T(I-A_{1\lambda}) &= (I-B_{1\lambda})^T(I-B_{1\lambda}) + S_\lambda^2\\ &\quad -(I-B_{1\lambda})^T S_\lambda - S_\lambda(I-B_{1\lambda}),\\ Z^T(I-B_{1\lambda})^T(I-B_{1\lambda})\mathbf{e} &= Z^T(I-B_{1\lambda})\mathbf{e} - Z^T B_{1\lambda}^T(I-B_{1\lambda})\mathbf{e}. \end{aligned}$$

Recall that $n^{-1}Z^T\mathbf{e} = O_p(n^{-1/2})$. Then

$$\begin{aligned} n^{-1}Z^T(I-B_{1\lambda})\mathbf{e} &= [A_1(\lambda) - n^{-1}Z^T(I-S_\lambda)X]A_1^{-1}(n^{-1}Z^T\mathbf{e})\\ &\quad - [n^{-1}Z^T(I-S_\lambda)X]A_1^{-1}(\lambda)\\ &\quad \times [n^{-1}H^T(I-S_\lambda)^2\mathbf{e} - n^{-1}Z^T(2S_\lambda - S_\lambda^2)\mathbf{e}]\\ &= o(R_{1n}(\lambda)), \end{aligned}$$

by Lemma 1(a), (47) and Lemma 3(c), (f) and (i). Using the same argument to derive (47), we have

$$(51)\quad A_1(\lambda) - n^{-1}X^T(I-S_\lambda)^2X = O_p(R_{1n}(\lambda)).$$

By (39), Lemma 1(b), Lemma 3(f) and (g) and (51), we have

$$\begin{aligned} n^{-1}Z^T B_{1\lambda}^T (I - B_{1\lambda})\mathbf{e} &= [n^{-1}Z^T(I - S_\lambda)^2 X] A_1^{-1}(\lambda) \\ &\quad \times [A_1(\lambda) - n^{-1}X^T(I - S_\lambda)^2 X] A_1^{-1}(\lambda) [n^{-1}X^T(I - S_\lambda)^2 \mathbf{e}] \\ &= o(R_{1n}(\lambda)). \end{aligned}$$

Hence,

$$n^{-1}Z^T(I - B_{1\lambda})^T(I - B_{1\lambda})\mathbf{e} = o(R_{1n}(\lambda)). \tag{52}$$

Then by Lemma 3(c) and (i), we have

$$\begin{aligned} n^{-1}Z^T(I - B_{1\lambda})^T S_\lambda \mathbf{e} &= n^{-1}Z^T S_\lambda \mathbf{e} - [n^{-1}Z^T(I - S_\lambda)^2 X] A_1^{-1}(\lambda) \\ &\quad \times [n^{-1}X^T(I - S_\lambda) S_\lambda \mathbf{e}] \\ &= o(R_{1n}(\lambda)). \end{aligned} \tag{53}$$

Thus by (52), (53) and Lemma 3(i) we have

$$Z^T(I - A_{1\lambda})^T(I - A_{1\lambda}\mathbf{e}) = o_p(R_{1n}(\lambda)). \tag{54}$$

Part(c) holds by (50) and (54). □

Acknowledgments. We would like to thank two referees, an Associate Editor and a former Editor (Arthur Cohen) for the careful review. The referees' constructive comments considerably improved the paper and are greatly appreciated. We would also like to thank Professor Paul Speckman for his comments on our earlier work, which led to the development of the two-stage spline smoothing method.

REFERENCES

CHEN, H. and SHIAU, J. H. (1991). A two-stage spline smoothing method for partially linear models. *J. Statist. Plann. Inference* **27** 187–201.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

DEMMLER, A. and REINSCH, C. (1975). Oscillation matrices with spline smoothing. *Numer. Math.* **24** 375–382.

DENBY, L. (1986). Smooth regression function. Statistical Research Report 26, AT&T Bell Laboratories.

ENGLE, R. F., GRANGER, C. W., RICE, J. and WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.

EUBANK, R. L. and SPECKMAN, P. L. (1991). A bias reduction theorem with applications in nonparametric regression. *Scand. J. Statist.* **18** 211–222.

HECKMAN, N. (1986). Spline smoothing in partly linear models. *J. Roy. Statist. Soc. Ser. B* **48** 244–248.

LI, K. C. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.* **14** 1101–1112.

MALLOWS, C. L. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

RICE, J. (1986). Convergence rates for partial splined models. *Statist. Probab. Lett.* **4** 203–208.

SERFLING, R. J. (1970). Moment inequalities for the maximum cumulative sum. *Ann. Math. Statist.* **41** 1227–1234.

SHIAU, J., WAHBA, G. and JOHNSON, D. R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. *Journal of Atmospheric and Ocean Technology* **3** 713–725.

SPECKMAN, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. Unpublished manuscript.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

WAHBA, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14** 661–667.

WAHBA, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. In *Statistical Analysis of Time Series, Proceeding of the Japan U.S. Joint Seminar, Tokyo* 319–329. Inst. Statist. Math., Tokyo.

WAHBA, G. (1986). Partial and interaction splines for the semiparametric estimation of functions of several variables. In *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface* (T. J. Boardman, ed.) 75–80. Amer. Statist. Assoc., Alexandria, VA.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK 11794-3600

INSTITUTE OF STATISTICS
NATIONAL CHIAO-TUNG UNIVERSITY
HSINCHU, TAIWAN 30050