Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Clustered/Longitudinal Data

BY XIHONG LIN
Department of Biostatistics, University of Michigan
Ann Arbor, MI 48109, USA
xlin@umich.edu

NAISYIN WANG
Department of Statistics, Texas A&M University
College Station, TX 77843-3143, USA
nwang@stat.tamu.edu

ALAN H. WELSH
Faculty of Mathematical Studies, University of Southampton
Southampton, Hants SO17 1BJ, UK
A.H.Welsh@maths.soton.ac.uk

RAYMOND J. CARROLL
Department of Statistics, Texas A&M University
College Station, TX 77843-3143, USA
carroll@stat.tamu.edu

SUMMARY

We compare spline and kernel methods for clustered/longitudinal data. For independent data, it is well known that kernel methods and spline methods are essentially asymptotically equivalent (Silverman, 1984). However, the recent work of Welsh, et al. (2002) shows that the same is not true for clustered/longitudinal data. First, conventional kernel methods fail to account for the within-cluster correlation, while spline methods are able to account for this correlation. Second, kernel methods and spline methods were found to have different local behavior, with conventional kernels being local and splines being non-local. To resolve these differences, we show that a smoothing spline estimator is asymptotically equivalent to a recently proposed seemingly unrelated kernel estimator of Wang (2003) for any working covariance matrix. To gain insight into this asymptotic equivalence, we show that both the seemingly unrelated kernel estimator and the smoothing spline estimator using any working covariance matrix can be obtained iteratively by applying conventional kernel or spline smoothing to pseudo-observations. This result allows us to study the asymptotic properties of the smoothing spline estimator by deriving its asymptotic bias and variance. We show that smoothing splines are asymptotically consistent for an arbitrary working covariance and have the smallest variance when assuming the true covariance. We further show that both the seemingly unrelated kernel estimator and the smoothing spline estimator are nonlocal (unless working independence is assumed) but have asymptotically negligible bias. Their finite sample performance is compared through simulations. Our results justify the use of efficient, non-local estimators such as smoothing splines for clustered/longitudinal data.

Some key words: Asymptotic bias and variance; Asymptotic equivalent kernels; Consistency; Kernel regression; Longitudinal data; Non-locality; Nonparametric regression; Smoothing splines.

Short title: Equivalence of Splines and Kernels For Clustered/Longitudinal Data

1 INTRODUCTION

Nonparametric regression for clustered/longitudinal data has attracted considerable recent interest. Kernel methods have been considered by Zeger & Diggle (1994), Hoover, et al. (1998), Fan & Zhang (2000), Lin & Ying (2001), all of whom ignored the within-cluster correlation, and Severini & Staniswalis (1994) and Ruckstuhl, et al. (2000), who incorporated the within-cluster correlation into their kernel estimator. Lin & Carroll (2000) showed that the most efficient conventional kernel estimator is obtained by ignoring the correlation. Spline methods for clustered/longitudinal data have been investigated by Brumback & Rice (1998), Wang (1998), Zhang, et al. (1998), Lin & Zhang (1999), Verbyla, et al., (1999), among others. Most of these authors incorporated the within-cluster correlation into the construction of their spline estimators. These authors further demonstrated that an attractive feature of spline smoothing in clustered/longitudinal data is that it can be obtained by fitting mixed effects models.

For independent data, it is well known (Silverman, 1984) that kernel and spline estimators are asymptotically equivalent and are local in the sense that the estimator at a point gives non-zero weights only to observations whose covariate is in a shrinking neighborhood of that point. However, the relationship between kernel and spline estimators is not as well understood for clustered/longitudinal data. Welsh, et al. (2002) recently found that conventional kernel and spline estimators behave completely differently for clustered/longitudinal data. While the most efficient conventional kernel estimator requires completely ignoring the within-cluster correlation, the spline estimator with the smallest variance for a fixed smoothing parameter is obtained by incorporating the within-cluster correlation. Further, conventional kernel estimators are local asymptotically but spline estimators are not.

These results suggest that Silverman's results about the asymptotic equivalence of spline and kernel estimators for independent data do not hold for clustered/longitudinal data. This raises challenging questions. (1) For clustered/longitudinal data, what is the relationship between spline and kernel methods? In particular, is there a kernel estimator outside the conventional local kernel paradigm that is asymptotically equivalent to a spline estimator? (2) It is widely believed in the nonparametric literature that consistency of a nonparametric estimator requires locality. Since spline estimators are not local for clustered data, can they still be consistent? (3) What are the asymptotic properties, e.g., the asymptotic bias and variance, of a spline estimator for clustered/longitudinal data?

We will investigate these issues in this paper. We first show that for any working covariance matrix, the spline estimator is asymptotically equivalent to a recently proposed seemingly unrelated kernel estimator (Wang, 2003), which is constructed iteratively in a non-traditional way and is shown to be more efficient for clustered/longitudinal data than the best conventional local kernel estimator. Here the asymptotic equivalence is in a similar sense to Silverman (1984), namely that the weights of the smoothing spline estimator asymptotically converge to the weights of the seemingly unrelated kernel estimator using Silverman's kernel function. To gain insight into this asymptotic equivalence and to study the asymptotic properties of a smoothing spline, we show that for any working covariance matrix, both the smoothing spline estimator and the seemingly unrelated kernel estimator can be obtained iteratively by applying conventional kernel or spline smoothing to pseudo-observations. This result allows us to derive the asymptotic bias and variance of a smoothing spline estimator. It is shown that a smoothing spline is consistent for any arbitrary working covariance matrix and has the smallest variance when assuming the true covariance.

The outline of this paper is as follows. §2 describes the spline and kernel methods. §3 shows that the smoothing spline estimator is asymptotically equivalent to the seemingly unrelated kernel estimator for clustered/longitudinal data in the sense that their weights are asymptotically equivalent. §4 shows that both the spline estimator and the seemingly unrelated kernel estimator can be obtained iteratively using pseudo-observations. §5 derives the asymptotic bias and variance of the spline estimator. §6 shows that both the seemingly unrelated kernel estimator and the spline estimator are non–local and explains how non–local estimators can be consistent. §7 contains simulation results illustrating the equivalence we have shown in finite samples, and §8 contains concluding remarks.

2 The Kernel and Spline Methods for Nonparametric Regression in Clustered/Longitudinal Data

Consider data from n clusters with m_i observations consisting of an outcome variable and a single covariate. Clustered data have different types. For example, for longitudinal data, a cluster refers to a subject and observations within each cluster refer to the repeated measures over time; for familial data, a cluster refers to a family and observations from different family members within the same family are obtained. For simplicity, we assume the number of observations per cluster m_i is the same as m in all clusters. The results are applicable when cluster sizes vary between clusters. Suppose that Y_{ij} and T_{ij} are the outcome and the covariate of the jth observation in the ith cluster

$(i = 1, \cdots, n; j = 1, \cdots, m)$, and Y_{ij} depends on T_{ij} through

$$Y_{ij} = \theta(T_{ij}) + \epsilon_{ij}, \tag{1}$$

where $\theta(t)$ is an unknown smooth function, the errors $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{im})^T$ are independent with mean zero and true covariance matrix Σ. The covariate T_{ij} varies within each cluster. For longitudinal data, T_{ij} is a time-varying covariate or time. We assume in our asymptotic study that the cluster size m is finite ($m < \infty$) while the number of clusters n goes to infinity. For practical applications of model (1), see Zeger & Diggle (1994) and Zhang, et al. (1998).

Estimation of the nonparametric function $\theta(t)$ can proceed using kernel methods or spline methods. It is of interest to construct nonparametric estimators of $\theta(t)$ that account for the within-cluster correlation. Two methods which incorporate the within-cluster correlation will be described in this section: the seemingly unrelated kernel method (Wang, 2003) and the smoothing spline method (Wang, 1998; Zhang, et al., 1998).

2.1 The Seemingly Unrelated Kernel Estimator

In view of the findings that the conventional local kernel generalized estimating equation method fails to account for the within-cluster correlation (Lin & Carroll, 2000), Wang (2003) recently proposed a seemingly unrelated kernel estimator that we now describe. Define $Y_i = (Y_{i1}, \cdots, Y_{im})^T$, $\theta(T_i) = \{\theta(T_{i1}), \cdots, \theta(T_{im})\}^T$, $Y = (Y_1, \cdots, Y_n)^T$, and T_i, T, and $\theta(T)$ similarly. Denote $K_h(s) = h^{-1}K(s/h)$, where $K(\cdot)$ is a mean zero kernel function, and h is a bandwidth. For an arbitrary matrix A, denote by a^{jk} the (j,k)th element of A^{-1}. Let V be a working covariance matrix (Liang & Zeger, 1986). Consider a qth order polynomial kernel estimator. If $\theta(t)$ is estimated at the lth iteration by $\widehat{\theta}_K^{(l)}(t)$, the updated estimator of $\theta(t)$ at the $(l+1)$th iteration is $\widehat{\theta}_K^{(l+1)}(t) = \widehat{\alpha}_0$, where $\widehat{\alpha} = (\alpha_0, \alpha_1, \cdots, \alpha_q)^T$ solves

$$\sum_{i=1}^{n}\sum_{j=1}^{m} K_h(T_{ij} - t) B_{ij}(t)^T V^{-1}\{Y_i - \mu_{i(j)}(t)\} = 0, \tag{2}$$

where $B_{ij}(t)$ is an $m \times (q+1)$ matrix of zeros except that the jth row is $[1, (T_{ij} - t), \cdots, (T_{ij} - t)^q]^T$,

$$\mu_{i(j)}(t) = \left[\widehat{\theta}_K^{(l)}(T_{i1}), \cdots, \widehat{\theta}_K^{(l)}(T_{i,j-1}), \sum_{k=0}^{q} \alpha_k (T_{ij} - t)^k, \widehat{\theta}_K^{(l)}(T_{i,j+1}), \cdots, \widehat{\theta}_K^{(l)}(T_{im})\right]^T.$$

The final kernel estimator at convergence is called the seemingly unrelated (SUR) kernel estimator of $\theta(t)$ and is denoted by $\widehat{\theta}_K(t)$.

Wang (2003) gave the asymptotic bias and variance of $\widehat{\theta}_K(t)$ as $n \to \infty$, $h \to 0$ and $nh \to \infty$ for the linear kernel estimator ($q = 1$) as

$$\begin{aligned} E\{\widehat{\theta}_K(t)\} - \theta(t) &\approx h^2 \phi b_K(t); & (3) \\ \mathrm{var}\{\widehat{\theta}_K(t)\} &\approx \frac{\gamma}{nh} \frac{\tau(t)}{\eta^2(t)}, & (4) \end{aligned}$$

where $\phi = \int s^2 K(s) ds$, $\gamma = \int K^2(s) ds$, $\eta(t) = \sum_{j=1}^m v^{jj} f_j(t)$, $\tau(t) = \sum_{j=1}^m c_{jj} f_j(t)$ with c_{jj} being the (j,j)th element of $C = V^{-1}\Sigma V^{-1}$, $f_j(t)$ is the marginal density of T_j, and $b_K(t)$ solves

$$\sum_{j=1}^m \sum_{k=1}^m v^{jk} E\{b_K(T_k)|T_j = t\} f_j(t) = \frac{1}{2} \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} \theta^{(2)}(t).$$

Equivalently, $b_K(t)$ satisfies the Fredholm integral equation of the second kind:

$$b_K(t) + \int b_K(u) \zeta(u,t) du = \frac{1}{2} \theta^{(2)}(t), \tag{5}$$

where $\zeta(u,t) = \sum_{j=1}^m \sum_{k \neq j} v^{jk} f_{jk}(u,t) \,/\, \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\}$ and $f_{jk}(u,t)$ is the bivariate joint density of (T_{ij}, T_{ik}); see eq (17) of Lin & Carroll (2001). Wang (2003) showed that $\widehat{\theta}_K(t)$ with the smallest variance for fixed bandwidth is obtained by accounting for the within-cluster correlation by setting $V = \Sigma$, and it is more efficient than the best local kernel estimator of Lin & Carroll (2000).

2.2 The Generalized Least Squares Smoothing Spline Estimator

An alternative method for estimating $\theta(t)$ nonparametrically is to use smoothing splines. Assuming a working covariance matrix V, the pth order smoothing spline estimator minimizes

$$\frac{1}{n} \sum_{i=1}^n \{Y_i - \theta(T_i)\}^T V^{-1} \{Y_i - \theta(T_i)\} + \lambda \int \{\theta^{(p)}(t)\}^2 dt,$$

where λ is a smoothing parameter controlling for the trade off between the goodness of fit and the smoothness of the curve. The resulting pth order smoothing spline estimator of $\theta(t)$ is

$$\widehat{\theta}_S(T) = (\widetilde{V}^{-1} + n\lambda\Psi)^{-1} \widetilde{V}^{-1} Y, \tag{6}$$

where Ψ is the smoothing matrix (Green & Silverman, 1994) and $\widetilde{V} = \mathrm{diag}(V, \cdots, V)$. For $p = 1$, we have the linear smoothing spline and for $p = 2$, we have the cubic smoothing spline. We call $\widehat{\theta}_S(t)$ the generalized least square (GLS) smoothing spline estimator.

The GLS smoothing spline has attracted considerable recent attention because of its close connection with mixed effects models (Zhang, et al., 1998; Wang, 1998; Brumback and Rice, 1998; Verbyla, et al., 1999). In particular, it has been shown that an attractive feature of the GLS smoothing spline is that it can be easily fit using mixed effect models by writing the smoothing spline estimator $\widehat{\theta}_S(T)$ as a linear combination of fixed effects and random effects and can be easily calculated using the best linear unbiased predictors (BLUPs) from mixed models. However, despite its computational appeal, the theoretical properties of the GLS smoothing spline are not known. We will investigate them in this paper.

3 Asymptotic Equivalence of the Spline and the Seemingly Unrelated Kernel Estimators in the Sense of Silverman

In this section, we show that the spline estimator $\widehat{\theta}_S(t)$ and the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ are asymptotically equivalent in the sense of Silverman (1984), i.e., the weight functions used to calculate them are asymptotically equivalent when Silverman's (1984) kernel function is used for the SUR kernel estimator.

It is difficult to relate $\widehat{\theta}_S(t)$ and $\widehat{\theta}_K(t)$ by comparing (6) and (2) directly. Our strategy is to obtain a closed form expression for the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$. We then use this expression to relate $\widehat{\theta}_S(t)$ and $\widehat{\theta}_K(t)$. The closed form expression for the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ is given in Proposition 1 and its proof is given in Appendix A. For simplicity, we state in Proposition 1 the results for average kernels ($q = 0$). Results for a general qth order polynomial kernel are similar and are briefly stated after Proposition 1.

Proposition 1 *(1) For any working covariance matrix, the seemingly unrelated average kernel estimator at convergence $\widehat{\theta}_K(t)$ has a closed form expression,*

$$\widehat{\theta}_K(t) = K_{wh}^T(t)\{I + (\widetilde{V}^{-1} - \widetilde{V}^d)K_w\}^{-1}\widetilde{V}^{-1}Y, \tag{7}$$

where $K_{wh}(t) = \left\{\sum_{i=1}^n \sum_{j=1}^m K_h(T_{ij} - t)v^{jj}\right\}^{-1} \{K_h(T_{11} - t), \cdots, K_h(T_{nm} - t)\}^T$ *denotes a* $nm \times 1$ *vector,* $K_w = \{K_{wh}(T_{11}), \cdots, K_{wh}(T_{nm})\}^T$ *is an* $nm \times nm$ *matrix,* $\widetilde{V}^d = diag(V^d, \cdots V^d)$ *and* $V^d = diag(V^{-1})$.

(2) Denote by $\widehat{\theta}_K(T)$ *the seemingly unrelated average kernel estimator at convergence* $\widehat{\theta}_K(t)$ *evaluated at the vector of all the design points* T *, i.e.,* $\widehat{\theta}_K(T) = \{\widehat{\theta}_K(T_{11}), \cdots, \widehat{\theta}_K(T_{nm})\}^T$. *Then* $\widehat{\theta}_K(T)$ *has the closed form expression*

$$\widehat{\theta}_K(T) = \{I + K_w(\widetilde{V}^{-1} - \widetilde{V}^d)\}^{-1}K_w\widetilde{V}^{-1}Y. \tag{8}$$

For the qth order polynomial SUR kernel estimator, one modifies $K_{wh}(t)$ in (7) as $K_{wh}(t)^T =$

$\delta_1^T \left\{\widetilde{T}(t)^T K_{dh}(t)\widetilde{V}^d\widetilde{T}(t)\right\}^{-1} \widetilde{T}(t)^T K_{dh}(t)$, where $\delta_1 = (1, 0, \cdots, 0)^T$, $\widetilde{T}(t)$ is a $nm \times (q+1)$ matrix with the $\{(n-1)i+j\}$th row $\{1, (T_{ij}-t), \cdots, (T_{ij}-t)^q\}$, $K_{dh}(t) = \text{diag}\{K_h(T_{11}-t), \cdots K_h(T_{nm}-t)\}$.

Now consider the smoothing spline estimator $\widehat{\theta}_S(T)$. For any working covariance matrix V, simple calculations show that we can rewrite the spline estimator $\widehat{\theta}_S(T)$ in (6) as

$$\widehat{\theta}_S(T) = \left\{I + (\widetilde{V}^d + n\lambda\Psi)^{-1}(\widetilde{V}^{-1} - \widetilde{V}^d)\right\}^{-1} (\widetilde{V}^d + n\lambda\Psi)^{-1}\widetilde{V}^{-1}Y. \tag{9}$$

A comparison of (8) and (9) suggests that to show the weight function of the spline estimator $\widehat{\theta}_S(t)$ is asymptotically equivalent to that of the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$, we need to show $K_w \approx (\widetilde{V}^d + n\lambda\Psi)^{-1}$. Noticing that $\widetilde{V}^d$ is a diagonal matrix, we just need to show that under working independence, a weighted smoothing spline estimator is asymptotically equivalent to a conventional (weighted) kernel estimator in the sense of Silverman (1984). This can be done by modifying the results in §6 of Silverman (1984). Proposition 2 formally states the asymptotic equivalence of the weight function of the smoothing spline estimator $\widehat{\theta}_S(t)$ and that of the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$. It also provides the specific kernel function, and the asymptotic relationship between the smoothing parameter λ and the kernel bandwidth h. Its proof is given in Appendix B.

Proposition 2 *For any working covariance matrix V, denote the smoothing spline estimator by $\widehat{\theta}_S(t) = n^{-1}\sum_{i=1}^n \sum_{j=1}^m W_{S,ij}(t,T)Y_{ij}$ and the seemingly unrelated kernel estimator by $\widehat{\theta}_K(t) = n^{-1}\sum_{i=1}^n \sum_{j=1}^m W_{K,ij}(t,T)Y_{ij}$. Then the smoothing spline estimator $\widehat{\theta}_S(t)$ with the smoothing parameter λ is asymptotically equivalent to the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ with the kernel function $K(t)$ solving the differential equation (Silverman, 1984)*

$$(-1)^p K^{(2p)}(t) + K(t) = \Delta(t) \tag{10}$$

where $\Delta(t)$ is the Dirac delta function and the effective bandwidth $h(t) = \left\{\lambda / \sum_{j=1}^m v^{jj} f_j(t)\right\}^{1/2p}$, in the sense that their weight functions $W_S(t,T)$ and $W_K(t,T)$ are asymptotically equivalent.

Remarks:

1. The asymptotic equivalence here is in the sense of Silverman (1984). The kernel function $K(t)$ in (10) is identical to that given by Silverman (1984). For clustered data, a smoothing spline is thus effectively a seemingly unrelated kernel estimator with a varying bandwidth that depends not only on the marginal densities of the T_j but also on the working covariance matrix V.

2. The kernel function $K(t)$ in (10) has Fourier transform $(1 + t^{2p})^{-1}$ (Silverman, 1984) and satisfies $\int K(t)dt = 1$, $\int t^q K(t)dt = 0$ $(0 < q < 2p)$ and $\int t^{2p} K(t)dt = (-1)^{p-1}$, i.e., $K(t)$ is a $2p$th order kernel. For $p = 1$ (a linear spline), $K(t)$ is the Laplace density $1/2\exp(-|t|)$ and is the traditional second order kernel. For $p = 2$ (a cubic spline),

$$K(t) = 1/2\exp(-|t|/\sqrt{2})\sin(|t|/\sqrt{2} + \pi/4), \tag{11}$$

which is a 4th order kernel.

4 Understanding the Relationship Between the Spline and Seemingly Unrelated Kernel Estimators

Although the results in Section 3 provide a direct formula-based relationship between the spline estimator and the seemingly unrelated kernel estimator, they do not give us a good insight into why this equivalence holds. In this section, we provide a transparent relationship between the spline estimator and the seemingly unrelated kernel estimator, which provides us with clear insight into why they are asymptotically equivalent. This relationship also allows us to derive the asymptotic bias and variance of the spline estimator later in Section 5.

We now derive a transparent relationship between the spline estimator and the seemingly unrelated kernel estimator. Specifically, we show that both estimators can be obtained in the same iterative fashion using pseudo-observations. They differ only in that standard (weighted) kernel smoothing is used at each iteration for the seemingly unrelated kernel estimator, while standard (weighted) spline smoothing is used at each iteration for the smoothing spline estimator.

We start with the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$. For simplicity, consider the average kernel $(p = 0)$. Using equation (2), for any working covariance matrix V, simple calculations show that the seemingly unrelated kernel estimator at the $(l + 1)$th iteration can be rewritten as

$$\widehat{\theta}_K^{(l+1)}(t) \quad = \quad \frac{n^{-1}\sum_{i=1}^n \sum_{j=1}^m K_h(T_{ij} - t)v^{jj}Y_{ij}^{(l+1)}}{n^{-1}\sum_{i=1}^n \sum_{j=1}^m v^{jj}K_h(T_{ij} - t)} \tag{12}$$

where

$$Y_{ij}^{(l+1)} = Y_{ij} + (v^{jj})^{-1}\sum_{k\neq j} v^{jk}\left\{Y_{ik} - \widehat{\theta}_K^{(l)}(T_{ik})\right\} \tag{13}$$

denotes the pseudo-observation at the $(l+1)$th iteration. Equation (12) shows that at each iteration, conventional (weighted) Nadaraya-Watson kernel smoothing is applied to the pseudo-observations $Y_{ij}^{(l+1)}$ with the weights $\{v^{jj}\}$.

We next consider the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$ given in (6). We show that $\widehat{\theta}_S(t)$ can also be obtained iteratively by applying conventional weighted spline smoothing to the pseudo-observations $Y_{ij}^{(l+1)}$ with the weights $\{v^{jj}\}$ at each iteration. Consider calculating a smoothing spline estimator $\widehat{\theta}_S^*(t)$ in the following fashion. Set the initial estimator $\widehat{\theta}_S^{(0)}(t)$ of $\theta(t)$ to be the standard pth order smoothing spline estimator obtained by assuming independence among all observations. If the smoothing spline estimator at the lth iteration is denoted by $\widehat{\theta}_S^{(l)}(t)$, the updated estimator of $\theta(t)$ at the $(l+1)$th iteration is obtained by minimizing with respect to $\theta(\cdot)$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} v^{jj}\left\{Y_{ij}^{(l+1)} - \theta(T_{ij})\right\}^2 + \lambda \int \left\{\theta^{(p)}(t)\right\}^2 dt,$$

where the $Y_{ij}^{(l+1)}$ are the pseudo-observations defined in (13) except $\widehat{\theta}_K^{(l)}(t)$ is replaced by $\widehat{\theta}_S^{(l)}(t)$. It follows that the resulting estimator at the $(l+1)$th iteration is the conventional weighted pth order smoothing spline estimator

$$\widehat{\theta}_S^{(l+1)}(T) = (\widetilde{V}^d + n\lambda\Psi)^{-1}\widetilde{V}^d Y^{(l+1)}, \tag{14}$$

where $Y^{(l+1)} = \{Y_1^{(l+1)^T}, \cdots, Y_n^{(l+1)^T}\}^T$ with $Y_i^{(l+1)} = \{Y_{i1}^{(l+1)}, \cdots, Y_{im}^{(l+1)}\}^T$. This means at the $(l+1)$th iteration, we update $\theta(t)$ by applying the conventional weighted pth order spline smoothing to the pseudo-observations $Y_{ij}^{(l+1)}$ with the weights v^{jj}. Following Silverman (1984) and Nychka (1995), one can easily further show that for any t, the estimator at the $(l+1)$th iteration can be written asymptotically as

$$\widehat{\theta}_S^{(l+1)}(t) = \left(n\sum_{j=1}^{m} v^{jj}\right)^{-1} \sum_{i=1}^{n}\sum_{j=1}^{m} G_{\lambda^*}(t, T_{ij}) v^{jj} Y_{ij}^{(l+1)} + o_p(1), \tag{15}$$

where $G_{\lambda^*}(t,s)$ is the Green's function defined in (A.10) with $\lambda^* = \lambda/\{\sum_{j=1}^m v^{jj}\}$ and $f(t) = \{\sum_{j=1}^m v^{jj} f_j(t)\}/\{\sum_{j=1}^m v^{jj}\}$. Denote by $\widehat{\theta}_S^*(t)$ the iterative weighted smoothing spline estimator at convergence. Proposition 3 shows that $\widehat{\theta}_S^*(t)$ equals the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$ in (6). The proof is given in Appendix C.

Proposition 3 *The iterative weighted smoothing spline estimator at convergence $\widehat{\theta}_S^*(t)$ has a closed form expression and equals the GLS smoothing spline estimator $\widehat{\theta}_S(t)$ in (6).*

Now the relationship between the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ and the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$ becomes transparent, and we can see why they are asymptotically equivalent. Both estimators can be obtained iteratively. At each iteration, the seemingly unrelated kernel estimator applies standard weighted kernel smoothing to the pseudo-observations $Y_{ij}^{(l+1)}$, while the GLS smoothing spline estimator applies standard weighted spline smoothing to the pseudo-observations $Y_{ij}^{(l+1)}$. Since standard weighted spline smoothing and standard weighted kernel smoothing are asymptotically equivalent (Silverman, 1984) at each iteration, they should be asymptotically equivalent at convergence.

5 The Asymptotic Bias and Variance of the Generalized Least Square Smoothing Spline Estimator

The results in Section 3 only show that the generalized least squares smoothing spline estimator is asymptotically equivalent to the seemingly unrelated kernel estimator in the sense that their weight functions are asymptotically equivalent. However, analogously to the independent data case pointed out by Nychka (1995), these results are too rough to establish the asymptotic bias and variance of the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$. These calculations are of substantial interest, since they provide the asymptotic properties of $\widehat{\theta}_S(t)$ and allow us to investigate whether $\widehat{\theta}_S(t)$ is consistent. We perform such calculate in this section.

It is difficult to study the asymptotic bias and variance of the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$ using its closed form expression in equation (6). We hence need to take a different route. The results in Proposition 3 show that $\widehat{\theta}_S(t)$ can be obtained iteratively with standard weighted spline smoothing at each iteration. This connection provides a much easier way to calculate the asymptotic bias and variance of $\widehat{\theta}_S(t)$ by iteratively applying the asymptotic bias and variance results of the smoothing spline estimator for independent data (Nychka, 1995) at each iteration. Proposition 4 provides the asymptotic bias and variance of the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$. Its proof is given in Appendix D.

Proposition 4 *Denote by $\widehat{\theta}_S(t)$ the pth order generalized least squares smoothing spline estimator given in (6) using any given working covariance matrix V. Assume the marginal densities $f_j(t)$ of the T_{ij} have uniformly continuous derivatives. We make the same asymptotic assumptions about n and λ as Nychka (1995) and assume Nychka's bias and variance results of the pth order spline in his equations (1.9) and (1.10) hold for independent data and the properties of Green's function in (A.11) and (A.14) hold. Then we have the following.*

(1) The asymptotic bias of $\widehat{\theta}_S(t)$ is

$$E\left\{\widehat{\theta}_S(t)\right\} - \theta(t) \approx (-1)^{p-1}\frac{\lambda}{\sum_{j=1}^m v^{jj} f_j(t)} b_S(t),$$

where $b_S(t)$ satisfies

$$\sum_{j=1}^{m}\sum_{k=1}^{m} v^{jk} E\{b_S(T_k)|T_j = t\} f_j(t) = \frac{1}{a_p}\left\{\sum_{j=1}^{m} v^{jj} f_j(t)\right\} \theta^{(2p)}(t),$$

where a_p is a constant. Equivalently, $b_S(t)$ solves the Fredholm integral equation of the second kind, with the right hand side of (5) replaced by $\theta^{(2p)}(t)/a_p$.

(2) The asymptotic variance of $\widehat{\theta}_S(t)$ is

$$\mathrm{var}\left\{\widehat{\theta}_S(t)\right\} \approx \frac{\gamma_p}{n}\left\{\frac{\lambda}{\sum_{j=1}^{m} v^{jj} f_j(t)}\right\}^{-1/2p} \frac{\tau(t)}{\eta^2(t)},$$

where γ_p is a constant and $\tau(t)$ and $\eta(t)$ are defined below equation (4). We believe that $\gamma_p = \int K^2(t)dt$ where $K(t)$ satisfies (10).

(3) While holding the effective bandwidth fixed as $h(t) = \left\{\lambda/\sum_{j=1}^{m} \sigma^{jj} f_j(t)\right\}^{1/2p}$, the generalized least squares smoothing spline estimator with the smallest variance is obtained by assuming the working covariance matrix V equals the true covariance Σ, i.e., $V = \Sigma$. Its variance is

$$\mathrm{var}_{min}\left\{\widehat{\theta}_S(t)\right\} \approx \frac{\gamma_p}{n}\left\{\frac{\lambda}{\sum_{j=1}^{m} \sigma^{jj} f_j(t)}\right\}^{-1/2p} \frac{1}{\sum_{j=1}^{m} \sigma^{jj} f_j(t)}.$$

(4) The generalized least squares smoothing spline estimator has the asymptotic expansion

$$\begin{aligned}\widehat{\theta}_S(t) - \theta(t) &= \frac{1}{n\sum_{j=1}^{m} v^{jj}} \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} G_{\lambda^*}(t, T_{ij}) v^{jk}\left\{Y_{ik} - \theta(T_{ik})\right\} + (-1)^{p-1} h^{2p}(t) b_S(t) \\ &\quad + o_p\left[\{nh(t)\}^{-1/2} + h^{2p}(t)\right], \qquad (16)\end{aligned}$$

where $G_{\lambda^*}(t,s)$ is the Green's function defined in (A.10) and $h(t) = \left[\lambda/\left\{\sum_{j=1}^{m} v^{jj} f_j(t)\right\}\right]^{1/2p}$ denotes the effective bandwidth (See Proposition 2).

Remarks:

1. The bias results in part (1) of Proposition 4 show that for any working covariance matrix V, the generalized least squares smoothing spline estimator is asymptotically consistent.

2. The expressions for the asymptotic bias and variance of the generalized least squares smoothing spline estimator closely resemble the forms of the asymptotic bias and variance of the seeming related kernel estimator given in (3) and (4). Specifically, setting the effective bandwidth $h(t) = \left[\lambda/\left\{\sum_{j=1}^{m} v^{jj} f_j(t)\right\}\right]^{1/2p}$ (Proposition 2), the asymptotic bias and variance of the generalized least squares smoothing spline estimator are

$$\begin{aligned} E\{\widehat{\theta}_S(t)\} - \theta(t) &= (-1)^{p-1} h^{2p}(t) b_S(t) + o_p\left\{h^{2p}(t)\right\}, \\ \mathrm{var}\{\widehat{\theta}_S(t)\} - \theta(t) &= \{nh(t)\}^{-1}\{\gamma_p \tau(t)/\eta^2(t)\} + o_p\left[\{nh(t)\}^{-1}\right], \end{aligned}$$

The above bias and variance calculations clearly show that the pth order generalized least squares smoothing spline estimator behaves asymptotically similar to the $q = (2p-1)$th order polynomial seemingly related kernel estimator with the kernel function being the $(2p)$th order kernel defined in (10). Note that Wang (2003) focused on the second order SUR kernels. Extensions of her results to higher SUR kernels are straightforward along the lines of Wand and Jones (1995, Section 5.4).

3. The asymptotic bias of the linear generalized least squares smoothing spline estimator ($p = 1$) is $h^2(t)b_S(t)$, where $b_S(t)$ solves (5) with the right-hand side of the equation equal to $\theta^{(2)}(t)/a_1$, where a_1 is some constant. Hence it corresponds to the second order SUR kernel. The asymptotic bias of the cubic generalized least squares smoothing spline is $-h^4(t)b_S(t)$, where $b_S(t)$ solves (5) with the right-hand side of the equation equal to $\theta^{(4)}(t)/a_2$, where a_2 is a constant. Hence it corresponds to a higher (4th) order SUR kernel.

4. When $m = 1$, we have cross-sectional independent data following $Y_i = \theta(T_i) + \epsilon_i$, where the ϵ_i are independent and identically distributed and follows $N(0, \sigma^2)$. The results in Proposition 4 reduce to those given in Nychka (1995) as

$$E\{\widehat{\theta}_S(t)\} - \theta(t) \approx (-1)^{p-1}h^{2p}(t)\theta^{(2p)}(t)/a_p, \tag{17}$$

$$\mathrm{var}\{\widehat{\theta}_S(t)\} \approx \{nh(t)\}^{-1}\{\gamma_p\sigma^2/f(t)\}, \tag{18}$$

where $h(t) = \{\lambda\sigma^2/f(t)\}^{1/2p}$ and $f(t)$ is the density of T_i.

5. For clustered/longitudinal data, the working independence smoothing spline estimator is calculated by treating the data to be independent with error variances σ_{jj}. Its asymptotic bias and variance takes the same form as (17) and (18) except that $\sigma^2/f(t)$ is replaced by $\left\{\sum_{j=1}^m \sigma_{jj}^{-1} f_j(t)\right\}^{-1}$ in (18) and $h(t)$ replaced by $h(t) = \left[\lambda/\left\{\sum_{j=1}^m \sigma_{jj}^{-1} f_j(t)\right\}\right]^{1/2p}$.

6 Non-locality and Consistency of Generalized Least Squares Splines and Seemingly Unrelated Kernels

The purpose of this section is to first demonstrate in Section 6.1 the non–locality of the generalized least squares spline and the seemingly unrelated kernel. It is widely believed in the nonparametric literature that a consistent nonparametric estimator requires locality, i.e., for any t, only observations whose covariate values are in the neighborhood of t are used asymptotically to estimate $\theta(t)$. A non-local nonparametric estimator is often inconsistent. For clustered/longitudinal data, both generalized least squares spline and seemingly unrelated kernel estimators are nonlocal, but from the results in Sections 2.1 and 4, they are consistent. It is hence of substantial interest to understand how non-local estimators can be consistent. We investigate this issue in §6.2.

6.1 Observation-Level Non–Locality

For independent data, both kernel and spline estimators are local. For clustered/longitudinal data, the same is true for working independence kernel and spline estimators (Welsh, et al., 2002). Assuming the number of observations per cluster m is finite, this means only observations from different clusters contribute to estimation of $\theta(t)$ at any t. Such locality is widely presumed to be necessary to ensure consistency of a nonparametric estimator. However, Welsh, et al. (2002) observed that the generalized least squares spline is not local. This non-locality of the spline is supported by the asymptotic spline expansion in (16). A similar asymptotic expansion holds for the seemingly unrelated kernel estimator by replacing $G_{\lambda_*}(t, T_{ij})$ in (16) by $K_h(T_{ij} - t)$ (Wang, 2003). It follows that the seemingly unrelated kernel estimator is not local either.

The non–locality can also be demonstrated for seemingly unrelated kernel estimators via (12) and for the generalized least squares spline estimators via (15). Here we see clearly that both methods are local in the pseudo observations (13): the seemingly unrelated kernel pseudo observation at convergence is $Y_{ij}^* = Y_{ij} + (v^{jj})^{-1} \sum_{j=1}^m \sum_{k \neq j} v^{jk} \{Y_{ik} - \widehat{\theta}_K(T_{ik})\}$, and a similar form holds for the generalized least squares spline. Clearly this term depends on all the responses in the cluster, not just Y_{ij}. These discussions suggest that if <u>any</u> observation in a cluster has an T_{ij} near the value at which the function is to be fit, then <u>all</u> the observations in the cluster contribute to the fit, i.e., both methods are not local at the observation-level.

To illustrate this non-locality, we provide a numerical example. For $n = 50$ and $m = 3$, we generated the covariate T_{ij} from the Uniform$(-2, 2)$ distribution and assumed Y_{ij} followed model (1) with $\theta(t) = \sin(2t)$ and an exchangeable covariance $\Sigma = 4(0.2I + 0.8J)$, where J is a $m \times m$ matrix of ones, i.e. the pair-wise within-cluster correlation is 0.8. Both the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ and the smoothing spline estimator $\widehat{\theta}_S(t)$ can be written in the form $\widehat{\theta}(t) = \sum_{i=1}^n \sum_{j=1}^m W_{ij}(t, T) Y_{ij}$, where the expression for the weights $W_{ij}(t, T)$ is given by equation (7) for the seemingly unrelated kernel estimator $\widehat{\theta}_K(t)$ and by equation (7) of Welsh, et al., (2002) for the smoothing spline estimator $\widehat{\theta}_S(t)$. Using these weight functions, we investigate how the observation at t is weighted when we estimate $\theta(s)$ for a series of values of s.

Figure 1(a) plots the weights for the SUR kernel estimator $\widehat{\theta}_K(t)$ at $t = 0.25$ with the kernel function (11) and bandwidth $h = 0.4$ for both working independence (which is identical to the classical kernel estimate) and assuming the true covariance. A similar plot for the cubic smoothing spline estimator $\widehat{\theta}_S(t)$ with $\lambda = 0.1$ is given in Figure 1(b). One can easily see that the working independence kernel and spline estimates are local, while the seemingly unrelated kernel estimate and the generalized least squares cubic smoothing spline estimate assuming the true covariance are not local. Further, since the kernel function (11) is the equivalent kernel of the cubic smoothing spline estimator, the shapes of the two weight curves are nearly identical. Figure 1 hence also supports the theoretical results in Sections 3 and 4 and provide numerical evidence that the asymptotic equivalence of the SUR kernel and the generalized least squares spline estimators works quite well in finite samples.

6.2 Cluster–Level Locality and Consistency

The discussions in Section 6.1 show that to estimate $\theta(t)$ at any t, all observations within the same cluster are used if any T_{ij} in a cluster is in the neighborhood of t. This sharing of information among members of the cluster is what makes the generalized least squares spline and seemingly unrelated kernel methods more efficient than their working independence versions. In other words, an efficient kernel/spline estimator that effectively accounts for the within-cluster correlation requires using all observations within the same cluster and hence has to be non-local at the observation–level.

This brings a question of substantial interest, i.e., how can a non-local estimator be consistent? Specifically, how could observations whose covariate values are outside the neighborhood of t still contribute to estimation of $\theta(t)$ without introducing asymptotic bias? This is often viewed as impossible for independent data.

We investigate this issue by examining the pseudo-observations used at each iteration of the seemingly unrelated kernel and generalized least squares spline estimators. Specifically, from equations (12) and (15), conventional (weighted) kernels and splines are applied to the pseudo-observations $Y_{ij}^{(l+1)}$ in (13) at each iteration with $\widehat{\theta}^{(l)}(t)$ estimated using the kernel method and the spline method at the previous iteration respectively. Hence both methods are local in the pseudo-observations. Each pseudo-observation $Y_{ij}^{(l+1)}$ is a weighted average of Y_{ij} and the residuals of the other observations within the same cluster $\{Y_{ik} - \widehat{\theta}^{(l)}(T_{ik})\}$ calculated by centering Y_{ik} around their means estimated consistently using either kernel or spline methods at the previous iteration $\widehat{\theta}^{(l)}(T_{ik})$. This means although both methods are non-local at the observation–level, they are local at the cluster–level. Further, although $Y_{ij}^{(l+1)}$ uses all observations within the same cluster, $Y_{ij}^{(l+1)}$

is asymptotically unbiased for $\theta(T_{ij})$, i.e., $E\left\{Y_{ij}^{(l+1)}\right\} \approx \theta(T_{ij})$. Hence by re-centering at each iteration, observations within the same cluster whose covariate values are not in the neighborhood of t are used to improve the efficiency of estimation of $\theta(t)$ without introducing bias. When these iterative procedures are used to calculate seemingly unrelated kernel and generalized least squares spline estimators, the re-centering is done by updating $\widehat{\theta}(T_{ij})$ at each iteration. When their closed form expressions in (7) and (6) are used, the re-centering is done implicitly for all design points simultaneously.

7 Simulation Study

In this section, we present simulation results which compare the finite sample efficiency of the seemingly unrelated kernel estimator and the smoothing spline estimator. We focus on using the true covariance in estimation; under working independence, the generalized least squares smoothing spline and seemingly unrelated kernel estimators reduce to conventional spline and kernel estimators respectively, and these estimators have already been compared in Welsh, et al. (2002). We consider the seemingly unrelated linear kernel estimator using the Epanechnikov kernel, and the cubic smoothing spline estimator.

We assumed in our simulation that the number of clusters $n = 50$ or $n = 100$, and the cluster size $m = 3$. The covariate T_{ij} was generated independently from the uniform[-2,2] distribution. The outcome Y_{ij} followed model (1) with $\theta(t)$ specified by each of the following 4 functions with different curvatures: (1) model 1: $\theta(t) = \sin(2t)$; (2) model 2: $\theta(t) = \sqrt{z(1-z)}\sin\{2\pi(1+2^{-3/5})/(z+2^{-3/5})\}$; (3) model 3: $\theta(t) = \sqrt{z(1-z)}\sin\{2\pi(1+2^{-7/5})/(z+2^{-7/5})\}$; (4) model 4: $\theta(t) = \sin(8z-4) + 2\exp\{-256(z-.5)^2\}$, where $z = (t+2)/4$. We assumed the marginal variances of the Y_{ij} were 1, and considered three true covariance structures: (1) exchangeable with common correlation $\rho = 0.6$; (2) autoregressive with correlation $\rho = 0.6$, (3) unstructured with $\rho_{12} = \rho_{23} = 0.8$ and $\rho_{13} = 0.5$.

For each configuration, we generated 200 simulated data sets and estimated $\theta(t)$ using the seeming unrelated kernel estimator and the generalized least squares cubic smoothing spline estimator, where the true covariance was assumed. For simplicity and for the sake of consistency when comparing different methods, for each simulated data set, we estimated the bandwidth parameter h for the SUR kernel estimator and the smoothing parameter λ for the smoothing spline estimator by minimizing the MSEs of the estimators $\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{\widehat{\theta}(t)-\theta(t)\right\}^2$ by using our knowledge of the actual function.

Table 1 compares the average mean squared error efficiencies comparing generalized least squares smoothing splines with seemingly unrelated kernels assuming the true covariance matrix. The results show that the relative efficiencies of generalized least squares spline estimates to seemingly unrelated kernel estimates are close to one. Hence generalized least squares spline estimates and seemingly unrelated kernel estimates have similar behavior in finite samples. These results are consistent with our theoretical findings that generalized least squares splines and seemingly unrelated kernels are asymptotically equivalent.

8 Discussion

Welsh, et al. (2002) showed that smoothing splines that incorporate the within-cluster correlation are not asymptotically equivalent to conventional kernel generalized estimating equation estimators for clustered/longitudinal data and that they have smaller variances than conventional kernel estimators for clustered data. We have shown, however, that these smoothing spline estimators are asymptotically equivalent to seemingly unrelated kernel estimators. Further, both seemingly unrelated kernels and smoothing splines can be obtained iteratively by applying standard weighted spline or kernel smoothing to pseudo-observations. This transparent connection allows us to study the asymptotic properties (biases and variances) of smoothing splines, where direct investigation is difficult. This result justifies the use of smoothing splines in clustered/longitudinal data. In addition, the pseudo observation insight shows that all the observations in a cluster contribute to the fit if any of the observations in the cluster have a value of T_{ij} near the point at which the fit is to be evaluated.

We have assumed in this paper that the number of observations per cluster is finite when the number of clusters goes to infinity. This assumption holds for most common longitudinal studies and studies involving clustered data such as familial studies and multi-center clinical trials. In some situations, however, both the number of observations and the number of clusters go to infinity, e.g., EEG recording is done every 30 seconds over night for patients in neurological research to study brain activities (Malow, et al., 1996). Examination of the proof in Appendix A shows that the asymptotic equivalence of smoothing splines and seemingly unrelated kernels in the sense of Silverman (1984) still holds even when both the number of observations and the number of clusters go to infinity. However, the asymptotic properties (biases and variances) of seemingly unrelated kernel estimators and smoothing splines in this situation are not well understood and are currently under investigation.

Acknowledgments

This work was supported in part by grants from the National Cancer Institute and the National Institute of Environmental Health Sciences.

APPENDIX

Appendix A Proof of Proposition 1

Consider the average kernel estimator ($q = 0$). Denote by $\widehat{\theta}_K^{(l)}(T) = \{\widehat{\theta}_K^{(l)}(T_{11}), \cdots, \widehat{\theta}_K^{(l)}(T_{nm})\}^T$ the seemingly unrelated kernel estimator at the lth iteration evaluated at the vector of all the design points $T = (T_{11}, \cdots, T_{nm})^T$. Using equation (2), some calculations give

$$\widehat{\theta}_K^{(l+1)}(t) = (1^T K_{dh} \widetilde{V}^d 1)^{-1} \left[1^T K_{dh} \widetilde{V}^d \widehat{\theta}_K^{(l)}(T) + 1^T K_{dh} \widetilde{V}^{-1} \left\{Y - \widehat{\theta}_K^{(l)}(T)\right\}\right], \tag{A.1}$$

where $K_{dh} = \mathrm{diag}\{K_h(T_{11} - t), \cdots K_h(T_{nm} - t)\}$. Denote by

$$K_{wh} = \{\sum_{i=1}^n \sum_{j=1}^m K_h(T_{11} - t) v^{jj}\}^{-1} \{K_h(T_{11} - t), \cdots, K_h(T_{nm} - t)\}^T$$

the vector of standardized kernel weights at t. Write $\widehat{\theta}_K^{(l)}(T) = K^{(l)} Y$. We need to find the relationship between $K^{(l+1)}$ and $K^{(l)}$. Using (A.1), simple calculations give

$$\widehat{\theta}_K^{(l+1)}(t) = K_{wh}(t)^T \{\widetilde{V}^d K^{(l)} + \widetilde{V}^{-1}(I - K^{(l)})\} Y. \tag{A.2}$$

It follows that

$$\widehat{\theta}_K^{(l+1)}(T) = K_w \{\widetilde{V}^d K^{(l)} + \widetilde{V}^{-1}(I - K^{(l)})\} Y, \tag{A.3}$$

and $K^{(l+1)} = K_w\{\widetilde{V}^d K^{(l)} + \widetilde{V}^{-1}(I - K^{(l)})\}$, where $K_w = \{K_{wh}(T_{11}), \cdots K_{wh}(T_{nm})\}^T$.

At convergence, $K^{(l+1)} = K^{(l)} = K_*$. Hence K_* solves $K_* = K_w\{\widetilde{V}^d K_* + \widetilde{V}^{-1}(I - K_*)\}$, i.e,

$$K_* = \{I + K_w(\widetilde{V}^{-1} - \widetilde{V}^d)\}^{-1} K_w \widetilde{V}^{-1}. \tag{A.4}$$

Substituting (A.4) into (A.2) and (A.3), some simple algebra give (7) and (8).

Appendix B Proof of Proposition 2

Using the results stated in the paragraph before Proposition 2, it suffices to show $K_w = (\widetilde{V}^d + n\lambda\Psi)^{-1}$ asymptotically. Since $\widetilde{V}^d$ is a diagonal matrix, we simply need to show that under working independence, the weighted spline estimator is asymptotically equivalent to the conventional kernel estimator. This can be easily shown using the results in §6 of Silverman (1984).

Specifically, for any weights w_{ij}, the weighted kernel estimator under working independence is $\widehat{\theta}_{WK}(t) = \{C_w(t)\}^{-1} \sum_{i=1}^n \sum_{j=1}^m w_{ij} K_h(T_{ij} - t) Y_{ij}$, where $C_w(t) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} K_h(T_{ij} - t)$ and the subscript WK denotes the weighted kernel under working independence. Denote by $C_w = \mathrm{diag}\{C_w(T_{11}), \cdots, C_w(T_{nm})\}$. Then $\widehat{\theta}_{WK}(t)$ evaluated at the vector of all design points T is

$$\widehat{\theta}_{WK}(T) = \{\widehat{\theta}_{WK}(T_{11}), \cdots, \widehat{\theta}_{WK}(T_{nm})\}^T = K_w W Y, \tag{A.5}$$

where $W = \mathrm{diag}(w_{11}, \cdots, w_{nm})$, $K_w = C_w^{-1} K_h$, K_h is a $nm \times nm$ matrix with elements $K_h(T_{ij} - T_{i'j'})$ $(i, i' = 1, \cdots, n,\ j, j' = 1, \cdots, m)$.

The weighted smoothing spline estimator under working independence evaluated at all the design points T is

$$\widehat{\theta}_{WS}(T) = (W + n\lambda\Psi)^{-1} W Y, \tag{A.6}$$

where the subscript WS denotes the weighted spline estimator under working independence. The results in §6 of Silverman (1984) show that the weights of $\widehat{\theta}_{WK}(T)$ and those of $\widehat{\theta}_{WS}(T)$ are asymptotically equivalent. Now let $W = \widetilde{V}^d$. A comparison between (A.5) with (A.6) gives $K_w = (\widetilde{V}^d + n\lambda\Psi)^{-1}$ asymptotically with the kernel function defined in (10).

We now study how the bandwidth h is related to the smoothing parameter λ. Following Silverman (1984), we first standardize the weights as $w_{ij} = v^{jj}/n \sum_{j=1}^m v^{jj}$ such that $\sum_{i=1}^n \sum_{j=1}^m w_{ij} = 1$, and calculate the weighted cumulative distribution function and its limit as

$$F_n^w(t) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} I(T_{ij} \le t) \to F(t) = \sum_{j=1}^m v^{jj} F_j(t) / \sum_{j=1}^m v^{jj}. \tag{A.7}$$

Using the results in Theorem A of Silverman (1984) and replacing λ by $\lambda / \sum_{j=1}^m v^{jj}$ and $f(t)$ by $\sum_{j=1}^m v^{jj} f_j(t) / \sum_{j=1}^m v^{jj}$, the bandwidth $h(t)$ is related to λ by

$$h(t) = \left(\frac{\lambda}{\sum_{j=1}^m v^{jj}} \right)^{1/2p} \left(\frac{\sum_{j=1}^m v^{jj} f_j(t)}{\sum_{j=1}^m v^{jj}} \right)^{-1/2p} = \left\{ \frac{\lambda}{\sum_{j=1}^m v^{jj} f_j(t)} \right\}^{1/2p}.$$

Appendix C Proof of Proposition 3

The pseudo-observations at the $(l+1)$th iteration $Y_{ij}^{(l+1)}$ in (13) can be written in a vector form as

$$Y^{(l+1)} = \left\{ \widetilde{V}^d \right\}^{-1} \left[\widetilde{V}^d \widehat{\theta}_S^{(l)}(T) + \widetilde{V}^{-1} \left\{ Y - \widehat{\theta}_S^{(l)}(T) \right\} \right]. \tag{A.8}$$

Plugging (A.8) into (14), some calculations give

$$\widehat{\theta}_S^{(l+1)}(T) = \left(\widetilde{V}^d + n\lambda\Psi \right)^{-1} \left[\left(\widetilde{V}^d - \widetilde{V}^{-1} \right) \widehat{\theta}_S^{(l)}(T) + \widetilde{V}^{-1} Y \right]. \tag{A.9}$$

Write $\widehat{\theta}_S^{(l)}(T) = S^{(l)}Y$. Then from (A.9), we have $S^{(l+1)} = \left(\widetilde{V}^d + n\lambda\Psi\right)^{-1}\left[(\widetilde{V}^d - \widetilde{V}^{-1})S^{(l)} + \widetilde{V}^{-1}\right]$. At convergence $S^{(l+1)} = S^{(l)} = S_*$. Hence S_* solves $S_* = \left(\widetilde{V}^d + n\lambda\Psi\right)^{-1}\left[(\widetilde{V}^d - \widetilde{V}^{-1})S_* + \widetilde{V}^{-1}\right]$. Simple calculations give $S_* = (\widetilde{V}^{-1} + n\lambda\Psi)^{-1}\widetilde{V}^{-1}$. This finishes the proof.

Appendix D Proof of Proposition 4

For independent data with $m = 1$, the model is $Y_i = \theta(T_i) + \epsilon_i$, where the ϵ_i are independent and follow $N(0, \sigma^2)$. Denote by $f(t)$ the density of T_i, which has a uniformly continuous derivative. Nychka (1995) stated the bias and variance of the $(2p)$th order smoothing spline estimator as

$$
\begin{aligned}
E\{\widehat{\theta}_S(t)\} - \theta(t) &\approx \frac{(-1)^{p-1}\lambda}{f(t)}\frac{\theta^{(2p)}(t)}{a_p} \\
\operatorname{var}\{\widehat{\theta}_S(t)\} &\approx \frac{\gamma_p}{n}\left\{\frac{\lambda}{f(t)}\right\}^{-1/2p}\frac{\sigma^2}{f(t)},
\end{aligned}
$$

where a_p and γ_p are some constants. He provided a rigorous proof of this result for the linear spline $(p = 1)$ and sketched an argument for the general pth $(p > 1)$ order spline. We assume in our proof that $n \to \infty$ and $\lambda \to 0$ and λ and t are chosen such that the above bias and variance results hold for independent data for any $p \geq 1$.

We first state a few facts about the Green's function, which are used extensively in the proof. The Green's function is associated with the solution of the differential equation of $\alpha(t)$

$$
\lambda^*(-1)^p\frac{d^{2p}\alpha(t)}{dt^{2p}} + f(t)\alpha(t) = f(t)g(t), \tag{A.10}
$$

for any functions $f(t)$ and $g(t)$ and any arbitrary positive constant λ^*. Then the solution of this differential equation is $\alpha(t) = \int G_{\lambda^*}(t,\tau)f(\tau)g(\tau)d\tau$, where $G_{\lambda^*}(t,\tau)$ is called the Green's function. The Green's function has the following properties

$$
\begin{aligned}
\int G_{\lambda^*}(t,\tau)g(\tau)f(\tau)d\tau - g(t) &= \frac{(-1)^{p-1}\lambda^*}{a_p f(t)}g^{(2p)}(t) + o\{\lambda^*/f(t)\} && \text{(A.11)}\\
\int G^2_{\lambda^*}(t,\tau)f(\tau)d\tau &= \frac{\gamma_p}{f(t)}\left\{\frac{\lambda^*}{f(t)}\right\}^{-1/2p} + o\left[\{\lambda^*/f(t)\}^{-1/2p}\right] && \text{(A.12)}\\
\int\int G_{\lambda^*}(t,\tau_1)G_{\lambda^*}(t,\tau_2)r(\tau_1,\tau_2)d\tau_1 d\tau_2 &= \frac{r(t,t)}{f^2(t)} + o(1) && \text{(A.13)}\\
\int d(\tau)G^2_{\lambda^*}(t,\tau)f(\tau)d\tau &= d(t)\frac{\gamma_p}{f(t)}\left\{\frac{\lambda^*}{f(t)}\right\}^{-1/2p} + o\left[\{\lambda^*/f(t)\}^{-1/2p}\right] && \text{(A.14)}
\end{aligned}
$$

where γ_p is some constant and $d(\tau)$ and $r(\tau_1, \tau_2)$ are arbitrary functions. Equations (A.11) and (A.12) are given in Nychka (1995), equation (A.13) can be easily shown using (A.10) and (A.11). We believe that (A.14) holds but have not proved it.

D.1 Asymptotic Bias and Variance of the Working Independence Estimator

Under model (1), we first derive an asymptotic expansion of the smoothing spline estimator assuming working independence $V = \text{diag}\{v^{jj}\}$. Denote by $\widehat{\theta}_{SI}(t)$ such an estimator. We can show that the results in Section 6 of Silverman (1984) still hold for $\widehat{\theta}_{SI}(t)$ and asymptotically, at any given t,

$$\widehat{\theta}_{SI}(t) = \frac{1}{\sum_{j=1}^m v^{jj}} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m G_{\lambda^*}(t, T_{ij}) v^{jj} Y_{ij} \right\} + o_p(1),$$

where $G_{\lambda^*}(t,s)$ is the Green's function associated with (A.10) with $\lambda^* = \lambda / \left(\sum_{j=1}^m v^{jj}\right)$ and $f(t) = \sum_{j=1}^m v^{jj} f_j(t) / \sum_{j=1}^m v^{jj}$.

We hence have

$$\begin{aligned} E\{\widehat{\theta}_{SI}(t)\} - \theta(t) &\approx \frac{1}{\sum_{j=1}^m v^{jj}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m v^{jj} G_{\lambda^*}(t, T_{ij}) \theta(T_{ij}) - \theta(t) \qquad \text{(A.15)} \\ &\approx \frac{1}{\sum_{j=1}^m v^{jj}} \sum_{j=1}^m v^{jj} \int G_{\lambda^*}(t,\tau)\theta(\tau) f_j(\tau) d\tau - \theta(t) \approx \int G_{\lambda^*}(t,\tau)\theta(\tau) f(\tau) d\tau - \theta(t). \end{aligned}$$

Using (A.11) and the definitions of $f(t)$ and λ^*, the bias of $\widehat{\theta}_{SI}(t)$ is

$$E\{\widehat{\theta}_{SI}(t)\} - \theta(t) = \frac{(-1)^{p-1}\lambda}{\sum_{j=1}^m v^{jj} f_j(t)} \frac{\theta^{(2p)}(t)}{a_p} + o_p\{h^{2p}(t)\}, \qquad \text{(A.16)}$$

where $h(t) = \left[\lambda / \left\{\sum_{j=1}^m v^{jj} f_j(t)\right\}\right]^{1/2p}$. Further $\text{var}\left\{\widehat{\theta}_{SI}(t)\right\} \approx A_{1n} + A_{2n}$, where

$$\begin{aligned} A_{1n} &= \left(n \sum_{j=1}^m v^{jj} \right)^{-2} \sum_{i=1}^n \sum_{j=1}^m \left(v^{jj}\right)^2 \sigma_{jj} G^2_{\lambda^*}(t, T_{ij}) \\ &= \left(n \sum_{j=1}^m v^{jj} \right)^{-1} \int \frac{\sum_{j=1}^m \left(v^{jj}\right)^2 \sigma_{jj} f_j(\tau)}{\sum_{j=1}^m v^{jj} f_j(\tau)} G^2_{\lambda^*}(t,\tau) f(\tau) d\tau + o_p(n^{-1}), \\ A_{2n} &= \left(n \sum_{j=1}^m v^{jj} \right)^{-2} \sum_{i=1}^n \sum_{j=1}^m \sum_{k \neq j}^m v^{jj} v^{kk} \sigma_{jk} G_{\lambda^*}(t, T_{ij}) G_{\lambda^*}(t, T_{ik}) \\ &= \left(\sum_{j=1}^m v^{jj} \right)^{-2} n^{-1} \sum_{j=1}^m \sum_{j \neq k}^m v^{jj} v^{kk} \sigma_{jk} \int G_{\lambda^*}(t,\tau) G_{\lambda^*}(t,s) f_{jk}(\tau, s) d\tau ds + o_p(n^{-1}). \end{aligned}$$

Here $f_{jk}(\tau, s)$ is the bivariate density of (T_{ij}, T_{ik}). Using (A.14) and the definitions of $f(t)$ and λ^*, one can calculate A_{1n} and using (A.13), $A_{2n} = O_p(1/n)$. It follows that

$$\text{var}\left\{\widehat{\theta}_{SI}(t)\right\} = \frac{\sum_{j=1}^m \left(v^{jj}\right)^2 \sigma_{jj} f_j(t)}{\left\{\sum_{j=1}^m v^{jj} f_j(t)\right\}^2} \frac{\gamma_p}{n} \left\{ \frac{\lambda}{\sum_{j=1}^m v^{jj} f_j(t)} \right\}^{-1/2p} + o_p\left[\{nh(t)\}^{-1}\right]. \qquad \text{(A.17)}$$

Some calculations further show that $\widehat{\theta}_{SI}(t)$ has the following asymptotic expansion

$$\widehat{\theta}_{SI}(t) - \theta(t) = \frac{1}{n\sum_{j=1}^{m} v^{jj}} \sum_{i=1}^{n}\sum_{j=1}^{m} v^{jj} G_{\lambda^*}(t, T_{ij})\{Y_{ij} - \theta(T_{ij})\} + a(t) + o_p\left[\{nh(t)\}^{-1/2} + h^{2p}(t)\right], \tag{A.18}$$

where

$$a(t) = \frac{(-1)^{p-1}\lambda}{\sum_{j=1}^{m} v^{jj} f_j(t)} \frac{\theta^{(2p)}(t)}{a_p}. \tag{A.19}$$

D.2 The Asymptotic Bias and Variance of the Generalized Least Squares Smoothing Spline Estimator

Using the results in Proposition 3, the generalized least squares smoothing spline estimator can be obtained equivalently by iteratively applying standard weighted spline smoothing to the pseudo-observations. We hence study the asymptotic properties of the generalized least squares smoothing spline estimator $\widehat{\theta}_S(t)$ through this iterative procedure by deriving the asymptotic bias and variance of $\widehat{\theta}_S^{(l)}(t)$ at each iteration and then those at convergence, which give the asymptotic bias and variance of $\widehat{\theta}_S(t)$. This asymptotic analysis strategy is advantageous since working independence spline smoothing is applied at each iteration and its properties have been derived in Section D.1.

The smoothing spline estimator $\widehat{\theta}_S^{(l+1)}(t)$ at the $(l+1)$th iteration is updated using (15) with the pseudo-observations $Y_{ij}^{(l+1)} = Y_{ij} + (v^{jj})^{-1} \sum_{k\neq j} v^{jk}\left\{Y_{ik} - \widehat{\theta}_S^{(l)}(T_{ik})\right\}$. Suppose we set the initial estimator of $\theta(t)$ using the working independence estimator described in Section D.1 and denote it by $\widehat{\theta}_S^{(0)}(t)$. We first study the asymptotic properties of the one-step estimator. By setting $l = 1$ and substituting $Y_{ij}^{(1)}$ into (15), at the first iteration, we have $\widehat{\theta}_S^{(1)}(t) = D_{1n} + D_{2n} + D_{3n} + o_p(1)$, where

$$D_{1n} = \frac{1}{n\sum_{j=1}^{m} v^{jj}} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{m} G_{\lambda^*}(t, T_{ij}) v^{ik}\{Y_{ik} - \theta(T_{ik})\} \tag{A.20}$$

$$D_{2n} = \frac{1}{n\sum_{j=1}^{m} v^{jj}} \sum_{i=1}^{n}\sum_{j=1}^{m} G_{\lambda^*}(t, T_{ij}) v^{jj}\theta(T_{ij}) \tag{A.21}$$

$$D_{3n} = -\frac{1}{n\sum_{j=1}^{m} v^{jj}} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k\neq j} v^{jk} G_{\lambda^*}(t, T_{ik})\left\{\widehat{\theta}_S^{(0)}(T_{ik}) - \theta(T_{ik})\right\}. \tag{A.22}$$

First examine D_{3n}. Using the asymptotic expansion $\widehat{\theta}_S^{(0)}(t)$ in (A.18), D_{3n} can be written as $D_{3n} = D_{3n,1} + D_{3n,2}$, where

$$\begin{aligned} D_{3n,1} &= -\frac{1}{\sum_{j=1}^{m} v^{jj}} \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k\neq j} v^{jk} G_{\lambda^*}(t, T_{ij}) \left[\frac{1}{\sum_{j=1}^{m} v^{jj}} \frac{1}{n} \sum_{r=1}^{n}\sum_{s=1}^{m} G_{\lambda^*}(T_{ik}, T_{rs}) v^{ss}\{Y_{rs} - \theta(T_{rs})\}\right] \\ &\quad + o_p\left[\{nh(t)\}^{-1/2}\right] \\ D_{3n,2} &= -\frac{1}{\sum_{j=1}^{m} v^{jj}} \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k\neq j} v^{jk} G_{\lambda^*}(t, T_{ij}) a(T_{ik}) + o_p\left\{h^{2p}(t)\right\}. \end{aligned}$$

Simple calculations give

$$D_{3n,1} = -\frac{1}{\left(\sum_{j=1}^m v^{jj}\right)^2} \frac{1}{n} \sum_{r=1}^n \sum_{s=1}^m v^{ss} \{Y_{rs} - \theta(T_{rs})\} \left\{ \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \sum_{k \neq j} v^{jk} G_{\lambda^*}(t, T_{ij}) G_{\lambda^*}(T_{ik}, T_{rs}) \right\} \tag{A.23}$$
$$+ o_p\left[\{nh(t)\}^{-1/2}\right]$$
$$= -\frac{1}{\left(\sum_{j=1}^m v^{jj}\right)^2} \frac{1}{n} \sum_{r=1}^n \sum_{s=1}^m v^{ss} \{Y_{rs} - \theta(T_{rs})\} \sum_{j=1}^m \sum_{k \neq j} v^{jk} \int \int G_{\lambda^*}(t, \tau) G_{\lambda^*}(T_{rs}, s) f_{jk}(\tau, s) d\tau ds$$
$$+ o_p\left[\{nh(t)\}^{-1/2}\right]$$

Using (A.13), one can easily see that $D_{3n,1} = o_p(1)$. This implies that $D_{3n,1}$ is negligible in the asymptotic expansion of $\widehat{\theta}_S^{(1)}(t)$ and in the asymptotic bias and variance calculations of $\widehat{\theta}_S^{(1)}(t)$.

The term $D_{3n,2}$ can be written as

$$\begin{aligned} D_{3n,2} &= -\frac{1}{\sum_{j=1}^m v^{jj}} \sum_{j=1}^m \sum_{k \neq j} v^{jk} \int G_{\lambda^*}(t, \tau) a(s) f_{jk}(\tau, s) d\tau ds + o_p\{h^{2p}(t)\} \\ &= -\frac{1}{\sum_{j=1}^m v^{jj}} \sum_{j=1}^m \sum_{k \neq j} v^{jk} \int G_{\lambda^*}(t, \tau) E\{a(T_k) | T_j = \tau\} f_j(\tau) d\tau + o_p\{h^{2p}(t)\} \\ &= -\sum_{j=1}^m \sum_{k \neq j} v^{jk} \int G_{\lambda^*}(t, \tau) \left[E\{a(T_k) | T_j = \tau\} \frac{f_j(\tau)}{\sum_{j=1}^m v^{jj} f_j(\tau)} \right\} f(\tau) d\tau + o_p\{h^{2p}(t)\} \end{aligned} \tag{A.24}$$

Setting $g(\tau) = E\{a(T_k)|T_j = \tau\} f_j(\tau)/f(\tau)$ and using (A.11), we have

$$D_{3n,2} = -\sum_{j=1}^m \sum_{k \neq j} v^{jk} E\{a(T_k) | T_j = t\} f_j(t) \Big/ \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} + o_p\{h^{2p}(t)\}. \tag{A.25}$$

This indicates that $D_{3n,2}$ contributes to the leading bias term of $\widehat{\theta}_S^{(1)}(t)$. It follows from the results of $D_{3n,1}$ and $D_{3n,2}$ that D_{3n} only contributes to the leading asymptotic bias of $\widehat{\theta}_S^{(1)}(t)$, but not to the leading asymptotic variance of $\widehat{\theta}_S^{(1)}(t)$.

One can easily see D_{2n} only contributes to the asymptotic bias of $\widehat{\theta}_S^{(1)}(t)$. Specifically $D_{2n} - \theta(t)$ takes exactly the same expression as the right-hand side of equation (A.15). It follows that

$$D_{2n} - \theta(t) = a(t) + o_p\{h^{2p}(t)\}, \tag{A.26}$$

where $a(t)$ is defined in (A.19).

Since D_{1n} has mean zero, it does not contribute to the bias. Combining equations (A.25) and (A.26), the asymptotic bias of $\widehat{\theta}^{(1)}(t)$ is $E\left\{\widehat{\theta}_S^{(1)}(t)\right\} - \theta(t) = b_1(t) + o_p\{h^{2p}(t)\}$, where

$$b_1(t) = a(t) - \left[\sum_{j=1}^m \sum_{k \neq j} v^{jk} E\{a(T_k) | T_j = t\} f_j(t) \Big/ \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} \right]. \tag{A.27}$$

Now calculate $\mathrm{var}\left\{\widehat{\theta}_S^{(l+1)}(t)\right\}$. The term D_{1n} determines the leading asymptotic variance of $\widehat{\theta}_S^{(1)}(t)$ and in fact is the leading term of the asymptotic expansion of $\widehat{\theta}_S^{(1)}(t)$. Rewrite D_{1n} as

$$D_{1n} = \frac{1}{n\sum_{j=1}^m v^{jj}} \sum_{i=1}^n G_{\lambda^*}(t, T_i)^T V^{-1}\{Y_i - \theta(T_i)\}, \tag{A.28}$$

where $G_{\lambda^*}(t, T_i) = \{G_{\lambda^*}(t, T_{i1}), \cdots, G_{\lambda^*}(t, T_{im})\}^T$. It follows that

$$\begin{aligned} \mathrm{var}\{\widehat{\theta}_S^{(1)}(t)\} &= \frac{1}{\left(n\sum_{j=1}^m v^{jj}\right)^2} \sum_{i=1}^n G_{\lambda^*}(t, T_i)^T V^{-1}\Sigma V^{-1} G_{\lambda^*}(t, T_i) + o_p[\{nh(t)\}^{-1}] \\ &= F_{1n} + F_{2n} + o_p[\{nh(t)\}^{-1}], \end{aligned}$$

where denoting $C = V^{-1}\Sigma V^{-1}$ and by c_{jk} the (j,k)th element of C,

$$\begin{aligned} F_{1n} &= \frac{1}{\left(n\sum_{j=1}^m v^{jj}\right)^2} \sum_{i=1}^n \sum_{j=1}^m G^2_{\lambda^*}(t, T_{ij}) c_{jj} \\ F_{2n} &= \frac{1}{\left(n\sum_{j=1}^m v^{jj}\right)^2} \sum_{i=1}^n \sum_{j=1}^m \sum_{k\neq j} G_{\lambda^*}(t, T_{ij}) G_{\lambda^*}(t, T_{ik}) c_{jk}. \end{aligned}$$

First consider the term F_{1n}, which can be calculated as

$$\begin{aligned} F_{1n} &= \frac{1}{n\left(\sum_{j=1}^m v^{jj}\right)^2} \sum_{j=1}^m c_{jj} \int G^2_{\lambda^*}(t, \tau) f_j(\tau) d\tau + o_p[\{nh(t)\}^{-1}] \\ &= \frac{1}{n\sum_{j=1}^m v^{jj}} \int \left\{\frac{\sum_{j=1}^m c_{jj} f_j(\tau)}{\sum_{j=1}^m v^{jj} f_j(\tau)}\right\} G^2_{\lambda^*}(t, \tau) f(\tau) d\tau + o_p[\{nh(t)\}^{-1}]. \end{aligned}$$

Using equality (A.14), we have

$$F_{1n} = \frac{\sum_{j=1}^m c_{jj} f_j(t)}{\left\{\sum_{j=1}^m v^{jj} f_j(t)\right\}^2} \frac{\gamma_p}{n} \left\{\frac{\lambda}{\sum_{j=1}^m v^{jj} f_j(t)}\right\}^{-1/2p} + o_p[\{nh(t)\}^{-1}]. \tag{A.29}$$

The term F_{2n} can be written as, using (A.13)

$$\begin{aligned} F_{2n} &= \frac{1}{n\left(\sum_{j=1}^m v^{jj}\right)^2} \sum_{j=1}^m \sum_{k\neq j} c_{jk} \int\int G_{\lambda^*}(t, \tau) G_{\lambda^*}(t, s) f_{jk}(\tau, s) d\tau ds + o_p(n^{-1}) \\ &= \frac{1}{n\left(f(t)\sum_{j=1}^m v^{jj}\right)^2} \sum_{j=1}^m \sum_{k\neq j} c_{jk} f_{jk}(t,t) + o_p(n^{-1}) = O_p(n^{-1}) \end{aligned}$$

It follows that

$$\mathrm{var}\left\{\widehat{\theta}_S^{(1)}(t)\right\} = \frac{\sum_{j=1}^m c_{jj} f_j(t)}{\left\{\sum_{j=1}^m v^{jj} f_j(t)\right\}^2} \frac{\gamma_p}{n} \left\{\frac{\lambda}{\sum_{j=1}^m v^{jj} f_j(t)}\right\}^{-1/2p} + o_p\left[\{nh(t)\}^{-1}\right] \tag{A.30}$$

The above calculations show that after the first iteration $\widehat{\theta}_S^{(1)}(t)$ has the asymptotic expansion

$$\begin{aligned}\widehat{\theta}_S^{(1)}(t) - \theta(t) &= \frac{1}{n\sum_{j=1}^m v^{jj}} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m G_{\lambda^*}(t, T_{ij}) v^{jk} \{Y_{ik} - \theta(T_{ik})\} \\ &\quad + b_1(t) + o_p[\{nh(t)\}^{-1/2} + h^{2p}(t)], \end{aligned} \tag{A.31}$$

where the bias term $b_1(t)$ is defined in (A.27).

Now consider the estimator of $\theta(t)$ at the second iteration $\widehat{\theta}_S^{(2)}(t)$. Using (A.9), we have $\widehat{\theta}_S^{(2)}(t) = D_{1n} + D_{2n} + D_{3n} + o_p(1)$, where D_{1n} and D_{2n} are the same as those given in equations (A.20) and (A.21), and D_{3n} is the same as (A.22) except that $\widehat{\theta}_S^{(0)}(T_{ik})$ is replaced by $\widehat{\theta}_S^{(1)}(T_{ik})$. Similarly, using the asymptotic expansion of $\widehat{\theta}_S^{(1)}(t)$ in (A.31), we write $D_{3n} = D_{3n,1} + D_{3n,2}$, where $D_{3n,1}$ is slightly different from (A.23)

$$\begin{aligned} D_{3n,1} &= -\frac{1}{\sum_{j=1}^m v^{jj}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \sum_{k \neq j} v^{jk} G_{\lambda^*}(t, T_{ij}) \\ &\quad \times \left[\frac{1}{\sum_{j=1}^m v^{jj}} \frac{1}{n} \sum_{r=1}^n \sum_{s=1}^m \sum_{\ell=1}^m G_{\lambda^*}(T_{ik}, T_{rs}) v^{s\ell} \{Y_{r\ell} - \theta(T_{r\ell})\} \right] + o_p\left[\{nh(t)\}^{-1/2}\right] \\ D_{3n,2} &= -\frac{1}{\sum_{j=1}^m v^{jj}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \sum_{k \neq j} v^{jk} G_{\lambda^*}(t, T_{ij}) b_1(T_{ik}) + o_p\left\{h^{2p}(t)\right\}. \end{aligned}$$

Similar calculations to those in (A.23) and (A.24) show that $D_{3n,1} = o_p(1)$ and

$$D_{3n,2} = -\sum_{j=1}^m \sum_{k \neq j} v^{jk} E\{b_1(T_k) | T_j = t\} f_j(t) \Big/ \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} + o_p\left\{h^{2p}(t)\right\}. \tag{A.32}$$

Hence the asymptotic variance of $\widehat{\theta}_S^{(2)}(t)$ takes exactly the same form as that given in (A.30) and the asymptotic bias of $\widehat{\theta}_S^{(2)}(t) = b_2(t) + o_p\{h^{2p}(t)\}$, where

$$b_2(t) = a(t) - \left[\sum_{j=1}^m \sum_{k \neq j} v^{jk} E\{b_1(T_k) | T_j = t\} f_j(t) \Big/ \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} \right].$$

It follows that the second iteration does not change the variance but only make a refinement of the bias. The asymptotic expansion of $\widehat{\theta}_S^{(2)}(t)$ takes the same form as the right-hand side of (A.31) except $b_1(t)$ is replaced by $b_2(t)$.

Using induction, one can easily see that the expansion of $\widehat{\theta}_S^{(l+1)}(t)$ at the $(l+1)$th iteration $(l > 2)$ is exactly the same as $\widehat{\theta}_S^{(2)}(t)$ except that $b_2(t)$ is replaced by $b_{l+1}(t)$, which satisfies

$$b_{l+1}(t) = a(t) - \left[\sum_{j=1}^m \sum_{k \neq j} v^{jk} E\{b_l(T_k) | T_j = t\} f_j(t) \Big/ \left\{ \sum_{j=1}^m v^{jj} f_j(t) \right\} \right].$$

Its asymptotic variance takes the same form as that in (4). In other words, further iterations only make refinements of the bias but not the variance.

At convergence $\widehat{\theta}_S^{(l+1)}(t)$ converges to $\widehat{\theta}_S(t)$ (Proposition 3), and $b_{l+1}(t) = b_l(t) = b_S(t)$, where $b_S(t)$ satisfies

$$b_S(t) = a(t) - \left[\sum_{j=1}^{m} \sum_{k \neq j} v^{jk} E\left\{b_S(T_k) | T_j = t\right\} f_j(t) \Big/ \left\{\sum_{j=1}^{m} v^{jj} f_j(t)\right\}\right].$$

The asymptotic bias of $\widehat{\theta}_S(t)$ is $b_S(t)$. The results of part (1) of Proposition 4 follows immediately. The asymptotic variance of $\widehat{\theta}_S(t)$ takes the same form as the right-hand side of equation (A.30). Hence the results in part (2) of Proposition 4 follow immediately. A direct application of the Cauchy-Schwartz inequality gives part (3). One can also easily see from the above calculations that the asymptotic expansion of $\widehat{\theta}_S(t)$ given in part (4) holds. It should be noted that these asymptotic expansion and asymptotic variance and bias do not depend on choice of the initial consistent estimator $\widehat{\theta}_S^{(0)}(t)$.

REFERENCES

Brumback, B. & Rice, J. A. (1998) Smoothing spline models for the analysis of nested and crossed samples of curves (with discussions). *J. Amer. Statist. Assoc.*, 93, 961-1006.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall, London.

Hoover, D. R., Rice, J. A., Wu, C. O. & Yang, Y. (1998). Nonparametric smoothing estimates of time–varying coefficient models with longitudinal data. *Biometrika*, 85, 809–822.

Fan, J. & Zhang, J.T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. Royal Statist. Soc., Series B*, 62, 303-322.

Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Lin, D. & Ying, Z. (2001) Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.*, 96, 103–126.

Lin, X. & Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Amer. Statist. Assoc.*, 95, 520-534.

Lin, X. & Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations, *J. Amer. Statist. Assoc.*, 96, 1045-1056.

Lin, X. & Zhang, D. (1999). Inference in generalized additive mixed model using smoothing splines. *J. Royal Statist. Soc., Series B*, 61, 381-400.

Malow, B. A., Kushwaha, R., Lin, X., Morton, K. L., Aldrich, M. S. (1997). Relationship of interictal epileptiform discharges to sleep depth in temporal lobe epilepsy. *EEG Clin. Neurophysiology*, 102, 20-26.

Nychka, D. (1995) Splines as local smoothers. *Ann. Statist.*, 23, 1175-1197.

Ruckstuhl, A., Welsh, A.H. & Carroll, R.J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statist. Sinica*, 10, 51–71.

Severini, T. A. & Staniswalis, J. G. (1994) Quasilikelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, 89, 501–511.

Silverman, B. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.*, 12, 898-916.

Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing.* Chapman and Hall.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *J. Royal Statist. Soc., Series B*, 60, 159-174.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, to appear.

Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussions)". *App. Statist.*, 48, 269-311.

Welsh, A. H., Lin, X. & Carroll, R. J. (2002). Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. *J. Amer. Statist. Assoc.*, 97, 482–493.

Zeger, S. L. & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, 50, 689-699.

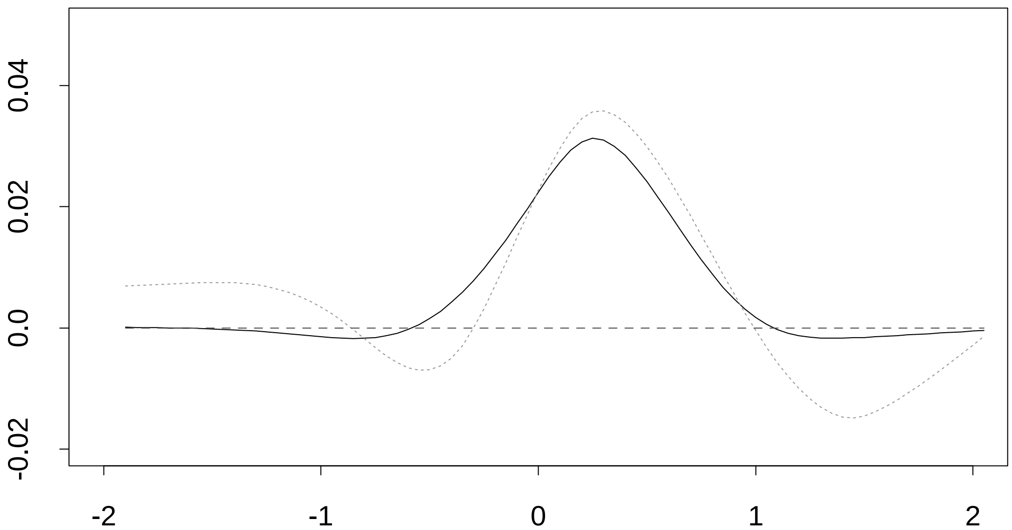
Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Amer. Statist. Assoc.*, 93, 710-719.

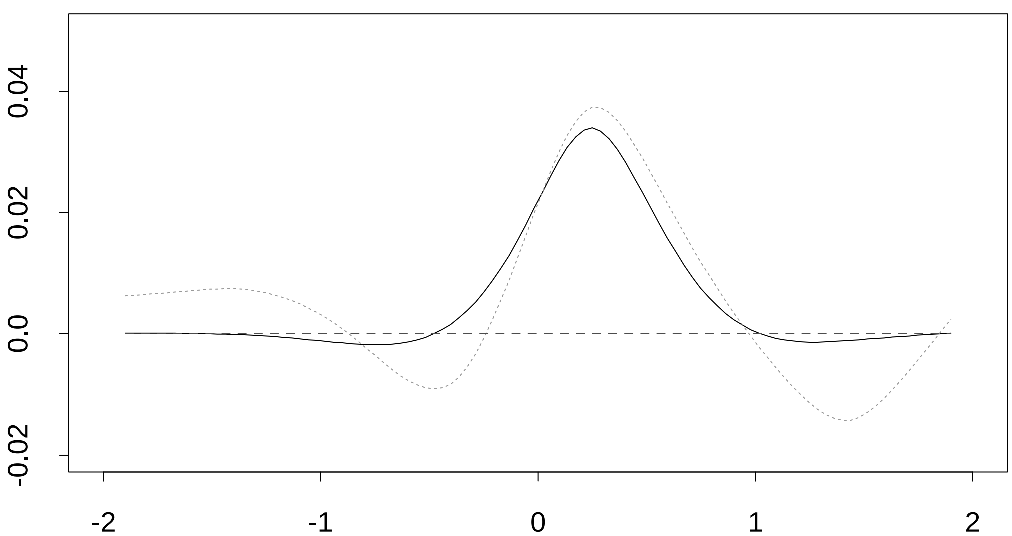
		MSE Efficiency of Splines Relative to SUR Kernels	
Model	Corr	$n = 50$	$n = 100$
1	1	1.15	1.22
1	2	1.16	1.21
1	3	0.99	1.07
2	1	0.99	1.03
2	2	0.98	1.01
2	3	0.94	0.99
3	1	0.98	1.00
3	2	0.98	1.00
3	3	0.92	0.98
4	1	1.09	1.07
4	2	1.08	1.06
4	3	1.25	1.13

Table 1: Table 1 Simulated relative efficiencies comparing smoothing splines with seemingly unrelated (SUR) kernels assuming the true covariance. The four models correspond to model 1: $\theta(t) = \sin(2t)$; model 2: $\theta(t) = \sqrt{z(1-z)}\sin\{2\pi(1+2^{-3/5})/(z+2^{-3/5})\}$; model 3: $\theta(t) = \sqrt{z(1-z)}\sin\{2\pi(1+2^{-7/5})/(z+2^{-7/5})\}$; model 4: $\theta(t) = \sin(8z-4) + 2\exp\{-256(z-.5)^2\}$. The three correlation structures correspond to Corr = 1: the autoregressive case with $\rho = 0.6$; Corr = 2: the exchangeable case with $\rho = 0.6$; Corr = 3: the unstructured case with $\rho_{12} = \rho_{23} = 0.8$ and $\rho_{13} = 0.5$.

FIGURE

Figure 1. The weights for the iterative kernel estimate and the smoothing spline estimate when $n = 50$ and $m = 3$ in the exchangeable case with $\rho = 0.8$. The solid line assumes working independence and dotted line assumes the true covariance: (a) the seemingly unrelated kernel estimate; (b) the smoothing spline estimate.


0.04
0.02
0.0
-0.02
-2
-1
0
1
2


0.04
0.02
0.0
-0.02
-2
-1
0
1
2