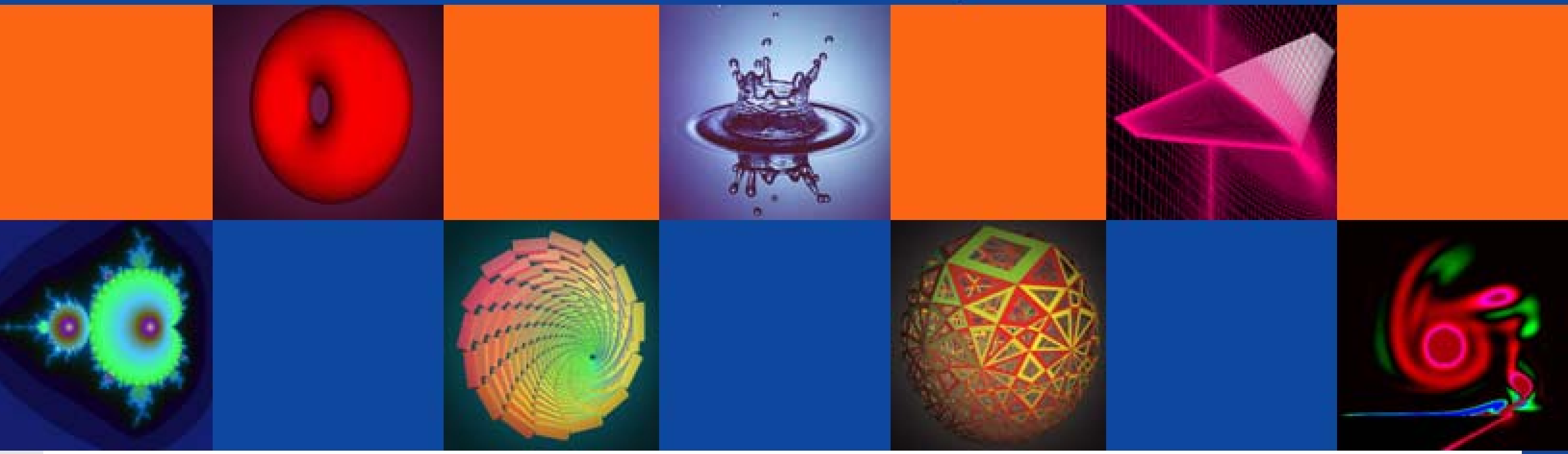
The University of Texas ARLINGTON

MATHEMATICS
Preprint Series



Variance Reduction in Smoothing Splines

Robert L. Paige
Shan Sun
Keyi Wang

Technical Report 2007-22

http://www.uta.edu/math/preprint/

Variance Reduction in Smoothing Splines

Robert L. Paige*, Texas Tech University   Shan Sun, University of Texas Arlington
Keyi Wang, Edwards Lifesciences

October 1, 2007

Abstract

We develop a variance reduction method for smoothing splines. We do this by showing that the quadratic interpolation method introduced in Cheng et al. (2006), for local linear estimators, also works for smoothing splines. For a given point of estimation, Cheng et al. (2006) define a variance-reduced local linear estimate as a linear combination of classical estimates at three nearby points. We use equivalent kernel function results from Nychka (1995) and Lin et al. (2004) in the development of our methodologies. First, we develop a variance reduction method for spline estimators in univariate regression models. Next, we develop an analogous variance reduction method for spline estimators in clustered/longitudinal models. Finally, simulation studies are performed which demonstrate the efficacy of our variance reduction methods in finite sample settings.

Keywords: Variance reduction; smoothing splines; clustered/longitudinal data; nonparametric regression.

1 Introduction

We generalize the variance reduction methodologies, for local linear estimators, introduced in Cheng et al. (2006) to smoothing spline estimators. For a given point of estimation, Cheng et al. (2006) define a variance-reduced local linear estimate as a linear combination of classical estimates at three nearby points. This linear combination is constructed in such a way to obtain maximal reduction in asymptotic variance while producing an asymptotic bias which is the same as that of the classical estimate. There are a few specific features of their variance reduced estimator which our variance reduced estimators will also possess; (i) global automatic smoothing parameter values can be easily obtained; (ii) the asymptotic mean squared error is often improved considerably; (iii) the amount of reduction is uniform across different locations, regression functions, designs and error distributions;

*To whom correspondence should be addressed.

(iv) the reductions in asymptotic variance are realized in finite sample cases; and (v) the estimators admit simple forms and only slightly increase the amount of computation time.

There is a large literature discussing improvements and modifications for kernel and local linear estimators, most of which have to do with bias reduction; see Cheng et al. (2006) for more details. There are very few variance reduction methods in print, besides those described in Cheng et al. (2006); two exceptions are Cheng and Hall (2003, 2004). In contrast there is a much smaller literature on improvements and modifications for spline estimators and results in print only address bias reduction; see for instance bias minimizing splines considered in Agarwal and Studden (1980), and the bias reduction lemma proven in Eubank (1999).

In section 2, we review the variance reduction method from Cheng et al. (2006). In section 3, we briefly review equivalent kernels for smoothing spline estimators. In section 4, we develop a variance reduced smoothing splines estimator for the classical nonparametric univariate regression model via Cheng et al.'s quadratic interpolation procedure. In section 5, we develop a reduced variance generalized least squares (GLS) smoothing spline estimator for longitudinal/clustered data. Here we show that quadratic interpolation again yields a variance reduced estimator. The GLS spline estimator is of particular interest since it achieves it smallest variance when it is constructed with the true within-subject correlation, as shown in Lin et al. (2004). In contrast, most efficient local kernel estimators ignore with-in subject correlation as shown in Lin and Carroll (2000). Other spline methods for clustered/longitudinal data have been investigated by Brumback and Rice (1998), Wang (2003), Zhang et al. (1998), Lin and Zhang (1999) and Verbyla et al. (1999). In most of these works, within-subject correlation was incorporated into the construction of spline estimators. Finally, in section 6, we present the results of simulation studies which demonstrate the efficacy of our variance reduction methods in finite sample settings.

2 Variance Reduction for Local Linear Estimators

Consider the nonparametric univariate regression model

$$Y_j = \theta(X_j) + \epsilon_j \tag{1}$$

where observation $Y = (Y_1, \ldots, Y_n)^T$ depends upon smooth function $\theta(x)$, and errors $\epsilon_1, \ldots . \epsilon_n$ which are assumed to be independent, possessing a zero mean and variance $\sigma^2(x)$. Cheng et al.

(2006) take as their classical estimator a version of the local linear regressor that admits asymptotic unconditional variance,

$$\hat{\theta}(x) = \sum_{j=1}^{n} w\left(x, x_j\right) Y_j \tag{2}$$

where $w\left(x, x_j\right)$ is a weight function given as

$$h \frac{S_{n,2}(x) K_h(x - x_j) - S_{n,1}(x)(x - x_j) K_h(x - x_j)}{S_{n,0}(x) S_{n,2}(x) - S_{n,1}(x) S_{n,1}(x) + n^{-2}}$$

where $S_{n,l}(x) = h \sum_{j=1}^{n} (x - x_j)^l K_h(x - x_j)$ for $l = 0, 1, 2,$ $K(\cdot)$ is a kernel function, $h > 0$ is the bandwidth and $K_h(t) = K(t/h)/h$.

They then define their variance reduced estimate for $\theta(x)$ as

$$\tilde{\theta}(x) = \sum_{i=0}^{2} C_i(r) \hat{\theta}(\alpha_{x,i}) \tag{3}$$

where $\alpha_{x,0}, \alpha_{x,1}$, and $\alpha_{x,2}$ is an equally spaced grid of points and $C_0(r) = r(r-1)/2$, $C_1(r) = (1-r^2)$, and $C_2(r) = r(r+1)/2$. Furthermore, the bin width of this grid of points is chosen so that for $j = 0, 1, 2,$ $\alpha_{x,j} = x - (r + 1 - j)\delta h$ where $r \in (-1, 1) \backslash \{0\}$ and constant $\delta > 0$.

Cheng et al. (2006) show that under the following set of assumptions: (i) $K(\cdot)$ is a symmetric density function with compact support; (ii) mean function $\theta(x)$ has a bounded second derivative $\theta^{(2)}(x)$; (iii) error density $f(x)$ satisfies $f(x) > 0$ and $|f(x) - f(y)| \leq c|x - y|^{\alpha}$ for some $0 < \alpha < 1$; (iv) $\sigma^2(x)$ is continuous and bounded and (v) $h \to 0$ and $nh \to \infty$ as $n \to \infty$, that

$$E\{\tilde{\theta}(x)\} = \theta(x) + \frac{1}{2}\theta^{(2)}(x) s_2 h^2 + o\{h^4 + (nh)^{-\frac{1}{2}}\}, \tag{4}$$

$$Var\{\tilde{\theta}(x)\} = \frac{\sigma^2(x)}{nhf(x)}\{\nu_{02} - r^2(1 - r^2)A(\delta)\} + o\{h^4 + (nh)^{-\frac{1}{2}}\}, \tag{5}$$

where $\nu_{02} = \int K(s)^2 \, ds$,

$$A(\delta) = \frac{3}{2}A(0, \delta) - 2A(\frac{1}{2}, \delta) + \frac{1}{2}A(1, \delta), \tag{6}$$

$$A(a, \delta) = \int K(t - a\delta) K(t + a\delta) dt \tag{7}$$

and $A(\delta)$ has the following properties: (i) for any symmetric kernel function $K(\cdot)$, $A(\delta) \geq 0$ for any $\delta \geq 0$; and (ii) if $K(\cdot)$ has unique maximum and is concave, then $A(\delta)$ is increasing in $\delta \geq 0$. From these results it follows that one obtains maximal reduction in asymptotic variance by setting $r = \pm 1/\sqrt{2}$.

3 Equivalent Kernel Functions

The pth order spline estimator for $\theta(\cdot)$, which we shall denote as $\hat{\theta}_\lambda(\cdot)$, is defined as the minimizer of

$$\frac{1}{n}\sum_{j=1}^{n}\{Y_j-\theta(x_j)\}^2+\lambda\int_{[0,1]}\left\{\theta^{(p)}(x)\right\}^2 dx \tag{8}$$

over all functions $\theta(\cdot)$ in the pth order Sobolev space, $W_2^p[0,1]$, defined as $\{\theta(\cdot):\theta(\cdot),\ldots,\theta^{(p-1)}(\cdot)$ absolutely continuous and $\theta^{(p)}(\cdot)\in L_2[0,1]\}$. Here $\theta^{(k)}(\cdot)$ denotes the kth derivative of $\theta(\cdot)$ for $k=1,\ldots,p$ and in equation (8) λ is the smoothing parameter. The pth order smoothing spline estimator, of $\hat{\theta}_\lambda(x)$ is given as

$$\hat{\theta}_\lambda(x)=(I+n\lambda S)^{-1}Y \tag{9}$$

where S is the smoothing matrix; see Green and Silverman (1994).

This smoothing spline estimator can be defined as weighted average of the observations, much like the classical estimator discussed in Cheng et al. (2006), for some equivalent "spline kernel" function; see Silverman (1984). The spline kernel function in general does not have a closed form. Nychka (1995) describes how the kernel function for a pth order spline may be approximated by the Green's function, $G_\lambda(x,\tau)$, which solves a particular $2p$th order differential equation and presents an exponential envelope bound on the absolute difference between the spline kernel function and the Green's function which approximates it; see Theorem 2.1, Nychka (1995). This bound and other bounds given in Nychka (1995) hold exactly for finite sample sizes and are sufficiently accurate for our work.

4 Variance Reduced Spline Estimators for Univariate Regression Models

Our variance reduced estimate for $\theta(x)$ in model (1), $\tilde{\theta}_\lambda(x)$, is defined in a fashion similar to Cheng et al.'s so that substitution of spline estimates for local linear estimates is made, i.e.,

$$\tilde{\theta}_\lambda(x)=\sum_{i=0}^{2}C_i(r)\hat{\theta}_\lambda(\alpha_{x,i}). \tag{10}$$

and where root $\gamma=\lambda^{1/2p}$ plays a role analogous to that played by bandwidth parameter h in kernel estimator and for $j=0,1,2$, $\alpha_{x,j}=x-(r+1-j)\delta\gamma$ where $r\in(-1,1)\backslash\{0\}$ and $\delta>0$.

For the development of asymptotic bias and variance formulae for $\tilde{\theta}_\lambda(x)$ for we assume the following sets of conditions from Nychka (1995) (i) $D_n = \sup_{[0,1]} |F_n(x) - F(x)|$ converges almost surely to zero where $F_n(\cdot)$ and $F(\cdot)$ are the empirical and true cumulative distribution functions, respectively, for the X_i's; (ii) the exponential envelope condition (assumptions A and B); (iv) the Holder condition on $\theta(\cdot)$; Nychka's pth order bias and variance formulae in his equations 1.9 and 1.10, and (iv) Requisite conditions for the $p > 1$ version of Nychka's Theorem 5.1 to hold (see section 7 of Nychka (1995) for details). Under this combined set of assumptions we prove the following theorem.

Theorem 1 *Assume that $\hat{\theta}_\lambda$ is a pth order smoothing spline estimate and the observation points are not equally spaced. Suppose that $\theta(\cdot) \in C_{2p}[0,1]$ and that f has a uniformly continuous derivative. Then we have*

$$E\left\{\tilde{\theta}_\lambda(x)\right\} - \theta(x) = \frac{(-1)^{p-1}\lambda}{f(x)}\theta^{(2p)}(x) + o(\lambda) \tag{11}$$

$$Var\left\{\tilde{\theta}_\lambda(x)\right\} = \frac{\sigma^2 f^{1/2p-1}(x)}{n}(\nu_{02} - r^2(1-r^2)A\left\{\kappa^{-1} f(x)^{1/2p}\delta\right\} + o_p\{(n\lambda^{1/2p})^{-1}\} \tag{12}$$

uniformly for x in the interior of $[0,1]$ where $\kappa = \int_{[0,1]} f(x)^{1/2p}\,dx$, $\delta > 0$ is a constant, and equivalent kernel $K(\cdot)$ solves equation (10) from Lin et al. (2004).

Remark The precise definition of the interior of $[0,1]$ is given in Nychka (1995) is that for some small $\Delta > 0$, $x \in [\Delta, 1-\Delta]$.

Remark Equation (10) from Lin et al. (2004) is

$$(-1)^p K^{(2p)}(x) + K(x) = \Delta(x)$$

where $\Delta(x)$ is the Dirac delta function.

A comparison of our results with equations 1.9 and 1.10 of Nychka (1995) shows that the asymptotic bias of $\tilde{\theta}_\lambda(x)$ and $\hat{\theta}_\lambda(x)$ is identical, while the leading term in the asymptotic variance of $\tilde{\theta}_\lambda(x)$ is smaller than that of $\hat{\theta}_\lambda(x)$ (gotten by letting $r = 1$) for any $r \in (-1,1)\backslash\{0\}$ since $r^2(1-r^2)A\left\{\kappa^{-1} f(x)^{1/2p}\delta\right\} \geq 0$ over this range of values. Furthermore, one obtains maximal reduction in asymptotic variance by setting $r = \pm 1/\sqrt{2}$.

Proof of Theorem 1. First we consider the asymptotic bias of $\tilde{\theta}_\lambda(x)$. Using stochastic Taylor series expansions we have

$$
\begin{aligned}
E\{\tilde{\theta}_\lambda(x)\} &= \sum_{i=0}^{2} C_i(r) E\{\hat{\theta}_\lambda(\alpha_{x,i})\} \\
&= \sum_{i=0}^{2} C_i(r) E\{\hat{\theta}_\lambda(x) + \hat{\theta}_\lambda^{(1)}(x)(\alpha_{x,i} - x) + \hat{\theta}_\lambda^{(2)}(x)(\alpha_{x,i} - x)^2/2 + O_p(\gamma^3)\} \\
&= \sum_{i=0}^{2} C_i(r) \{E\{\hat{\theta}_\lambda(x)\} - E\{\hat{\theta}_\lambda^{(1)}(x)\}(r+1-i)\delta\gamma + E\{\hat{\theta}_\lambda^{(2)}(x)\}(r+1-i)^2\,(\delta\gamma)^2\,/2 + O(\gamma^3)\} \\
&= \sum_{i=0}^{2} C_i(r) \{\theta(x) + \frac{(-1)^{p-1}\,\lambda}{f(x)} \theta^{(2p)}(x) + o(\lambda)\} - E\{\hat{\theta}_\lambda^{(1)}(x)\}\,(\delta\gamma) \sum_{i=0}^{2} C_i(r)(r+1-i) \\
&+ \left[E\{\hat{\theta}_\lambda^{(2)}(x)\}(\delta\gamma)^2/2\right] \sum_{i=0}^{2} C_i(r)(r+1-i)^2 + O(\lambda^{3/2}) \\
&= \theta(x) + \frac{(-1)^{p-1}\,\lambda}{f(x)} \theta^{(2p)}(x) + o(\lambda)
\end{aligned}
$$

Therefore the asymptotic bias remains unchanged by the construction of our variance reduced spline estimator. Note that in above derivation we make use of the following identity:

$$
C_0(r)(-1-r)^j + C_1(r)(-r)^j + C_2(r)(1-r)^j = \delta_{0,j} \text{ for } j = 0, 1, 2. \tag{13}
$$

Next, we consider the asymptotic variance of $\tilde{\theta}_\lambda(x)$. To do this, however, we must first consider the covariance of spline estimates $\hat{\theta}_\lambda(u)$ and $\hat{\theta}_\lambda(v)$ at points $u, v \in [\alpha_{x,0}, \alpha_{x,2}]$;

$$
\begin{aligned}
Cov(\hat{\theta}_\lambda(u), \hat{\theta}_\lambda(v)) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w(u, x_i) w(v, x_j) Cov(Y_i, Y_j) \\
&= \frac{\sigma^2}{n^2} \sum_{i=1}^{n} w(u, x_i) w(v, x_i) \\
&= \frac{\sigma^2}{n} \int_{[0,1]} w(u, \tau) w(v, \tau) dF_n(\tau) \\
&= \frac{\sigma^2}{n} (B_1 + B_2 + B_3)
\end{aligned}
$$

where

$$
\begin{aligned}
B_1 &= \int_{[0,1]} G_\lambda(u, \tau) G_\lambda(v, \tau) dF(\tau), \\
B_2 &= \int_{[0,1]} (w(u, \tau) w(v, \tau) - G_\lambda(u, \tau) G_\lambda(v, \tau)) dF(\tau), \\
B_3 &= \int_{[0,1]} w(u, \tau) w(v, \tau) d(F - F_n)(\tau).
\end{aligned}
$$

For the B_2 term we have,

$$\begin{aligned}
B_2 &= \int_{[0,1]} (w(u,\tau)w(v,\tau) - G_\lambda(u,\tau)G_\lambda(v,\tau))f(\tau)d\tau \\
&= \int_{[0,1]} [w(v,\tau)(w(u,\tau) - G_\lambda(u,\tau)) + G_\lambda(u,\tau)(w(v,\tau) - G_\lambda(v,\tau))]f(\tau)d\tau \\
&\leq \int_{[0,1]} (|w(v,\tau)||w(u,\tau) - G_\lambda(u,\tau)| + |G_\lambda(u,\tau)||w(v,\tau) - G_\lambda(v,\tau)|)f(\tau)d\tau \\
&\leq sup_{[0,1]} f(\tau) \left\{ \int_{[0,1]} |w(v,\tau)||w(u,\tau) - G_\lambda(u,\tau)|d\tau + \int_{[0,1]} |G_\lambda(u,\tau)||w(v,\tau) - G_\lambda(v,\tau)|)d\tau \right\}
\end{aligned}$$

Using Nychka (1995)'s exponential envelope conditions we have for some positive constants $C, \alpha < \infty$

$$\begin{aligned}
\int_{[0,1]} |w(v,\tau)||w(u,\tau) - G_\lambda(u,\tau)|d\tau &\leq \sup_{[0,1]} |w(v,\tau)| \int_{[0,1]} |w(u,\tau) - G_\lambda(u,\tau)|d\tau \\
&= \frac{C\delta_n}{(1-\delta_n)\gamma} \int_{[0,1]} \exp\left(-\alpha|u-\tau|/\gamma\right) d\tau \\
&= \frac{C_2\delta_n}{(1-\delta_n)} = O(\delta_n) = O(D_n/\lambda^{1/2p}).
\end{aligned}$$

where positive constant $C_2 < \infty$.

Similarly,

$$\int_{[0,1]} |G_\lambda(u,\tau)||w(v,\tau) - G_\lambda(v,\tau)|)d\tau \ = O(D_n/\lambda^{1/2p})$$

which yields

$$B_2 = O(D_n/\lambda^{1/2p}).$$

For the B_3 term;

$$\begin{aligned}
B_3 &= \int_{[0,1]} w(u,\tau)w(v,\tau)d(F - F_n)(\tau) \\
&\leq \left| \int_{[0,1]} w(u,\tau)w(v,\tau)dF(\tau) - \int_{[0,1]} w(u,\tau)w(v,\tau)dF_n(\tau) \right|.
\end{aligned}$$

According to Nychka (1995) Lemma 4.1

$$\begin{aligned}
B_3 &\leq \sup_{[0,1]} |F_n - F| \int_{[0,1]} \left| \frac{dw(u,\tau)w(v,\tau)}{d\tau} \right| f(\tau) d\tau \\
&= D_n \int_{[0,1]} |w_\tau^{'}(u,\tau)w(v,\tau) + w(u,\tau)w_\tau^{'}(v,\tau)| f(\tau) d\tau \\
&= D_n \sup_{[0,1]} |f(\tau)| \int_{[0,1]} |w_\tau^{'}(u,\tau)w(v,\tau) + w(u,\tau)w_\tau^{'}(v,\tau)| d\tau \\
&\leq D_n \sup_{[0,1]} |f(\tau)| \left\{ \int_{[0,1]} |w_\tau^{'}(u,\tau)w(v,\tau)| d\tau + \int_{[0,1]} |w(u,\tau)w_\tau^{'}(v,\tau)| d\tau \right\}.
\end{aligned}$$

By exponential envelope conditions in Nychka (1995) we have for some positive constants $C, \alpha < \infty$

$$\begin{aligned}
\int_{[0,1]} |w_\tau^{'}(u,\tau)w(v,\tau)| d\tau \leq \sup_{[0,1]} |w(v,\tau)| &\int_{[0,1]} |w_\tau^{'}(u,\tau)| d\tau \\
&\frac{C}{(1-\delta_n)\gamma^2} \int_{[0,1]} \exp\left(-\alpha|u-\tau|/\gamma\right) d\tau \\
&\leq \frac{C_2}{(1-\delta_n)\gamma}.
\end{aligned}$$

for some positive finite constant C_2. Using similar arguments we find that

$$\int_{[0,1]} |w(u,\tau)w_\tau^{'}(v,\tau)| d\tau \leq \frac{C_3}{(1-\delta_n)\gamma}$$

for some positive finite constant C_3. Finally we have

$$B_3 \leq \frac{D_n\,(C_2 + C_3)}{(1-\delta_n)\gamma} = O(D_n/\lambda^{1/2p}).$$

For the B_1 term, we need a technique outlined in Nychka (1995, section 7) by which equivalent kernel function $G_\lambda(x,\tau)$ for a nonuniform design is approximated, as $\gamma \to 0$, by $G^U_{\lambda/\kappa^2}(\Gamma\,(t)\,, \Gamma\,(\tau))\zeta\,(\tau)\,/f\,(\tau)$ where $\zeta\,(t) = \Gamma\,(t)^{'} = (1/\kappa)\,f(t)^{1/2p}$. Letting $h_t(\tau) = G_\lambda(t,\tau)$ and $\bar{h}_t(\tau) = G^U_{\lambda/\kappa^2}(\Gamma\,(t)\,, \Gamma\,(\tau))\zeta\,(\tau)\,/f\,(\tau)$ consider,

$$\begin{aligned}
\left| \int_{[0,1]} h_u h_v f\,(\tau)\, d\tau - \int_{[0,1]} \bar{h}_u \bar{h}_v f\,(\tau)\, d\tau \right| &= \left| \int_{[0,1]} \left\{ h_u h_v - h_u \bar{h}_v + h_u \bar{h}_v - \bar{h}_u \bar{h}_v \right\} f\,(\tau)\, d\tau \right| \\
&= \left| \int_{[0,1]} \left\{ h_u \left[h_v - \bar{h}_v \right] + \bar{h}_v \left[h_u - \bar{h}_u \right] \right\} f\,(\tau)\, d\tau \right| \\
&\leq \int_{[0,1]} \left\{ |h_u| \left| h_v - \bar{h}_v \right| + \left| \bar{h}_v \right| \left| h_u - \bar{h}_u \right| \right\} f\,(\tau)\, d\tau
\end{aligned}$$

Consider the first term in the above integral, from Theorem 5.1 of Nychka (1995), there exists $C, \alpha < \infty$ for which

$$\begin{aligned}\int_{[0,1]} |h_u| \left|h_v - \bar{h}_v\right| f(\tau)\, d\tau &\le \sup_{[0,1]} \{|h_u|\, f(\tau)\} \int_{[0,1]} \left|h_v - \bar{h}_v\right| f(\tau)\, d\tau \\ &\le C \int_{[0,1]} \exp\left(-\alpha\, |v-\tau|\,/\gamma\right) d\tau \\ &= C_2 \gamma = o(\gamma)\,.\end{aligned}$$

The second term can be shown to be $o(\gamma)$ in a similar fashion where boundedness of $\left|\bar{h}_v\right|$ proven by with the following bound:

$$\left|\bar{h}_v\right| = \left|\bar{h}_v - h_v + h_v\right| \le \left|\bar{h}_v - h_v\right| + |h_v|\,.$$

Therefore, the leading term in the covariance will be determined by

$$\int_{[0,1]} \kappa^{-2} G^U_{\lambda/\kappa^2}(\Gamma(u), \Gamma(\tau)) G^U_{\lambda/\kappa^2}(\Gamma(v), \Gamma(\tau)) f(\tau)^{1/p-1}\, d\tau$$

as $\lambda \to 0$. Letting $\bar{y} = \Gamma(y)$ this integral may written as

$$\int_{[0,1]} \kappa G^U_{\lambda/\kappa^2}(\bar{u}, \bar{\tau}) G^U_{\lambda/\kappa^2}(\bar{v}, \bar{\tau}) f\left(\Gamma^{-1}(\bar{\tau})\right)^{1/2p-1} d\bar{\tau}.$$

For small λ, $f\left(\Gamma^{-1}(\bar{\tau})\right)^{1/2p-1}$ will be dominated by the product Green's functions in the integrand. This yields

$$Cov\left\{\hat{\theta}_\lambda(u), \hat{\theta}_\lambda(v)\right\} = \frac{\sigma^2 f(x)^{1/2p-1}}{n} \int_{[0,1]} G^U_{\lambda/\kappa^2}(\bar{u}, t) G^U_{\lambda/\kappa^2}(\bar{v}, t) dt + o(\gamma) \tag{14}$$

for u and v in the interior of $[0,1]$.

Next, consider

$$Var\{\tilde{\theta}_\lambda(x)\} = \sum_{i=0}^{2} C_i^2(r) Var\{\hat{\theta}_\lambda(\alpha_{x,i})\} + 2 \sum_{i=0}^{2} \sum_{j=i+1}^{2} C_i(r) C_j(r) Cov\{\hat{\theta}_\lambda(\alpha_{x,i}), \hat{\theta}_\lambda(\alpha_{x,j})\}. \tag{15}$$

Taylor series expansion yields

$$\Gamma(\alpha_{x,j}) = \Gamma(x) - \kappa^{-1} f^{1/2p}(x) \left[(r+1-j)\,\gamma\delta\right] + O(\gamma^2) \tag{16}$$

for $j = 1, 2, 3$. Expression (12) is may be obtained by plugging (16) and (14) into (15). Silverman (1984) shows that $G^U_\lambda(t, \tau)$ may be uniformly approximated by $K\left(|t-\tau|\,/\gamma\right)/\gamma$ for an equivalent kernel $K(\cdot)$; which yields, after a bit of algebra, expression (12). Furthermore, by Proposition 1 of Cheng et al. (2006), we have that $A\left\{\kappa^{-1} f(x)^{1/2p}\, \delta\right\} > 0$ and as such we obtain maximal reduction in variance for $r = \pm 1/\sqrt{2}$. $\square$

5 Variance Reduced Spline Estimators for Clustered\Longitudinal Models

In what follows we assume our clustered data comes from n clusters where the ith cluster consists of m_i observations. Furthermore, we assume that each observation of a response variable is observed with the value of exactly one covariate. To simplify our development, we assume that $m_i = m$ for all i. We do this without lack of generality, since our results will be applicable in data sets with non-constant cluster sizes. The nonparametric model for clustered data is

$$Y_{ij} = \theta(X_{ij}) + \epsilon_{ij} \tag{17}$$

where Y_{ij} and $X_{ij} \in [0,1]$ are the jth observed response and covariate values in the ith cluster for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Here, $\theta(x)$ is an unknown but smooth regression function, and errors $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im})^T$ are independent with mean zero and covariance matrix Σ.

Lin et al. (2004) define the (pth order) generalized least squares (GLS) smoothing spline estimator of $\theta(x)$ as the function which minimizes

$$\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \theta(X_i)\}^T W^{-1}\{Y_i - \theta(X_i)\} + \lambda \int_{[0,1]} \{\theta^{(p)}(x)\}^2 dx \tag{18}$$

over $\theta \in W_2^m[0,1]$, where $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im})$, $X_i = (X_{i1}, X_{i2}, \ldots, X_{im})$, W denotes a working covariance matrix, λ is a smoothing parameter, and $\theta^{(p)}(x)$ is the pth derivative of spline estimator for $\theta(x)$. The resulting GLS smoothing spline estimator, of $\hat{\theta}_\lambda(x)$ is given as

$$\hat{\theta}_\lambda(x) = (\tilde{W}^{-1} + n\lambda S)^{-1}\tilde{W}^{-1}Y \tag{19}$$

where $\tilde{W}^{-1} = diag\{W^{-1}, \cdots, W^{-1}\}$. For a fixed value of λ, $var\{\hat{\theta}_\lambda(x)\}$ achieves its minimal value for $W = \Sigma$ as shown in Welsh et al. (2002).

Our variance reduced GLS spline estimator for $\theta(x)$ is given as

$$\tilde{\theta}_\lambda(x) = \sum_{i=0}^{2} C_i(r)\hat{\theta}_\lambda(\alpha_{x,i}) \tag{20}$$

where $\alpha_{x,0}, \alpha_{x,1}, \alpha_{x,2}$, and $C_0(r)$, $C_1(r)$ and $C_2(r)$ are defined as in section 4.

In the development of asymptotic bias and variance formulae for $\tilde{\theta}_\lambda(x)$ for we assume (i) all sets of conditions given in section 4 and (ii) Green's function properties (A11) through (A14) from Lin et al. (2004).

Theorem 2 *Denote by $\tilde{\theta}_\lambda(X)$ the pth order variance reduced GLS spline estimator given in (19) using any given working covariance matrix W . Assume that the marginal densities $f_j(x)$ of the X_{ij} have uniformly continuous derivatives then we have for x in the interior of $[0,1]$;*

$$E\left\{\tilde{\theta}_\lambda(x)\right\} - \theta(x) = (-1)^{p-1}\left(\frac{\lambda}{\eta(x)}\right) b_s(x) + o(\lambda) \tag{21}$$

where $\eta(x) = \sum_{j=1}^m v^{jj} f_j(x)$, v^{jj} is the (j,j)th element of $V^{-1}$and $b_s(x)$ satisfies

$$\sum_{j=1}^m \sum_{k=1}^m v^{jk} E\{b_s(X_k)|X_j = x\} f_j(x) = \frac{1}{a_p}\eta(x)\theta^{(2p)}(x) \tag{22}$$

and where a_p is a constant and

$$Var\left\{\tilde{\theta}_\lambda(x)\right\} = \frac{1}{n}\left\{\frac{\lambda}{\eta(x)}\right\}^{-\frac{1}{2p}} \frac{\upsilon(x)}{\eta^2(x)}\left(\upsilon_{02} - r^2(1-r^2)A\left\{\kappa^{-1} f\,(x)^{1/2p}\,\delta\right\}\right) + o_p\{(n\lambda^{1/2p})^{-1}\}. \tag{23}$$

where equivalent kernel $K(\cdot)$ solves equation (10) in Lin et al. (2004), and $\upsilon(x) = \sum_{j=1}^m c_{jj} f_j(x)$ with c_{jj} being the (j,j)th element of $C = V^{-1}\Sigma V^{-1}$.

A comparison of these results with those from Proposition 4 of Lin et al. (2004) shows that the asymptotic bias of $\tilde{\theta}_\lambda(x)$ and $\hat{\theta}_\lambda(x)$ are the same, while the leading term in the asymptotic variance for $\tilde{\theta}_\lambda(x)$ is smaller than that of $\hat{\theta}_\lambda(x)$ (which corresponds to the $r = 1$ case) for any $r \in (-1,1)\backslash\{0\}$ since $r^2(1-r^2)A\left\{\kappa^{-1} f\,(x)^{1/2p}\,\delta\right\} \geq 0$ over this range of values. As in section 4, one obtains maximal reduction in asymptotic bias with $r = \pm 1/\sqrt{2}$.

Proposition 2 of Lin et al. (2004) describes the asymptotic equivalence between the GLS spline estimator with smoothing parameter λ and the seeming unrelated kernel estimator from Wang (2003) with effective bandwidth $h(x) = \{\lambda / \sum_{j=1}^m \sigma^{jj} f_j(x)\}^{1/2p}$. From this it follows that the variance reduced GLS spline estimator with the smallest variance is obtained by assuming that the working covariance matrix V equals the true covariance Σ, i.e., $V = \Sigma$. Its variance is

$$Var_{min}\left\{\tilde{\theta}_\lambda(x)\right\} = \frac{1}{n}\left\{\frac{\lambda}{\eta(x)}\right\}^{-\frac{1}{2p}} \frac{\tau(x)}{\eta^2(x)} \frac{\left[\upsilon_{02} - r^2(1-r^2)A\left\{\kappa^{-1} f\,(x)^{1/2p}\,\delta\right\}\right]}{\sum_{j=1}^m \sigma^{jj} f_j(x)} + o_p\{(n\lambda^{1/2p})^{-1}\}. \tag{24}$$

Proof of Theorem 2. It straightforward to show with stochastic Taylor series expansions that the leading term in the bias of the $\tilde{\theta}_\lambda(x)$ is the same as that of $\hat{\theta}_\lambda(x)$. Lin et al. (2004) present in part (iv) of proposition 4 the following asymptotic expansion for the GLS spline estimator:

$$\hat{\theta}_\lambda(x) - \theta\,(x) = D\,(x) + (-1)^{p-1}\,h^{2p}\,(x)\,b(x) + o_p\left[\{nh\,(x)\}^{-1/2} + h^{2p}\,(x)\right]$$

where

$$D(x) = \left(n \sum_{j=1}^{m} v^{jj} \right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{m} G_{\tilde{\lambda}}(x, X_{ik}) v^{jk} \{Y_{ik} - \theta(X_{ik})\},$$

$\tilde{\lambda} = \lambda / \sum_{j=1}^{m} v^{jj}$, effective bandwidth $h(x) = \left[\lambda / \left\{ \sum_{j=1}^{m} v^{jj} f_j(x) \right\} \right]^{1/(2p)}$, and $b(x)$ satisfies a condition stated in part (i) of their proposition 4.

To derive the asymptotic variance of $\tilde{\theta}_\lambda(x)$ we need to first study the asymptotic covariance of GLS spline estimates $\hat{\theta}_\lambda(u)$ and $\hat{\theta}_\lambda(v)$ at points $u, v \in [\alpha_{x,0}, \alpha_{x,2}]$. The $D(x)$ term may be written in matrix-vector form as

$$D(x) = \left(n \sum_{j=1}^{m} v^{jj} \right)^{-1} \sum_{i=1}^{n} G_{\tilde{\lambda}}(x, X_i)^T V^{-1} \Xi_i \tag{25}$$

where $G_{\tilde{\lambda}}(x, X_i) = \{G_{\tilde{\lambda}}(x, X_{i1}), G_{\tilde{\lambda}}(x, X_{i2}), \cdots, G_{\tilde{\lambda}}(x, X_{im})\}^T$ and $\Xi_i = Y_i - \theta(X_i)$. It follows that

$$\begin{aligned} Cov\left\{ \hat{\theta}_\lambda(u), \hat{\theta}_\lambda(v) \right\} &= Cov[D(u), D(v)] + o(1) \\ &= \left(n \sum_{j=1}^{m} v^{jj} \right)^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov[G_{\tilde{\lambda}}(u, X_i)^T V^{-1} \Xi_i, G_{\tilde{\lambda}}(v, X_j)^T V^{-1} \Xi_j] + o(1) \\ &= T_1(u, v) + T_2(u, v) + o(1) \end{aligned}$$

where

$$T_1(u, v) = \left(n \sum_{j=1}^{m} v^{jj} \right)^{-2} \sum_{i=1}^{n} \sum_{j=1}^{m} G_{\tilde{\lambda}}(u, X_{ij}) c_{jj} G_{\tilde{\lambda}}(v, X_{ij}),$$

$$T_2(u, v) = \left(n \sum_{j=1}^{m} v^{jj} \right)^{-2} \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k \neq j} G_{\tilde{\lambda}}(u, X_{ij}) c_{jk} G_{\tilde{\lambda}}(v, X_{ik}).$$

Consider the T_2 term first

$$\begin{aligned} T_2(u, v) &= \left(\sum_{j=1}^{m} v^{jj} \right)^{-2} \frac{1}{n} \sum_{j=1}^{m} \sum_{k \neq j} c_{jk} \frac{1}{n} \sum_{i=1}^{n} G_{\tilde{\lambda}}(u, X_{ij}) G_{\tilde{\lambda}}(v, X_{ik}) \\ &= \left(\sum_{j=1}^{m} v^{jj} \right)^{-2} \frac{1}{n} \sum_{j=1}^{m} \sum_{k \neq j} c_{jk} \int_{[0,1]^2} G_{\tilde{\lambda}}(u, X_{ij}) G_{\tilde{\lambda}}(v, X_{ik}) dF_n(\tau, s) \\ &= \left(\sum_{j=1}^{m} v^{jj} \right)^{-2} \frac{1}{n} \sum_{j=1}^{m} \sum_{j=1}^{m} \sum_{k \neq j} c_{jk} \int_{[0,1]^2} G_{\tilde{\lambda}}(u, \tau) G_{\tilde{\lambda}}(v, s) f_{jk}(\tau, s) d\tau ds + o_p(n^{-1}). \end{aligned}$$

Taylor series expansion yields, for some v^* between u and v,

$$G_{\tilde{\lambda}}(v,s) = G_{\tilde{\lambda}}(u,s) + G'_{\tilde{\lambda}}(v^*,s)\,(u-v)\,.$$

As such the above integral may be written as the sum of two integrals:

$$\int_{[0,1]^2} G_{\tilde{\lambda}}(u,\tau)G_{\tilde{\lambda}}(u,s)f_{jk}(\tau,s)d\tau ds = \frac{f_{ij}\,(u,u)}{f\,(u)^2} + o\,(1)$$

by equality $A14$ in Lin et al. (2004); and

$$(u-v)\int_{[0,1]^2} G_{\tilde{\lambda}}(u,\tau)G'_{\tilde{\lambda}}(v^*,s)f_{jk}(\tau,s)d\tau ds = O\,(\gamma)$$

since $(u-v) = O\,(\gamma)$ and the boundedness of $G_{\tilde{\lambda}}(u,\tau)G'_{\tilde{\lambda}}(v^*,s)$ is guaranteed by the exponential envelope conditions of Nychka (1995). As result $T_2\,(u,v) = O_p\left(n^{-1}\right)$

For the T_1 term we have

$$\begin{aligned}
T_1\,(u,v) &= \left(\sum_{j=1}^m \upsilon^{jj}\right)^{-2} \frac{1}{n}\sum_{j=1}^m c_{jj}\int_{[0,1]} G_{\tilde{\lambda}}(u,\tau)G_{\tilde{\lambda}}(v,\tau)f_j(\tau)d\tau + o_p(1) \\
&= \left(\sum_{j=1}^m \upsilon^{jj}\right)^{-1} \frac{1}{n}\int_{[0,1]} \frac{\upsilon(\tau)}{\eta(\tau)} G_{\tilde{\lambda}}(u,\tau)G_{\tilde{\lambda}}(v,\tau)f(\tau)d\tau + o_p(1) \\
&= \left(\sum_{j=1}^m \upsilon^{jj}\right)^{-1} \frac{1}{n}\int_{[0,1]} \frac{\upsilon(\tau)}{\kappa^2\eta(\tau)} G^U_{\tilde{\lambda}/\kappa^2}(\Gamma\,(u)\,,\Gamma\,(\tau))G^U_{\tilde{\lambda}/\kappa^2}(\Gamma\,(v)\,,\Gamma\,(\tau))f\,(\tau)^{1/p-1}\,d\tau + o_p(1)
\end{aligned}$$

as $\lambda \to 0$. Letting $\bar{u} = \Gamma\,(u)\,,\ \bar{v} = \Gamma\,(v)$ and $\bar{\tau} = \Gamma\,(\tau)$, $T_{1,n}$ may written as

$$\left(\sum_{j=1}^m \upsilon^{jj}\right)^{-1} \frac{1}{n}\int_{[0,1]} \frac{\upsilon(\Gamma^{-1}\,(\bar{\tau}))}{\kappa\eta(\Gamma^{-1}\,(\bar{\tau}))} G^U_{\tilde{\lambda}/\kappa^2}(\bar{u},\bar{\tau})G^U_{\tilde{\lambda}/\kappa^2}(\bar{v},\bar{\tau})f\left(\Gamma^{-1}\,(\bar{\tau})\right)^{1/2p-1}\,d\bar{\tau}.$$

As $\gamma \to 0$ $\frac{\upsilon(\Gamma^{-1}(\bar{\tau}))}{\eta(\Gamma^{-1}(\bar{\tau}))} f^{1/2p-1}(\Gamma^{-1}\,(\bar{\tau}))$ will be dominated by the product Green's functions in the integrand. It follows that

$$\begin{aligned}
Cov\left\{\hat{\theta}_\lambda(u),\hat{\theta}_\lambda(v)\right\} &= \left(\sum_{j=1}^m \upsilon^{jj}\right)^{-1} \frac{\upsilon(x)f\,(x)^{1/2p-1}}{n\kappa\eta(x)}\int_{[0,1]} G^U_{\tilde{\lambda}/\kappa^2}(\bar{u},t)G^U_{\tilde{\lambda}/\kappa^2}(\bar{v},t)dt + o_p(1) \\
&= \frac{1}{n}\left\{\frac{\lambda}{\eta(x)}\right\}^{-\frac{1}{2p}} \frac{\upsilon(x)}{\eta^2(x)}\int_{[0,1]} G^U_{\tilde{\lambda}/\kappa^2}(\bar{u},t)G^U_{\tilde{\lambda}/\kappa^2}(\bar{v},t)dt + o_p(1) \qquad (26)
\end{aligned}$$

uniformly for u and v in the interior of $[0,1]$.

For the variance of $\tilde{\theta}_\lambda(x)$, we have

$$Var\left\{\tilde{\theta}_\lambda(x)\right\} = \sum_{i=0}^{2} C_i^2(r) Var\left\{\hat{\theta}_\lambda(\alpha_{x,i})\right\} + 2\sum_{i=0}^{2}\sum_{j=i+1}^{2} C_i(r)C_j(r)Cov\left\{\hat{\theta}_\lambda(u), \hat{\theta}_\lambda(v)\right\}. \qquad (27)$$

Expression (23) is now easily derived.

6 Simulation Studies

In this section we discuss the results of simulation studies, performed in FORTRAN, to evaluate the finite sample performance of our proposed variance reduced (cubic) smoothing spline estimators. Here we use an averaged estimator $\tilde{\theta}_\lambda^{avg}(x) = \left\{\tilde{\theta}_\lambda^{-1/\sqrt{2}}(x) + \tilde{\theta}_\lambda^{1/\sqrt{2}}(x)\right\}/2$ where $\tilde{\theta}_\lambda^{\pm 1/\sqrt{2}}(x)$ denotes the variance reduced estimator with r set equal to $\pm 1/\sqrt{2}$. Cheng et al. (2006) show that the averaged variance reduced local linear estimator can be expected to have smaller bias, than their $r = -1/\sqrt{2}$ or $r = 1/\sqrt{2}$ estimators, under fairly general conditions. This result is easily shown to also hold for our variance reduced spline estimators.

We simulated data from several univariate regression and clustered/longitudinal models. For each dataset, the smoothing parameter was chosen by GCV (Generalized Cross Validation), parameter δ was taken to be 0.6, 1.0 and 1.4, to test the sensitivity to the choice of δ, the mean integral squared errors (MISE's) were approximated by the average, of approximate integrals of the squared error, over the 200 replicates.

The three univariate regression models we considered were: (i) $0.3\exp\{-16(x-0.25)^2\} + 0.7\exp\{-64(x-0.75)^2\}$; (ii) $2-5x+\exp\{-400(x-0.5)^2\}$ and (iii) $\sin(5\pi x)$. In each case X-values were generated according to a uniform $(0,1)$ design, random errors was taken to be independent $N(0,1)$, and the sample size was taken to be 100. The ratio of MISE's for the classical smoothing spline (9) and variance reduced smoothing spline (10) were computed. Results are summarized in the Table 1. From our simulations, we see that the variance reduced spline estimator has smaller variance than that of the classical estimator and this variance is not sensitive to the choice of δ.

Four clustered/longitudinal models were considered: (i) $\sin(2w)$ where $w = 4x-2$; (ii) $\sqrt{x(1-x)}\sin\{2\pi(1+2^{-3/5})/(x+2^{-3/5})\}$; (iii) $\sqrt{x(1-x)}\sin\{2\pi(1+2^{-7/5})/(x+2^{-7/5})\}$ and (iv) $\sin(8x-4)+2\exp\{-256(x-0.5)^2\}$. The number of subjects was taken to be $n=50$ or $n=100$, the cluster size was set at $m=3$ and covariate X_{ij} was generated independently from the uniform

	$\delta = 0.6$	$\delta = 1.0$	$\delta = 1.4$
Model 1	1.0619	1.1480	1.1663
Model 2	1.0268	1.1124	1.2332
Model 3	1.0139	1.0629	1.1394

Table 1: Ratio of MISE's for the Univariate Regression Models.

$(0, 1)$ distribution. We assumed that the marginal variances of the $Y_{ij}'s$ were 1, and considered three hypothetical covariance structures: (i) Exchangeable with common correlation of 0.6; (ii) Autoregressive with correlation 0.6 and (iii) Unstructured with $\rho_{12} = \rho_{23} = 0.8$ and $\rho_{13} = 0.5$. For each configuration, we generated 200 simulated datasets and estimated $\theta(x)$ using classical GLS smoothing spline estimator (19) and variance reduced GLS smoothing spline estimator (20) with $W = \Sigma$. The ratio of MISE's for the classical and variance reduced estimators were computed. Results are summarize in Tables 2, 3 and 4.

	Exchang. $n = 50$	Exchang. $n = 100$	$AR(1)$ $n = 50$	$AR(1)$ $n = 100$	Unstruct. $n = 50$	Unstruct. $n = 100$
Model 1	1.1260	1.2291	1.1213	1.2159	1.1559	1.3059
Model 2	1.1613	1.2381	1.1489	1.2172	1.2321	1.3252
Model 3	1.0339	1.0134	1.0168	0.9825	1.1706	1.2201
Model 4	1.0132	1.0759	1.0131	1.0726	1.0138	1.0773

Table 2: Ratio of MISE's for the Longitudinal Models, $\delta = 0.6$

	Exchang. $n = 50$	Exchang. $n = 100$	$AR(1)$ $n = 50$	$AR(1)$ $n = 100$	Unstruct. $n = 50$	Unstruct. $n = 100$
Model 1	1.4157	1.5682	1.4015	1.5356	1.4682	1.6613
Model 2	1.3925	1.5130	1.3677	1.4620	1.5018	1.5875
Model 3	0.9394	0.8854	0.9422	0.8896	1.2089	1.1313
Model 4	1.0605	1.2459	1.0603	1.2370	1.0616	1.2447

Table 3: Ratio of MISE's for the Longitudinal Models, $\delta = 1.0$

From these simulations, we see that the variance reduced spline estimator generally has smaller

	Exchang. $n=50$	Exchang. $n=100$	$AR(1)$ $n=50$	$AR(1)$ $n=100$	Unstruct. $n=50$	Unstruct. $n=100$
Model 1	1.7124	1.8287	1.6883	1.7828	1.7532	1.9095
Model 2	1.5152	1.6158	1.4845	1.5481	1.6559	1.7167
Model 3	0.9394	0.8859	0.9422	0.8897	1.0377	1.0858
Model 4	1.1351	1.4362	1.1341	1.4221	1.1359	1.4279

Table 4: Ratio of MISE's for the Longitudinal Models, $\delta = 1.4$

variance than that of the classical estimator and this variance is sensitive to the choice of δ particularly when the regression function is highly oscillatory. In such situations, it is best to use smaller values of δ. This phenomenon has also been noted in section 4.2 of Cheng et al. (2006) were they suggest that $\delta = 1$ is a good default value for variance reduced local linear estimators in univariate regression problems. In contrast, our results suggest that $\delta = 0.6$ is a good default value for variance reduced smoothing spline estimators in clustered or longitudinal data problems.

References

[1] Agarwal, G.G. and Studden, W.J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines, Ann. Statist., 8, 1307-1325.

[2] Brumback, B. and Rice, J.A. (1998). Smoothing spline regression models for the analysis of nested and crossed samples of curves (with Discussion), J. Am. Statist. Assoc., 93, 961-1006.

[3] Cheng, M., Peng, L. and Wu, J. (2006). Reducing variance in univariate smoothing, Ann. Statist., accepted.

[4] Eubank, R.L. (1999). Nonparametric Regression and Spline Smoothing. New York: Marcel Dekker.

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

[6] Lin, X. and Carroll, R. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error, J. Am. Statist. Assoc., 95, 520-34.

[7] Lin, X., Wang, N., Welsh, A. and Carroll, R. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data, Biometrika, 91, 177-193.

[8] Lin, X. and Zhang, D. (1999). Inference in generalised additive mixed models using smoothing splines, J. R. Statist. Soc. B, 61, 381-400.

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

[10] Silverman, B. W. (1984). Spline smoothing: The equivalent variable kernel method, Ann. Statist., 12, 898-916.

[11] Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with Discussion), Appl. Statist., 48, 269-311.

[12] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation, Biometrika, 90, 43-52.

[13] Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data, J. Am. Statist. Assoc., 93, 710-9.