Finite Sample Performance of Backfitting, Marginal Integration and Two Stage Estimators under Common Bandwidth Selection Criterion

Carlos Martins-Filho
Department of Economics
Oregon State University
Ballard Hall 303
Corvallis, OR 97331-3612 USA
email: carlos.martins@orst.edu
Voice: + 1 541 737 1476
Fax: + 1 541 737 5917

and

Ke Yang
Department of Economics
Oregon State University
Ballard Hall 303
Corvallis, OR 97331-3612 USA
email: yangk@onid.orst.edu
Voice: + 1 541 737 2321
Fax: + 1 541 737 5917

September, 2004

Abstract. In this paper we investigate the finite sample performance of four estimators that are currently available for additive nonparametric regression models - the backfitting B-estimator, the marginal integration M-estimator and two versions of a two stage 2S-estimator, the first proposed by Kim, Linton and Hengartner (1999) and the second which we propose in this paper. We derive the conditional bias and variance of the 2S estimators and suggest a procedure to obtain optimal bandwidths that minimize an asymptotic approximation of the mean average squared error (AMASE). We are particularly concerned with the performance of these estimators when bandwidth selection is done based on data driven methods, since in this case even the asymptotic properties of the estimators are currently unavailable. We compare the estimators' performance based on various bandwidth selection procedures currently available in the literature as well as with the procedures proposed herein *via* a Monte Carlo study.

Keywords and Phrases. additive nonparametric regression; local linear estimation; backfitting estimation; marginal integration; two-stage additive estimation..

JEL classifications. C14; C15.

1 Introduction

Let (Y, X) be a random vector with joint density f, $Y \in \Re$ and $X \in \Re^d$, d a finite positive integer, and suppose there is a random sample of size n, denoted by $\{(y_i, x_i)\}_{i=1}^n$ available for a researcher whose objective is the estimation of the nonparametric regression $E(Y|X = x) = m(x)$, where $x' = (x_1, ..., x_d)$. Stone(1980) showed that the best rate obtainable in the estimation of $m(x)$ is $n^{s/(2s+d)}$ where s is the degree of smoothness of the function m. The fact that the optimal rate depends inversely on d is known as the *curse of dimensionality* in nonparametric regression estimation. However, as shown by Stone(1985), if an additivity constraint is imposed on $m(x)$, i.e.,

$$E(Y|X = x) = \alpha + \sum_{\delta=1}^{d} m_\delta(x_\delta) \tag{1}$$

with $E(m_\delta(x_\delta)) = 0$, each of the directional regressions $m_\delta(\cdot)$ can be estimated at an optimal rate $n^{s/(2s+1)}$ which does not depend on d. This circumvention of the curse of dimensionality has contributed to the popularity of additive nonparametric regression models in both the theoretical and applied literatures.[1]

Currently three estimators have emerged as viable alternatives for the regression model in (1): the *Backfitting* estimator (B-estimator), proposed by Buja et al.(1989); the *Marginal Integration* estimator (M-estimator), proposed by Newey (1994), Tjøstheim and Auestad (1994) and Linton and Nielsen(1995); and a two stage estimator (2S-estimator), proposed by Linton (1997) and Kim et al.(1999). The estimators differ in how the additivity constraint is used to produce final estimators of m_δ, but they all share the use of kernel based univariate nonparametric estimation methods, such as Nadaraya-Watson or local polynomial fitting in intermediate stages.[2]

Asymptotic properties of the B-estimator have been studied by Buja et al. (1989), Opsomer and Ruppert (1997) and Opsomer (2000). Under some conditions, asymptotic approximations for the conditional bias and variance of the B-estimator under local polynomial fitting have been obtained but unfortunately asymptotic distributions are still unavailable, rendering impossible the construction of pointwise confidence intervals

[1]See, *inter alia* Hastie and Tibshirani(1990) and Pagan and Ullah(1999).

[2]Alternative nonparametric smoothing methods could potentially be used, but such methods have not received the attention given to kernel based methods. See, e.g., Wahba(1990).

even asymptotically. The asymptotic properties of the M-estimator were studied by Linton and Nielsen(1995) when $d = 2$ and by Linton and Härdle(1996) for $d > 2$. They show that if regressor specific bandwidths used in estimation converge to zero at a suitable rate the M-estimator for m_δ converges to a normal distribution at the optimal rate $n^{s/(2s+1)}$. The 2S-estimator for m_δ proposed by Kim et al.(1999) is also $n^{s/(2s+1)}$ asymptotically normal under some conditions on the rate of convergence of the bandwidths.

Unfortunately, all asymptotic properties obtained for these estimators rely on the bandwidths being *non-stochastic* . For all estimators asymptotic properties and distributions are still unknown when the bandwidths are stochastic. In practice bandwidths are chosen by data driven methods, such as *cross validation,* and various *plug in* methods such as those proposed by Silverman(1986) and Opsomer and Ruppert(1998). The difficulty in obtaining previous asymptotic normality results in this setting is two-fold. First, data driven bandwidths are stochastic sequences that may interact in a pernicious way with regressors and regressand creating added difficulty in establishing asymptotic normality. Second, they are chosen to minimize some criterion function (loss or risk). For the most widely used criterion functions, the resulting optimal sequence of bandwidths may not converge to zero at the specified rate necessary to obtain asymptotic normality of the M or 2S estimators. In summary, to render the asymptotic distributional results cited above useful it is necessary to adapt them to the case where bandwidths are a data dependent stochastic sequence. Thus, as a practical matter, little is known about *both* asymptotic and finite sample distributional properties of the above mentioned estimators.

As an alternative to asymptotics, experimental evidence on the finite sample performance of the estimators based on data driven bandwidth selection methods can be provided. Hence, one of our objectives in this paper is to conduct a Monte Carlo investigation to reveal some of the finite sample characteristics of the distributions of B, M and 2S estimators for an additive nonparametric regression model. To provide a fair assessment of the performance of the estimators under study we present a unified method for bandwidth selection. For B-estimation, Opsomer and Ruppert (1998) proposed an automated direct plug-in (DPI) bandwidth selection method based on the minimization of an asymptotic approximation of the estimator's mean average squared error. Their DPI bandwidth outperformed a cross validation bandwidth in finite sample simulations. In this paper we propose a novel DPI method for M and 2S estimation that is inspired by the DPI method of Opsomer and Ruppert. To this end we obtain conditional squared errors for the M and 2S estimators by deriving their conditional bias and variance. Hence, our simulations compare the performance of the B, M and 2S estimators under the DPI methods proposed in this paper but also under the bandwidth selection rules proposed in the previous literature, including those suggested by Linton and Nielsen(1995) and Kim et al.(1999). We are also particularly interested in the impact of different degrees of regressor dependency on the estimation of m_δ.[3] Ultimately, our objective is to provide applied researchers with information that allows for a more accurate comparison of these three competing estimation alternatives in a finite sample setting. Besides this introduction the paper has 5 more sections. Section 2 describes in a unified notation the estimators under study and their properties. Section 3 provides asymptotic conditional bias and variance for the M and $2S$ estimators, a plug-in method to select their bandwidths and a description of the bandwidth selection method for the B-estimator. Section 4 presents the data generation processes used in the simulation and Section 5 discusses the results and makes some recommendations. Section 6 provides a brief conclusion with some directives for future research.

2 Estimators under Study

Since the data generating processes (DGP) used in our simulations (section 4) are for a bivariate regression model, in this section we provide a unified notation for the estimators under study for the case where $d = 2$. We assume that a random sample of size n - $\{y_t, x_t, z_t\}_{t=1}^n$ - is available on the random vector (Y, X, Z) with joint density $f(y, x, z)$ such that $E(Y|X = x, Z = z) = m(x, z) = \alpha + m_1(x) + m_2(z)$, $V(Y|X = x, Z = z) = \sigma^2$ and $E(m_1(X)) = E(m_2(Z)) = 0$. Here, α and σ^2 are unknown parameters, $m_1(\cdot)$ and $m_2(\cdot)$ are real valued functions with regularity properties that will be made explicit in the next section. We assume that the primary interest is on the estimation of the unknown parameter α and the functions m_1 and m_2. We will denote the marginal densities of X and Z by $f_X(x)$ and $f_Z(z)$ respectively, and the joint marginal density of (X, Z) by $f_{XZ}(x, z)$. For expositional purposes we define the following vectors: $\vec{y} =$

[3] Previous experimental work on the performance of B and M estimators under fairly simple bandwidth selection rules was provided by Sperlich et al.(2000).

$(y_1, ..., y_n)'$, $\vec{x} = (x_1, ..., x_n)'$, $\vec{z} = (z_1, ..., z_n)'$, $\vec{m}_1(\vec{x}) = (m_1(x_1), ...m_1(x_n))'$, $\vec{m}_2(\vec{z}) = (m_2(z_1), ...m_2(z_n))'$, $e_t^k = (0, ..., 1, ..., 0)'$ is a vector of length k, with the t^{th} element equal to 1, and for any constant c, $\vec{c}_k = (c, c, ..., c)'$ is a vector of length k. We denote by $K_d : \Re^d \rightarrow \Re$ a d-variate symmetric kernel function with $d = 1, 2$ and by h_1 and h_2 the bandwidths associated with the estimation of m_1 and m_2, respectively. Furthermore, since we focus on local linear estimation in defining B,M and $2S$ estimation we define the following weighting matrices,

$$S_1 = \begin{pmatrix} s_1(x_1) \\ \vdots \\ s_1(x_n) \end{pmatrix}, \text{ and } S_2 = \begin{pmatrix} s_2(z_1) \\ \vdots \\ s_2(z_n) \end{pmatrix},$$

where $s_1(x), s_2(z) : \Re \rightarrow \Re^n$ such that

$$s_1(x) = e_1^{2'}(R_X(x)'W_X(x)R_X(x))^{-1}R_X(x)'W_X(x) \text{ and}$$

$$s_2(z) = e_1^{2'}(R_Z(z)'W_Z(z)R_Z(z))^{-1}R_Z(z)'W_Z(z),$$

$W_X(x) = diag\left\{K_1\left(\frac{x_t - x}{h_1}\right)\right\}_{t=1}^n$, $W_Z(z) = diag\left\{K_1\left(\frac{z_t - z}{h_2}\right)\right\}_{t=1}^n$, $R_X(x) = (\vec{1}_n, \vec{x} - \vec{1}_n x)$ and $R_Z(z) = (\vec{1}_n, \vec{z} - \vec{1}_n z)$.

2.1 Backfitting Estimator

The B-estimator is based on the fact that under suitable conditions[4]

$$E(Y - \alpha - m_2(Z)|X) = m_1(X) \text{ and } E(Y - \alpha - m_1(X)|Z) = m_2(Z). \tag{2}$$

Given that $E(Y) = \alpha$ and the equations in (2), a method of moments estimation suggests $\alpha^B = n^{-1}1_n'\vec{y} = \bar{y}$ as an estimator for α and $\vec{m}_1^B(\vec{x})$, $\vec{m}_2^B(\vec{z})$ as estimators for $\vec{m}_1(\vec{x})$ and $\vec{m}_2(\vec{z})$, where

$$\begin{pmatrix} I_n & S_1^* \\ S_2^* & I_n \end{pmatrix} \begin{pmatrix} \vec{m}_1^B(\vec{x}) \\ \vec{m}_2^B(\vec{z}) \end{pmatrix} = \begin{pmatrix} S_1^* \\ S_2^* \end{pmatrix} \vec{y}, \tag{3}$$

I_n is an identity matrix of dimension n, $S_d^* = \left(I_n - \frac{1}{n}\vec{1}_n\vec{1}_n'\right) S_d = D_n S_d$ for $d = 1, 2$. As discussed by Opsomer and Ruppert(1997), for $d = 2$ explicit expressions for $\vec{m}_1^B(\vec{x})$, $\vec{m}_2^B(\vec{z})$ exist and can be written as

$$\vec{m}_1^B(\vec{x}) = \left(I_n - (I_n - S_1^* S_2^*)^{-1}(I_n - S_1^*)\right)\vec{y} \tag{4}$$

$$\vec{m}_2^B(\vec{z}) = \left(I_n - (I_n - S_2^* S_1^*)^{-1}(I_n - S_2^*)\right)\vec{y} \tag{5}$$

[4]See Buja et al.(1989).

provide $||S_1^* S_2^*|| < 1$ for some matrix norm $|| \cdot ||$.

2.2 Marginal Integration Estimator

The M-estimator for a bivariate model was proposed by Linton and Nielsen(1995) and is based on the fact that for any function $Q(x) : \Re \to \Re$ such that $\int dQ(x) = 1$ we have

$$\int m(x,z)dQ(z) = \int (\alpha + m_1(x) + m_2(z))dQ(z) = c_1 + m_1(x) \tag{6}$$

$$\int m(x,z)dQ(x) = \int (\alpha + m_1(x) + m_2(z))dQ(x) = c_2 + m_2(z) \tag{7}$$

where $c_1 = \alpha + \int m_2(z)dQ(z)$ and $c_2 = \alpha + \int m_1(x)dQ(x)$ are constants with respect to z and x, respectively. Hence, taking $Q(\cdot)$ to be the empirical distribution function, the M-estimator is defined by first obtaining an estimator for $m(x,z)$, in this case a bivariate local linear estimator evaluated at $X = x, Z = z$ defined as $\hat{m}(x,z;h_1,h_2) = e_1^{3'} \left(X(x,z)'W(x,z)X(x,z)\right)^{-1} X(x,z)'W(x,z)\vec{y}$, where $X(x,z) = \left(\vec{1}_n, \vec{x} - \vec{1}_n x, \vec{z} - \vec{1}_n z\right)$ and $W(x) = diag\left\{K_2\left(\frac{1}{h_1}(x_t - x), \frac{1}{h_2}(z_t - z)\right)\right\}_{t=1}^{n}$. We then define the matrix

$$\hat{m}(\vec{x},\vec{z}) = \begin{pmatrix} \hat{m}(x_1,z_1) & \hat{m}(x_1,z_2) & \cdots & \hat{m}(x_1,z_n) \\ \hat{m}(x_2,z_1) & \hat{m}(x_2,z_2) & \cdots & \hat{m}(x_2,z_n) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{m}(x_n,z_1) & \hat{m}(x_n,z_2) & \cdots & \hat{m}(x_n,z_n) \end{pmatrix}$$

and the M-estimators are given by $\vec{m}_1^M(\vec{x}) = \frac{1}{n}\hat{m}(\vec{x},\vec{z})\vec{1}_n - \bar{y}$ and $\vec{m}_2^M(\vec{z}) = \frac{1}{n}\hat{m}(x,z)'\vec{1}_n - \bar{y}$ together with $\alpha^M = \bar{y}$.

2.3 Two-Stage Estimators

The $2S$-estimators we consider are inspired by Kim et al.(1999). They are based on the fact that for a function $w(X,Z) : \Re^2 \to \Re$ where $w(X,Z) = \frac{f_X(X)f_Z(Z)}{f_{XZ}(X,Z)}$ and the additive structure in (1), we have that when $d = 2$

$$E(Yw(X,Z) - \alpha w(X,Z)|X = x) = m_1(x) \tag{8}$$

and

$$E(Yw(X,Z) - \alpha w(X,Z)|Z = z) = m_2(z). \tag{9}$$

The conditional moments described in (8) and (9) suggest two feasible regression procedures for the estimation of m_1 and m_2. The first, proposed by Kim et al.(1999), uses $E(w(X,Z)|X=x) = E(w(X,Z)|Z=z) = 1$ to write (8) and (9) as,

$$E(Yw(X,Z)|X=x) = \alpha + m_1(x) \equiv \gamma_1(x) \tag{10}$$

and

$$E(Yw(X,Z)|Z=z) = \alpha + m_2(z) \equiv \gamma_2(z). \tag{11}$$

They propose a two-step estimation procedure where first $\gamma_1(x)$ and $\gamma_2(z)$ are estimated based on nonparametric regressions of $y_t\hat{w}(x_t,z_t)$ on x_t and z_t respectively, with $\hat{w}(x_t,z_t) = \frac{\hat{f}_X(x_t)\hat{f}_Z(z_t)}{\hat{f}_{XZ}(x_t,z_t)}$, $\hat{f}_X(x) = \frac{1}{ng_1}\sum_{k=1}^n K_1\left(\frac{x_k-x}{g_1}\right)$, $\hat{f}_Z(z) = \frac{1}{ng_2}\sum_{k=1}^n K_1\left(\frac{z_k-x}{g_2}\right)$ and $\hat{f}_{XZ}(x,z) = \frac{1}{ng_1g_2}\sum_{k=1}^n K_1\left(\frac{x_k-x}{g_1}\right)K_1\left(\frac{z_k-z}{g_2}\right)$, where g_1 and g_2 are bandwidths associated with X and Z respectively. Using an internalized Nadaraya-Watson estimator[5] with bandwidths g_1 and g_2 in this first step we obtain

$$\vec{\gamma}_1^E(\vec{x}) = \begin{pmatrix} \gamma_1^E(x_1) \\ \gamma_1^E(x_2) \\ \vdots \\ \gamma_1^E(x_n) \end{pmatrix} \text{ and } \vec{\gamma}_2^E(\vec{z}) = \begin{pmatrix} \gamma_2^E(z_1) \\ \gamma_2^E(z_2) \\ \vdots \\ \gamma_2^E(z_n) \end{pmatrix}$$

where $\gamma_1^E(x_i) = \frac{1}{ng_1}\sum_{j=1}^n K_1\left(\frac{x_j-x_i}{g_1}\right)\frac{\hat{f}_Z(z_j)}{\hat{f}_{XZ}(x_j,z_j)}y_j$, $\gamma_2^E(z_i) = \frac{1}{ng_2}\sum_{j=1}^n K_1\left(\frac{z_j-z_i}{g_2}\right)\frac{\hat{f}_X(x_j)}{\hat{f}_{XZ}(x_j,z_j)}y_j$ and $i = 1,...,n$. These first stage estimators are then used in a second stage one-step backfitting. Using local linear estimation, the final estimators are given by,

$$\vec{m}_1^{2S1}(\vec{x}) = S_1(\vec{y} - \vec{\gamma}_2^E(\vec{z})) \text{ and } \vec{m}_2^{2S1}(\vec{z}) = S_2(\vec{y} - \vec{\gamma}_1^E(\vec{x})).$$

with $\alpha^{2S} = \bar{y}$. Here we propose a new alternative two-stage estimation procedure based directly on moment conditions (8) and (9). First, preliminary estimators for $m_1(\cdot)$ and $m_2(\cdot)$ are obtained by a nonparametric regression of $\frac{\hat{f}_X(x_t)\hat{f}_Z(z_t)}{\hat{f}_{XZ}(x_t,z_t)}(y_t - \alpha^{2S})$ on x_t and z_t respectively. Using the same internalized Nadaraya-Watson estimator described above, we obtain

$$\vec{m}_1^p(\vec{x}) = \begin{pmatrix} m_1^p(x_1) \\ m_1^p(x_2) \\ \vdots \\ m_1^p(x_n) \end{pmatrix} \text{ and } \vec{m}_2^p(\vec{z}) = \begin{pmatrix} m_2^p(z_1) \\ m_2^p(z_2) \\ \vdots \\ m_2^p(z_n) \end{pmatrix}$$

[5]See Jones, Davies and Park(1994) and Kim et al.(1999)

where $m_1^p(x_i) = \frac{1}{ng_1}\sum_{j=1}^n K_1\left(\frac{x_j-x_i}{g_1}\right)\frac{\hat{f}_Z(z_j)}{\hat{f}_{XZ}(x_j,z_j)}(y_j-\bar{y})$, $m_2^p(z_i) = \frac{1}{ng_2}\sum_{j=1}^n K_1\left(\frac{z_j-z_i}{g_2}\right)\frac{\hat{f}_X(x_j)}{\hat{f}_{XZ}(x_j,z_j)}(y_j-\bar{y})$ for $i=1,...,n$. In the second stage, a one step backfitting procedure is implemented and we define,

$$\vec{m}_1^{2S2}(\vec{x}) = S_1(\vec{y}-\bar{y}-\vec{m}_2^p(\vec{z})) \text{ and } \vec{m}_2^{2S2}(\vec{z}) = S_2(\vec{y}-\bar{y}-\vec{m}_1^p(\vec{x})).$$

3 Asymptotic Approximations and Bandwidth Selection

The *plug in* bandwidth selection methods we consider for all estimators depend on obtaining suitable asymptotic approximations for the conditional squared error of the estimators. To this end we make the following general assumptions that are sufficient to obtain the conditional bias and variance of the 2S estimators under study. These assumptions are similar to those assumed by Opsomer and Ruppert(1997), Linton and Nielsen(1995) and Kim et al.(1999).

Assumption 1 *The kernel K_1 is such that $K_1 : [-1,1] \to [0, B_K)$ for some finite $B_K > 0$, $K_1(x) = K_1(-x)$ for $x \in \Re$, $\mu_1 = \int xK_1(x)dx = 0$, $\mu_2 = \int x^2K_1(\psi)d\psi > 0$ and there exists a constant c such that $|K_1(u) - K_1(v)| \leq c|u-v|$ for all $u,v \in \Re$, $\int K_1^2(v)dv$ exists and we write $R_K = \int K_1^2(v)dv$. $K_2(u,v) = K_1(u)K_1(v)$.*

Assumption 2 *The functions $m_\delta(\cdot)$ for $\delta = 1,2$, $f_X(x)$, $f_Z(z)$, $f_{XZ}(x,z)$ are twice continuously differentiable and we write $d^k m_\delta(x) = \frac{d^k m_\delta}{dx^k}(x)$, $d^k f_X(x) = \frac{d^k f_X}{dx^k}(x)$, $d^k f_Z(z) = \frac{d^k f_Z}{dz^k}(z)$ for $k=1,2$, $\frac{\partial f_{XZ}(x,z)}{\partial_d}$ is the partial derivative of f_{XZ} with respect to its d^{th} argument, $\frac{\partial^2 f_{XZ}(x,z)}{\partial_d \partial_\delta}$ is the second derivative of f_{XZ} with respect to its d^{th} and δ^{th} argument. Such derivatives are continuous in their compact support given by S_X, S_Z and $S_X \times S_Z$. We assume further that there exist generic constants $0 < b_f < B_f$ that are respectively lower and upper bounds on f_X,f_Z and f_{XZ} and $|dm_\delta| < B_{dm}$.*

Assumption 3 *We assume that there exist nonstochastic bandwidths g_1, h_1 and g_2, h_2 associated with regression directions $m_1(\cdot)$ and $m_2(\cdot)$ respectively. We assume that these bandwidths are such that $g_1, h_1, g_2, h_2 \to 0$, $nh_1h_2, ng_1g_2 \to \infty$ as $n \to \infty$, and that $g_d \sim h_d$ (same order) for $d = 1,2$.*

3.1 Backfitting Estimator

Opsomer and Ruppert(1997) show that when $\frac{nh_1}{logn}, \frac{nh_2}{logn} \to \infty$ it is possible to obtain asymptotic approximations for the conditional bias and variance of $\vec{m}_1^B(x)$ and $\vec{m}_2^B(z)$. These asymptotic approximations are of limited use for applied researchers as they convey very little information on the distributional properties of the estimator. Their most common use has been in obtaining an asymptotic approximation for the estimator's mean average squared error (AMASE),which is a widely used criterion for obtaining *plug in* bandwidths. By definition, for a generic estimator $\hat{m}(x,z)$ for $m(x,z)$

$$MASE(\hat{m}|\vec{x},\vec{z}) = \frac{1}{n}\sum_{i=1}^{n}(E(\hat{m}(x_i,z_i)|\vec{x},\vec{z}) - m(x_i,z_i))^2 + \frac{1}{n}\sum_{i=1}^{n}V(\hat{m}(x_i,z_i)|\vec{x},\vec{z})$$

Since conditional on the regressors MASE depends on h_1 and h_2, it has served as an optimand for obtaining data dependent expressions for h_1 and h_2 that minimize an asymptotic approximation for MASE (AMASE). Since AMASE is highly nonlinear on the bandwidths, the optimization can only be accomplished by a numerical procedure. However, under the assumption that the regressors are independent and that the observations lie in the interior of the support of the joint density of the regressors, it is possible to obtain an analytical solution for the optimization problem. Expressions for h_1 and h_2 that are optimal in the sense that they minimize the AMASE for a B-estimator are given by:

$$h_1^B = \left(\frac{\sigma^2 R_K}{n\mu_2^2\theta_{11}} n^{-1}\sum_{t=1}^{n}\frac{1}{f_X(x_t)}\right)^{1/5} \text{ and } h_2^B = \left(\frac{\sigma^2 R_K}{n\mu_2^2\theta_{22}} n^{-1}\sum_{t=1}^{n}\frac{1}{f_Z(z_t)}\right)^{1/5} \tag{12}$$

where $\theta_{11} = n^{-1}\sum_{t=1}^{n}(d^2m_1(x_t) - E(d^2m_1(x_t)))^2$, $\theta_{22} = n^{-1}\sum_{t=1}^{n}(d^2m_2(z_t) - E(d^2m_2(z_t)))^2$. A few points are worth noting regarding the practical use of these expressions: a) θ_{11}, θ_{22}, f_X, and f_Z are unknown, rendering h_1^B and h_2^B inadequate for producing feasible B-estimators; b) their relatively simple analytical form comes from the strong assumption of independence among the regressors; c) to render the expressions in (12) useful in practice, the unknown quantities θ_{11}, θ_{22}, f_X, and f_Z must be estimated. This is standard procedure for all *plug-in* bandwidth selection methods. Hence, feasible B-estimator are nonlinear (in the regressand) with conditional biases and variances that are in general different from those used to obtain AMASE. Opsomer and Ruppert(1998) propose an estimation procedure for the unknowns in (12) which we adopt in our simulations.

3.2 Marginal Integration Estimator

Linton and Nielsen(1995) show that when $nh_1h_2^2, nh_2h_1^2 \to \infty$ then $\sqrt{nh_1}(m_1^M(x) - E(m_1^M(x)))$ and $\sqrt{nh_2}(m_2^M(z) - E(m_2^M(z)))$ are asymptotically normal. Unfortunately, as in the case of the B-estimator, these results in general do not hold when the sequence of bandwidths is stochastic, as when they are chosen *via* any type of data dependent procedure such as cross validation or plug-in methods. Hence, for applied researchers the current asymptotic normality results for the M-estimator are of limited, if any, use. Furthermore, the AMASE for the M-estimator, even under regression independence does not produce closed analytical expressions for optimal bandwidths similar to those for B-estimator. The AMASE for the M estimator and the optimal bandwidths that minimize AMASE are presented in Theorem 1 whose proof is straightforward from the results in Linton and Nielsen(1995) and is omitted. For simplicity in obtaining optimal values for h_1 and h_2 below, we assume X and Z are independent.

Theorem 1 *Let $(x, z) \in S_X \times S_Z$ the support of f_{XZ} and assume that X and Z are independent. Assume that Assumptions 1-3 are holding and that $nh_1h_2^2, nh_2h_1^2 \to \infty$.*

(i) The conditional bias and variance of $m_1^M(x)$ for $x \in S_X$ are given by,

$$E(m_1^M(x) - m_1(x)|\vec{x}, \vec{z}) = \frac{1}{2}h_1^2\mu_2 d^2 m_1(x) + \frac{1}{2}h_2^2\mu_2 E(d^2 m_2(Z)) + o_p(h_1^2 + h_2^2) \tag{13}$$

and

$$V(m_1^M(x)|\vec{x}, \vec{z}) = \frac{\sigma^2 R_K}{nh_1} f_X(x)^{-1} + o_p((nh_1)^{-1}). \tag{14}$$

Mutatis mutandis identical expressions for $m_2^M(z)$ are obtained. For $m^M(x, z) = \bar{y} + m_1^M(x) + m_2^M(z)$ we have,

$$\begin{aligned} E(m^M(x, z) - m(x, z)|\vec{x}, \vec{z}) &= \frac{1}{2}h_1^2\mu_2(d^2 m_1(x) + E(d^2 m_1(X)) + \frac{1}{2}h_2^2\mu_2(d^2 m_2(z) + E(d^2 m_2(z)) \\ &+ o_p(h_1^2 + h_2^2) \end{aligned}$$

and

$$V(m^M(x,z)|\vec{x}, \vec{z}) = \frac{\sigma^2 R_K}{n}((h_1 f_X(x))^{-1} + (h_2 f_Z(z))^{-1}) + o_p\left(\frac{1}{nh_1} + \frac{1}{nh_2}\right). \tag{15}$$

(ii) The conditional MASE for the M-estimator is given by,

$$\begin{aligned} MASE &= \frac{1}{4}h_1^4\mu_2^2\psi_{11} + \frac{1}{2}h_1^2h_2^2\mu_2^2\psi_{12} + \frac{1}{4}h_2^4\mu_2^2\psi_{22} \\ &+ \frac{\sigma^2 R(K)}{n}\left(\frac{1}{nh_1}\sum_{t=1}^{n}(f_X(x_t))^{-1} + \frac{1}{nh_2}\sum_{t=1}^{n}(f_Z(z_t))^{-1}\right) + o_p\left(h_1^4 + h_2^4 + \frac{1}{nh_1} + \frac{1}{nh_2}\right) \end{aligned}$$

where $\psi_{d\delta} = \frac{1}{n}\sum_{t=1}^{n}(d^2m_d(x_t) + E(d^2m_d(x_t))(d^2m_\delta(z_t) + E(d^2m_\delta(z_t)).$

(iii) The bandwidths that minimize the conditional MASE, disregarding the term $o_p(\cdot)$*, denoted by* h_1^M, h_2^M*, must satisfy,*

$$(h_1^M)^5\mu_2^2\psi_{11} + (h_1^M)^3(h_2^M)^2\mu_2^2\psi_{12} = \frac{\sigma^2 R_K}{n}\left(\frac{1}{n}\sum_{t=1}^{n}(f_X(x_t))^{-1}\right) \quad (16)$$

$$(h_2^M)^5\mu_2^2\psi_{22} + (h_2^M)^3(h_1^M)^2\mu_2^2\psi_{12} = \frac{\sigma^2 R_K}{n}\left(\frac{1}{n}\sum_{t=1}^{n}(f_Z(z_t))^{-1}\right) \quad (17)$$

The proof of part (i) is a direct consequence of the theorem in Linton and Nielsen (1995) using the empirical cumulative distribution function as an estimate for the cumulative marginal distribution function of Z. (ii) follows from (i), and (iii) is obtained by setting the partial derivatives of (ii) with respect to h_1 and h_2 equal to zero. As in the case of B-estimation these optimal bandwidths depend on unknown quantities in (16) and (17) that have to be estimated to render them useful. Specifically it is necessary to estimate $\phi_{d\delta}$, f_X and f_Z. Hence, the stochastic nature of the estimates of h_1^M and h_2^M and their dependence on the regressand produce the same nonlinearities and difficulties that were alluded to when discussing B-estimation.

3.3 Two Stage Estimator

In this section we obtain the conditional MASE for the two stage estimators described in subsection 2.3. The conditional bias and variance for m^{2S1} are provided in Theorem 2, while Theorem 3 provides similar results for m^{2S2}. These results are then used to construct conditional MASE and to obtain optimal bandwidths for the two stage estimators. The proofs depend on the auxiliary Lemma 1 in the appendix which establishes uniform convergence of certain bounded functions of X and Z. As in the case of B and M estimation, certain requirements on the speed of convergence to zero of the bandwidths are necessary.

Theorem 2 *Suppose that Assumptions 1-3 hold, that* $ng_1^3(ln(g_1))^{-1} \to \infty$ *and that* $n(g_1g_2)^3(ln(g_1g_2))^{-1} \to$

∞. *Then, the conditional bias of $m_1^{2S1}(x)$ for $x \in S_X$ is given by,*

$$\begin{aligned}
E\left(\hat{m}_1^{2S1}(x) - m_1(x)|\vec{x},\vec{z}\right) &= \frac{h_1^2}{2}\mu_2 d^2 m_1(x) - \frac{1}{2}g_2^2\mu_2 E(d^2 m_2(Z)|\vec{x}) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int d^2 f_X(v) m(v,z_i)dv|\vec{x}\right) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int m(v,Z) f_X(v) f_{XZ}^{-1}(v,Z)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(v,Z)}{\partial_d\partial_d}dv|\vec{x}\right) + o_p(h_1^2) + o_p(g_2^2)
\end{aligned}$$

and

$$V(m_1^{2S1}(x_i)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x_i)} + o_p((nh_1)^{-1}) \tag{18}$$

Mutatis mutandis similar expressions are obtained for m_2. The conditional bias and variance of $m^{2S1}(x_i,z_i)$ are

$$\begin{aligned}
E(m^{2S1}(x_i,z_i) - m(x_i,z_i)|\vec{x},\vec{z}) &= \frac{h_1^2}{2}\mu_2 d^2 m_1(x_i) - \frac{1}{2}g_2^2\mu_2 E(d^2 m_2(z_i)|\vec{x}) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int d^2 f_X(v) m(v,z_i)dv|\vec{x}\right) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int m(v,z_i) f_X(v) f_{XZ}^{-1}(v,z_i)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_d\partial_d}dv|\vec{x}\right) \\
&+ \frac{h_2^2}{2}\mu_2 d^2 m_2(z_i) - \frac{1}{2}g_1^2\mu_2 E(d^2 m_1(x_i)|\vec{z}) \\
&- \frac{1}{2}g_1^2\mu_2 E\left(\int d^2 f_Z(v) m(x_i,v)dv|\vec{z}\right) \\
&- \frac{1}{2}g_1^2\mu_2 E\left(\int m(x_i,v) f_Z(v) f_{XZ}^{-1}(x_i,v)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(x_i,v)}{\partial_d\partial_d}dv|\vec{z}\right) \\
&+ o_p(h_1^2) + o_p(g_2^2) + o_p(h_2^2) + o_p(g_1^2)
\end{aligned}$$

and

$$V(m^{2S1}(x_i,z_i)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x_i)} + \frac{1}{nh_2}\frac{\sigma^2 R_K}{f_Z(z_i)} + o_p\left(\frac{1}{nh_1}\right) + o_p\left(\frac{1}{nh_2}\right) \tag{19}$$

Proof See appendix.

Theorem 3 *Suppose that Assumptions 1-3 hold, that $ng_1^3(ln(g_1))^{-1} \to \infty$ and that $n(g_1g_2)^{2p+1}(ln(g_1g_2))^{-1} \to \infty$ and let $\mu(x_i,z_i) = m_1(x_i) + m_2(z_i)$. Then, the conditional bias of $m_1^{2S2}(x_i)$ is given by,*

$$\begin{aligned}
E\left(\hat{m}_1^{2S2}(x_i) - m_1(x_i)|\vec{x},\vec{z}\right) &= \frac{h_1^2}{2}\mu_2 d^2 m_1(x_i) - \frac{1}{2}g_2^2\mu_2 E(d^2 m_2(z_i)|\vec{x}) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int d^2 f_X(v)\mu(v,z_i)dv|\vec{x}\right)
\end{aligned}$$

$$- \frac{1}{2}g_2^2\mu_2 E\left(\int \mu(v,z_i) f_X(v) f_{XZ}^{-1}(v,z_i)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_d \partial_d} dv|\vec{x}\right) + o_p(h_1^2) + o_p(g_2^2)$$

and

$$V(m_1^{2S2}(x_i)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x_i)} + o_p((nh_1)^{-1}) \tag{20}$$

Mutatis mutandis similar expressions are obtained for m_2. *The conditional bias and variance of* $m^{2S2}(x_i,z_i)$ *are*

$$\begin{aligned}
E(m^{2S2}(x_i,z_i) - m(x_i,z_i)|\vec{x},\vec{z}) &= \frac{h_1^2}{2}\mu_2 d^2 m_1(x_i) - \frac{1}{2}g_2^2\mu_2 E(d^2 m_2(z_i)|\vec{x}) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int d^2 f_X(v)\mu(v,z_i)dv|\vec{x}\right) \\
&- \frac{1}{2}g_2^2\mu_2 E\left(\int f_X(v) f_{XZ}^{-1}(v,z_i)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_d\partial_d}dv|\vec{x}\right) \\
&+ \frac{h_2^2}{2}\mu_2 d^2 m_2(z_i) - \frac{1}{2}g_1^2\mu_2 E(d^2 m_1(x_i)|\vec{z}) \\
&- \frac{1}{2}g_1^2\mu_2 E\left(\int d^2 f_Z(v)\mu(x_i,v)dv|\vec{z}\right) \\
&- \frac{1}{2}g_1^2\mu_2 E\left(\int f_Z(v) f_{XZ}^{-1}(x_i,v)\sum_{d=1}^{2}\frac{\partial^2 f_{XZ}(x_i,v)}{\partial_d\partial_d}dv|\vec{z}\right) \\
&+ o_p(h_1^2) + o_p(g_2^2) + o_p(h_2^2) + o_p(g_1^2)
\end{aligned}$$

and

$$Var(m^{2S2}(x_i,z_i)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x_i)} + \frac{1}{nh_2}\frac{\sigma^2 R_K}{f_Z(z_i)} + o_p\left(\frac{1}{nh_1}\right) + o_p\left(\frac{1}{nh_2}\right) \tag{21}$$

Proof See appendix.

From Theorems 2 and 3 we can obtain expressions for the conditional MASE for the two stage estimators we consider. Corollaries 1 and 2 provide conditional MASE for the two stage estimators we consider under the assumption that X and Z.

Corollary 1 *Let* X *and* Z *be independent and assume that the bandwidths used in the first stage -* g_1, g_2 *- are identical to* h_1, h_2 *used in the second stage of the estimation. Then,*

a) the conditional bias for $m_1^{2S1}(x)$ *is given by,*

$$\begin{aligned}
E(m_1^{2S1}(x) - m_1(x)|\vec{x},\vec{z}) &= \frac{1}{2}h_1^2\mu_2 d^2 m_1(x) - \frac{1}{2}h_2^2\mu_2 E(d^2 m_2(Z)) \\
&+ \frac{1}{2}h_2^2\mu_2\int d^2 f_Z(v)\gamma_2(z)dz + o_p(h_1^2 + h_2^2)
\end{aligned}$$

and the conditional variance is given by

$$V(m_1^{2S1}(x)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x)} + o_p\left(\frac{1}{nh_1}\right).$$

Mutatis mutandis, similar expressions for $m_2^{2S1}(x)$ are obtained.

b) The conditional bias and variance for $m^{2S1}(x,z)$ are given by,

$$\begin{aligned} E(m^{2S1}(x,z) - m(x,z)|\vec{x},\vec{z}) &= \frac{1}{2}h_1^2\mu_2\left(d^2m_1(x) - E(d^2m_2(Z)) + \int d^2f_Z(v)\gamma_2(v)dv\right) \\ &+ \frac{1}{2}h_2^2\mu_2\left(d^2m_2(z) - E(d^2m_1(X)) + \int d^2f_X(v)\gamma_1(v)dv\right) + o_p(h_1^2+h_2^2) \end{aligned}$$

and

$$V(m^{2S1}(x,z)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x)} + \frac{1}{nh_2}\frac{\sigma^2 R_K}{f_z(z)} + o_p\left(\frac{1}{nh_1}) + \frac{1}{nh_2}\right)$$

c) From b) we obtain

$$\begin{aligned} MASE(m^{2S1}(x,z)|\vec{x},\vec{z}) &= \frac{1}{4}h_1^4\mu_2^2\phi_{11} + \frac{1}{4}h_2^4\mu_2^2\phi_{22} + \frac{1}{2}h_1^2h_2^2\mu_2\phi_{12} + \sigma^2R_K\left(\frac{1}{nh_1}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_X(x_t)}\right. \\ &+ \left.\frac{1}{nh_2}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_Z(z_t)}\right) + o_p(h_1^4+h_2^4) + o_p\left(\frac{1}{nh_1}+\frac{1}{nh_1}\right) \end{aligned}$$

where

$$\begin{aligned} \phi_{11} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_1(x_t) - E(d^2m_2(Z)) + \int d^2f_Z(v)\gamma_2(v)dv\right)^2 \\ \phi_{22} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_2(z_t) - E(d^2m_1(X)) + \int d^2f_X(v)\gamma_1(v)dv\right)^2 \\ \phi_{12} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_1(x_t) - E(d^2m_2(Z)) + \int d^2f_Z(v)\gamma_2(v)dv\right)\left(d^2m_2(z_t) - E(d^2m_1(X)) + \int d^2f_X(v)\gamma_1(v)dv\right) \end{aligned}$$

Corollary 2 *Let X and Z be independent and assume that the bandwidths used in the first stage - g_1, g_2 - are identical to h_1, h_2 used in the second stage of the estimation. Then,*

a) the conditional bias for $m_1^{2S2}(x)$ is given by,

$$\begin{aligned} E(m_1^{2S2}(x) - m_1(x)|\vec{x},\vec{z}) &= \frac{1}{2}h_1^2\mu_2 d^2m_1(x) - \frac{1}{2}h_2^2\mu_2E(d^2m_2(Z)) \\ &+ \frac{1}{2}h_2^2\mu_2\int d^2f_Z(v)m_2(z)dz + o_p(h_1^2+h_2^2) \end{aligned}$$

and the conditional variance is given by

$$V(m_1^{2S2}(x)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x)} + o_p\left(\frac{1}{nh_1}\right).$$

Mutatis mutandis, similar expressions for $m_2^{2S2}(x)$ are obtained.

b) The conditional bias and variance for $m^{2S2}(x,z)$ are given by,

$$\begin{aligned} E(m^{2S2}(x,z) - m(x,z)|\vec{x},\vec{z}) &= \frac{1}{2}h_1^2\mu_2\left(d^2m_1(x) - E(d^2m_2(Z)) + \int d^2 f_Z(v)m_2(v)dv\right) \\ &+ \frac{1}{2}h_2^2\mu_2\left(d^2m_2(z) - E(d^2m_1(X)) + \int d^2 f_X(v)m_1(v)dv\right) + o_p(h_1^2 + h_2^2) \end{aligned}$$

and

$$V(m^{2S2}(x,z)|\vec{x},\vec{z}) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x)} + \frac{1}{nh_2}\frac{\sigma^2 R_K}{f_z(z)} + o_p\left(\frac{1}{nh_1}) + \frac{1}{nh_2}\right)$$

c) From b) we obtain

$$\begin{aligned} MASE(m^{2S2}(x,z)|\vec{x},\vec{z}) &= \frac{1}{4}h_1^4\mu_2^2\phi_{11} + \frac{1}{4}h_2^4\mu_2^2\phi_{22} + \frac{1}{2}h_1^2h_2^2\mu_2\phi_{12} + \sigma^2 R_K\left(\frac{1}{nh_1}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_X(x_t)}\right. \\ &+ \left.\frac{1}{nh_2}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_Z(z_t)}\right) + o_p(h_1^4 + h_2^4) + o_p\left(\frac{1}{nh_1} + \frac{1}{nh_1}\right) \end{aligned}$$

where

$$\begin{aligned} \chi_{11} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_1(x_t) - E(d^2m_2(Z)) + \int d^2 f_Z(v)m_2(v)dv\right)^2 \\ \chi_{22} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_2(z_t) - E(d^2m_1(X)) + \int d^2 f_X(v)m_1(v)dv\right)^2 \\ \chi_{12} &= \frac{1}{n}\sum_{t=1}^{n}\left(d^2m_1(x_t) - E(d^2m_2(Z)) + \int d^2 f_Z(v)m_2(v)dv\right)\left(d^2m_2(z_t) - E(d^2m_1(X)) + \int d^2 f_X(v)m_1(v)dv\right) \end{aligned}$$

We note that $m_d^{2S1}(x)$ and $m_d^{2S2}(x)$ for $d = 1,2$ have conditional variances that are equal to that of B and M estimators and a univariate local linear estimator. The expressions for the bias of $m_d^{2S1}(x)$ and $m_d^{2S2}(x)$ have three extra terms if compared to the univariate local linear estimator. Kim et al.(1999) are able to eliminate these extra terms if there is undersmoothing of the first stage estimation, i.e. letting g_1, g_2 degenerate at a faster speed relative to h_1, h_2. Note that this *oracle* property of their estimation procedure can be obtained in the context of backfitting by choosing bandwidths that undersmooth at various steps of the backfitting algorithm. When X and Z are independent and under local linear estimation the B-estimator's conditional bias and variance can be written as

$$E\left(m_1^B(x) - m_1(x)|\vec{x},\vec{z}\right) = \frac{1}{2}h_1^2\mu_2\left(d^2m_1(x) - \int d^2m_1(x)f_X(v)dv\right) + o_p\left(\frac{1}{\sqrt{n}}\right) + o_p(h_1^2+h_2^2) \quad (22)$$

and

$$V\left(\hat{m}_1(x)|\vec{x},\vec{z}\right) = \frac{1}{nh_1}\frac{\sigma^2 R_K}{f_X(x)} + o_p(\frac{1}{nh_1}). \quad (23)$$

Hence, in B-estimation the bias of the estimator depends only on the total curvature of m_1, weighted by the density. The bias of the 2S estimator as well as that of the M-estimator do depend on the total curvature of the other variable even when X and Z are independent.

Given the AMASE results from Corollaries 1 and 2 the optimal bandwidths that minimize the conditional AMASE for 2S1 and 2S2 must satisfy the following two sets of equations

$$(h_1^{2S1})^5\mu_2^2\phi_{11} + (h_2^{2S1})^2(h_1^{2S1})^3\mu_2\phi_{12} = \sigma^2 R_K\left(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_X(x_t)}\right) \quad (24)$$

$$(h_2^{2S1})^5\mu_2^2\phi_{22} + (h_1^{2S1})^2(h_2^{2S1})^3\mu_2\phi_{12} = \sigma^2 R_K\left(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_Z(z_t)}\right) \quad (25)$$

and

$$(h_1^{2S2})^5\mu_2^2\chi_{11} + (h_2^{2S2})^2(h_1^{2S2})^3\mu_2\chi_{12} = \sigma^2 R_K\left(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_X(x_t)}\right) \quad (26)$$

$$(h_2^{2S2})^5\mu_2^2\chi_{22} + (h_1^{2S2})^2(h_2^{2S1})^3\mu_2\chi_{12} = \sigma^2 R_K\left(\frac{1}{n}\sum_{t=1}^{n}\frac{1}{f_Z(z_t)}\right). \quad (27)$$

3.4 Data Driven Bandwidth Selection

The choice of data driven bandwidth selectors for the Monte Carlo experiments was based on two considerations. First, we want to have a bandwidth selector that interfered minimally with the performance of the estimators. By this, we mean a bandwidth estimator that transferred minimal noise from the estimation of f_X, f_Z, $\theta_{d\delta}$, $\psi_{d,\delta}$, $\phi_{d,\delta}$ and $\chi_{d\delta}$ for $d,\delta = 1,2$, $\int d^2f_X(v)\gamma_1(v)dv$, $\int d^2f_Z(v)\gamma_2(v)dv$, $\int d^2f_X(v)m_1(v)dv$ and $\int d^2f_Z(v)m_2(v)dv$ to the estimation of m_1 and m_2. This provides an ideal setting to compare the performance of the estimators, as any differences can be attributed to the structure of the estimators themselves and not to the estimation of the unknowns in the expressions for the optimal bandwidth. Second, we want to compare the performance of the estimators when using bandwidth selectors proposed in the previous section and those already proposed in the literature. Since cross-validation methods have been shown to possess several undesirable properties[6] we concentrate on plug-in bandwidth selectors.

3.5 True Bandwidths

Elimination of the noise that is generated by the estimation of the parameters in the expression for optimal bandwidths - equation (12) for B-estimator, equations (16) and (17) for M-estimation, (24) and (25) for 2S1-estimator and (26) and (27) for 2S2-estimator - can be accomplished in a Monte Carlo study setting since the true values of these unknowns can be obtained directly from the specification of the DGP. Hence, the first set of bandwidths that we select are based on complete information about the normally unknown functionals that appear on the specification of the optimal bandwidths.[7] In this case the only difficulty involves the evaluation of the integrals that define the expectations that appears in $\psi_{d,\delta}$, $\phi_{d,\delta}$ and $\chi_{d\delta}$ for $d, \delta = 1, 2$ and $\int d^2 f_X(v)\gamma_1(v)dv$, $\int d^2 f_Z(v)\gamma_2(v)dv$, $\int d^2 f_X(v)m_1(v)dv$ and $\int d^2 f_Z(v)m_2(v)dv$. These expectations can be difficult to compute depending on the nature of m_d. In our study all integrals were calculated numerically using the Gauss-Legendre quadrature method. These bandwidths are reported on Tables 7.1 - 7.3.

3.6 Estimated Bandwidths

The estimated bandwidths for the B-estimator were obtained using the procedure recently proposed by Opsomer and Ruppert(1998) to estimate θ_{11}, θ_{22}, σ^2. We assumed that f_X and f_Z were uniform densities over a compact support and estimated their inverses by $max_t(x_t) - min_t(x_t)$ and $max_t(z_t) - min_t(z_t)$, respectively, where $max_t(x_t)$ and $min_t(x_t)$ are the maximum and minimum sample values for X.

We consider two different estimated bandwidths for M-estimation. The first, which produces an M-estimator we label M_L in the tables with simulation results, was proposed by Linton and Nielsen(1995). The estimated bandwidths are given by,

[6]See Park, Byeong and Marron(1990), Simonoff (1996) and Opsomer and Ruppert (1998).
[7]Note that the true optimal bandwidths are different across samples since MASE is evaluated at sample points.

$$\ddot{h}_1 = \left(\frac{\ddot{\sigma}^2 R_K(max_t(x_t) - min_t(x_t))}{n\mu_2^2(\hat{\beta}_1 + \hat{\beta}_2)^2} \right)^{1/5} \text{ and } \ddot{h}_2 = \left(\frac{\ddot{\sigma}^2 R_K(max_t(z_t) - min_t(z_t))}{n\mu_2^2(\hat{\beta}_1 + \hat{\beta}_2)^2} \right)^{1/5}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS estimates of the parameters associated with $x_t^2/2$ and $z_t^2/2$ of a regression of y_t on a constant, $x_t^2/2$, $z_t^2/2$, x_t, z_t and $x_t z_t$. $\ddot{\sigma}^2$ is the typical estimate for the variance in classical linear regression models. The second procedure involves the numerical solution of equations (16) and (17). Once again, we assumed that f_X and f_Z were uniform densities over a compact support and estimated their inverses by $max_t(x_t) - min_t(x_t)$ and $max_t(z_t) - min_t(z_t)$. $\psi_{d\delta}$ were estimated using the same procedure for the estimation of $\theta_{d\delta}$ proposed by Opsomer and Ruppert with the necessary sign change inside the summations. The estimated bandwidths that result from the numerical solution of (16) and (17) produce a set of estimators that we label M in the tables with the simulation results.

We also consider two different estimated bandwidths for 2S1 estimation. The first is the simple rule of thumb proposed in Kim et al.(1999) in which h_1 and h_2 are selected as follows,

$$h_1^K = n^{-1/5}\frac{1}{2}\hat{\sigma}_X \text{ and } h_2^K = n^{-1/5}\frac{1}{2}\hat{\sigma}_Z$$

where $\hat{\sigma}_X = \sqrt{\frac{1}{n}\sum_{t=1}^n (x_t - \bar{x})^2}$ and $\hat{\sigma}_Z = \sqrt{\frac{1}{n}\sum_{t=1}^n (z_t - \bar{z})^2}$. These estimated bandwidths produce an estimator that we label $2S1_K$ in the tables. The second bandwidth selection procedure we consider for the 2S1 estimator is based on the numerical solution of equations (24) and (25). To this end the unknown quantities $\phi_{d\delta}$ must be estimated together with f_X, f_Z and σ^2. The estimation of $\phi_{d\delta}$ depends on the estimation of two parts - $d^2m_d(x_t) - E(d^2m_\delta)$ and $\int d^2 f_X(v)\gamma_1(v)dv$ (or $\int d^2 f_Z(v)\gamma_2(v)dv$). The first term is estimated as in the case of B-estimation, the second term can be interpreted as $E(\gamma_1(v)\frac{d^2 f_X(v)}{f_X(v)})$ which is estimated by $n^{-1}\sum_{t=1}^n \hat{\gamma}_1(x_t)\frac{d^2\hat{f}_X(x_t)}{\hat{f}_X(x_t)}$, where $\hat{\gamma}_1$ comes from a preliminary B-estimator and $\hat{f}$ is estimated by a kernel density estimator with a Silverman rule-of-thumb bandwidth. Finally, σ^2 is estimated as in the case for B-estimation.

Finally, the estimated bandwidths used to produce the 2S2 estimator is the result of the numerical solution for equations (26) and (27). As in the case for 2S1 the unknowns that appear in equations (26) and (27), i.e., $\chi_{d\delta}$ must be estimated together with f_X, f_Z and σ^2. We follow the same estimation procedure described above for 2S1 with the exception that $\hat{\gamma}_d$ is substituted by $\hat{m}_d$.

4 The Data Generating Process - DGP

4.1 Pseudorandom Number Generation

The generation of simulated random numbers which have distributional properties similar to those assumed for the random variables in the model under study is a crucial step in Monte Carlo experimentation. Generating multivariate normal variables with a general covariance matrix, as required below, involves several algorithms. All univariate normal random number generators use a uniform generator and a transformation of the simulated uniform random numbers to produce simulated standard normal random numbers. Multivariate normal generators add an additional step by transforming the simulated univariate normal random numbers into simulated multivariate normal random vectors that possess the desired covariance structure. Hence, three algorithms are involved and problems can arise at any step in the process.

Noticeable improvements in the methodology for all three of these steps have appeared in the statistical computing literature. Unfortunately, these advances have not made their way into the econometrics literature. In particular, to obtain the initial set of uniformly distributed pseudorandom numbers, most studies rely on multiplicative or mixed congruential pseudorandom number generators. This practice continues despite the demonstration by Marsaglia(1968) that congruential generators suffer from serious autocorrelation of various orders, and the simple illustrative example of this problem provided by Kennedy and Gentle(1980). Generalized Feedback Shift Register (GFSR) uniform pseudorandom number generators have emerged as an alternative to congruential generators. Fushimi(1990) proposed a recurrence formula that produces uniform pseudorandom numbers that perform well in a series of tests for randomness, including a test suggested by Marsaglia(1985). Hence, we use Fushimi's algorithm with a seed value of 1589. This seed value was used successfully by Fushimi in testing the algorithm.

We use the pseudorandom numbers obtained from the implementation of Fushimi's method in conjunction with the algorithm suggested by Kinderman and Ramage(1976) to obtain univariate standard normal pseudorandom numbers. Following Graybill(1969) and Barr and Slezak(1972), we use a triangular factorization (Cholesky's) method to generate multivariate normal pseudorandom vectors with the desired covariance structure.

4.2 Regression Model Design and Estimation

Using the procedures described above the data used in this study is generated by a fully specified bivariate additive nonparametric regression model. First, we generate a sequence of bivariate normally distributed random variables $\{x_t, z_t\}_{t=1}^n$, with joint density given by

$$\begin{pmatrix} x_t \\ z_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1/9 & c/9 \\ c/9 & 1/9 \end{pmatrix}\right)$$

where $c = 0, 0.25, 0.75$, gives the desired correlation between the random variables. We allow for different correlation values because one of our objectives is to evaluate how regressor dependency impacts the performance of the estimators. One of the assumptions required to obtain expressions for the conditional mean and variance of the estimators under study is that f_{XZ} have compact support. To satisfy this assumption we discard every generated data point that is outside $[0, 1]$ and resample until all generated pseudorandom numbers are within this interval. The regression error ϵ_t is generated independently as a standard normal, and the regressands y_t are obtained in accordance with three models

$$y_t = m_1(x_t) + m_2(z_t) + \epsilon_t \text{ which we label Model 1,} \tag{28}$$

$$y_t = m_1(x_t) + m_3(z_t) + \epsilon_t \text{ which we label Model 2,} \tag{29}$$

$$y_t = m_2(x_t) + m_3(z_t) + \epsilon_t \text{ which we label Model 3,} \tag{30}$$

where $m_1(x) = 1 - 6x + 36x^2 - 53x^3 + 22x^5$, $m_2(x) = sin(5\pi x)$ and $m_3(x) = exp(3x)$ for $t = 1, \cdots, n$. We chose functions that have very different curvatures making the use of a common bandwidth inadequate. Figure 1 provides graphs of m_1, m_2 and m_3 over the relevant range of X and Z.

We chose to consider samples of relatively small size for two reasons. First, the small sample sizes reduce the computational burden in the Monte Carlo. Second, we wanted to evaluate the estimators performance under fairly undesirable conditions. We generate samples of size $n = 200, 350, 500$ and for all sample sizes we generated 100 samples.

4.3 Estimation

Computer codes for the estimation were written in *GAUSS* software version 5.0 and estimation and random number generation was done on a PC running on an 3.1 Mhz - Intel Pentium IV processor. Computing time in implementing nonparametric estimators is always an important concern, even more so in a Monte Carlo experiment. However, we chose not to use any of the time saving procedures currently available in the literature. The most prominent of these is binning, as described by Fan and Marron(1994). Although estimators based on binning provide extraordinary savings in computing time, they remain *approximations* for the unbinned estimator. Fan and Marron (1994) provide experimental evidence that they are extremely good approximations for the unbinned estimator, but their evidence is limited to one dimensional estimation. As they observe, in higher dimensions - as the case in this paper - accuracy questions with regards to binning are still wide open. Hence, we decided to do all computations using the full form of the estimators and not their binned versions.

As expected, computing time for all estimators is extremely sensitive to sample size, but it is not significantly affected by bandwidth estimation. The M estimator is by far the most demanding with regards to computing time of all estimators under study. For this reason we chose to limit our simulations of M estimation to the case where $n = 200$. All other estimators can be implemented at fairly equivalent time with the estimated B-estimator using the procedure of Opsomer and Ruppert(1998) being the fastest to implement. The $2S1$ and $2S2$ estimators when estimated using the bandwidth selection procedure proposed in section 3.4 are slower to implement than the B-estimator as the procedure requires the numerical solution for a nonlinear system of equations. However, once bandwidths are chosen the 2S1 and 2S2 estimators are faster to implement than all other procedures. The computational time for all estimators increases significantly with the number of regressors.

We used a Gaussian kernel to construct the estimators. A more desirable choice would be to use an Epanechnikov kernel, or any kernel with compact support, since it would satisfy the compact support assumption necessary to obtain some of the theoretical results regarding these estimators. Unfortunately, even when true bandwidths are being used in the estimation, the M estimator is frequently not defined due to singularity of the $W(x, z)$ matrix. Clearly, the problem emerges because the bandwidths are in some samples numerically too small. It is reasonable to expect that for larger samples this problem would disappear. Whether or not such problem has been encountered by users of the M estimator is unclear. However, the application of the M estimator in Linton and Nielsen(1995) and Linton and Härdle(1996) uses a Gaussian kernel even though they have an explicit assumption on kernel support compactness.

5 Results

The analysis of the experimental results focuses on the average squared error of the estimators, their bias, variance and on the estimation of the bandwidths. Since some preliminary finite sample experimental evidence on the performance of the B-estimator and M-estimators are already available (Opsomer and Ruppert,1998 and Sperlich et al.,1999), we are primarily interested on the performance of the 2S estimators and on the *relative* performance of all estimators.

Tables 3 and 4 provide Average Squared Errors (ASE) across experiments using true and estimated bandwidths respectively for all estimators and for the different sample sizes and correlation levels. Some general regularities are promptly identified. First, increases in sample sizes reduce ASE for all estimators and across all correlation levels with true and estimated bandwidths.[8]

The effects of increased correlation on the ASE of the estimators are quite different. For the B-estimator ASE is similar across correlation levels for each sample size, but they do differ across models. In some cases the results even show mild decrease in ASE as correlation increases. These regularities are true when true and estimated bandwidths are used. Results are quite different for M and 2S estimation. All estimators seem to be impacted by increased regression correlation with ASE increasing as c grows when true or estimated bandwidths are used. It is apparent, however that ASE is not significantly affected by mild correlation among the regressors. The increase is significant, however, when the correlation moves from low levels 0 or

[8]The results on M estimation reported are for sample size $n = 200$ only. However, experiments for sample sizes $n = 100, 150$ were also implemented and the regularities reported here regarding the impact of sample size increase on other estimators were observed also in M estimation.

0.25 to 0.75. The impact of increases in c on ASE do vary across models, but increases in n do not seem to reduce the disparity in ASE. Finally, the M estimator seems to be the most sensitive of the estimators considered to increases in c.

One should also observe that, as expected, across all experiments and estimators the reported ASEs increase from Table 3 to Table 4, confirming that in finite samples the noise introduced by estimated bandwidths impacts the performance of the estimators. Also as expected, increased sample size dampens this impact.

The most noticeable result from Tables 3 and 4 is that, as measured by ASE, the B-estimator is superior to all estimators across all correlation levels, sample sizes and models. The second best is the 2S2 estimator we propose followed in order by M and 2S1 estimation. From Table 4 we observe that when we use our proposed bandwidth estimation procedure ($2S1$) the estimator improves relative to its implementation using the bandwidth selection procedure proposed by Kim et al.(1999)($2S1$) across all experiments. The same can be said about M and M_L except when the correlation level is at $c = 0.75$.

Tables 1 and 2 provide average bias for all estimators across experiments using true and estimated bandwidths, respectively. As in the case of ASE, some general regularities can be noticed. With almost no exception (normally involving 2S2) the B-estimator shows smallest bias across estimators followed by 2S2 when bandwidths are estimated. Table 2 reveals that the largest bias occur with M estimation. As in the case of ASE our proposed bandwidth selection methods reduces the bias of the 2S1 estimator but surprisingly this does not occur with M estimation.

Tables 1 and 2 also reveal that the bias increases with c across all experiments and estimators. Again, this is particularly noticeable when $c = 0.75$. The impact of sample size on bias when true bandwidths are used is different across estimators. For B and 2S1 no discernible pattern is observed, but for 2S2 the bias falls with sample size. When bandwidths are estimated the bias falls for all estimators and models as n increases except for B. Combining the results from Tables 1 and 2 with Tables 3 and 4 we can conclude that the variance of the estimators decreases with sample size across all experiments for all estimators.

Tables 5.1 -5.3 and 6.1 - 6.3 provide average squared errors for the estimation of m_d for $d = 1, 2, 3$ for all correlation levels and sample sizes using true and estimated bandwidths. The general regularities observed for ASE in Tables 3 and 4 seem to apply in each regression direction. What these directional regression results do suggest is that the ASE per direction is impacted significantly by the curvature of the functions being estimated and that the curvature of one function impacts the ASE properties of the other regression direction for all estimators.

Tables 7.1-7.3 and 8.1-8.3 provide the average *true* and estimated bandwidths across experiments for different sample sizes, correlations and models. Tables 7.1-7.3 and 8.1-8.3 reveal that true and estimated bandwidths for all estimators are quite insensitive to correlation levels. They do however noticeably change across models. Although expected gains from increased sample size do not appear dramatic for the samples sizes considered in this experiment, our proposed bandwidth estimation procedure produces bandwidths that are much closer to the true bandwidths than those produced by the procedures suggested by Linton and Nielsen(1995) and Kim et al.(1999) for M and 2S1 estimation respectively. In addition, the true bandwidths are identical (up to two decimal points) for all estimators, across all models and experiments.All estimated bandwidths for models 1 and 2 undersmooth if compared to the true bandwidths reported in Tables 5.1-5.2. For model 3 bandwidths oversmooth if compared to the true bandwidths reported in Table 5.3. How much under or over smoothing occurs depends largely on the degree of curvature of the m_d that compose the models. When there is more curvature, as in the case of models 1 and 3 the degree of under and oversmoothing seems to increase, indicating that increased curvature makes for more difficult bandwidth estimation.

6 Conclusion

Additive nonparametric regression models have gained increased popularity by their easy of interpretation and the fact that these models allow for the circumvention of the curse of dimensionality. Backfitting, marginal integration and two stage estimators have recently emerged as viable alternatives for the estimation of additive nonparametric regression models. Little is known about the finite and asymptotic properties of all estimators when bandwidths are selected by data driven procedures. Applied researchers are not only uninformed about the estimators' properties but are also unaware of their relative performance. In this paper we provided experimental evidence on the finite properties of these estimators and on their relative performance. We also propose a modification of the two stage estimator first introduced by Kim et al.(1999) that outperforms the original two stage estimator. However, the backfitting estimator seems to emerge as the best estimator among those currently available in the literature. This superiority is based on an evaluation of the estimators' ASE under estimated and true bandwidths. Separate evidence on their bias and variance is also provided to support this conclusion. Although Monte Carlo studies suffer from the problem of specificity, we believe that the results here are strong enough to recommend the use of backfitting estimation.

Figure 1: Directional Regressions m_1, m_2 and m_3

$$m_1(x) = 1 - 6x + 36x^2 - 53x^3 + 22x^5$$

5in1.75inc:/papers/martins-yang(04)/m1.eps

$$m_2(x) = sin(5\pi x)$$

5in1.75inc:/papers/martins-yang(04)/m2.eps

$$m_3(x) = exp(3x)$$

5in1.75inc:/papers/martins-yang(04)/m3.eps

7 Tables

TABLE 1

Average Bias using *True* Bandwidth			
	$n = 200$ (B,M,2S1,2S2)	$n = 350$ (B,2S1,2S2)	$n = 500$ (B,2S1,2S2)
	Model 1		
$c = 0$	(0.010, -0.011,0.164,-0.008)	(0.005,0.119,-0.010)	(0.005,0.100,-0.006)
$c = 0.25$	(-0.000,-0.030,0.168,-0.012)	(0.003,0.126,-0.006)	(0.004,0.108,-0.001)
$c = 0.75$	(-0.024,-0.031,0.192, -0.085)	(-0.016,0.163,-0.064)	(-0.012,0.146,-0.055)
	Model 2		
$c = 0$	(-0.028,0.046,1.510,0.149)	(-0.007,1.149,0.151)	(-0.010,0.923,0.109)
$c = 0.25$	(0.018,0.086,1.601,0.196)	(0.020,1.222,0.172)	(0.030,1.030,0.168)
$c = 0.75$	(0.044,0.012,2.824,0.396)	(0.052,2.394,0.360)	(0.060,2.160,0.328)
	Model 3		
$c = 0$	(-0.022,0.097,2.082,0.240)	(-0.014,1.595,0.201)	(-0.006,1.302,0.173)
$c = 0.25$	(-0.020,0.074,2.160,0.277)	(-0.011,1.609,0.204)	(0.007,1.391,0.219)
$c = 0.75$	(0.152,0.239,3.302,0.603)	(0.160,2.893,0.591)	(0.136,2.633,0.563)

TABLE 2

Average Bias using *Estimated* Bandwidth			
	$n = 200$ (B,M,2S1,M_L,2$S1_K$,2S2)	$n = 350$ (B,2S1,2$S1_K$,2S2)	$n = 500$ (B,2S1,2$S1_K$,2S2)
	Model 1		
$c = 0$	(0.010,-0.459,0.176,0.037,0.213,-0.003)	(0.005,0.118,0.155,-0.007)	(0.005,0.097,0.128,-0.004)
$c = 0.25$	(-0.000,-0.481,0.192,0.016,0.215,-0.007)	(0.003,0.126,0.161,-0.004)	(0.004,0.105,0.138,0.000)
$c = 0.75$	(-0.024,-0.439,0.197,0.005,0.247,-0.084)	(-0.016,0.162,0.209,-0.061)	(-0.012,0.140,0.185,-0.053)
	Model 2		
$c = 0$	(-0.028,5.378,1.976,0.142,2.573,0.150)	(-0.007,1.329,1.968,0.155)	(-0.010,1.022,1.612,0.108)
$c = 0.25$	(0.018,5.398,2.136,0.182,2.652,0.192)	(0.020,1.469,2.055,0.170)	(0.030,1.174,1.720,0.166)
$c = 0.75$	(0.044,5.294,3.411,0.091,3.823,0.348)	(0.052,2.724,3.230,0.333)	(0.060,2.383,2.912,0.300)
	Model 3		
$c = 0$	(-0.022,5.895,2.377,0.206,2.427,0.221)	(-0.014,1.613,1.866,0.184)	(-0.006,1.270,1.532,0.161)
$c = 0.25$	(-0.020,5.842,2.424,0.185,2.524,0.264)	(-0.011,1.617,1.900,0.184)	(0.007,1.359,1.630,0.201)
$c = 0.75$	(0.152,5.944,3.743,0.322,3.698,0.556)	(0.160,3.106,3.222,0.564)	(0.136,2.760,2.926,0.538)

TABLE 3

Average Squared Error using *True* Bandwidth			
	$n = 200$ (B,M,2S1,2S2)	$n = 350$ (B,2S1,2S2)	$n = 500$ (B,2S1,2S2)
	Model 1		
$c = 0$	(0.084,0.127,0.120,0.089)	(0.054,0.071,0.056)	(0.039,0.051,0.041)
$c = 0.25$	(0.090,0.152,0.125,0.095)	(0.056,0.075,0.059)	(0.042,0.056,0.045)
$c = 0.75$	(0.083,1.440,0.150,0.106)	(0.051,0.098,0.068)	(0.040,0.076,0.055)
	Model 2		
$c = 0$	(0.061,0.126,2.550,0.142)	(0.039,1.441,0.090)	(0.028,0.933,0.055)
$c = 0.25$	(0.055,0.122,2.749,0.165)	(0.036,1.561,0.093)	(0.028,1.094,0.073)
$c = 0.75$	(0.062,1.162,8.877,0.783)	(0.038,6.289,0.580)	(0.028,5.037,0.463)
	Model 3		
$c = 0$	(0.079,0.272,4.671,0.235)	(0.051,2.721,0.142)	(0.039,1.809,0.101)
$c = 0.25$	(0.086,0.228,5.095,0.310)	(0.055,2.812,0.161)	(0.040,2.045,0.128)
$c = 0.75$	(0.079,1.630,11.111,0.933)	(0.052,8.330,0.751)	(0.040,6.925,0.648)

TABLE 4

Average Squared Error using *Estimated* Bandwidth			
	$n = 200$ (B,M,2S1,M_L,$2S1_K$,2S2)	$n = 350$ (B,2S1,$2S1_K$,2S2)	$n = 500$ (B,2S1,$2S1_K$,2S2)
	Model 1		
$c = 0$	(0.097,0.150,0.139,0.233,0.150,0.102)	(0.063,0.080,0.088,0.066)	(0.048,0.059,0.063,0.049)
$c = 0.25$	(0.099,0.185,0.147,0.229,0.156,0.104)	(0.065,0.085,0.094,0.067)	(0.047,0.061,0.069,0.050)
$c = 0.75$	(0.095,2.462,0.174,0.397,0.195,0.119)	(0.059,0.106,0.124,0.075)	(0.045,0.080,0.095,0.060)
	Model 2		
$c = 0$	(0.073,0.219,4.572,0.109,7.035,0.180)	(0.045,2.012,4.05,0.105)	(0.031,1.199,2.720,0.062)
$c = 0.25$	(0.069,0.184,5.040,0.108,7.284,0.216)	(0.042,2.339,4.32,0.112)	(0.031,1.458,2.999,0.085)
$c = 0.75$	(0.077,1.847,12.970,0.413,15.677,0.915)	(0.044,8.131,11.099,0.648)	(0.032,6.143,8.950,0.505)
	Model 3		
$c = 0$	(0.092,0.202,6.335,0.195,6.247,0.250)	(0.061,2.905,3.673,0.142)	(0.046,1.808,2.465,0.101)
$c = 0.25$	(0.101,0.406,6.694,0.212,6.819,0.328)	(0.062,2.967,3.853,0.158)	(0.047,2.033,2.772,0.124)
$c = 0.75$	(0.097,1.661,14.472,0.605,13.800,1.025)	(0.063,9.731,10.274,0.778)	(0.049,7.683,8.507,0.656)

TABLE 5.1 - MODEL 1

Average Squared Error by Regression Direction using *True* Bandwidth				
	B-(m_1, m_2)	M-(m_1, m_2)	2S1-(m_1, m_2)	2S2-(m_1, m_2)
		n=200		
$c = 0$	(0.030,0.056)	(0.047,0.079)	(0.038,0.066)	(0.032,0.060)
$c = 0.25$	(0.034,0.060)	(0.063,0.092)	(0.043,0.069)	(0.037,0.063)
$c = 0.75$	(0.033,0.060)	(0.562,0.902)	(0.051,0.087)	(0.046,0.078)
		n=350		
$c = 0$	(0.018,0.036)	-	(0.023,0.041)	(0.019,0.038)
$c = 0.25$	(0.021,0.036)	-	(0.026,0.042)	(0.022,0.038)
$c = 0.75$	(0.019,0.037)	-	(0.033,0.058)	(0.029,0.051)
		n=500		
$c = 0$	(0.012,0.027)	-	(0.016,0.03)	(0.013,0.028)
$c = 0.25$	(0.015,0.028)	-	(0.018,0.03)	(0.015,0.029)
$c = 0.75$	(0.017,0.029)	-	(0.028,0.04)	(0.025,0.040)

TABLE 5.2 - MODEL 2

Average Squared Error by Regression Direction using *True* Bandwidth				
	B-(m_1, m_3)	M-(m_1, m_3)	2S1-(m_1, m_3)	2S2-(m_1, m_3)
		n=200		
$c = 0$	(0.032,0.107)	(0.053,0.142)	(0.650,0.725)	(0.073,0.142)
$c = 0.25$	(0.027,0.103)	(0.049,0.144)	(0.700,0.867)	(0.090,0.150)
$c = 0.75$	(0.036,0.114)	(0.455,0.818)	(2.984,2.668)	(0.915,0.486)
		n=350		
$c = 0$	(0.020,0.071)	-	(0.363,0.436)	(0.040,0.092)
$c = 0.25$	(0.018,0.063)	-	(0.393,0.499)	(0.048,0.087)
$c = 0.75$	(0.021,0.066)	-	(2.106,1.972)	(0.665,0.360)
		n=500		
$c = 0$	(0.014,0.052)	-	(0.231,0.285)	(0.025,0.062)
$c = 0.25$	(0.013,0.042)	-	(0.272,0.364)	(0.036,0.063)
$c = 0.75$	(0.016,0.048)	-	(1.670,1.621)	(0.534,0.300)

TABLE 5.3 - MODEL 3

Average Squared Error by Regression Direction using *True* Bandwidth				
	B-(m_2, m_3)	M-(m_2, m_3)	2S1-(m_2, m_3)	2S2-(m_2, m_3)
		n=200		
$c = 0$	(0.052,0.134)	(0.134,0.235)	(1.205,1.300)	(0.144,0.191)
$c = 0.25$	(0.055,0.121)	(0.113,0.194)	(1.334,1.377)	(0.191,0.188)
$c = 0.75$	(0.057,0.141)	(0.838,0.918)	(3.683,3.466)	(1.082,0.510)
		n=350		
$c = 0$	(0.033,0.071)	-	(0.692,0.758)	(0.077,0.103)
$c = 0.25$	(0.034,0.084)	-	(0.711,0.786)	(0.091,0.120)
$c = 0.75$	(0.036,0.091)	-	(2.785,2.683)	(0.842,0.424)
		n=500		
$c = 0$	(0.025,0.055)	-	(0.456,0.524)	(0.054,0.079)
$c = 0.25$	(0.026,0.056)	-	(0.515,0.600)	(0.066,0.089)
$c = 0.75$	(0.027,0.062)	-	(2.326,2.213)	(0.702,0.352)

TABLE 6.1 - MODEL 1

Average Squared Error by Regression Direction using *Estimated* Bandwidth						
	B-(m_1, m_2)	M-(m_1, m_2)	$2S1-(m_1, m_2)$	M_L-(m_1, m_2)	$2S1_K$-(m_1, m_2)	2S2-(m_1, m_2)
			n=200			
$c = 0$	(0.036,0.064)	(0.062,0.087)	(0.046,0.075)	(0.036,0.202)	(0.053,0.073)	(0.038,0.068)
$c = 0.25$	(0.039,0.064)	(0.080,0.108)	(0.052,0.076)	(0.043,0.196)	(0.057,0.077)	(0.042,0.067)
$c = 0.75$	(0.039,0.068)	(1.037,1.376)	(0.056,0.095)	(0.123,0.273)	(0.064,0.098)	(0.051,0.086)
			n=350			
$c = 0$	(0.020,0.044)	-	(0.025,0.049)	-	(0.031,0.045)	(0.021,0.046)
$c = 0.25$	(0.024,0.043)	-	(0.029,0.049)	-	(0.034,0.047)	(0.025,0.044)
$c = 0.75$	(0.021,0.043)	-	(0.034,0.064)	-	(0.041,0.067)	(0.031,0.056)
			n=500			
$c = 0$	(0.014,0.034)	-	(0.018,0.037)	-	(0.022,0.033)	(0.015,0.035)
$c = 0.25$	(0.016,0.032)	-	(0.020,0.036)	-	(0.024,0.036)	(0.017,0.033)
$c = 0.75$	(0.018,0.033)	-	(0.028,0.050)	-	(0.034,0.053)	(0.026,0.044)

TABLE 6.2 - MODEL 2

Average Squared Error by Regression Direction using *Estimated* Bandwidth						
	B-(m_1, m_3)	M-(m_1, m_3)	$2S1-(m_1, m_3)$	M_L-(m_1, m_3)	$2S1_K$-(m_1, m_2)	2S2-(m_1, m_2)
			n=200			
$c = 0$	(0.036,0.115)	(0.106,0.184)	(1.174,1.179)	(0.047,0.116)	(1.813,1.805)	(0.102,0.156)
$c = 0.25$	(0.033,0.112)	(0.074,0.177)	(1.300,1.435)	(0.043,0.126)	(1.883,2.017)	(0.133,0.166)
$c = 0.75$	(0.041,0.123)	(0.843,1.108)	(4.225,3.655)	(0.208,0.307)	(5.033,4.387)	(1.210,0.527)
			n=350			
$c = 0$	(0.022,0.075)	-	(0.51,0.571)	-	(1.038,1.080)	(0.050,0.096)
$c = 0.25$	(0.020,0.067)	-	(0.59,0.689)	-	(1.109,1.203)	(0.063,0.093)
$c = 0.75$	(0.022,0.070)	-	(2.66,2.440)	-	(3.557,3.217)	(0.808,0.392)
			n=500			
$c = 0$	(0.016,0.054)	-	(0.301,0.349)	-	(0.689,0.723)	(0.030,0.064)
$c = 0.25$	(0.015,0.044)	-	(0.368,0.451)	-	(0.763,0.859)	(0.045,0.066)
$c = 0.75$	(0.016,0.050)	-	(2.001,1.901)	-	(2.861,2.647)	(0.628,0.327)

TABLE 6.3 - MODEL 3

Average Squared Error by Regression Direction using *Estimated* Bandwidth						
	B-(m_2, m_3)	M-(m_2, m_3)	$2S1-(m_2, m_3)$	M_L-(m_2, m_3)	$2S1_K$-(m_2, m_3)	2S2-(m_3, m_3)
			n=200			
$c = 0$	(0.058,0.143)	(0.105,0.186)	(1.644,1.687)	(0.120,0.147)	(1.614,1.678)	(0.159,0.195)
$c = 0.25$	(0.065,0.129)	(0.151,0.334)	(1.755,1.735)	(0.129,0.150)	(1.782,1.789)	(0.210,0.192)
$c = 0.75$	(0.063,0.152)	(0.726,1.042)	(4.691,4.272)	(0.337,0.397)	(4.477,4.128)	(1.249,0.484)
			n=350			
$c = 0$	(0.039,0.074)	-	(0.749,0.785)	-	(0.941,0.983)	(0.079,0.103)
$c = 0.25$	(0.040,0.086)	-	(0.760,0.822)	-	(0.983,1.035)	(0.093,0.120)
$c = 0.75$	(0.041,0.096)	-	(3.203,3.016)	-	(3.367,3.175)	(0.907,0.413)
			n=500			
$c = 0$	(0.029,0.057)	-	(0.462,0.515)	-	(0.628,0.681)	(0.055,0.077)
$c = 0.25$	(0.031,0.058)	-	(0.519,0.586)	-	(0.706,0.777)	(0.068,0.087)
$c = 0.75$	(0.033,0.066)	-	(2.553,2.385)	-	(2.800,2.612)	(0.742,0.343)

TABLE 7.1 - MODEL 1
AVERAGE *True* BANDWIDTHS

	B	M	2S1	2S2
	$n = 200$			
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
$c = 0$	(0.062,0.036)	(0.060,0.036)	(0.061,0.036)	(0.062,0.036)
$c = 0.25$	(0.062,0.036)	(0.061,0.036)	(0.062,0.036)	(0.062,0.036)
$c = 0.75$	(0.063,0.036)	(0.061,0.036)	(0.063,0.036)	(0.063,0.036)
	$n = 350$			
$c = 0$	(0.055,0.032)	-	(0.055,0.032)	(0.055,0.032)
$c = 0.25$	(0.055,0.032)	-	(0.055,0.032)	(0.055,0.032)
$c = 0.75$	(0.056,0.032)	-	(0.056,0.032)	(0.056,0.032)
	$n = 500$			
$c = 0$	(0.051,0.030)	-	(0.051,0.030)	(0.051,0.030)
$c = 0.25$	(0.051,0.030)	-	(0.051,0.030)	(0.051,0.030)
$c = 0.75$	(0.052,0.030)	-	(0.052,0.030)	(0.052,0.030)

TABLE 7.2 - MODEL 2
AVERAGE *True* BANDWIDTHS

	B	M	2S1	2S2
	$n = 200$			
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
$c = 0$	(0.061,0.066)	(0.057,0.040)	(0.055,0.059)	(0.061,0.066)
$c = 0.25$	(0.062,0.067)	(0.057,0.040)	(0.055,0.059)	(0.062,0.066)
$c = 0.75$	(0.063,0.067)	(0.058,0.040)	(0.056,0.059)	(0.063,0.067)
	$n = 350$			
$c = 0$	(0.055,0.059)	-	(0.049,0.052)	(0.055,0.059)
$c = 0.25$	(0.055,0.059)	-	(0.049,0.052)	(0.055,0.059)
$c = 0.75$	(0.056,0.060)	-	(0.050,0.053)	(0.056,0.060)
	$n = 500$			
$c = 0$	(0.051,0.055)	-	(0.046,0.049)	(0.051,0.055)
$c = 0.25$	(0.051,0.055)	-	(0.046,0.049)	(0.051,0.055)
$c = 0.75$	(0.052,0.056)	-	(0.046,0.049)	(0.052,0.056)

TABLE 7.3 - MODEL 3
AVERAGE *True* BANDWIDTHS

	B	M	2S1	2S2
	$n = 200$			
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
$c = 0$	(0.036,0.066)	(0.037,0.044)	(0.035,0.061)	(0.036,0.066)
$c = 0.25$	(0.036,0.066)	(0.037,0.044)	(0.035,0.062)	(0.036,0.066)
$c = 0.75$	(0.036,0.068)	(0.037,0.044)	(0.035,0.062)	(0.036,0.067)
	$n = 350$			
$c = 0$	(0.032,0.059)	-	(0.032,0.055)	(0.032,0.059)
$c = 0.25$	(0.032,0.059)	-	(0.032,0.055)	(0.032,0.059)
$c = 0.75$	(0.032,0.060)	-	(0.031,0.056)	(0.032,0.060)
	$n = 500$			
$c = 0$	(0.030,0.055)	-	(0.029,0.051)	(0.030,0.055)
$c = 0.25$	(0.030,0.055)	-	(0.029,0.051)	(0.030,0.055)
$c = 0.75$	(0.030,0.056)	-	(0.029,0.052)	(0.030,0.056)

TABLE 8.1 - MODEL 1

Average *Estimated* Bandwidths						
	B	M	2S1	M_L	$2S1_K$	2S2
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
			$n = 200$			
$c = 0$	(0.054,0.042)	(0.053,0.041)	(0.054,0.042)	(0.095,0.095)	(0.042,0.042)	(0.054,0.042)
$c = 0.25$	(0.054,0.040)	(0.052,0.039)	(0.054,0.040)	(0.093,0.093)	(0.042,0.042)	(0.054,0.040)
$c = 0.75$	(0.057,0.043)	(0.055,0.041)	(0.057,0.043)	(0.098,0.098)	(0.041,0.041)	(0.057,0.043)
			$n = 350$			
$c = 0$	(0.051,0.039)	-	(0.051,0.039)	-	(0.038,0.038)	(0.051,0.0394)
$c = 0.25$	(0.050,0.038)	-	(0.050,0.038)	-	(0.038,0.038)	(0.050,0.0388)
$c = 0.75$	(0.052,0.038)	-	(0.052,0.038)	-	(0.036,0.036)	(0.052,0.0389)
			$n = 500$			
$c = 0$	(0.047,0.037)	-	(0.047,0.037)	-	(0.035,0.035)	(0.047,0.037)
$c = 0.25$	(0.048,0.035)	-	(0.048,0.035)	-	(0.035,0.035)	(0.048,0.035)
$c = 0.75$	(0.049,0.036)	-	(0.049,0.036)	-	(0.034,0.034)	(0.049,0.036)

TABLE 8.2 - MODEL 2

Average *Estimated* Bandwidths						
	B	M	2S1	M_L	$2S1_K$	2S2
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
			$n = 200$			
$c = 0$	(0.054,0.048)	(0.051,0.036)	(0.054,0.048)	(0.069,0.069)	(0.042,0.042)	(0.054,0.048)
$c = 0.25$	(0.054,0.046)	(0.051,0.035)	(0.054,0.046)	(0.066,0.066)	(0.042,0.042)	(0.054,0.046)
$c = 0.75$	(0.055,0.041)	(0.052,0.035)	(0.054,0.041)	(0.068,0.068)	(0.041,0.041)	(0.055,0.041)
			$n = 350$			
$c = 0$	(0.050,0.047)	-	(0.050,0.047)	-	(0.038,0.038)	(0.050,0.0476)
$c = 0.25$	(0.049,0.045)	-	(0.049,0.045)	-	(0.038,0.038)	(0.049,0.0450)
$c = 0.75$	(0.051,0.040)	-	(0.051,0.040)	-	(0.036,0.036)	(0.051,0.0401)
			$n = 500$			
$c = 0$	(0.047,0.045)	-	(0.047,0.045)	-	(0.035,0.035)	(0.047,0.045)
$c = 0.25$	(0.047,0.043)	-	(0.047,0.043)	-	(0.035,0.035)	(0.047,0.043)
$c = 0.75$	(0.049,0.038)	-	(0.049,0.038)	-	(0.034,0.034)	(0.049,0.038)

TABLE 8.3 - MODEL 3

Average *Estimated* Bandwidths						
	B	M	2S1	M_L	$2S1_K$	2S2
	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)	(h_1, h_2)
			$n = 200$			
$c = 0$	(0.041,0.047)	(0.042,0.039)	(0.041,0.047)	(0.064,0.064)	(0.042,0.042)	(0.041,0.047)
$c = 0.25$	(0.043,0.048)	(0.042,0.040)	(0.043,0.047)	(0.066,0.066)	(0.042,0.042)	(0.043,0.048)
$c = 0.75$	(0.043,0.040)	(0.042,0.037)	(0.043,0.040)	(0.065,0.065)	(0.041,0.041)	(0.043,0.040)
			$n = 350$			
$c = 0$	(0.039,0.047)	-	(0.039,0.047)	-	(0.038,0.038)	(0.039,0.047)
$c = 0.25$	(0.039,0.046)	-	(0.039,0.046)	-	(0.038,0.038)	(0.039,0.046)
$c = 0.75$	(0.039,0.040)	-	(0.039,0.040)	-	(0.036,0.036)	(0.039,0.040)
			$n = 500$			
$c = 0$	(0.036,0.045)	-	(0.036,0.045)	-	(0.035,0.035)	(0.036,0.045)
$c = 0.25$	(0.036,0.045)	-	(0.036,0.045)	-	(0.035,0.035)	(0.036,0.045)
$c = 0.75$	(0.036,0.039)	-	(0.036,0.039)	-	(0.034,0.034)	(0.036,0.039)

8 Appendix

Theorem 1. *Proof* Let $\epsilon' = (\epsilon_1, \cdots, \epsilon_n)$ where $\epsilon_t = y_t - \alpha - m_1(x_t) - m_2(z_t)$. By construction,

$$\begin{aligned}
\hat{m}_1^{2S1}(x) &= s_1(x)(\vec{y} - \vec{\gamma}_2^E(\vec{z})) \\
&= s_1(x)(1_n\alpha + \vec{m}_1(\vec{x}) + \vec{m}_2(\vec{z}) + \epsilon - \vec{\gamma}_2^E(\vec{z})) \\
&= s_1(x)(\vec{m}_1(\vec{x}) + \epsilon) + s_1(x)(\vec{\gamma}_2(\vec{z}) - \vec{\gamma}_2^E(\vec{z})) \qquad (31)
\end{aligned}$$

where $\vec{\gamma}_2(\vec{z})' = (\gamma_2(z_1), \cdots, \gamma_2(z_n))$. Under our assumptions, using the results of Fan (1992) for local linear estimation,

$$E\left(s_1(x)(\vec{m}_1(\vec{x}) + \epsilon)\,|\vec{x}, \vec{z}\right) = s_1(x)\vec{m}_1(\vec{x}) = m_1(x) + \frac{h_1^2}{2}\mu_2 d^2 m_1(x) + o_p(h_1^2). \qquad (32)$$

We now look at the second term in (31). Note that the i^{th} element of $-(\vec{\gamma}_2(\vec{z}) - \vec{\gamma}_2^E(\vec{z}))$ is

$$\begin{aligned}
\gamma_2^E(z_i) - \gamma_2(z_i) &= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}y_k - \gamma_2(z_i) \\
&= \hat{L}_{1n}(z_i) + \hat{L}_{2n}(z_i) + \hat{L}_{3n}(z_i)
\end{aligned}$$

where

$$\begin{aligned}
\hat{L}_{1n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}\epsilon_k \\
\hat{L}_{2n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}(m_2(z_k) - m_2(z_i)) \\
\hat{L}_{3n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n}\left\{\frac{1}{g_2}K_1(\frac{z_k - z_i}{g_2})\frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}m(x_k, z_i) - \gamma_2(z_i)\right\}.
\end{aligned}$$

If, $\hat{L}_1' = (\hat{L}_{1n}(z_1), \cdots, \hat{L}_{1n}(z_n))$, $\hat{L}_2' = (\hat{L}_{2n}(z_1), \cdots, \hat{L}_{2n}(z_n))$, $\hat{L}_3' = (\hat{L}_{3n}(z_1), \cdots, \hat{L}_{3n}(z_n))$, then the last term in (31) can be written as

$$s_1(x)(\vec{\gamma}_2(\vec{z}) - \vec{\gamma}_2^E(\vec{z})) = -s_1(x)(\hat{L}_1 + \hat{L}_2 + \hat{L}_3). \qquad (33)$$

and $E(s_1(x)(\vec{\gamma}_2(\vec{z}) - \vec{\gamma}_2^E(\vec{z}))|\vec{x}, \vec{z}) = -s_1(x)(E(\hat{L}_1|\vec{x}, \vec{z}) + E(\hat{L}_2|\vec{x}, \vec{z}) + E(\hat{L}_3|\vec{x}, \vec{z}))$. By assumption $E(\hat{L}_1|\vec{x}, \vec{z}) = 0$, we now treat $\hat{L}_2$ and $\hat{L}_3$ separately. In what follows we define $\bar{f}_X(x) = E(\hat{f}_X(x)) = g_1^{-1}\int K_1\left(\frac{v-x}{g_1}\right)f_X(v)dv$ and $\bar{f}_{XZ}(x, z) = E(\hat{f}_{XZ}(x, z)) = (g_1 g_2)^{-1}\int\int K_1\left(\frac{v-x}{g_1}\right)K_1\left(\frac{u-z}{g_2}\right)f_{XZ}(u, v)dudv$, and

$$L_{2n}(z_i) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)}(m_2(z_k) - m_2(z_i)).$$

Given that there exists $0 < B_{dm_2}$ such that $|dm_2(z)| < B_{dm_2}$ for all $z \in S_Z$ and S_Z is compact, we have that by using the Mean Value Theorem

$$|\hat{L}_{2n}(z_i) - L_{2n}(z_i)| \leq B_{dm_2} B_Z sup_{(x_k,z_k)\in S_X \times S_Z} \left| \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)} - \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} \right| \hat{f}_Z(z_i) \tag{34}$$

for a bound B_Z on $|z_k - z_i|$. Hence, given that there exists $0 < B_1, B_{f_Z}, B_{f_X}$ such that $|f_Z(z)| < B_{f_Z}$ for all $z \in S_Z$, $|f_X(x)| < B_{f_X}$ for all $x \in S_X$, and $|f_{XZ}|^{-1} < B_1$ for all $(x, z) \in S_X \times S_Z$ which follows from Assumption 2, we have

$$\begin{aligned}
sup_{z_i \in S_Z} |\hat{L}_{2n}(z_i) - L_{2n}(z_i)| &\leq B_{dm_2} B_Z sup_{(x_k,z_k)\in S_X \times S_Z} \left| \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)} - \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} \right| \\
&\times \left(sup_{z_i \in S_Z} |\hat{f}_Z(z_i) - \bar{f}_Z(z_i)| + B_{f_Z} \right) \\
&\leq \left(sup_{z_i \in S_Z} |\hat{f}_Z(z_i) - \bar{f}_Z(z_i)| + B_{f_Z} \right) B_{dm_2} B_Z \Big(sup_{x_k \in S_X} |\hat{f}_X(x_k) - \bar{f}_X(x_k)| \\
&\times sup_{(x_k,z_k)\in S_X \times S_Z} |\hat{f}^{-1}_{XZ}(x_k, z_k) - \bar{f}^{-1}_{XZ}(x_k, z_k)| + B_1 sup_{x_k \in S_X} |\hat{f}_X(x_k) - \bar{f}_X(x_k)| \\
&+ B_{f_X} sup_{(x_k,z_k)\in S_X \times S_Z} |\hat{f}^{-1}_{XZ}(x_k, z_k) - \bar{f}^{-1}_{XZ}(x_k, z_k)| \Big)
\end{aligned}$$

Let

$$l^n_{21}(z_i) = dm_2(z_i) \frac{1}{n} \sum_{k=1}^{n} \frac{1}{g_2} K_1 \left(\frac{z_k - z_i}{g_2} \right) \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} (z_k - z_i),$$

$$l^n_{22}(z_i) = \frac{1}{2} d^2 m_2(z_i) \frac{1}{n} \sum_{k=1}^{n} \frac{1}{g_2} K_1 \left(\frac{z_k - z_i}{g_2} \right) \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} (z_k - z_i)^2,$$

$b^n_2(z_i) = \frac{\delta}{2} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{g_2} K_1 \left(\frac{z_k - z_i}{g_2} \right) \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} (z_k - z_i)^2$ and $a^n(z_i) = \frac{\mu_2}{2} g_2^2 d^2 m_2(z_i)$. Then, for any $\delta > 0$ by Taylor's Theorem we can write, $|L_{2n}(z_i) - l^n_{21}(z_i) - l^n_{22}(z_i)| < \frac{\delta}{2} b^n_2(z_i)$ and by the triangle inequality we obtain, $|L_{2n}(z_i) - a^n(z_i)| < |l^n_{21}(z_i)| + |l^n_{22}(z_i) - a^n(z_i)| + \frac{\delta}{2} |b^n_2(z_i) - g_2^2 \mu_2| + \frac{\delta}{2} g_2^2 \mu_2$ hence, we have

$$sup_{z_i \in S_Z} |L_{2n}(z_i) - a^n(z_i)| < sup_{z_i \in S_Z} |l^n_{21}(z_i)| + sup_{z_i \in S_Z} |l^n_{22}(z_i) - a^n(z_i)| + \frac{\delta}{2} sup_{z_i \in S_Z} |b^n_2(z_i) - g_2^2 \mu_2|$$

and

$$P\left(sup_{z_i \in S_Z} |L_{2n}(z_i) - a^n(z_i)| > \epsilon\right) < P\left(sup_{z_i \in S_Z} |l^n_{21}(z_i)| > \epsilon\right) + P\left(sup_{z_i \in S_Z} |l^n_{22}(z_i) - a^n(z_i)| > \epsilon\right) +$$

$$P\left(sup_{z_i \in S_Z} |b^n_2(z_i) - g_2^2 \mu_2| > \frac{2\epsilon}{\delta} \right).$$

Finally, since $sup_{z_i \in S_Z}|\hat{L}_{2n}(z_i) - a^n(z_i)| \leq sup_{z_i \in S_Z}|\hat{L}_{2n}(z_i) - L_{2n}(z_i)| + sup_{z_i \in S_Z}|L_{2n}(z_i) - a^n(z_i)|$ we conclude that,$\hat{L}_{2n}(z_i) = \frac{\mu_2}{2} g_2^2 d^2 m_2(z_i) + o_p(g_2^2)$ uniformly in S_Z.

Let $L_{3n}(z_i) = \frac{1}{n}\sum_{k=1}^{n}\{\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}(x_k,z_k)}m(x_k,z_i) - \gamma_2(z_i)\}$, then if there exists $0 < B_m$ such that $m(x,z) < B_m$ for all $(x,z) \in S_X \times S_Z$ we have

$$|\hat{L}_{3n}(z_i) - L_{3n}(z_i)| \leq B_m sup_{(x_k,z_k)\in S_X \times S_Z}\left|\frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k,z_k)} - \frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}\right| sup\hat{f}_Z(z_i)$$

which is similar in structure to inequality (34). Hence, using the same arguments we have that $\hat{L}_{3n}(z_i) = L_{3n}(z_i) + o_p(g_2^2)$ uniformly in S_Z. Let

$$A_1^n(z_i) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)} \text{ and}$$

$$A_2^n(z_i) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}m_1(x_k).$$

Now note that,

$$\begin{aligned} L_3^n(z_i) &= \alpha A_1^n(z_i) + A_2^n(z_i) + m_2(z_i)A_1^n(z_i) - \gamma_2(z_i) \\ &= \gamma_2(z_i)(A_1^n(z_i) - 1) + A_2^n(z_i) \end{aligned}$$

and consequently $E(L_3^n(z_i)) = \gamma_2(z_i)(E(A_1^n(z_i)) - 1) + E(A_2^n(z_i))$. We look at each expectation separately.

$$\begin{aligned} E(A_1^n(z_i)) &= \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}f_{XZ}(x_k,z_k)dx_k dz_k \\ &= \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\bar{f}_X(x_k)dx_k dz_k - \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}(\bar{f}_{XZ}(x_k,z_k) \\ &- f_{XZ}(x_k,z_k))dx_k dz_k \\ &= \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)(\bar{f}_X(x_k) - f_X(x_k))dx_k dz_k + \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)f_X(x_k)dx_k dz_k \\ &- \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}(\bar{f}_{XZ}(x_k,z_k) - f_{XZ}(x_k,z_k))dx_k dz_k \\ &= 1 + \int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)(\bar{f}_X(x_k) - f_X(x_k))dx_k dz_k - \left(\int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\right. \\ &\times \left(\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)} - \frac{f_X(x_k)}{f_{XZ}(x_k,z_k)}\right)(\bar{f}_{XZ}(x_k,z_k) - f_{XZ}(x_k,z_k))dx_k dz_k \\ &+ \left.\int\int \frac{1}{g_2}K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{f_X(x_k)}{f_{XZ}(x_k,z_k)}(\bar{f}_{XZ}(x_k,z_k) - f_{XZ}(x_k,z_k))dx_k dz_k\right) \\ &= 1 + c_1^n(z_i) - (c_2^n(z_i) + c_3^n(z_i)). \end{aligned}$$

Also, using the fact that $E(m_1(X)) = 0$ we can similarly write,

$$
\begin{aligned}
E(A_2^n(z_i)) &= \int\int \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right)(\bar{f}_X(x_k) - f_X(x_k))m_1(x_k)dx_k dz_k \\
&+ \int\int \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right)\frac{f_X(x_k)}{f_{XZ}(x_k, z_k)}(\bar{f}_{XZ}(x_k, z_k) - f_{XZ}(x_k, z_k))m_1(x_k)dx_k dz_k \\
&- \int\int \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right)\left(\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k, z_k)} - \frac{f_X(x_k)}{f_{XZ}(x_k, z_k)}\right)(\bar{f}_{XZ}(x_k, z_k) - f_{XZ}(x_k, z_k))dx_k dz_k \\
&= D_1^n(z_i) + D_2^n(z_i) + D_3^n(z_i)
\end{aligned}
$$

By Taylor's Theorem, for all $(x, z) \in S_X \times S_Z$ and $\delta > 0$

$$
-g_1^2 \frac{1}{2} f_X^{(2)} \mu_2 - \frac{1}{2}\mu_2 g_1^2 \delta < \bar{f}_X(x) - f_X(x) < g_1^2 \frac{1}{2} d^2 f_X(x)\mu_2 + \frac{1}{2}\mu_2 g_1^2 \delta
$$

and

$$
-\frac{\mu_2}{2}\sum_{i=1}^{2}\frac{\partial^2 f_{XZ}(x, z)}{\partial_i \partial_i} g_i^2 - \frac{\mu_2}{2}\sum_{i=1}^{2} g_i^2 \delta < \bar{f}_{XZ}(x, z) - f_{XZ}(x, z) < \frac{\mu_2}{2}\sum_{i=1}^{2}\frac{\partial^2 f_{XZ}(x, z)}{\partial_i \partial_i} g_i^2 + \frac{\mu_2}{2}\sum_{i=1}^{2} g_i^2 \delta.
$$

Therefore, given Assumption 2 and provided that S_X is bounded

$$
\frac{C_1^n(z_i)}{g_2^2} = \frac{\mu_2}{2}\int d^2 f_X(v)dv + o(1)
$$

uniformly in S_Z.

$$
\frac{C_3^n(z_i)}{g_2^2} = \frac{\mu_2}{2}\int f_X(v)\frac{\partial^2 f_{XZ}(v, z_i)}{\partial_1 \partial_1}\frac{1}{f_{XZ}(v, z_i)}dv + \frac{\mu_2}{2}\int f_X(v)\frac{\partial^2 f_{XZ}(v, z_i)}{\partial_2 \partial_2}\frac{1}{f_{XZ}(v, z_i)}dv + o(1)
$$

uniformly in S_Z. We ignore $C_2^n(z_i)$ as it is of order smaller than $C_1^n(z_i)$ and $C_3^n(z_i)$. Also,

$$
\frac{D_1^n(z_i)}{g_2^2} = \frac{\mu_2}{2}\int f_X^{(2)}(v)m_1(v)dv + o(1)
$$

uniformly in S_Z.

$$
\frac{D_2^n(z_i)}{g_2^2} = \frac{\mu_2}{2}\int f_X(v)m_1(v)\frac{\partial^2 f_{XZ}(v, z_i)}{\partial_1 \partial_1}\frac{1}{f_{XZ}(v, z_i)}dv + \frac{\mu_2}{2}\int f_X(v)m_1(v)\frac{\partial^2 f_{XZ}(v, z_i)}{\partial_2 \partial_2}\frac{1}{f_{XZ}(v, z_i)}dv + o(1)
$$

uniformly in S_Z. As above, we ignore $D_3^n(z_i)$ as it is of order smaller than $D_1^n(z_i)$ and $D_3^n(z_i)$. Now, note that

$$
sup_{z_i \in S_Z}|L_{3n}(z_i) - E(L_{3n}(z_i))| \leq B_m sup_{z_i \in S_Z}|A_1^n(z_i) - E(A_1^n(z_i))| + sup_{z_i \in S_Z}|A_2^n(z_i) - E(A_2^n(z_i))|.
$$

By Lemma 1

$$\frac{1}{g_2^2} sup_{z_i \in S_X} |L_{3n}(z_i) - E(L_{3n}(z_i))| = o_p(1) \tag{35}$$

given that $\frac{\bar{f}_X(x_k)}{\bar{f}_{XZ}(x_k,z_k)}$ is bounded. Now, let $\tau_n(z_i) = \gamma_2(z_i)(T_{1n}(z_i) - 1) - T_{2n}(z_i)$, where

$$T_{1n}(z_i) = 1+\frac{\mu_2}{2}g_2^2 \int d^2 f_X(v)dv - \frac{\mu_2}{2}g_2^2 \int f_X(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_1\partial_1}\frac{1}{f_{XZ}(v,z_i)}dv + \frac{\mu_2}{2}g_2^2 \int f_X(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_2\partial_2}\frac{1}{f_{XZ}(v,z_i)}dv$$

$$\begin{aligned} T_{2n}(z_i) &= \frac{\mu_2}{2}g_2^2 \int m_1(v) d^2 f_X(v)dv + \frac{\mu_2}{2}g_2^2 \int f_X(v)m_1(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_1\partial_1}\frac{1}{f_{XZ}(v,z_i)}dv \\ &+ \frac{\mu_2}{2}g_2^2 \int f_X(v)m_1(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_2\partial_2}\frac{1}{f_{XZ}(v,z_i)}dv \end{aligned}$$

then,

$$\frac{1}{g_2^2} sup_{z_i \in S_X} |E(L_{3n}(z_i)) - \tau_n(z_i)| \le \frac{1}{g_2^2} sup_{z_i \in S_X} |E(A_1^n(z_i)) - T_{1n}(z_i)| + \frac{1}{g_2^2} sup_{z_i \in S_X} |E(A_2^n(z_i)) - T_{2n}(z_i)| \tag{36}$$

Hence,

$$\begin{aligned} \frac{1}{g_2^2} sup_{z_i \in S_X} |\hat{L}_{3n}(z_i) - \tau_n(z_i)| &\le \frac{1}{g_2^2} sup_{z_i \in S_X} |\hat{L}_{3n}(z_i) - L_{3n}(z_i)| + \frac{1}{g_2^2} sup_{z_i \in S_X} |L_{3n}(z_i) - E(L_{3n}(z_i))| \\ &+ \frac{1}{g_2^2} sup_{z_i \in S_X} |E(L_{3n}(z_i)) - \tau_n(z_i)| \end{aligned}$$

and combining equations (34),(35) and (36) we obtain

$$\begin{aligned} \hat{L}_{3n}(z_i) &= \frac{\mu_2}{2}g_2^2 \int d^2 f_X(v)m(v,z_i)dv + \frac{\mu_2}{2}g_2^2 \left(\int f_X(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_1\partial_1} f_{XZ}^{-1}(v,z_i)dv \right. \\ &+ \left. \int f_X(v)\frac{\partial^2 f_{XZ}(v,z_i)}{\partial_2\partial_2} f_{XZ}^{-1}(v,z_i)dv \right) + o_p(g_2^2) \end{aligned}$$

uniformly in S_Z. Hence, combining the approximations for $\hat{L}_{2n}(z_i)$ and $\hat{L}_{3n}(z_i)$ we have

$$s_1(x)\hat{L}_2 = \frac{1}{2}g_2^2\mu_2 E(d^2 m_2(z_i)|\vec{x}) + o_p(h_1^2) + o_p(g_2^2)$$

and

$$\begin{aligned} s_1(x)\hat{L}_3 &= \frac{1}{2}g_2^2\mu_2 E\left(\int d^2 f_X(v)m(v,z_i)dv|\vec{x} \right) \\ &+ \frac{1}{2}g_2^2\mu_2 E\left(\int f_X(v) f_{XZ}^{-1}(v,z_i) \sum_{d=1}^{2} \frac{\partial^2 f_{XZ}(v,z_i)}{\partial_d\partial_d}dv|\vec{x} \right) + o_p(h_1^2) + o_p(g_2^2) \end{aligned}$$

which completes the proof of part a) of the theorem.

b) Let $[a_{ij}]_{i=1,j=1}^{m,p}$ denote an $m \times p$ matrix with typical element a_{ij}. We write $\vec{y} - \vec{\gamma}_2^E(\vec{z}) = (I - \frac{1}{ng_2}B_n)\vec{y}$ where

$$B_n = \left[K_1\left(\frac{z_j - z_i}{g_2}\right) \frac{\hat{f}_X(x_j)}{\hat{f}_{XZ}(x_j, z_j)} \right]_{i=1,j=1}^{n,n}.$$

Hence, $E(\hat{m}_1^{2S1}(x)|\vec{x}, \vec{z}) = s_1(x)\left(I - \frac{1}{ng_2}B_n\right)\vec{m}$ and

$$\begin{aligned}
nh_1 V\left(\hat{m}_1^{2S1}(x)|\vec{x}, \vec{z}\right) &= nh_1\sigma^2 s_1(x)\left(I - \frac{1}{ng_2}B_n\right)\left(I - \frac{1}{ng_2}B_n\right)' s_1(x)' && (37)\\
&= \sigma^2\left(nh_1 s_1(x)s_1(x)' - \frac{nh_1}{ng_2}s_1(x)B's_1(x)' - \frac{nh_1}{ng_2}Bs_1(x)'\right. && (38)\\
&+ \left.\frac{nh_1}{n^2g_2^2}s_1(x)BB's_1(x)'\right) && (39)\\
&= \sigma^2(V_{1n}(x) + V_{2n}(x) + V_{3n}(x) + V_{4n}(x)) && (40)
\end{aligned}$$

From Fan(1992) $V_{1n}(x) \xrightarrow{p} \frac{1}{f_X(x)}\int K^2(v)dv$. Now,

$$V_{2n}(x) = e_1^2\left(\frac{R_X'(x)W_X(x)R_X(x)}{nh_1}\right)^{-1}\frac{R_X'(x)W_X(x)B_nW_X(x)R_X(x)}{n^2g_2h_1}\left(\frac{R_X'(x)W_X(x)R_X(x)}{nh_1}\right)^{-1}e_1^{2'},$$

since $\left(\frac{R_X'(x)W_X(x)R_X(x)}{nh_1}\right)^{-1}$ converges in probability to a finite matrix we focus on the matrix that appears in the middle, which can be written as

$$\frac{R_X'(x)W_X(x)B_nW_X(x)R_X(x)}{n^2g_2h_1} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix},$$

where

$$m_{11} = h_1\frac{1}{nh_1}\sum_{i=1}^n K_1\left(\frac{x_i - x}{h_1}\right)\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i, z_i)}\frac{1}{nh_1g_2}\sum_{j=1}^n K_1\left(\frac{z_i - z_j}{g_2}\right)K_1\left(\frac{x_j - x}{h_1}\right),$$

$$m_{12} = h_1^2\frac{1}{nh_1}\sum_{i=1}^n K_1\left(\frac{x_i - x}{h_1}\right)\frac{x_i - x}{h_1}\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i, z_i)}\frac{1}{nh_1g_2}\sum_{j=1}^n K_1\left(\frac{z_i - z_j}{g_2}\right)K_1\left(\frac{x_j - x}{h_1}\right),$$

$$m_{21} = h_1^2\frac{1}{nh_1}\sum_{i=1}^n K_1\left(\frac{x_i - x}{h_1}\right)\frac{x_i - x}{h_1}\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i, z_i)}\frac{1}{nh_1g_2}\sum_{j=1}^n K_1\left(\frac{z_i - z_j}{g_2}\right)K_1\left(\frac{x_j - x}{h_1}\right)\frac{x_j - x}{h_1},$$

$$m_{22} = h_1^3\frac{1}{nh_1}\sum_{i=1}^n \frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i, z_i)}K_1\left(\frac{x_i - x}{h_1}\right)\frac{x_i - x}{h_1}\frac{1}{nh_1g_2}\sum_{j=1}^n K_1\left(\frac{z_i - z_j}{g_2}\right)K_1\left(\frac{x_j - x}{h_1}\right)\frac{x_j - x}{h_1}$$

We now show that $m_{ij} = o_p(1)$ for all i, j. First,

$$\begin{aligned} m_{11} &= h_1 \frac{1}{nh_1} \sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right)(\hat{f}_{XZ}(x,z_i) - f_{XZ}(x,z_i))\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)} - \frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right) \\ &+ h_1 \frac{1}{nh_1} \sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right)(\hat{f}_{XZ}(x,z_i) - f_{XZ}(x,z_i))\frac{f_X(x_i)}{f_{XZ}(x_i,z_i)} \\ &+ h_1 \frac{1}{nh_1} \sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right) f_{XZ}(x,z_i)\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)} - \frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right) \\ &+ h_1 \frac{1}{nh_1} \sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right) f_{XZ}(x,z_i)\frac{f_X(x_i)}{f_{XZ}(x_i,z_i)} \\ &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

and $|m_{11}| \leq |\alpha_1| + |\alpha_2| + |\alpha_3| + |\alpha_4|$. Now we note that

$$|\alpha_1| \leq h_1 sup_{z_i \in S_Z}|\hat{f}_{XZ}(x,z_i) - f_{XZ}(x,z_i)| sup_{x_i,z_i \in S_X \times S_Z}\left|\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)} - \frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right| \frac{1}{nh_1}\sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right),$$

$$|\alpha_2| \leq h_1 B_{f_X} B_{f_{X,Z}}^{-1} sup_{z_i \in S_Z}|\hat{f}_{XZ}(x,z_i) - f_{XZ}(x,z_i)| \frac{1}{nh_1}\sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right),$$

$$|\alpha_3| \leq h_1 B_{f_{X,Z}} sup_{x_i,z_i \in S_X \times S_Z}\left|\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)} - \frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right| \frac{1}{nh_1}\sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right),$$

$$|\alpha_4| \leq h_1 B_{f_X} \frac{1}{nh_1}\sum_{i=1}^{n} K_1\left(\frac{x_i - x}{h_1}\right).$$

Hence, by Lemma 1 all expressions to the left of the inequalities converge in probability to zero, and consequently $m_{11} \xrightarrow{p} 0$. Now, note that m_{12} (m_{21} is identical in structure) is identical to m_{11} except for the presence in the summand of $\frac{x_i - x}{h_1}$, but since $K(\cdot) = 0$ outside of its compact support and S_X compact we have by Lemma 1 that $m_{12} \xrightarrow{p} 0$. The same argument is also applied to show that $m_{22} \xrightarrow{p} 0$ and therefore $V_{2n}, V_{3n} \xrightarrow{p} 0$.

$$V_{4n} = e_1^{2'}\left(\frac{R'_X(x)W_X(x)R_X(x)}{nh_1}\right)^{-1} \frac{R'_X(x)W_X(x)B_nB'_nW_X(x)R_X(x)}{n^3g_2^2h_1}\left(\frac{R'_X(x)W_X(x)R_X(x)}{nh_1}\right)^{-1} e_1^{2'}$$

and as in the case of V_{2n} we focus on showing that the matrix $\frac{R'_X(x)W_X(x)B_nB'_nW_X(x)R_X(x)}{n^3g_2^2h_1}$ converges in probability to zero. Note that,

$$\frac{R'_X(x)W_X(x)B_nB'_nW_X(x)R_X(x)}{n^3g_2^2h_1} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

where

$$u_{11} = g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \frac{1}{ng_2^2}\sum_{l=1}^{n} K_1\left(\frac{x_l-x}{h_1}\right) K_1\left(\frac{z_i-z_l}{g_2}\right) \frac{1}{nh_1g_2}\sum_{j=1}^{n} K_1\left(\frac{z_i-z_j}{g_2}\right) K_1\left(\frac{x_j-x}{h_1}\right),$$

$$\begin{aligned} u_{12} &= h_1 g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \frac{1}{ng_2^2}\sum_{l=1}^{n} K_1\left(\frac{x_l-x}{h_1}\right) K_1\left(\frac{z_i-z_l}{g_2}\right) \\ &\times \frac{1}{nh_1g_2}\sum_{j=1}^{n} K_1\left(\frac{z_i-z_j}{g_2}\right) K_1\left(\frac{x_j-x}{h_1}\right)\frac{x_j-x}{h_1}, \end{aligned}$$

$$\begin{aligned} u_{21} &= h_1 g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \frac{1}{ng_2^2}\sum_{l=1}^{n} K_1\left(\frac{x_l-x}{h_1}\right) \frac{x_l-x}{h_1} K_1\left(\frac{z_i-z_l}{g_2}\right) \\ &\times \frac{1}{nh_1g_2}\sum_{j=1}^{n} K_1\left(\frac{z_i-z_j}{g_2}\right) K_1\left(\frac{x_j-x}{h_1}\right), \end{aligned}$$

$$\begin{aligned} u_{22} &= h_1^2 g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \frac{1}{ng_2^2}\sum_{l=1}^{n} K_1\left(\frac{x_l-x}{h_1}\right) \frac{x_l-x}{h_1} K_1\left(\frac{z_i-z_l}{g_2}\right) \\ &\times \frac{1}{nh_1g_2}\sum_{j=1}^{n} K_1\left(\frac{z_i-z_j}{g_2}\right) K_1\left(\frac{x_j-x}{h_1}\right)\frac{x_j-x}{h_1}. \end{aligned}$$

We will argue that $u_{ij} \xrightarrow{p} 0$ for all i, j. First, we observe that

$$\begin{aligned} |u_{11}| &= g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \left|\hat{f}_{XZ}^2(x,z_i) - f_{XZ}^2(x,z_i)\right| + g_2 \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 f_{XZ}^2(x,z_i) \\ &\leq g_2\left(sup_{z_i\in S_Z}\left|\hat{f}_{XZ}^2(x,z_i) - f_{XZ}^2(x,z_i)\right| + B_{f_{XZ}}^2\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 \\ &\leq g_2\left(sup_{z_i\in S_Z}\left|\hat{f}_{XZ}^2(x,z_i) - f_{XZ}^2(x,z_i)\right| + B_{f_{XZ}}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 - \left(\frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right)^2\right| + B\right) \\ &\leq g_2\left(sup_{z_i\in S_Z}\left|\hat{f}_{XZ}^2(x,z_i) - f_{XZ}^2(x,z_i)\right| + B_{f_{XZ}}^2\right) \\ &\times \left(sup_{(x_i,z_i)\in S_X\times S_Z}\left|\left(\frac{\hat{f}_X(x_i)}{\hat{f}_{XZ}(x_i,z_i)}\right)^2 - \left(\frac{f_X(x_i)}{f_{XZ}(x_i,z_i)}\right)^2\right| + B\right) \end{aligned}$$

and since by Lemma 1 $\hat{f}_{XZ}(x,z) - f_{XZ}(x,z) = o_p(1)$ and $\hat{f}_X(x) - f_X(x) = o_p(1)$ uniformly $u_{11} = o_p(1)$. We note also that u_{21}, u_{12} and u_{22} differ from u_{11} only in that $\frac{x_j-x}{h_1}$ and $\frac{x_l-x}{h_1}$ appear in the summands. Again, since $K(\cdot) = 0$ outside of its compact support and S_X compact, we have by Lemma 1 that $u_{21}, u_{12}, u_{22} = o_p(1)$ and consequently $V_{4n} = o_p(1)$.

Theorem 2. *Proof* Let $\epsilon' = (\epsilon_1, \cdots, \epsilon_n)$ where $\epsilon_t = y_t - \alpha - m_1(x_t) - m_2(z_t)$. By construction,

$$\begin{aligned} \hat{m}_1^{2S2}(x) &= s_1(x)(\vec{y} - \vec{1}_n \bar{y} - \vec{m}_2^p(\vec{z})) \\ &= s_1(x)\vec{1}_n(\alpha - \bar{y}) + s_1(x)(\vec{m}_1(\vec{x}) + \epsilon) + s_1(x)(\vec{m}_2(\vec{z}) - \vec{m}_2^p(\vec{z})). \end{aligned} \tag{41}$$

Note that for the first term it is easy to show that $E(\alpha - \bar{y}|\vec{x}, \vec{z}) = O_p(n^{-1/2})$, the second term is identical to the first term that appeared in equation (31) in the proof of Theorem 1. Now we look at the i^{th} element of $-(\vec{m}_2(\vec{z}) - \vec{m}_2^p(\vec{z}))$ is

$$\begin{aligned} m_2^p(z_i) - m_2(z_i) &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right) \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}(y_k - \bar{y}) - m_2(z_i) \\ &= \hat{L}_{0n}(z_i) + \hat{L}_{1n}(z_i) + \hat{L}_{2n}(z_i) + \hat{L}_{3n}(z_i) \end{aligned}$$

where

$$\begin{aligned} \hat{L}_{0n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right) \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}(\alpha - \bar{y}) = A_1^n(z_i)(\alpha - \bar{y}) \\ \hat{L}_{1n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right) \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}\epsilon_k \\ \hat{L}_{2n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n} \frac{1}{g_2} K_1\left(\frac{z_k - z_i}{g_2}\right) \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}(m_2(z_k) - m_2(z_i)) \\ \hat{L}_{3n}(z_i) &= \frac{1}{n}\sum_{k=1}^{n} \left\{ \frac{1}{g_2} K_1(\frac{z_k - z_i}{g_2}) \frac{\hat{f}_X(x_k)}{\hat{f}_{XZ}(x_k, z_k)}\mu(x_k, z_i) - m_2(z_i) \right\} = A_{2n}(z_i) + m_2(z_i)(A_{1n}(z_i) - 1). \end{aligned}$$

where $A_{1n}(z_i)$, $A_{2n}(z_i)$, $\hat{L}_{1n}(z_i)$, $\hat{L}_{2n}(z_i)$ are as defined in the proof of Theorem 1. Hence, using the convergence results of Theorem 1 we obtain the desired expression for the conditional bias of $\hat{m}_1^{2S2}(x)$. For the conditional variance we note that,

$$\hat{m}_1^{2S2}(x) - E(\hat{m}_1^{2S2}(x)|\vec{x}, \vec{z}) = s_1(x)\left(I - \frac{1}{ng_2}B_n\right)(\epsilon - \bar{\epsilon})$$

and consequently,

$$\begin{aligned} V(\hat{m}_1^{2S2}(x)|\vec{x}, \vec{z}) &= s_1(x)\left(I - \frac{1}{ng_2}B_n\right) E\left((\epsilon - \bar{\epsilon})(\epsilon - \bar{\epsilon})'|\vec{x}, \vec{z}\right)\left(I - \frac{1}{ng_2}B_n\right)' s_1(x)' \\ &= \sigma^2 s_1(x)\left(I - \frac{1}{ng_2}B_n\right)\left(I - \frac{1}{ng_2}B_n\right)' s_1(x)' \\ &- \sigma^2 s_1(x)\left(I - \frac{1}{ng_2}B_n\right)\frac{\vec{1}_n\vec{1}_n'}{n}\left(I - \frac{1}{ng_2}B_n\right)' s_1(x)' = V_1 - V_2 \\ nh_1 V(\hat{m}_1^{2S2}(x)|\vec{x}, \vec{z}) &= nh_1 V_1 - nh_1 V_2. \end{aligned}$$

The first term in the conditional variance expression is identical to equation (37) in the proof of Theorem 1, and the second term can easily be shown to be $o_p(1)$.

Lemma 1 *Assume Assumptions 1-3 and suppose that* $\phi(x,z) : \Re^2 \to \Re$ *is a continuous function in* G_1 *a compact subset of* $\Re$ *with* $|\phi(x,z)| < B_\phi < \infty$. *Let*

$$s_j(x) = (ng_1)^{-1} \sum_{t=1}^{n} K\left(\frac{x_t - x}{g_1}\right)\left(\frac{x_t - x}{g_1}\right)^j \phi(x_t, z_t) \text{ with } j = 0, 1, 2.$$

a) If $ng_1^{2p+1}(ln(g_1))^{-1} \to \infty$ *for* $p > 0$, *then* $sup_{x\in G}|s_j(x) - E(s_j(x))| = o_p(g_1^p)$

b) Let G_2 *be a compact subset of* $\Re^2$ *and*

$$\hat{s}(x,z) = (ng_1g_2)^{-1} \sum_{t=1}^{n} K\left(\frac{x_t - x}{g_1}\right)\left(\frac{z_t - z}{g_2}\right)^j \phi(x_t, z_t)$$

If $n(g_1g_2)^{2p+1}(ln(g_1g_2))^{-1} \to \infty$ *for* $p > 0$, *then* $sup_{(x,z)\in G_2}|\hat{s}(x,z) - E(\hat{s}(x,z))| = o_p((g_1g_2)^p)$

Proof a) We prove the case where $j = 0$. Similar arguments can be used for $j = 1, 2$. Let $B(x_0, r) = \{x \in \Re : |x - x_0| < r\}$ for $r \in \Re^+$. G_1 compact implies that there exists $x_0 \in G_1$ such that $G_1 \subseteq B(x_0, r)$. Therefore for all $x, x' \in G_1$ $|x - x'| < 2r$. Let $g_1 > 0$ be a sequence such that $g_1 \to 0$ as $n \to \infty$ where $n \in \{1, 2, 3\cdots\}$. For any n, by the Heine-Borel theorem there exists a finite collection of sets $\{B(x_k, g_1^a)\}_{k=1}^{l_n}$ such that $G_1 \subset \cup_{k=1}^{l_n} B(x_k, g_1^a)$ for $x_k \in G_1$ with $l_n < g_1^{-a}r$ for $a \in (0, \infty)$. By assumption $|s_0(x) - s_0(x_k)| \le (nh_n)^{-1}\sum_{t=1}^{n} m|g_1^{-1}(x_k - x)|B_\phi < B_\phi m h_n^{a-2}$ for $x \in B(x_k, g_1^a)$. Similarly, $|E(s_0(x_k)) - E(s_0(x))| < B_\phi m g_1^{a-2}$ for $x \in B(x_k, g_1^a)$. Hence, $|s_0(x) - E(s_0(x))| \le |s_0(x_k) - E(s_0(x_k))| + 2B_\phi m g_1^{a-2}$ for $x \in B(x_k, g_1^a)$ and

$$sup_{x\in G_1}|s_0(x) - E(s_0(x))| \le max_{1\le k\le l_n}|s_0(x_k) - E(s(x_k))| + 2B_\phi m g_1^{a-2}.$$

To show that $lim_{n\to\infty}P\left(sup_{x\in G_1}|s_0(x) - E(s_0(x))| \ge g_1^p\epsilon\right) = 0$ for $p > 0$ we need $g_1^{a-p-2} \to 0$ as $n \to \infty$ and $lim_{n\to\infty}P\left(max_{1\le k\le l_n}|s_0(x) - E(s_0(x))| \ge g_1^p\epsilon\right) = 0$. But

$$P\left(max_{1\le k\le l_n}|s_0(x_k) - E(s_0(x_k))| \ge g_1^p\epsilon\right) \le \sum_{k=1}^{l_n} P\left(|s_0(x_k) - E(s_0(x_k))| \ge g_1^p\epsilon\right)$$

Using Bernstein's inequality, we have

$$P\left(|s_0(x_k) - E(s_0(x_k))| \geq g_1^p\epsilon\right) < 2exp\left(\frac{-nh_n^{2p}\epsilon^2}{2(\bar{\sigma}^2 + B_W\frac{g_1^p\epsilon}{3})}\right),$$

where $\bar{\sigma}^2 = n^{-1}\sum_{t=1}^n V(W_{tn}) = h_n^{-2}E\left(K^2\left(\frac{X_t - x_k}{h_n}\right)f^2(X_t, R_t)\right) - \left(h_n^{-1}E\left(K\left(\frac{X_t - x_k}{h_n}\right)f(X_t, R_t)\right)\right)^2$. Under assumptions A1 and A3 and the fact that $f(x,z)$ and $f_{X,Z}(x,z)$ are continuous in G_1 we have that $g_1\bar{\sigma}^2 = O(1)$. Hence for the desired result the righthand side of the inequality must approach zero as $n \to \infty$. It suffices to show that $\frac{nh_n^{2p}\epsilon^2}{2\bar{\sigma}^2 + 2/3B_W g_1^p\epsilon} + aln(g_1) \to \infty$, which given that $g_1\bar{\sigma}^2 = O(1)$ will result if $\frac{nh_n^{2p+1}}{ln(g_1)} \to \infty$.

b) Let $\theta = (x,z)'$ a typical element in Re^2. Let $B(\theta_0, r) = \{\theta \in \Re : |\theta - \theta_0| < r\}$ for $r \in \Re^+$. G_2 compact implies that there exists $\theta_0 \in G_2$ such that $G_2 \subseteq B(\theta_0, r)$. Therefore for all $\theta, \theta' \in G_2$ $|\theta - \theta'| < 2r$. Let $g_1, g_2 > 0$ be a sequence such that $g_1, g_2 \to 0$ as $n \to \infty$ where $n \in \{1, 2, 3\cdots\}$. For any n, by the Heine-Borel theorem there exists a finite collection of sets $\{B(\theta_k, r_n)\}_{k=1}^{l_n}$ such that $G_2 \subset \cup_{k=1}^{l_n} B(\theta_k, r_n)$ for $\theta_k \in G_2$ with $l_n < r_n^{-1}r^2$, $r_n = (g_1g_2)^a$ for $a \in (0,\infty)$. For $\theta \in B(\theta_k, r_n)$, $|\hat{s}(\theta) - \hat{s}(\theta_k)| \leq B_K m(g_1 + g_2)(g_1g_2)^{a-2}$. Similarly, $|E(\hat{s}(\theta_k)) - E(\hat{s}(\theta))| < B_K m(g_1 + g_2)(g_1g_2)^{a-2}$. Hence,

$$sup_{\theta \in G_2}|\hat{s}(\theta) - E(\hat{s}(\theta))| \leq max_{1 \leq k \leq l_n}|\hat{s}(\theta_k) - E(s(\theta_k))| + 2B_K m(g_1 + g_2)(g_1g_2)^{a-2}.$$

To show that $lim_{n\to\infty}P\left(sup_{\theta \in G_2}|\hat{s}(\theta) - E(\hat{s}(\theta))| \geq (g_1g_2)^p\epsilon\right) = 0$ for $p > 0$ it suffices to have $(g_1g_2)^{a-p-2} = O(1)$ and $lim_{n\to\infty}P\left(max_{1\leq k \leq l_n}|\hat{s}(\theta) - E(\hat{s}(\theta))| \geq (g_1g_2)^p\epsilon\right) = 0$. But

$$P\left(max_{1 \leq k \leq l_n}|\hat{s}(\theta_k) - E(\hat{s}(\theta_k))| \geq (g_1g_2)^p\epsilon\right) \leq \sum_{k=1}^{l_n} P\left(|\hat{s}(\theta_k) - E(\hat{s}(\theta_k))| \geq (g_1g_2)^p\epsilon\right)$$

Using Bernstein's inequality, we have

$$P\left(|\hat{s}(\theta_k) - E(\hat{s}(\theta_k))| \geq (g_1g_2)^p\epsilon\right) < 2exp\left(\frac{-n(g_1g_2)^{2p+1}\epsilon^2}{2g_1g_2\bar{\sigma}^2 + 2B_W\frac{(g_1g_2)^p\epsilon}{3})}\right).$$

where $\bar{\sigma}^2 = \frac{1}{n}\sum_{t=1}^n V(W_{tn})$, and

$$W_{tn} = \frac{1}{g_1g_2}K_1\left(\frac{x_t - x}{g_1}\right)K_1\left(\frac{z_t - z}{g_2}\right) - E\left(\frac{1}{g_1g_2}K_1\left(\frac{x_t - x}{g_1}\right)K_1\left(\frac{z_t - z}{g_2}\right)\right).$$

Hence for the desired result the righthand side of the inequality must approach zero as $n \to \infty$. For this it suffices to have $\frac{n(g_1g_2)^{2p+1}}{ln(g_1g_2)} \to \infty$.

References

1. Barr, D.R. and N.L. Slezak, 1972, A comparison of multivariate normal generators, Communications of the Association for Computing Machinery, 15, 1048-1049.

2. Buja, A., Hastie, T.J. and R.J. Tibshirani, 1989, Linear smoothers and additive models, Annals of Statistics, 17, 453-555.

3. Friedman, J.H. and W. Stuetzel, 1981, Projection pursuit regression, Journal of the American Statistical Association, 76, 817-823.

4. Fan, J. and S.J. Marron, 1994, Fast implementation of nonparametric curve estimators, Journal of Computational and Graphical Statistics, 3, 35-56.

5. Fushimi, M., 1990, Random number generation with the recursion $X_t = X_{t-3p} \oplus X_{t-3q}$, Journal of Computational and Applied Mathematics, 31, 105-118.

6. Graybill, F.A., 1969, Introduction do matrices with applications in statistics, Wadsworth: Belmont, CA.

7. Jones, M.C., S.J. Davies and B.U. Park, 1994, Versions of kernel type regression estimators, Journal of the American Statistical Association, 89, 825-832.

8. Kinderman, A. and J. Ramage, 1976, Computer generation of normal random variables, Journal of the American Statistical Association, 71, 893-896.

9. Kennedy, W. and J. Gentle, 1980, Statistical computing, Marcel Dekker: New York.

10. Härdle, W. and O. Linton, 1994, Applied nonparametric methods, in Robert Engle and Daniel McFadden eds. Handbook of Econometrics, v.4 . North Holland: Amsterdam.

11. Hastie, T.J. and R.J. Tibshirani, 1986, Generalized additive models, Statistical Science, 1, 297-318.

12. Hastie, T.J. and R.J. Tibshirani, 1990, Generalized additive models, Chapman and Hall: New York.

13. Kim, W., O. Linton and N. Hengartner, 1999, A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals, Journal of Computational and Graphical Statistics, 8, 278-297.

14. Linton, O., 1997, Efficient estimation of additive nonparametric regression models, Biometrika, 84, 469-474.

15. Linton, O. and W. Härdle, 1996, Estimation of additive regression models with known links, Biometrika, 83, 529-540.

16. Linton, O. and J.P. Nielsen, 1995, A Kernel method of estimating structured nonparametric regression based on marginal integration, Biometrika, 82, 93-100.

17. Marsaglia, G., 1968, Random numbers fall mainly in the planes, Proceedings of the National Academy of Sciences, 61, 25-28.

18. Marsaglia, G., 1985, A current view of random number generators, in L. Billard editor Computer Science and Statistics: Proceedings of the 16th Symposium on the Interface, North Holland: Amsterdam.

19. Newey, W., 1994, Kernel estimation of partial means, Econometric Theory, 10, 233-253.

20. Opsomer, J., 2000, Asymptotic properties of backfitting estimators, Journal of Multivariate Analysis, 73, 166-179.

21. Opsomer, J. and D. Ruppert, 1997, Fitting a bivariate additive model by local polynomial regression, Annals of Statistics, 25, 186-211.

22. Opsomer, J. and D. Ruppert, 1998, A fully automated bandwidth selection method for fitting additive models, Journal of the American Statistical Association, 93, 605-619.

23. Pagan, A. and A. Ullah, 1999, Nonparametric econometrics. Cambridge: Cambridge University Press.

24. Park, B. and J.S. Marron, 1990, Comparison of data-driven bandwidth selectors, Journal of the American Statistical Association, 85, 66-72.

25. Ruppert, D., S.J. Sheather, and M.P. Wand, 1995, An effective bandwidth selector for least squares regression, Journal of the American Statistical Association, 90, 1257-1270.

26. Sheather, S.J. and M.C. Jones, 1991, A reliable data-based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society B, 53, 683-690.

27. Silverman, B.W., 1986, Density estimation for statistics and data analysis, Chapman and Hall: New York.

28. Simonoff, J.S., 1996, Smoothing methods in statistics, Springer: New York.

29. Sperlich, S., O. Linton, W. Härdle, 1999, Integration and backfitting methods in additive models - finite sample properties and comparison, TEST, 8, 1-39.

30. Staniswalis, J.G. and Severini, T.A, 2000, Fitting the additive model by recursion on dimension, Communications in Statistics, 29, 689-701.

31. Stone, C.,1985, Additive regression and other nonparametric models, Annals of Statistics, 6, 689-705.

32. Stone, C.,1990, Optimal rates of convergence for nonparametric estimators, Annals of Statistics, 8, 1348-1360.

33. Tjøstheim, D. and B. Auestad, 1994, Nonparametric identification of nonlinear time series projections, Journal of the American Statistical Association, 89, 1398-1409.

34. Wahba, G., 1990, Spline models for observational data. SIAM: Philadelphia, PA.