THE UNIVERSITY OF BRITISH COLUMBIA
DEPARTMENT OF STATISTICS

TECHNICAL REPORT # 222

MAY 2006

INFERENCE IN PARTIALLY LINEAR MODELS WITH CORRELATED ERRORS

BY

ISABELLA R. GHEMENT

NANCY E. HECKMAN

Inference in Partially Linear Models with Correlated Errors

Isabella GHEMENT and Nancy HECKMAN

Keywords and phrases: backfitting, correlated errors, inference, partially linear model, smoothing parameter choice
MSC 2000: Primary 62G08 (nonparametric regression); Secondary 62P12 (applications to environmental and related topics).

Abstract: We study statistical inference for the linear effects in partially linear regression models with correlated errors. To estimate these effects, we introduce several local linear backfitting estimators. Our asymptotic mean squared error calculations establish that these estimators achieve $\sqrt{n}$–consistency provided we use smoothing parameters that undersmooth the estimated smooth effect. We introduce three methods for choosing smoothing parameters from the data. Two of these methods are modifications of the Empirical Bias Bandwidth Selection method of Opsomer and Ruppert (1999). The third method is a non–asymptotic plug–in method. Using our estimators and smoothing parameter methods, we construct approximate confidence intervals for the linear effects. Based on a simulation study, we recommend one estimator and two smoothing parameter choices. We illustrate our recommended inferential procedures with an analysis of data collected in a community–level time series study of the health effects of air pollution.

1. INTRODUCTION

Partially linear models combine the ease of interpretation of linear regression models with the modelling flexibility of non-parametric regression models. They generalize linear regression models by allowing one or more of the covariate effects to be smooth, of unknown form, while keeping other covariate effects linear. Determining how much to smooth the estimator of the non-parametric component is crucial, with the amount of smoothing depending on the goal. Here our goal is inference on the linear component of our model when errors are possibly correlated. We study backfitting estimation procedures and show how to choose the amount of smoothing and how to construct confidence intervals for the linear component.

We consider partially linear models with one smooth covariate effect. Specifically suppose we observe the data $(Y_i, X_{ij}, Z_i),\ i = 1, \ldots, n, j = 1, \ldots, p$, where Y_i is a continuous response, $X_{i1}, \ldots, X_{ip}$ are measurements on p covariates and the Z_i's are fixed design points which, without loss of generality, we assume lie in the unit interval. Our model for these data is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + m(Z_i) + \epsilon_i, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is an unknown parameter vector, m is an unknown, smooth function and the ϵ_i's are error terms. To ensure identifiability, we assume that m satisfies the integral restriction $\int_0^1 m(z) f(z) dz = 0$. In practice, we replace the integral by the summation restriction

$$\sum_{i=1}^{n} m(Z_i) = 0. \tag{2}$$

We assume that the errors are such that $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma_\epsilon^2 > 0$ and $Corr(\epsilon_i, \epsilon_j) = \boldsymbol{\Psi}_{ij}$, with $\boldsymbol{\Psi}$ denoting the $n \times n$ error correlation matrix. Härdle et al. (2000) provide a detailed exposition of partially linear models.

In this paper, we are interested in the linear effects $\boldsymbol{\beta}$, treating m and the correlation between the ϵ_i's as nuisances. In particular, we wish to address the following two questions. How should we choose the degree of smoothness of the estimator of m to accurately estimate $\boldsymbol{\beta}$? How can we construct valid confidence intervals for linear combinations of $\boldsymbol{\beta}$'s components?

Model (1) has been investigated extensively under the assumption of independent, identically distributed errors. A common theme of the published results has been determining if $\boldsymbol{\beta}$ can be estimated $\sqrt{n}$–consistently. Speckman (1988) showed that his partial residual based estimator of $\boldsymbol{\beta}$ can achieve $\sqrt{n}$–consistency if m is estimated at the 'usual' nonparametric rate of $n^{-4/5}$. Opsomer and Ruppert (1999) established that their backfitting estimator of

$\boldsymbol{\beta}$ can also achieve $\sqrt{n}$–consistency provided m is estimated at a slower rate than the 'usual' nonparametric rate.

Most of the literature concerning partially linear models with correlated errors has focused on Speckman-type estimators of $\boldsymbol{\beta}$. In the first of a series of theoretical papers, Aneiros-Pérez and Quintela-del-Río showed that their modified Speckman estimator of $\boldsymbol{\beta}$ is $\sqrt{n}$–consistent when the errors are α-mixing, stationary (2001a). To choose the amount of smoothing needed for accurate estimation of m, they proposed using cross-validation, modified for correlated errors (2001b). To choose the amount of smoothing needed for accurate estimation of $\boldsymbol{\beta}$, they proposed and studied an asymptotic plug-in method (2002), but did not implement it. Their theory requires that the errors follow a stationary autoregressive process of known finite order. For the case of α–mixing, non-stationary errors, You and Chen (2004) proposed two methods for constructing confidence intervals for $\boldsymbol{\beta}$ using the usual Speckman estimator. One method is based on asymptotic normality of the estimator; the other is based on a block bootstrap. You et al. (2005) constructed confidence intervals and tests of hypotheses for functions of $\boldsymbol{\beta}$ based on a jackknife estimator for $\boldsymbol{\beta}$, which they obtained from the usual Speckman estimator; their inferential procedures are valid under the assumption that the errors follow a stationary moving average process of possibly infinite order.

Backfitting-type estimators in partially linear models are popular with practitioners in a variety of research areas. They are easy to compute, through the *gam* function in the statistical packages R and Splus, and they allow for easy interpretation of the effect of the covariate of interest. They have been used extensively in the analysis of the health effects of air pollution. See, for example, Dominici et. al. (2002) and references there-in. However, little is known of their theoretical properties in the presence of error correlation. Even practical implementation for uncorrelated errors has had problems, in particular the implementation in the S-Plus function *gam*. Dominici et al. (2002) showed how to correct problems with *gam*'s default parameters, in order to improve estimators of $\boldsymbol{\beta}$'s components and their standard errors. Ramsay et al. (2003) pointed to a further problem with *gam*'s approximate standard error calculation. Dominici et al. (2004) resolved this problem by providing software for exact standard error calculation.

In this paper, we fill in the gap in the literature about the use of backfitting estimators when errors are correlated. We derive conditions under which these estimators can achieve $\sqrt{n}$–consistency. We introduce three data-driven methods for choosing the degree of smoothness of the estimator of m when accurate estimation of $\boldsymbol{c}^\top\boldsymbol{\beta}$ is desired, with $\boldsymbol{c}$ being user-specified. Two of these methods are based on the Empirical Bias Bandwidth Selection method of Opsomer and Ruppert (1999), modified to account for possible error correlation. The third method is a new non-asymptotic plug-in method. We also propose methods for constructing approximate confidence intervals for $\boldsymbol{c}^\top\boldsymbol{\beta}$ and we investigate their finite sample properties by means of a simulation study. Our simulation study includes a comparison with confidence intervals constructed from the usual Speckman estimator. Finally, we apply our preferred inferential procedures to the analysis of data collected in a community-level time series study of the health effects of air pollution.

The rest of the paper is organized as follows. Section 2 introduces three backfitting estimators of $\boldsymbol{\beta}$, and their asymptotic properties are given in Section 3, along with the asymptotic properties of our estimator of m. Section 4 presents our data-driven methods for choosing the appropriate amount of smoothing for accurate estimation of $\boldsymbol{c}^\top\boldsymbol{\beta}$. These methods require preliminary estimation of the non-parametric component m and the error correlation structure, topics discussed in Section 5. Section 6 introduces standard and bias-adjusted confidence intervals for $\boldsymbol{c}^\top\boldsymbol{\beta}$. A simulation study is carried out in Section 7. Section 8 presents the results of our data analysis. Some conclusions are drawn in Section 9. The proofs of the main results are provided in Section 10, with lemmas presented in the Appendix.

2. BACKFITTING ESTIMATORS

We now provide a formal definition for the generic backfitting estimators of the unknowns $\boldsymbol{\beta}$ and $\boldsymbol{m} = (m(Z_1), \ldots, m(Z_n))^\top$ and then restrict discussion to three particular types: the usual, the modified and the estimated modified backfitting estimators.

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$, $\boldsymbol{X}$ be the $n \times (p+1)$ design matrix corresponding to the parametric part of model (1), $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top$, and $\boldsymbol{\Omega}$ be an $n \times n$ matrix of weights such that the $(p+1) \times (p+1)$ matrix $\boldsymbol{X}^\top \boldsymbol{\Omega} \boldsymbol{X}$ is invertible. The choice of $\boldsymbol{\Omega}$ determines the particular type of backfitting estimator. Let $\boldsymbol{S}_h$ be an $n \times n$ smoother matrix depending on a smoothing parameter h and let $\boldsymbol{S}_h^c$ be the centred version of $\boldsymbol{S}_h$, obtained as

$$\boldsymbol{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^\top / n)\boldsymbol{S}_h \tag{3}$$

where $\mathbf{1}$ is an n-vector of 1's.

The generic backfitting estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$ of $\boldsymbol{\beta}$ and $\boldsymbol{m}$ are defined as the fixed points of the backfitting equations

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} &= (\boldsymbol{X}^\top \boldsymbol{\Omega} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega} (\boldsymbol{Y} - \widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}) \\ \widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h} &= \boldsymbol{S}_h^c (\boldsymbol{Y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}). \end{aligned}$$

Use of the matrix $\boldsymbol{S}_h^c$ instead of $\boldsymbol{S}_h$ ensures that $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$ satisfies the identifiability condition (2).

In practice, one could solve the backfitting equations for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$ iteratively by employing a modification of the backfitting algorithm of Buja et. al. (1989), who consider the case $\boldsymbol{\Omega} = \boldsymbol{I}$. However, we need not iterate to find $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$, since we can derive explicit expressions for both. Provided the matrix $\boldsymbol{X}^\top \boldsymbol{\Omega} (\boldsymbol{I} - \boldsymbol{S}_h^c) \boldsymbol{X}$ is invertible, algebraic manipulations of the backfitting equations yield

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} = (\boldsymbol{X}^\top \boldsymbol{\Omega} (\boldsymbol{I} - \boldsymbol{S}_h^c) \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega} (\boldsymbol{I} - \boldsymbol{S}_h^c) \boldsymbol{Y} \equiv \boldsymbol{M}_{\boldsymbol{\Omega},h} \boldsymbol{Y} \tag{4}$$

and

$$\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h} = \boldsymbol{S}_h^c (\boldsymbol{I} - \boldsymbol{X} \boldsymbol{M}_{\boldsymbol{\Omega},h}) \boldsymbol{Y}. \tag{5}$$

We consider three specifications of $\boldsymbol{\Omega}$: $\boldsymbol{\Omega} = \boldsymbol{I}$, $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$, where $\widehat{\boldsymbol{\Psi}}$ estimates $\boldsymbol{\Psi}$. Taking $\boldsymbol{\Omega} = \boldsymbol{I}$ in (4) yields the usual backfitting estimator of $\boldsymbol{\beta}$. We would expect that this estimator is not efficient as it does not account for the error correlation. When $\boldsymbol{\Psi}$ is known, taking $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ yields the modified backfitting estimator, which we would expect to be more efficient. In the usual case that $\boldsymbol{\Psi}$ is unknown, we use $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$ and refer to the resulting estimator as the estimated modified backfitting estimator.

We compute these backfitting estimators of $\boldsymbol{\beta}$ by taking $\boldsymbol{S}_h$ to be the $n \times n$ local linear smoother matrix, whose (i,j)th element is defined as:

$$S_{ij} = \frac{w_j^{(Z_i)}}{\sum_{k=1}^n w_k^{(Z_i)}} \ ,$$

where

$$w_k^{(z)} = K\Big(\frac{z - Z_k}{h}\Big)\Big[\mathcal{S}_{n,2}(z) - (z - Z_k)\mathcal{S}_{n,1}(z)\Big], \ k = 1, \ldots, n.$$

Here, K is a user-specified kernel function and

$$\mathcal{S}_{n,l}(z) = \sum_{k=1}^n K\Big(\frac{z - Z_k}{h}\Big)(z - Z_k)^l, \ l = 1, 2.$$

Local linear smoothing is an effective smoothing method in nonparametric regression; see Fan and Gijbels (1992) and Fan (1993).

3. ASYMPTOTICS

In this section, we determine the asymptotic conditional bias, variance and mean squared error of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$, given $\boldsymbol{X}$. We establish that, for these estimators to converge to $\boldsymbol{\beta}$ at the usual parametric rate of $1/n$, we should choose the bandwidth h to be of order $n^{-\alpha}$, $\alpha \in [1/4, 1/3)$. We also calculate the asymptotic bias and variance of our estimators of m and show that optimal estimation of m in our context requires $h = O(n^{-1/5})$, a much larger value of h. We therefore conclude that, to accurately estimate $\boldsymbol{\beta}$, we should undersmooth the estimated m. Finally, we give conditions on $\widehat{\boldsymbol{\Psi}}$ that ensure that $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ is close to $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$. The proofs of the theorems appear in Section 10.

Throughout this section, the X_{ij}'s are random and the Z_i's are fixed. In addition, we will assume, as in Speckman (1988), that the X_{ij}'s and Z_i's are related via the nonparametric regression model

$$X_{ij} = g_j(Z_i) + \eta_{ij},\ i = 1, \ldots, n, j = 1, \ldots, p, \tag{6}$$

where the g_j's are smooth, unknown functions and the η_{ij}'s are unobserved error terms having mean 0. We also assume the following.

(A1) *The bandwidth $h = h_n$ satisfies $h \to 0$ and $nh^3 \to \infty$ as $n \to \infty$.*

(A2) *The functions m and $g_1, \ldots, g_p$ have three continuous derivatives.*

(A3) *The Z_i's follow a regular design, i.e. there exists a strictly positive, twice continuously differentiable density f on $[0, 1]$ with*

$$\int_0^{Z_i} f(z)dz = \frac{i}{n+1},\ i = 1, \ldots, n.$$

(A4) *The kernel function K is a probability density function symmetric about zero and Lipschitz continuous, with compact support $[-1, 1]$.*

(A5) *The random vectors $\{(\eta_{i1}, \ldots, \eta_{ip})^\top, i = 1, \ldots, n\}$ are independent, identically distributed, having mean 0 and covariance matrix $\mathbf{\Sigma}$. Furthermore, they are independent of the errors $\{\epsilon_j,\ j = 1, \ldots, n\}$.*

(A6) *The ϵ_i's are realizations of a mean 0, covariance stationary process with $Var(\epsilon_i) = \sigma_\epsilon^2$ and $Cov(\epsilon_i, \epsilon_{i+k}) = \sigma_\epsilon^2 \Psi_{i,i+k} \equiv \sigma_\epsilon^2 \rho(k)$.*

In Theorems 1, 2 and 3 below, we place additional restrictions on $\mathbf{\Psi}$, the correlation matrix of the ϵ_i's. In Theorem 1, which provides asymptotics for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, the conditions are quite mild. In Theorem 2, which provides asymptotics for $\widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h}$, we impose the restriction that the ϵ_i's are from an autoregressive process of known finite order. This allows us to find an explicit expression for $\mathbf{\Psi}^{-1}$. In Theorem 3, where we study the asymptotic behaviour of $\widehat{m}_{\boldsymbol{I},h}$ and $\widehat{m}_{\mathbf{\Psi}^{-1},h}$, we assume that $\sum_1^\infty k|\rho(k)| < \infty$, a condition satisfied by autoregressive processes of finite order.

Before stating the main results in this section, we introduce some useful notation. Set $g_0(z) \equiv 1$ and let $\boldsymbol{G}$ be an $n \times (p+1)$ matrix and $\boldsymbol{g}^*$ be a $(p+1)$-vector with

$$G_{ij} = g_{j-1}(Z_i) \text{ and } g_j^* = \int_0^1 g_{j-1}(z)f(z)dz.$$

Also, let $\boldsymbol{w}$ be the $(p+1)$-vector with

$$w_i = \int_0^1 g_{i-1}(z)m''(z)f(z)dz - \int_0^1 g_{i-1}(z)f(z)dz \cdot \int_0^1 m''(z)f(z)dz.$$

Define the $(p+1) \times (p+1)$ matrix $\boldsymbol{\Sigma}^{(0)}$ as

$$\boldsymbol{\Sigma}^{(0)} = \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{pmatrix}.$$

Finally, set $\nu_j(K) = \int_{-1}^1 u^j K(u)du$.

THEOREM 1. *In addition to assumptions (A1)-(A6), assume the following.*

(A7) $\boldsymbol{\Psi}$ *has a bounded spectral norm as* $n \to \infty$.

(A8) There exists a $p \times p$ *matrix* $\boldsymbol{\Phi}$ *with* $\sum_{i,i'=1}^n \eta_{ij}\eta_{i'j'}\boldsymbol{\Psi}_{ii'}/n$ *converging to* $\boldsymbol{\Phi}_{jj'}$ *in probability as* $n \to \infty$.

Let $\boldsymbol{V_I} = \boldsymbol{\Sigma}^{(0)} + \boldsymbol{g}^*\boldsymbol{g}^{*\top}$ *and*

$$\boldsymbol{\Phi}^{(0)} = \begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{\Phi} \end{pmatrix}.$$

If the identifiability condition (2) *holds, then*

$$(i) \quad Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}) = -\frac{h^2}{2}\nu_2(K)\boldsymbol{V_I}^{-1}\boldsymbol{w} + \boldsymbol{o}_P(h^2) = \boldsymbol{O}_P(h^2),$$

$$(ii) \quad Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}) = \frac{\sigma_\epsilon^2}{n}\boldsymbol{V_I}^{-1}\boldsymbol{\Phi}^{(0)}\boldsymbol{V_I}^{-1} + \frac{\sigma_\epsilon^2}{n^2}\boldsymbol{V_I}^{-1}\boldsymbol{G}^\top(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top\boldsymbol{G}\boldsymbol{V_I}^{-1} + \boldsymbol{o}_P\left(n^{-1}\right)$$
$$= \boldsymbol{O}_P\left(n^{-1}\right)$$

and

$$(iii) \quad E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h} - \boldsymbol{\beta}||_2^2\,\middle|\boldsymbol{X}\right) = \frac{h^4}{4}\nu_2^2(K)\boldsymbol{w}^\top\boldsymbol{V_I}^{-2}\boldsymbol{w} + \frac{\sigma_\epsilon^2}{n}trace\left\{\boldsymbol{V_I}^{-1}\boldsymbol{\Phi}^{(0)}\boldsymbol{V_I}^{-1}\right\}$$
$$+ \frac{\sigma_\epsilon^2}{n^2}trace\left\{\boldsymbol{V_I}^{-1}\boldsymbol{G}^\top(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top\boldsymbol{G}\boldsymbol{V_I}^{-1}\right\} + \boldsymbol{o}_P(h^4) + \boldsymbol{o}_P\left(n^{-1}\right)$$
$$= O_P(h^4) + O_P\left(n^{-1}\right).$$

Theorem 2. *In addition to assumptions (A1)-(A5), assume that*

(A9) the ϵ_i's represent n consecutive realizations from a covariance stationary autoregressive process of finite order R having mean 0, finite, non-zero variance σ_ϵ^2 and satisfying

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \cdots + \phi_R\epsilon_{t-R} + u_t,\ t = 0, \pm 1, \pm 2, \ldots$$

with $\{u_t\},\ t = 0, \pm 1, \pm 2, \ldots$ being independent, identically distributed random variables having mean 0 and finite, non-zero variance σ_u^2.

Set

$$\boldsymbol{V}_{\boldsymbol{\Psi}} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 + \sum_{k=1}^{R}\phi_k^2\right)\boldsymbol{\Sigma}^{(0)} + \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^{R}\phi_k\right)^2 \boldsymbol{g}^*\boldsymbol{g}^{*\top}.$$

If the identifiability condition (2) *holds, then*

$$(i)\quad Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X}) = -\frac{h^2}{2}\nu_2(K)\frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^{R}\phi_k\right)^2 \boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{w} + \boldsymbol{o}_P(h^2) = \boldsymbol{O}_p(h^2),$$

$$(ii)\quad Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X}) = \frac{1}{n}\cdot\frac{\sigma_\epsilon^4}{\sigma_u^2}\left(1 + \sum_{k=1}^{R}\phi_k^2\right)\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{\Sigma}^{(0)}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}$$
$$+ \frac{\sigma_\epsilon^2}{n^2}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{G}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top\boldsymbol{\Psi}^{-1}\boldsymbol{G}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} + \boldsymbol{o}_P\left(n^{-1}\right) = \boldsymbol{O}_P\left(n^{-1}\right)$$

and

$$(iii)\quad E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h} - \boldsymbol{\beta}||_2^2\middle|\boldsymbol{X}\right) = \frac{h^4}{4}\nu_2^2(K)\cdot\frac{\sigma_\epsilon^4}{\sigma_u^4}\left(1 - \sum_{k=1}^{R}\phi_k\right)^4 \boldsymbol{w}^\top\boldsymbol{V}_{\boldsymbol{\Psi}}^{-2}\boldsymbol{w}$$
$$+ \frac{1}{n}\cdot\frac{\sigma_\epsilon^4}{\sigma_u^2}\cdot\left(1 + \sum_{k=1}^{R}\phi_k^2\right) trace\left\{\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{\Sigma}^{(0)}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\right\}$$
$$+ \frac{\sigma_\epsilon^2}{n^2} trace\left\{\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{G}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top\boldsymbol{\Psi}^{-1}\boldsymbol{G}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\right\}$$
$$+ o_P(h^4) + o_P\left(n^{-1}\right) = O_P(h^4) + o_P\left(n^{-1}\right).$$

Comment 1. *The matrix $\boldsymbol{V}_{\boldsymbol{\Psi}}$ is strictly positive definite since it is non-negative definite and, if $\boldsymbol{v}^\top\boldsymbol{V}_{\boldsymbol{\Psi}}\boldsymbol{v} = 0$, then we can easily show that $\boldsymbol{v}$ is the zero vector.*

Comment 2. *Theorem 2's conditions are a special case of Theorem 1's. To see this, suppose that (A9) holds. Then clearly (A6) holds. Assumption (A7) holds by Lemma 3 (i) in the Appendix. To see that (A8) holds with* $\mathbf{\Phi} = \mathbf{\Sigma}$, *write*

$$\frac{1}{n}\sum_{i,i'=1}^{n} \eta_{ij}\eta_{i'j'}\mathbf{\Psi}_{ii'} = \frac{1}{n}\sum_{i=1}^{n} \eta_{ij}\eta_{ij'} + \frac{1}{n}\sum_{i\neq i'} \rho(|i-i'|)\eta_{ij}\eta_{i'j'} = \frac{1}{n}\sum_{i=1}^{n} \eta_{ij}\eta_{ij'} + \frac{1}{n}\sum_{k=1}^{n-1} \rho(k)S_k^n$$

where $S_k^n = n^{-1}\big[\sum_{i=1}^{n-k} \eta_{ij}\eta_{i+k,j'} + \sum_{i=k+1}^{n} \eta_{ij}\eta_{i-k,j'}\big]$. *By the weak law of large numbers,* $\sum_i \eta_{ij}\eta_{ij'}/n$ *converges to its expectation, namely* $\Sigma_{jj'}$. *To show that* $\sum_{k=1}^{n-1}\rho(k)S_k^n/n$ *converges to 0 in probability, first note that the summands in each of the two terms in* S_k^n *are* k*-dependent and so, using the weak law of large numbers for* k*-dependent sequences,* S_k^n *converges to 0 in probability as* $n \to \infty$. *Clearly,* $|E(S_k^n)| \leq B$ *for some* B *independent of* k *and* n. *Thus for fixed* k_0, $\sum_{k=1}^{k_0}\rho(k)S_k^n$ *converges to 0 in probability and*

$$E\left(\left|\sum_{k=k_0+1}^{n-1} \rho(k)S_k^n\right|\right) \leq B \sum_{k=k_0+1}^{\infty} |\rho(k)|$$

which can be made arbitrarily small by taking k_0 *large. Therefore* $\sum_{k=1}^{n-1}\rho(k)S_k^n/n$ *converges to 0 in probability and (A8) holds.*

Under what conditions on h are the estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, $\widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\mathbf{\Psi}}^{-1},h}$ $\sqrt{n}$-consistent? To answer this, we first use Theorems 1 and 2, letting $\widehat{\boldsymbol{\beta}}_{\mathbf{\Omega},h}$ denote either $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ or $\widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h}$. If h is of order $n^{-\alpha}$, $\alpha \in [1/4, 1/3)$, then we have the required $h \to 0$, $nh^3 \to \infty$ and $E(||\widehat{\boldsymbol{\beta}}_{\mathbf{\Omega},h} - \boldsymbol{\beta}||^2|\boldsymbol{X}) = O_P(1/n)$. Thus $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ and $\widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h}$ are $\sqrt{n}$-consistent. If one assumes that $\boldsymbol{X}^\top(\widehat{\mathbf{\Psi}}^{-1} - \mathbf{\Psi}^{-1})(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}/n = \boldsymbol{o}_p(1)$ and $\boldsymbol{X}^\top(\widehat{\mathbf{\Psi}}^{-1} - \mathbf{\Psi}^{-1})(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})/\sqrt{n} = \boldsymbol{o}_p(1)$ then one directly shows that $\widehat{\boldsymbol{\beta}}_{\widehat{\mathbf{\Psi}}^{-1},h} - \widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h} = \boldsymbol{o}_P(1/\sqrt{n})$. Thus $\widehat{\boldsymbol{\beta}}_{\widehat{\mathbf{\Psi}}^{-1},h}$ is $\sqrt{n}$-consistent whenever $\widehat{\boldsymbol{\beta}}_{\mathbf{\Psi}^{-1},h}$ is. These technical conditions are similar to conditions (A.15) and (A.16) imposed by Aneiros Pérez and Quintela del Río (2001a) for studying the effect of estimation of $\mathbf{\Psi}$ on the rate of convergence of their modified Speckman estimator. Unfortunately, such conditions are difficult to check.

The following theorem provides explicit expressions for the asymptotic bias and variance for components of $\widehat{\boldsymbol{m}}_{\mathbf{\Omega},h}$ where $\mathbf{\Omega} = \boldsymbol{I}$ or $\mathbf{\Psi}^{-1}$. From these, we see that the h that minimizes $[\widehat{\boldsymbol{m}}_{\mathbf{\Omega},h}]_i$'s asymptotic conditional mean squared error is of order $n^{-1/5}$.

THEOREM 3. *Let* $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h} = \boldsymbol{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. *Suppose either (i)* $\boldsymbol{\Omega} = \boldsymbol{I}$ *and the conditions of Theorem 1 hold, or (ii)* $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ *and the conditions of Theorem 2 hold. Suppose in addition that* $\sum_{k=1}^{\infty} k|\rho(k)| < \infty$. *Let the asymptotic bias of* $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ *from (i) of Theorem 1 or 2 be written as* $h^2\nu_2(K)\boldsymbol{b}/2 + \boldsymbol{o}_p(h^2)$ *and let* $\boldsymbol{G}_{i\cdot}^\top$ *denote the ith row of* $\boldsymbol{G}$. *Then*

$$Bias([\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}]_i|\boldsymbol{X}) = h^2\frac{\nu_2(K)}{2}\left[m''(Z_i) - \int m''(z)\ f(z)\ dz - (\boldsymbol{G}_{i\cdot} - \boldsymbol{g}^*)^\top\boldsymbol{b}\right] + o_p(h^2)$$

uniformly in i such that $Z_i \in [h, 1-h]$, *and*

$$var([\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}]_i|\boldsymbol{X}) = \frac{\sigma_\epsilon^2}{f(Z_i)\ nh}\ \nu_0(K^2)\ \left(\rho(0) + 2\sum_1^\infty \rho(k)\right) + o_p(1/(nh))$$

uniformly in i such that $Z_i \in [2h, 1-2h]$. *Furthermore,* $Bias([\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}]_i|\boldsymbol{X}) = O_p(h^2)$ *and* $var([\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}]_i|\boldsymbol{X}) = O_p(1/(nh))$ *uniformly in all* Z_i.

In summary, an h that is optimal for estimating m will not be optimal for estimating $\boldsymbol{\beta}$ and vice versa. To accurately estimate m, we should use h of order $n^{-1/5}$, but this will lead to $\widehat{\boldsymbol{\beta}}$ converging to $\boldsymbol{\beta}$ at rate only $n^{-2/5}$. Conversely if we use an h that is good for estimating $\boldsymbol{\beta}$, namely h of order $n^{-\alpha}$ with $\alpha \in [1/4, 1/3)$, we will undersmooth the estimated m.

4. SMOOTHING PARAMETER SELECTION FOR ESTIMATING $\boldsymbol{c}^\top\boldsymbol{\beta}$

In this section, we develop data-driven methods for choosing h to obtain accurate backfitting estimators of $\boldsymbol{c}^\top\boldsymbol{\beta}_{\boldsymbol{\Omega},h}$, where $\boldsymbol{c} = (c_0, c_1, \ldots, c_p)^\top$ is user-specified. We choose h to minimize an estimator of the conditional mean squared error of $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}$. Our minimization is always via a grid search over a specified grid $\mathcal{H} = \{h_1, \ldots, h_N\}$. We first discuss the case when $\boldsymbol{\Omega} = \boldsymbol{I}$ or $\boldsymbol{\Psi}^{-1}$.

To estimate the conditional variance of $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, we use (4) and estimators of σ_ϵ^2 and $\boldsymbol{\Psi}$ (described in Section 5 below), so that

$$\widehat{Var}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X}) = \widehat{\sigma}_\epsilon^2\boldsymbol{c}^\top\boldsymbol{M}_{\boldsymbol{\Omega},h}\widehat{\boldsymbol{\Psi}}\boldsymbol{M}_{\boldsymbol{\Omega},h}^\top\boldsymbol{c}, \tag{7}$$

We consider three methods to estimate the bias: a local empirical method, a global empirical method and a non-asymptotic plug-in method. The local empirical method is essentially the Empirical Bias Bandwidth Selection (EBBS) method devised by Opsomer and Ruppert (1999) for uncorrelated errors. The global empirical method, a modification of EBBS, is new. In our simulations we find it performs better than the local empirical method.

Both empirical methods estimate $Bias(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X})$ using the approximation

$$E(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X}) \approx a_0 + \sum_{t=2}^{T} a_t h^t, \tag{8}$$

valid as $h \to 0$, where $a_0 = \boldsymbol{c}^\top \boldsymbol{\beta}$ and $a_t, t = 2, \ldots, T$, are unknown asymptotic constants. This approximation can be obtained by a more delicate Taylor series analysis than that used to obtain the bias expressions of Theorems 1 and 2.

The two empirical methods estimate $Bias(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X})$ using ordinary least squares to fit model (8). Both methods require a grid of h's and calculations of $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ for h in the grid. For convenience, we use the same grid $\mathcal{H}$ as for the grid search. The global empirical method uses the entire 'data set' $\{(h, \boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}), h \in \mathcal{H}\}$ to fit (8) by ordinary least squares, yielding $\widehat{a}_0, \widehat{a}_2, \ldots, \widehat{a}_T$ and an estimator of the bias as a function of h:

$$\begin{aligned} \widehat{Bias}(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h^*}|\boldsymbol{X}) &= \widehat{E}(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h^*}|\boldsymbol{X}) - \widehat{a}_0 \\ &= \sum_{t=2}^{T} \widehat{a}_t h^{*t}. \end{aligned} \tag{9}$$

The local empirical method uses subsets of the entire data set, with a different subset and so a different least squares fit for each h. Specifically, fix $h^* = h_j \in \mathcal{H}$ and fit model (8) using the smaller 'data set' $\Big\{(h_k, \boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h_k}) : \ k = j - k_1, \ldots, j + k_2\}$, where k_1 and k_2 are user-defined tuning parameters satisfying $k_1 + k_2 \geq T$. The bias of $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h^*}$ is then estimated as the least squares fit $\sum_{t=2}^{T} \widehat{a}_t h^{*t}$.

The non-asymptotic plug-in method we propose for estimating $Bias(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X})$ uses equation (4) and an estimator of $\boldsymbol{m}$ defined in Section 5 below, yielding

$$\widehat{Bias}(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X}) = \boldsymbol{c}^\top \boldsymbol{M}_{\boldsymbol{\Omega},h} \widehat{\boldsymbol{m}}. \tag{10}$$

The formulae for the bias and variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ depend on $\boldsymbol{\Omega}$ being non-random. For $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$, we ignore the fact that $\widehat{\boldsymbol{\Psi}}$ is random and proceed in the obvious way. We estimate the variance of $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ via (7) with $\widehat{\boldsymbol{\Psi}}^{-1}$ replacing $\boldsymbol{\Omega}$. For our empirical and plug-in bias estimation methods, we again replace $\boldsymbol{\Omega}$ with $\widehat{\boldsymbol{\Psi}}^{-1}$.

5. ESTIMATING $\boldsymbol{m}$, σ_ϵ^2 AND $\boldsymbol{\Psi}$

In order to calculate (7) and (10), we need to estimate the nonparametric component $\boldsymbol{m}$, the error variance σ_ϵ^2 and the error correlation matrix $\boldsymbol{\Psi}$. Our estimator of $\boldsymbol{m}$ depends on a bandwidth b. This bandwidth is not to be confused with the bandwidth h in Section 4 in that b is targeted at producing an accurate estimator of $\boldsymbol{X\beta} + \boldsymbol{m}$. This bandwidth b gives us estimators of $\boldsymbol{m}$ and $\boldsymbol{\beta}$, namely $\widehat{\boldsymbol{m}}_{\boldsymbol{I},b}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$. Our estimators of σ_ϵ^2 and $\boldsymbol{\Psi}$ depend on $\widehat{\boldsymbol{m}}_{\boldsymbol{I},b}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$.

5.1. Estimating $\boldsymbol{m}$.

We estimate $\boldsymbol{m}$ via $\widehat{\boldsymbol{m}}_{\boldsymbol{I},b}$, with b chosen from the data via cross-validation, modified to account for possible error correlation. Specifically, if l is a non-negative integer that quantifies our belief in the extent of the serial correlation in the ϵ_i's, we choose b to minimize

$$MCV_l(b) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b} - \widehat{m}_{\boldsymbol{I},b}^{(-i;\,l)}(Z_i) \right)^2. \tag{11}$$

Here, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$ is the usual local linear backfitting estimator of $\boldsymbol{\beta}$ using all of the data and $\widehat{m}_{\boldsymbol{I},b}^{(-i;\,l)}$ is a centred locally linear estimator of m computed as follows from the partial residuals $r_{j,b} = Y_j - \boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$, $j = 1, \ldots, n$, $|i-j| > l$. We use the 'data points' $(Z_j, r_{j,b})$, $|i-j| > l$, in a usual local linear regression to obtain $\widehat{m}_{\boldsymbol{I},b}^{*(-i;\,l)}(z)$, $z \in [0,1]$ and then centre $\widehat{m}_{\boldsymbol{I},b}^{*(-i;\,l)}$ for our 'data', so that $\sum_{j:\,|i-j|>l} \widehat{m}_{\boldsymbol{I},b}^{(-i;l)}(Z_j) = 0$:

$$\widehat{m}_{\boldsymbol{I},b}^{(-i;\,l)}(z) = \widehat{m}_{\boldsymbol{I},b}^{*(-i;\,l)}(z) - \frac{1}{\#\{j:\ |i-j| > l\}} \sum_{j:\,|i-j|>l} \widehat{m}_{\boldsymbol{I},b}^{*(-i;\,l)}(Z_j).$$

It is easy to see that the modified cross-validation criterion in (11) attempts to estimate the conditional mean average squared error of $\widehat{\boldsymbol{m}}_{\boldsymbol{I},b}$, given $\boldsymbol{X}$.

Note that we use all of the data to compute $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$ rather than computing both $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}^{(-i;\, l)}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{I},b}^{(-i;\, l)}(Z_i)$ for each i. We believe that our computational simplification will not affect to a great degree the estimation of $\boldsymbol{m}$. A similar simplification was used by Aneiros-Pérez and Quintela-del-Río (2001b) for their modified cross-validation method.

How does one choose l? In theory, l should be large enough so that there is negligible correlation between Y_i and $(Y_j, X_{j1}, \ldots, X_{jp}, Z_j)$ with $|i-j| > l$. But choosing l too large may yield a highly variable MCV_l, since $\widehat{m}_{\boldsymbol{I},b}^{(-i,l)}$ will depend on very few data points. In practice, we have found that the final results of our methodology are stable over a reasonable range of l values. Our simulation study provides further evidence suggesting that the procedure is insensitive to the choice of l, as long as l is not too close to zero.

5.2. Estimating σ_ϵ^2 and $\boldsymbol{\Psi}$.

To estimate the variance σ_ϵ^2 and correlation matrix $\boldsymbol{\Psi}$ of the errors, we assume (A9), that the model errors follow a stationary autoregressive process of unknown, but finite, order. This assumption will clearly not be appropriate for all applications. However, we expect it to cover those situations where the errors are realizations of a stationary stochastic process. In practice, finite order autoregressive processes are sufficiently accurate because higher order parameters tend to become small and not significant for estimation (Bos et. al., 2002).

First, we estimate $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ by the model residuals $\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{I},b} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b} - \widehat{\boldsymbol{m}}_{\boldsymbol{I},b}$, where b is chosen by leave-$(2l+1)$-out cross-validation, as described in Section 5.1. Next, we use $\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{I},b}$ to estimate the order R via the finite sample criterion developed by Broersen (2000). We then estimate σ_ϵ^2 and the unknown parameters in $\boldsymbol{\Psi}$ via Burg's algorithm, described, for instance, in Brockwell and Davis (1991). Finally, we estimate $\boldsymbol{\Psi}$ by replacing unknown parameters with their estimated values.

If assumption (A9) is questionable, we suggest using the $\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{I},b}$'s $\equiv \widehat{\boldsymbol{\epsilon}}$'s to estimate σ_ϵ^2 and

$\boldsymbol{\Psi}$ as follows:

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{t=1}^{n}\widehat{\epsilon}_t^2$$

$$\widehat{\Psi}_{i,j} = \frac{1}{\widehat{\sigma}_\epsilon^2}\left(\frac{1}{n}\sum_{t=1}^{n-|i-j|}\widehat{\epsilon}_t\widehat{\epsilon}_{t+|i-j|}\right), \text{ for } i \neq j.$$

However, we do not study this approach in this paper.

6. Confidence Intervals for $\boldsymbol{c}^\top\boldsymbol{\beta}$

To conduct inference on $\boldsymbol{c}^\top\boldsymbol{\beta}$, we use approximate $100(1-\alpha)\%$ confidence intervals constructed from the estimators $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$ and $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$, possibly adjusted for finite-sample bias, and their estimated standard errors. We choose h via the methods in Section 4. Recall that each of these methods produces an estimator of the bias. We will use these bias estimators to bias-adjust our confidence intervals. We use standard normal probability cut-offs to construct all intervals. The simulation results presented in Section 7 provide support for the use of these intervals, although we have not yet demonstrated the asymptotic normality of our estimators.

The standard $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^\top\boldsymbol{\beta}$ constructed from $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ is

$$\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h} \pm z_{\alpha/2}\widehat{SE}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}), \tag{12}$$

where $z_{\alpha/2}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution and $\widehat{SE}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h})$ is the square root of the estimator of $Var(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X})$ in (7).

Since $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ may be biased in finite samples, the standard confidence interval in (12) may not be correctly centred. Therefore, we propose the bias-adjusted version:

$$\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h} - \widehat{Bias}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}) \pm z_{\alpha/2}\widehat{SE}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}). \tag{13}$$

Here $\widehat{Bias}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h})$ is calculated using the same bias estimator used in choosing h.

The length of the bias-adjusted confidence interval for $\boldsymbol{c}^\top\boldsymbol{\beta}$ in (13) is the same as that of the standard confidence interval in (12). The coverage properties of the bias-adjusted confidence interval may, however, be better than those of the standard confidence interval.

Standard and bias-adjusted versions for the confidence intervals relying on $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$ and $\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ are obtained in a similar fashion.

7. A SIMULATION STUDY

In our simulation study, we consider model (1) with $p = 1$ and we conduct inference on β_1. We compare the mean squared error (MSE) of our estimators of β_1 and the coverage and length of our corresponding confidence intervals. We also include a comparison with the usual Speckman estimator. We take $n = 100$, $\beta_0 = \beta_1 = 0$ and consider two m functions, $m_1(z) = 2sin(3z) - 2(cos(0) - cos(3))/3$ and $m_2(z) = 2sin(6z) - 2(cos(0) - cos(6))/6$. We take $Z_i = i/(n+1)$ and $X_i = g(Z_i) + \eta_i$, with $g(z) = 0.4z + 0.3$ and η_i uniformly distributed on $(-0.3, 0.3)$. We assume the ϵ_i's follow a stationary $AR(1)$ model generated by

$$\epsilon_1 = u_1/\sqrt{1-\rho^2}, \qquad \epsilon_i = \rho\epsilon_{i-1} + u_i \quad i = 2, \ldots, n \tag{14}$$

with $\rho = 0$, 0.2, 0.4, 0.6 and 0.8. The u_i's, independent of the ϵ_i's, are independent, identically distributed normal random variables having mean 0 and standard deviation $\sigma_u = 0.5$. For each of the ten model configurations, we generate 500 data sets.

Opsomer and Ruppert (1999) used essentially the same simulation settings for their simulation study, but they only considered the case $\rho = 0$, for independent errors.

Figure 1 displays data generated from our model for $\rho = 0, 0.4, 0.8$ and $m = m_1$. The responses Y_i are qualitatively different for different values of ρ. For $\rho = 0$, the responses vary randomly about the m curve. As ρ increases from 0.4 to 0.8, the autocorrelation-induced structure in the variation of the Y_i's about the curve makes it virtually impossible to distinguish the signal from the autoregressive noise.

The estimators of β_1 we consider are of the form $\widehat{\beta}_1 = \boldsymbol{c}^\top\widehat{\boldsymbol{\beta}} \equiv (0,1)\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is: (i) the usual local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, (ii) the estimated modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ or (iii) the usual Speckman estimator $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top,h}$. We choose the smoothing parameters of our two backfitting estimators via the methods introduced in Section 4: the global empirical method, the local empirical method and the non-asymptotic plug-in method. We choose the smoothing parameter of the usual Speckman estimator as the minimizer of a modified cross-validation criterion, obtained from (11) by replacing $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},b}$ with $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top,h}$ and $\widehat{m}_{\boldsymbol{I},b}^{(-i;l)}(Z_i)$ with $\widehat{m}_{(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top,h}^{(-i;l)}(Z_i)$. All methods use a grid $\mathcal{H} = \{0.01, 0.02, \ldots, 0.5\}$.

The methods require the specification of various tuning parameters. All require an l for the MCV criterion (11). We consider $l = 0, 1, \ldots, 10$. In the global and local empirical methods, we take $T = 3$ for the polynomial expansion of the bias in (9). We take $k_1 = k_2 = 5$ in the local empirical method.

We made pairwise comparisons of the MSE's of our estimators computed with our bandwidth choice methods by examining boxplots of differences in log MSE's and by conducting level 0.05 paired t-tests of the null hypothesis of equality of expected log MSE's. Specifically we first made separate comparisons of the bandwidth choice methods for the usual and for the estimated modified backfitting estimators, to determine our preferred bandwidth choices for each estimator. We then compared the estimators using our preferred bandwidth choice methods to the usual Speckman estimator with h chosen by modified cross-validation. For $\rho = 0$, the MSE results were mixed. For $\rho > 0$ and $l \geq 4$, however, the results were conclusive. In this latter case, the MSE's of the usual backfitting estimators tended to be smallest when using h selected by either the global empirical method or by the non-asymptotic plug-in method. The MSE's of the estimated modified backfitting estimators tended to be smallest when using the global empirical method. The MSE's of these three preferred methods of estimation were comparable, and tended to be lower than the MSE's of the Speckman estimator.

We studied coverage of both standard and bias-adjusted confidence intervals for β_1 obtained from $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ with h chosen by our preferred methods. We also studied standard confidence intervals based on our Speckman-type estimator. We did not consider bias adjustments for these confidence intervals because $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top,h}$ has small bias both when $\rho = 0$ (Speckman, 1988) and when $\rho > 0$ (Aneiros-Pérez and Quintela-del-Río, 2001a). To assess a method's coverage we calculated $\widehat{p}$, the proportion of the 500 resulting confidence intervals which contained the true value of β_1. If $\widehat{p}$ fell within $1.96\sqrt{\widehat{p}(1-\widehat{p})/500}$ of 0.95, we deemed that the confidence interval procedure achieved nominal coverage.

We found no advantage to bias-correction and therefore now just discuss standard confidence intervals. Intervals based on $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ with h chosen by our preferred method had very poor coverage properties and so we eliminated this method from further study. The remaining three methods, namely those based on $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ with the global empirical and non-asymptotic plug-in choices of h and $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{S}_h^c)^\top,h}$ produced intervals with much better coverage. The methods based on $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ achieved nominal coverage for $\rho > 0$ and $l \geq 4$. The Speckman-based intervals achieved coverage close to the nominal for small values of l, but tended to undercover slightly for large values of l and $\rho > 0$. This is illustrated in Figure 2 for selected simulation settings. Depicted are 95% confidence intervals for the true coverage probability.

We next compared the log lengths of the standard confidence intervals produced by our three remaining methods. We used level 0.05 two-sided paired t-tests and visual assessments of boxplots of differences in log lengths. We found that the Speckman-based confidence intervals tended to be wider, particularly when $\rho > 0$. Figure 3, showing average confidence interval lengths for these three methods, illustrates this for selected simulation settings.

In summary, in the context of our simulation study, the backfitting estimators with h selected by either the non-asymptotic plug-in method or the global empirical method provide accurate estimators of β_1 and produce narrow confidence intervals that achieve nominal coverage. We recommend their use, with values of l that are large enough, that is, at least 4. The usual Speckman estimator with h selected via modified cross-validation is competitive and tends to be superior for small values of l. However, this estimator sometimes has large log MSE and produces slightly longer confidence intervals that sometimes suffer from slight undercoverage. We do not recommend choosing h by the local EBBS method nor do we recommend inference using the estimated modified back-fitting estimator of β_1. Both methods performed poorly in our simulation study.

8. AN APPLICATION TO AN AIR POLLUTION STUDY

We apply our preferred inferential methodology to the analysis of mortality and air pollution data collected in Mexico City between January 1, 1994, and December 31, 1996. The data consist of daily counts of non-accidental deaths, temperature (^oC), relative humidity (%), and daily levels of ambient concentration of PM10 - airborne particulate matter less than 10 microns in diameter ($10\mu g/m^3$). Our goal is to determine if the pollutant PM10 has a significant short-term effect on the non-accidental death rate in Mexico City after controlling for long-term temporal trends and meteorological confounding.

Pairwise scatter plots of the data are shown in Figure 4. The most striking features in these plots are the strong annual cycles in the daily levels of mortality, PM10, temperature and relative humidity. The annual cycles in the mortality levels are possibly produced by unobserved seasonal factors such as influenza and respiratory infections. Our analysis of the health effects of PM10 must account for the potential confounding effect of these annual cycles on the association between PM10 and log mortality.

Let D_i denote the observed number of non-accidental deaths in Mexico City on day i, and let P_i, T_i and H_i denote the daily measures of PM10, temperature and relative humidity, respectively. A model that provides an adequate description of the variability in the log mortality counts is:

$$log(D_i) = \beta_0 + \beta_1 P_i + \beta_2 T_i + \beta_3 H_i + m(i) + \epsilon_i, i = 1, 2, \ldots, 1096. \tag{15}$$

Here, m is an unknown, smooth univariate function representing a long-term temporal trend.

We estimate β_1 via $(0,1,0,0)\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h} \equiv \boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$. To choose h, we use our two preferred methods: non-asymptotic plug-in and global empirical. For both methods, we use a grid $\mathcal{H} = \{7, 8, \ldots, 90\}$. For the global empirical method, we take $T = 4$ in equation (8). We allow the tuning parameter l of the leave-$(2l+1)$-out cross-validation to take on the values $0, 1, \ldots, 10$ to investigate sensitivity to the choice of l. See Ghement et al. (2006) for more details of the data analysis, including model selection and choice of T and $\mathcal{H}$.

Figure 5 shows the corresponding 95% confidence intervals for β_1 obtained from formula

(12). The two methods produce remarkably similar confidence intervals. Further, the choice of l has little influence on these intervals, likely because the extent of the residual correlation present in the data is modest. From Figure 5, there is no conclusive proof that β_1, the short-term PM10 effect on log mortality, is significantly different from 0.

9. FINAL REMARKS

We have provided asymptotic theory for backfitting estimates of $\boldsymbol{\beta}$ and m in the partially linear model (1) with correlated errors. Our theory shows that the smoothing parameter appropriate for inference for $\boldsymbol{\beta}$ is different in order of magnitude from the smoothing parameter appropriate for inference for m. We propose and study several methods for smoothing parameter choice and for constructing confidence intervals for linear combinations of the components of $\boldsymbol{\beta}$. Our simulation study indicates that the popular Speckman estimate performs reasonably well, except when data are highly correlated. We also find that the EBBS method for smoothing choice does not perform well, but a modification, a global EBBS does perform well. We also found that a non-asymptotic plug-in method did well.

We hoped to improve efficiency of our estimators by a generalized least squares approach, that is, via our estimator $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$. However, our simulation studies indicated that such an approach was unsuccessful, even when our errors followed a simple $AR(1)$ process. Thus, we recommend using $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$. We also considered adjusting our confidence intervals to account for the bias in $\widehat{\boldsymbol{\beta}}$. These bias adjustments yielded no improvement in the context of our simulation.

10. PROOFS OF THEOREMS

Throughout we use the notation that, for a matrix $\boldsymbol{A}$, $\boldsymbol{A}_{\cdot j}$ denotes the jth column. We let $\boldsymbol{\eta}$ be the $n \times (p+1)$ matrix with first column entries equal to 0 and, for $j = 2, \ldots, p+1$, $[\boldsymbol{\eta}]_{ij} = \eta_{i,j-1}$ where $\eta_{i,j-1}$ is as in assumption (A5). For the proofs of both theorems, we repeatedly use the bounds $||\boldsymbol{\eta}_{\cdot j}||_2 = O_P(n^{1/2})$, $||\boldsymbol{G}_{\cdot j}||_2 = O(n^{1/2})$ and $||\boldsymbol{\Psi}||_S = O(1)$. The proofs of the first two bounds are straightforward. For Theorem 1, the last bound is an assumption. For Theorem 2, we prove the last bound in Lemma 3. This lemma also provides us with the bound $||\boldsymbol{\Psi}^{-1}||_S = O(1)$. All lemmas appear in the Appendix. We now prove Theorem 2. The proof of Theorem 1 is simpler and is sketched afterwards.

Proof of Theorem 2.

Proof of (i): From (4) with $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$,

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h} &= \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\cdot\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{Y}\\ &\equiv \left(\boldsymbol{B}_{\boldsymbol{\Psi}^{-1},h}\right)^{-1}\cdot\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{Y}.\end{aligned} \tag{16}$$

We first show that

$$\boldsymbol{B}_{\boldsymbol{\Psi}^{-1},h} = \boldsymbol{V}_{\boldsymbol{\Psi}} + \boldsymbol{o}_P(1). \tag{17}$$

Rewriting $\boldsymbol{B}_{\boldsymbol{\Psi}^{-1},h}$ using $\boldsymbol{X} = \boldsymbol{G}+\boldsymbol{\eta}$ from equation (6), we obtain

$$\boldsymbol{B}_{\boldsymbol{\Psi}^{-1},h} = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G} + \frac{1}{n}\boldsymbol{G}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta} + \frac{1}{n}\boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta} - \frac{1}{n}\boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c\boldsymbol{\eta}. \tag{18}$$

By Lemma 5 with $\boldsymbol{v}_n = \boldsymbol{\eta}_{\cdot i}$ and $\boldsymbol{w}_n = \boldsymbol{\eta}_{\cdot j}$,

$$\frac{1}{n}\boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\left(1+\sum_{k=1}^R\phi_k^2\right)\boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1). \tag{19}$$

To complete the proof of (17), we show that the second and last terms in (18) are $\boldsymbol{o}_P(1)$ and that

$$\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\left(1-\sum_{k=1}^R\phi_k\right)^2\boldsymbol{g}^*\boldsymbol{g}^{*\top} + \boldsymbol{o}_P(1). \tag{20}$$

Consider the (i,j)th component of the left side of (20), namely $\boldsymbol{X}_{\cdot i}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}/n = (\boldsymbol{G}+\boldsymbol{\eta})_{\cdot i}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}/n$. To analyze $\boldsymbol{G}_{\cdot i}^\top\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}/n$, we apply Lemma 4 with $\boldsymbol{v}_n = \boldsymbol{G}_{\cdot i}$ and $\boldsymbol{w}_n = (\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}$. By assumptions (A2) and (A3), $\boldsymbol{v}_n$ satisfies the conditions of Lemma 4. The components of $\boldsymbol{w}_n$ are bounded by (v) of Lemma 1 with $\boldsymbol{v} = \boldsymbol{G}_{\cdot j}$, so $\boldsymbol{w}_n$ also satisfies the conditions of Lemma 4. Thus

$$\begin{aligned}\frac{1}{n}\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j} &= \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^R \phi_k\right)^2 \frac{1}{n}\boldsymbol{G}_{\cdot i}^\top(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j} + O\left(n^{-1}\right) \\ &= \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^R \phi_k\right)^2 \frac{1}{n^2}\boldsymbol{G}_{\cdot i}^\top \boldsymbol{1}\boldsymbol{1}^\top \boldsymbol{G}_{\cdot j} + o(1) \\ &= \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^R \phi_k\right)^2 g_i^* g_j^* + o(1)\end{aligned}$$

where the second line follows from (iii) of Lemma 2 with $m_1 = g_{i-1}$ and $m_2 = g_{j-1}$ and the last line follows from a Riemann integration argument. We now show that $\boldsymbol{\eta}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}/n$ is $o_P(1)$. When $i = 1$, this component is 0. For $i = 2, \ldots, p+1$, this component has mean 0 and variance

$$\begin{aligned}Var\left(\frac{1}{n}\boldsymbol{\eta}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}\right) &= \frac{\Sigma_{i-1,i-1}}{n^2} \cdot ||\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}||_2^2 \\ &\leq \frac{\Sigma_{i-1,i-1}}{n^2} \cdot ||\boldsymbol{\Psi}^{-1}||_S^2 \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{\cdot j}||_2^2 = \frac{1}{n^2} \cdot O(1) \cdot O(n) = o(1) \qquad (21)\end{aligned}$$

by (v) of Lemma 1 with $\boldsymbol{v} = \boldsymbol{G}_{\cdot j}$. Thus (20) holds.

To show that the second term in (18) is $o_P(1)$, consider $\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta}_{\cdot j}/n$. When $j = 1$, this component is 0. For $j = 2, \ldots, p+1$, this component has mean 0 and variance $\Sigma_{j-1,j-1} \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}||_2^2/n^2$. By Lemma 3 (iii), the components of $\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}$ are bounded, and so the variance is $O(1/n) = o(1)$ by (v) of Lemma 1 with $\boldsymbol{v} = \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}$.

Finally consider the (i, j)th component of the last term in (18). Since this component is 0 whenever $i = 1$ or $j = 1$, we consider $i, j = 2, \ldots, p+1$ and write

$$\left|\frac{1}{n}\boldsymbol{\eta}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot j}\right| \leq \frac{1}{n}||\boldsymbol{\eta}_{\cdot i}||_2 \cdot ||\boldsymbol{\Psi}^{-1}||_S \cdot ||\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot j}||_2 = \frac{1}{n} \cdot O_P(n^{1/2}) \cdot O(1) \cdot O_P(h^{-1/2}) = o_P(1).$$

For the above, we used the bound $||\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot j}||_2 = O_P(h^{-1/2})$ from Lemma 6. This completes our proof of (17).

Thus, on an $\boldsymbol{X}$-set whose probability converges to 1,

$$Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X}) = \left(\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} + \boldsymbol{o}_P(1)\right) \cdot \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}$$

The formula for the asymptotic bias follows by studying $\boldsymbol{X}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}/n$ by the same argument as in the proof of (20), but retaining the $O(h^2)$ term in Lemma 2 (*iii*).

Proof of (ii): From (16) and (17), we have, on an $\boldsymbol{X}$-set whose probability converges to 1,

$$Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X}) = \frac{\sigma_\epsilon^2}{n}\left(\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} + \boldsymbol{o}_p(1)\right) \cdot (\boldsymbol{X}^\top \boldsymbol{\mathcal{W}} \boldsymbol{X}/n) \cdot \left(\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} + \boldsymbol{o}_p(1)\right),$$

where $\boldsymbol{\mathcal{W}} = \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{\Psi}^{-1}$. The proof of *(ii)* follows if we show that

$$\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\mathcal{W}} \boldsymbol{X} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 + \sum_1^R \phi_k^2\right)\boldsymbol{\Sigma}^{(0)} + \frac{1}{n}\boldsymbol{G}^\top \boldsymbol{\mathcal{W}} \boldsymbol{G} + \boldsymbol{o}_P(1)$$

and that $\boldsymbol{G}^\top \boldsymbol{\mathcal{W}} \boldsymbol{G}/n = \boldsymbol{O}(1)$. To show the first, use $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$, and write

$$\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\mathcal{W}} \boldsymbol{X} = \frac{1}{n}\boldsymbol{G}^\top \boldsymbol{\mathcal{W}} \boldsymbol{G} + \frac{1}{n}\boldsymbol{G}^\top \boldsymbol{\mathcal{W}} \boldsymbol{\eta} + \frac{1}{n}\boldsymbol{\eta}^\top \boldsymbol{\mathcal{W}} \boldsymbol{G} + \frac{1}{n}\boldsymbol{\eta}^\top(\boldsymbol{\mathcal{W}} - \boldsymbol{\Psi}^{-1})\boldsymbol{\eta} + \frac{1}{n}\boldsymbol{\eta}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\eta}.$$

Since (19) holds for the last term, we must only establish that $\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\mathcal{W}} \boldsymbol{\eta}_{\cdot j}/n$ and $\boldsymbol{\eta}_{\cdot i}^\top \left(\boldsymbol{\mathcal{W}} - \boldsymbol{\Psi}^{-1}\right)\boldsymbol{\eta}_{\cdot j}/n$ are $o_P(1)$ for all $i, j = 1, \ldots, p+1$. When $j = 1$, these terms are 0.

Consider $\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\mathcal{W}} \boldsymbol{\eta}_{\cdot j}/n$, $j > 1$. This term is $o_P(1)$ since it has mean 0 and variance

$$\begin{aligned} Var\left(\frac{1}{n}\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\mathcal{W}} \boldsymbol{\eta}_{\cdot j}\right) &= \frac{\Sigma_{j-1,j-1}}{n^2} \cdot ||\boldsymbol{\mathcal{W}} \boldsymbol{G}_{\cdot i}||_2^2 \le \frac{\Sigma_{j-1,j-1}}{n^2} \cdot ||\boldsymbol{\mathcal{W}}||_S^2 \cdot ||\boldsymbol{G}_{\cdot i}||_2^2 \\ &\le \frac{\Sigma_{j-1,j-1}}{n^2} \cdot ||\boldsymbol{\Psi}^{-1}||_S^4 \cdot ||\boldsymbol{\Psi}||_S^2 \cdot (1 + ||\boldsymbol{S}_h^c||_F)^4 \cdot ||\boldsymbol{G}_{\cdot i}||_2^2 = \frac{1}{n^2}O(h^{-2}) \cdot O(n) \end{aligned}$$

by (iv) of Lemma 1, and thus the variance converges to 0.

Consider $\boldsymbol{\eta}_{\cdot i}^\top(\boldsymbol{\mathcal{W}} - \boldsymbol{\Psi}^{-1})\boldsymbol{\eta}_{\cdot j}/n$. When $i = 1$, this term is 0. When $i = 2, \ldots, p+1$, these terms are $o_P(1)$ since they can be bounded as

$$\begin{aligned} \left|\frac{1}{n}\boldsymbol{\eta}_{\cdot i}^\top \left(\boldsymbol{\mathcal{W}} - \boldsymbol{\Psi}^{-1}\right)\boldsymbol{\eta}_{\cdot j}\right| &= \left|\frac{1}{n}\boldsymbol{\eta}_{\cdot i}^\top \left(-\boldsymbol{S}_h^{c\top}\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c + \boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c \boldsymbol{\Psi} \boldsymbol{S}_h^{c\top}\boldsymbol{\Psi}^{-1}\right)\boldsymbol{\eta}_{\cdot j}\right| \\ &\le \frac{1}{n}||\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot i}||_2 \cdot ||\boldsymbol{\Psi}^{-1}||_S \cdot ||\boldsymbol{\eta}_{\cdot j}||_2 + \frac{1}{n}||\boldsymbol{\eta}_{\cdot i}||_2 \cdot ||\boldsymbol{\Psi}^{-1}||_S \cdot ||\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot j}||_2 \\ &\quad + \frac{1}{n}||\boldsymbol{S}_h^{c\top}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_{\cdot i}||_2 \cdot ||\boldsymbol{\Psi}||_S \cdot ||\boldsymbol{S}_h^{c\top}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_{\cdot j}||_2 \\ &= \frac{1}{n}O_P(n^{1/2})O_P(h^{-1/2}) + \frac{1}{n}O_P(h^{-1}) = O_P(n^{-1/2}h^{-1/2}). \end{aligned}$$

Here, we used Lemma 6's bounds on $||\boldsymbol{S}_h^{c\top}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_{\cdot i}||_2$ and $||\boldsymbol{S}_h^c \boldsymbol{\eta}_{\cdot i}||_2$. Thus, the first equality in (ii) holds.

To establish the second equality in (ii), write

$$|\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\mathcal{W}} \boldsymbol{G}_{\cdot j}| = |\boldsymbol{G}_{\cdot i}^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot j}|$$
$$\leq ||(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}||_2 \cdot ||\boldsymbol{\Psi}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot j}||_2 = O(n)$$

by applying Lemma 1 (v) in the Appendix twice, once with $\boldsymbol{v} = \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}$ and once with $\boldsymbol{v} = \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot j}$. Both $\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot i}$ and $\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{\cdot j}$ satisfy the conditions of Lemma 1 (v), as their components are bounded by Lemma 3 (iii).

Proof of (iii) : The proof follows easily from parts (i) and (ii), by writing $E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h} - \boldsymbol{\beta}||_2^2 \,\middle|\boldsymbol{X}\right) = ||Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X})||_2^2 + trace\left\{Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}|\boldsymbol{X})\right\}$.

Proof of Theorem 1. This proof is similar to that of Theorem 2, so is just sketched. The modification of the proof of *(i)* is straightforward. For the first equality in *(ii)*, one shows that, on an $\boldsymbol{X}$-set whose probability converges to 1, $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}) = (\sigma_\epsilon^2/n)\left(\boldsymbol{V}_{\boldsymbol{I}}^{-1} + \boldsymbol{o}_p(1)\right)(\boldsymbol{X}^\top \boldsymbol{\mathcal{W}}_{\boldsymbol{I}}\boldsymbol{X}/n) \cdot \left(\boldsymbol{V}_{\boldsymbol{I}}^{-1} + \boldsymbol{o}_p(1)\right)$, where $\boldsymbol{\mathcal{W}}_{\boldsymbol{I}} = (\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top$. As in Theorem 2, we can show that $\boldsymbol{X}^\top \boldsymbol{\mathcal{W}}_{\boldsymbol{I}}\boldsymbol{X}/n = \boldsymbol{G}^\top \boldsymbol{\mathcal{W}}_{\boldsymbol{I}}\boldsymbol{G}/n + \boldsymbol{\Phi}^{(0)} + \boldsymbol{o}_P(1)$. The only difference in the proof is that here we study the terms $\boldsymbol{\eta}^\top(\boldsymbol{\mathcal{W}}_{\boldsymbol{I}} - \boldsymbol{\Psi})\boldsymbol{\eta}$ and $\boldsymbol{\eta}^\top\boldsymbol{\Psi}\boldsymbol{\eta}$ instead of $\boldsymbol{\eta}^\top(\boldsymbol{\mathcal{W}} - \boldsymbol{\Psi}^{-1})\boldsymbol{\eta}$ and $\boldsymbol{\eta}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}$. The proof of the second equality in (ii) and that of (iii) are clear-cut modifications of the corresponding proofs in Theorem 2.

Proof of Theorem 3. Dropping $\widehat{\boldsymbol{m}}$'s and $\widehat{\boldsymbol{\beta}}$'s subscript of $\boldsymbol{\Omega}, h$ write

$$\widehat{\boldsymbol{m}} = \boldsymbol{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) - \boldsymbol{S}_h^c\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{22}$$

Then the conditional bias of $\widehat{\boldsymbol{m}}$ given $\boldsymbol{X}$ is

$$E(\widehat{\boldsymbol{m}}|\boldsymbol{X}) - \boldsymbol{m} = -(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} - \boldsymbol{S}_h^c\boldsymbol{X} \times Bias(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}).$$

We study the ith component of this conditional bias. By Lemma 2 (i)

$$-[(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}]_i = h^2\ \frac{\nu_2(K)}{2}\ \left[m''(Z_i) - \int_0^1 m''(z)\ f(z)\ dz\right] + o(h^2)$$

for $Z_i \in [h, 1-h]$ and is $O(h^2)$ for $Z_i \notin [h, 1-h]$. To analyze the ith component in the second term in the conditional bias expression, we show that

$$[\boldsymbol{S}_h^c\boldsymbol{X}]_{ij} = G_{ij} - g_j^* + o_p(1) \qquad j = 1, \ldots, p+1 \tag{23}$$

to get

$$[\boldsymbol{S}_h^c \boldsymbol{X} \times Bias(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})]_i = h^2 \frac{\nu_2(K)}{2} (\boldsymbol{G}_{i\cdot} - \boldsymbol{g}^*)^\top \boldsymbol{b} + o_p(h^2).$$

To prove (23) write

$$E[\boldsymbol{S}_h^c \boldsymbol{X}]_{ij} = [\boldsymbol{S}_h^c \boldsymbol{G}]_{ij} = G_{ij} - [(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}]_{ij} = G_{ij} - [\mathbf{1}\mathbf{1}^\top \boldsymbol{G}_{\cdot j}/n]_i + O(h^2) = G_{ij} - g_j^* + o(1) + O(h^2)$$

by first applying Lemma 2 (i) with $\boldsymbol{m}_2 = \boldsymbol{G}_{\cdot j}$ and then using Riemann integration. We now show that the variance of $[\boldsymbol{S}_h^c \boldsymbol{X}]_{ij}$ converges to 0 uniformly in i and j. Write

$$var\left([\boldsymbol{S}_h^c \boldsymbol{X}]_{ij}\right) = var\left(\sum_{k=1}^n [\boldsymbol{S}_h^c]_{ik} \boldsymbol{\eta}_{kj}\right) = \sum_{k=1}^n [\boldsymbol{S}_h^c]_{ik}^2 \, \Sigma_{jj} \leq \Sigma_{jj} \, \mathcal{C}_0 \sum_{k=1}^n \left[\frac{1}{nh} \, I\{|i-k| \leq \mathcal{C}_1 nh\} + \frac{1}{n}\right]^2$$

for some constants $\mathcal{C}_0$ and $\mathcal{C}_1$, by (iii) of Lemma 1. Therefore, the variance of $[\boldsymbol{S}_h^c \boldsymbol{X}]_{ij}$ is $O(1/(nh)) = o(1)$. This completes the proof of (23), and thus, the proof of the statements concerning the asymptotic conditional bias of $[\widehat{\boldsymbol{m}}]_i$.

We now study the conditional variance of $[\widehat{\boldsymbol{m}}]_i$ via the conditional variances of the ith components of the two terms in (22). To study the conditional variance of the first term's ith component, we use Corollary 1 of Francisco-Fernández and Vilar-Fernández (2001), which states that

$$var([\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_i | \boldsymbol{X}) = \frac{\sigma_\epsilon^2}{f(Z_i) \; nh} \; \nu_0(K^2) \; \left(\rho(0) + 2\sum_1^\infty \rho(k)\right) + o_p\left((nh)^{-1}\right).$$

Studying their proof, we see that this holds uniformly in i with $Z_i \in [2h, 1-2h]$ and that we can bound $var([\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_i|\boldsymbol{X})$ by a constant over nh uniformly in i with $Z_i \notin [2h, 1-2h]$. Both of these results require the additional condition that $\sum_{k=1}^\infty k|\rho(k)| < \infty$. To use these results for $\boldsymbol{S}_h^c$ instead of $\boldsymbol{S}_h$ write

$$var\left([\boldsymbol{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_i \middle| \boldsymbol{X}\right) = var\left([\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_i - \frac{1}{n}\sum_j [\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_j \middle| \boldsymbol{X}\right).$$

We now show that the conditional variance of $\sum_j [\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_j/n = O_p(1/n)$, and thus this term is negligible. Write

$$var\left(\frac{1}{n}\sum_j [\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_j \middle| \boldsymbol{X}\right) = \frac{1}{n^2}\sum_{j,j'} cov\left([\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_j, [\boldsymbol{S}_h(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_{j'} \middle| \boldsymbol{X}\right)$$

$$= \frac{1}{n^2} \sum_{j,j'} \sum_{l,l'} cov \left([\boldsymbol{S}_h]_{jl} [(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_l, [\boldsymbol{S}_h]_{j'l'} [(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})]_{l'} \middle| \boldsymbol{X} \right)$$

$$= \frac{\sigma^2}{n^2} \sum_{j,j'} \sum_{l,l'} [\boldsymbol{S}_h]_{jl} [\boldsymbol{S}_h]_{j'l'} \rho(|l - l'|) \leq \frac{\sigma^2}{n^2} C \sum_{l,l'} \left| \rho(|l - l'|) \right|$$

for some C not depending on i, by (ii) of Lemma 1 and assumption (A3) on the design density of the Z_j's. But

$$\frac{1}{n^2} \sum_{l,l'} \left| \rho(|l - l'|) \right| \leq \frac{2}{n^2} \sum_{l=1}^{n} \sum_{k=0}^{\infty} |\rho(k)| = O(n^{-1}) = o((nh)^{-1}).$$

We complete the analysis of the conditional variance of $[\widehat{\boldsymbol{m}}]_i$ by showing that the conditional variance of $[\boldsymbol{S}_h^c \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]_i$ is $O_p(1/n)$, and thus negligible. First we write the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ from Theorem 1 or 2 as $\sigma_\epsilon^2 \boldsymbol{V}_n/n + \boldsymbol{o}_p(1/n)$ and let $\boldsymbol{s}_i^\top$ denote the ith row of $\boldsymbol{S}_h^c \boldsymbol{X}$. By (23), $\boldsymbol{s}_i = \boldsymbol{O}_p(1)$ and so

$$var \left([\boldsymbol{S}_h^c \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})]_i | \boldsymbol{X} \right) = var \left[\boldsymbol{s}_i^\top \widehat{\boldsymbol{\beta}} | \boldsymbol{X} \right] = \boldsymbol{s}_i^\top \left[\frac{\boldsymbol{V}_n}{n} + \boldsymbol{o}_p \left(n^{-1} \right) \right] \boldsymbol{s}_i = O_p(n^{-1}).$$

APPENDIX

Throughout the Appendix, suppose that $S_{ij},\ i,j = 1, \ldots, n$, are the local linear smoothing weights defined at the end of Section 2. Also, let K be a kernel function as defined in (A4). For $z \in [0,1]$ and $h \in (0, 1/2]$, let:

$$\nu_l(K, z, h) = \int_{-z/h}^{(1-z)/h} s^l K(s) ds$$

and

$$V_l(K, z, h) = \frac{\nu_l(K, z, h)}{\nu_2(K, z, h)\nu_0(K, z, h) - \nu_1(K, z, h)^2}.$$

LEMMA 1. *Suppose that $h = h_n$ satisfies assumption (A1) and that assumptions (A3) and (A4) hold. Then the following hold.*

(i) *For $l = 1, 2, 3$, $\sup_{h \in (0,1/2]} \sup_{z \in [0,1]} |V_l(K, z, h)| < \infty$.*

(ii) There exists $\mathcal{C} > 0$ so that, for all n sufficiently large, $|S_{ij}| \leq \mathcal{C} \cdot I(|Z_i - Z_j| \leq h)/(nh)$ for all $i, j = 1, \ldots, n$.

(iii) There exists $\mathcal{C}_0$ and $\mathcal{C}_1$ independent of i, k and n such that

$$|[\boldsymbol{S}_h^c]_{ik}| \leq \mathcal{C}_0 \left[\frac{1}{nh} I\{|i-k| \leq \mathcal{C}_1 nh\} + \frac{1}{n} \right].$$

(iv) The Frobenius norm of $\boldsymbol{S}_h^c$ satisfies $||\boldsymbol{S}_h^c||_F = O(h^{-1/2})$.

(v) Given $\mathcal{C}_2 > 0$, there exists $\mathcal{C}_3 > 0$ such that for any $\boldsymbol{v} = (v_1, \ldots, v_n)^\top$ with $|v_j| \leq \mathcal{C}_2$, we have $|[(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{v}]_j| \leq \mathcal{C}_3$ and $|[(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{v}]_j| \leq \mathcal{C}_3$ for all $j = 1, \ldots, n$. Thus $||(\boldsymbol{I} - \boldsymbol{S}_h^c)^\top \boldsymbol{v}||_2^2 \leq n\mathcal{C}_3^2$ and $||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{v}||_2^2 \leq n\mathcal{C}_3^2$.

Proof. To establish (i), we bound $V_l(K, \cdot, h)$ on each of the intervals $[0, h]$, $[h, 1-h]$ and $[1-h, 1]$. First consider $z \in [0, h]$. Then

$$\sup_{z \in [0,h]} |V_l(K, z, h)| = \sup_{\alpha \in [0,1]} |V_l(K, \alpha h, h)| = \sup_{\alpha \in [0,1]} \frac{\int_{-\alpha}^1 s^l K(s) ds}{\int_{-\alpha}^1 s^2 K(s) ds \int_{-\alpha}^1 K(s) ds - \left(\int_{-\alpha}^1 s K(s) ds \right)^2}$$

$$= \sup_{\alpha \in [0,1]} \frac{1}{\int_{-\alpha}^1 K(s) ds} \cdot \frac{E(X_\alpha^l)}{Var(X_\alpha^l)},$$

where X_α has density $K(s)/\int_{-\alpha}^1 K(u) du$, $s \in [-\alpha, 1]$. Clearly, $\inf_{\alpha \in [0,1]} \int_{-\alpha}^1 K(s) ds > 0$. Also, $\sup_{\alpha \in [0,1]} E(X_\alpha^l) < \infty$ and $\inf_{\alpha \in [0,1]} Var(X_\alpha^l) > 0$. Thus, $\sup_{h \in (0,1/2]} \sup_{z \in [0,h]} |V_l(K, z, h)| < \infty$. Similarly, $\sup_{h \in (0,1/2]} \sup_{z \in [1-h,1]} |V_l(K, z, h)| < \infty$. For $z \in [h, 1-h]$, $V_l(K, z, h) = E(X_1^l)/Var(X_1)$, which is finite.

The proof of (ii) follows directly from (i), the compactness of the support of K, and the following standard result in smoothing. Uniformly in $i, j = 1, \ldots, n$,

$$S_{ij} = \frac{1}{f(Z_i)(n+1)h} \cdot \left[V_2(K, Z_i, h) - V_1(K, Z_i, h) \frac{Z_i - Z_j}{h} + o(1) \right] \cdot K\left(\frac{Z_i - Z_j}{h} \right).$$

The proof of (iii) follows from (ii) and assumption (A3) on the design density of the Z_i's. Statements (iv) and (v) follow directly from (iii).

LEMMA 2. *Let m_1 and m_2 be arbitrary functions having three continuous derivatives and define $\boldsymbol{m}_l = (m_l(Z_1), \ldots, m_l(Z_n))^\top$, $l = 1, 2$. Suppose that $h = h_n$ satisfies assumption (A1) and assumptions (A3) and (A4) are fulfilled. Then the following hold.*

(i) Let $B_m(K,z,h) = (m''(z)/2)\left[\nu_2(K,z,h)V_2(K,z,h) - \nu_3(K,z,h)V_1(K,z,h)\right]$. *Then uniformly in* $j = 1, \ldots, n$,

$$[(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}_2]_j = -h^2\left[B_{m_2}(K, Z_j, h) - \frac{1}{2}\nu_2(K)\cdot\int_0^1 m_2''(z)f(z)dz\right] + o(h^2) + \frac{11^\top}{n}\boldsymbol{m}_2 = O(h^2) + \frac{11^\top}{n}\boldsymbol{m}_2.$$

For j *with* $Z_j \in [h, 1-h]$,

$$[(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}_2]_j = -h^2\frac{\nu(K)}{2}\left[m_2''(Z_j) - \int_0^1 m_2''(z)f(z)dz\right] + o(h^2) + \frac{11^\top}{n}\boldsymbol{m}_2.$$

(ii) If $\boldsymbol{1}^\top\boldsymbol{m}_1 = 0$, *then* $||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}_1||_2^2 = O(nh^4)$.

(iii) $\boldsymbol{m}_1^\top(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}_2/n = -(h^2/2)\nu_2(K)[\int_0^1 m_1(z)m_2''(z)f(z)dz - \int_0^1 m_1(z)f(z)dz \cdot \int_0^1 m_2''(z)f(z)dz] + o(h^2) + \boldsymbol{m}_1^\top\boldsymbol{1}\boldsymbol{1}^\top\boldsymbol{m}_2/n^2$.

Proof. For convenience, drop the subscript on m_l. Using *(i)* of Lemma 1, we see that B_m is uniformly bounded in the sense that there exists $C^* > 0$ so that $\sup_{h\in(0,1/2]}\sup_{z\in[0,1]}|B_m(K,z,h)| \leq C^*$.

Write

$$(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} = \left[\boldsymbol{I} - (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n)\boldsymbol{S}_h\right]\boldsymbol{m} = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n)(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} + \frac{\boldsymbol{1}\boldsymbol{1}^\top}{n}\boldsymbol{m}.$$

By standard results in smoothing (see, for instance, the proofs of Theorems 1 and 4 in Fan and Gijbels, 1992), $[(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}]_k = -h^2B_m(K, Z_k, h) + o(h^2)$, uniformly in $k = 1, \ldots, n$. So

$$\left[(\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^\top/n)(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}\right]_j = -h^2B_m(K, Z_j, h) + h^2\frac{1}{n}\sum_{k=1}^n B_m(K, Z_k, h) + o(h^2).$$

Since $B_m(K, Z_j, h) = (1/2)\nu_2(K)m''(z)$ for $z \in [h, 1-h]$ and is uniformly bounded, a Riemann integration argument yields

$$\frac{1}{n}\sum_{k=1}^n B_m(K, Z_k, h) = \frac{\nu_2(K)}{2}\cdot\int_0^1 m''(z)f(z)dz + o(1).$$

Thus, uniformly in $j = 1, \ldots, n$,

$$[(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}_2]_j = -h^2\left[B_m(K, Z_j, h) - \frac{1}{2}\nu_2(K)\cdot\int_0^1 m''(z)f(z)dz\right] + o(h^2) + \frac{11^\top}{n}\boldsymbol{m}$$

and statement (i) follows.

The proofs of (ii) and (iii) are straightforward, using (i).

For the rest of the Appendix, we let $\epsilon_1, \ldots, \epsilon_n$ be successive observations from an AR process of finite order R satisfying assumption (A9). We denote the correlation matrix of $\epsilon_1, \ldots, \epsilon_n$ by $\mathbf{\Psi}$. The lemma below shows that the spectral norms of $\mathbf{\Psi}$ and $\mathbf{\Psi}^{-1}$ are bounded. Furthermore, the lemma provides an explicit formula for $\mathbf{\Psi}^{-1}$, taken from Lemma 1 in David and Bastin, 2001.

LEMMA 3.

(i) The spectral norm of $\mathbf{\Psi}$ is bounded as $n \to \infty$.

(ii) The inverse of $\mathbf{\Psi}$ exists and is given by:

$$\mathbf{\Psi}^{-1} = \frac{\sigma_\epsilon^2}{\sigma_u^2} \left[\boldsymbol{\mathcal{U}}^\top \boldsymbol{\mathcal{U}} - \boldsymbol{\mathcal{V}}^\top \boldsymbol{\mathcal{V}} \right],$$

where $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ are $n \times n$ lower triangular matrices defined as

$$\boldsymbol{\mathcal{U}} = \begin{pmatrix} 1 & & & & & \\ -\phi_1 & \cdot & & & & \\ & \cdot\ \cdot & & & & \\ -\phi_R & \cdot & \cdot & & & \\ 0 & \cdot & \cdot & \cdot & & \\ & \cdot\ \cdot & & \cdot & \cdot & \\ 0 & 0 & -\phi_R & & -\phi_1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mathcal{V}} = \begin{pmatrix} 0 & & & & \\ & \cdot & & & \\ 0 & & \cdot & & \\ -\phi_R & \cdot & & \cdot & \\ & \cdot & \cdot & & \cdot \\ -\phi_1 & & -\phi_R & 0 & 0 \end{pmatrix}.$$

(iii) For any n-vector $\boldsymbol{v}$, $[\mathbf{\Psi}^{-1}\boldsymbol{v}]_j$ is a linear combination of at most $1 + 2R + R^2$ elements of $\boldsymbol{v}$, with coefficients depending on $\phi_1, \ldots, \phi_R$ but not on n.

(iv) The spectral norm of $\mathbf{\Psi}^{-1}$ is bounded as $n \to \infty$.

Proof. Recall that $\Psi_{i,j} = \rho(i-j)$. To prove the boundedness of $||\mathbf{\Psi}||_S$ use the symmetry of $\mathbf{\Psi}$ and a well-known result on spectral norms to get:

$$||\mathbf{\Psi}||_S \leq \left[\max_{1 \leq i \leq n} \sum_{j=1}^{n} |\Psi_{i,j}| \right]^2 = \left[\max_{1 \leq i \leq n} \sum_{h=1-i}^{n-i} |\rho(h)| \right]^2 \leq \sum_{-\infty}^{\infty} |\rho(h)| < \infty$$

since there exist constants $\mathcal{C} > 0$ and $s \in (0,1)$ so that $|\rho(h)| \le \mathcal{C}s^{|h|}$. Thus (i) follows. The proof of (iii) follows by direct calculation. The boundedness of $||\boldsymbol{\Psi}^{-1}||_S$ follows easily by using the explicit expression for $\boldsymbol{\Psi}^{-1}$.

In the next two lemmas, we will find it useful to write $\boldsymbol{\mathcal{U}}$ as

$$\boldsymbol{\mathcal{U}} = \boldsymbol{I} - \phi_1 \boldsymbol{\mathcal{U}}_{(1)} - \cdots - \phi_R \boldsymbol{\mathcal{U}}_{(R)} \quad \text{where} \quad \left[\boldsymbol{\mathcal{U}}_{(k)}\right]_{i,j} = I\{i = j + k, 1 \le j \le n - k\}. \tag{24}$$

LEMMA 4. *Let* $\{\boldsymbol{v}_n = (v_{n,1}, \ldots, v_{n,n})^\top, n \ge 1\}$ *and* $\{\boldsymbol{w}_n = (w_{n,1}, \ldots, w_{n,n})^\top, n \ge 1\}$ *be two sequences of n-vectors. Suppose that there exists* $\mathcal{C}$ *such that the following conditions hold.*

(i) $\max_{1 \le j \le n} |v_{n,j}| \le \mathcal{C}$ *for all* n *sufficiently large.*

(ii) $\max_{1 \le j \le n-1} |v_{n,j} - v_{n,j+1}| \le \mathcal{C}/n$ *for all* n *sufficiently large.*

(iii) $\max_{1 \le j \le n} |w_{n,j}| \le \mathcal{C}$ *for all* n *sufficiently large.*

Then, as $n \to \infty$,

$$\frac{1}{n} \boldsymbol{v}_n^\top \boldsymbol{\Psi}^{-1} \boldsymbol{w}_n = \frac{\sigma_\epsilon^2}{\sigma_u^2} \left(1 - \sum_{k=1}^{R} \phi_k\right)^2 \frac{\boldsymbol{v}_n^\top \boldsymbol{w}_n}{n} + O\left(n^{-1}\right).$$

Proof. For notational simplicity, we omit the subscript n. By (ii) of Lemma 3, we have

$$\frac{1}{n} \boldsymbol{v}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{w} = \frac{\sigma_\epsilon^2}{\sigma_u^2} \cdot \frac{(\boldsymbol{\mathcal{U}} \boldsymbol{v})^\top (\boldsymbol{\mathcal{U}} \boldsymbol{w})}{n} - \frac{\sigma_\epsilon^2}{\sigma_u^2} \cdot \frac{(\boldsymbol{\mathcal{V}} \boldsymbol{v})^\top (\boldsymbol{\mathcal{V}} \boldsymbol{w})}{n}.$$

Clearly, the second term is $O(1/n)$, by the sparseness of $\boldsymbol{\mathcal{V}}$ and the boundedness of the components of $\boldsymbol{v}$ and $\boldsymbol{w}$. Using (24) to study the first term, write

$$\begin{aligned}
(\boldsymbol{\mathcal{U}} \boldsymbol{v})^\top (\boldsymbol{\mathcal{U}} \boldsymbol{w}) &= \left[(\boldsymbol{I} - \phi_1 \boldsymbol{\mathcal{U}}_{(1)} - \cdots - \phi_R \boldsymbol{\mathcal{U}}_{(R)}) \boldsymbol{v}\right]^\top \left[\left(\boldsymbol{I} - \phi_1 \boldsymbol{\mathcal{U}}_{(1)} - \cdots - \phi_R \boldsymbol{\mathcal{U}}_{(R)}\right) \boldsymbol{w}\right] \\
&= \boldsymbol{v}^\top \boldsymbol{w} - \sum_{k=1}^{R} \phi_k \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w} - \sum_{l=1}^{R} \phi_l \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w} + \sum_{k,l=1}^{R} \phi_k \phi_l \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w}.
\end{aligned}$$

We now show that $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} + O(1)$, $k = 1, \ldots, R$. One can easily see that

$$\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w} = \sum_{t=1}^{n-k} v_t \cdot w_{t+k} = \boldsymbol{v}^\top \boldsymbol{w} + \sum_{t=1}^{n-k} (v_t - v_{t+k}) \cdot w_{t+k} - \sum_{t=1}^{k} v_t \cdot w_t.$$

The last term is $O(1)$. The second term is also $O(1)$ since $|w_{t+k}|$ is bounded and $|v_t - v_{t+k}| \leq \mathcal{C}/n$. A similar argument shows that $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} + O(1)$ and $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w} = \boldsymbol{v}^\top \boldsymbol{w} + O(1)$, completing the proof of the lemma.

LEMMA 5. *Let (v_i, w_i), $i \geq 1$ be a sequence of independent and identically distributed bivariate random vectors with $E(v_i) = E(w_i) = 0$, the variances of v_i and w_i finite, and $\sigma_{vw} \equiv$ the covariance between v_i and w_i. Let $\boldsymbol{v}_n = (v_1, \ldots, v_n)^\top$ and $\boldsymbol{w}_n = (w_1, \ldots, w_n)^\top$. Then*

$$\frac{1}{n} \boldsymbol{v}_n^\top \boldsymbol{\Psi}^{-1} \boldsymbol{w}_n = \frac{\sigma_\epsilon^2}{\sigma_u^2} \left(1 + \sum_{k=1}^{R} \phi_k^2\right) \sigma_{vw} + o_P(1).$$

Proof. We drop the subscripts n. Using Lemma 3 (ii), the sparseness of $\boldsymbol{\mathcal{V}}$ and (24), write

$$\begin{aligned} \boldsymbol{v}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{w} &= \frac{\sigma_\epsilon^2}{\sigma_u^2} (\boldsymbol{\mathcal{U}} \boldsymbol{v})^\top (\boldsymbol{\mathcal{U}} \boldsymbol{w}) + O_p(1) \\ &= \frac{\sigma_\epsilon^2}{\sigma_u^2} \left[\boldsymbol{v}^\top \boldsymbol{w} - \sum_{k=1}^{R} \phi_k \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w} - \sum_{l=1}^{R} \phi_l \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w} + \sum_{k,l=1}^{R} \phi_k \phi_l \boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w} \right] + O_p(1). \end{aligned}$$

One can easily see that $\boldsymbol{v}^\top \boldsymbol{w}/n$ and $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{\mathcal{U}}_{(k)} \boldsymbol{w}/n = \sum_{k+1}^{n} v_t w_t / n$, $k = 1, \ldots, R$, converge in probability to σ_{vw} as $n \to \infty$. Furthermore, $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{w}/n$, $k = 1, \ldots, R$, and $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w}/n$, $l = 1, \ldots, R$, and $\boldsymbol{v}^\top \boldsymbol{\mathcal{U}}_{(k)}^\top \boldsymbol{\mathcal{U}}_{(l)} \boldsymbol{w}/n$, $k, l = 1, \ldots, R$, $k \neq l$, converge in probability to 0. Combining these results completes the proof.

LEMMA 6. *In addition to assumption (A9), suppose assumptions (A1) and (A3)-(A5) hold. Then, for any $i = 1, \ldots, p$, $||\boldsymbol{S}_h^{c\top} \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_{\cdot i}||_2 = \mathcal{O}_P(h^{-1/2})$ and $||\boldsymbol{S}_h^{c} \boldsymbol{\eta}_{\cdot i}||_2 = \mathcal{O}_P(h^{-1/2})$.*

Proof. By Markov's Theorem,

$$\begin{aligned} ||\boldsymbol{S}_h^{c\top} \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_{\cdot i}||_2^2 &= O_P\left(E(\boldsymbol{\eta}_{\cdot i}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{S}_h^c \boldsymbol{S}_h^{c\top} \boldsymbol{\Psi}^{-1} \boldsymbol{\eta}_{\cdot i})\right) \\ &= \Sigma_{ii}\ O_P\left(trace(\boldsymbol{\Psi}^{-1} \boldsymbol{S}_h^c \boldsymbol{S}_h^{c\top} \boldsymbol{\Psi}^{-1})\right) \\ &= O_P\left(||\boldsymbol{\Psi}^{-1} \boldsymbol{S}_h^c||_F^2\right) = O(||\boldsymbol{S}_h^c||_F^2) \end{aligned}$$

by (iv) of Lemma 3 applied to $\boldsymbol{v}$ equal to a column of $\boldsymbol{S}_h^c$. The first bound in the lemma now follows directly from (iv) of Lemma 1. The second bound of the lemma is proven similarly.

ACKNOWLEDGEMENTS

The authors thank Professor Sverre Vedal, Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, and Dr. Eduardo Hernández-Garduño, Division of Tuberculosis Control, British Columbia Centre for Disease Control, Vancouver, for kindly providing the Mexico City data. We also thank our colleague, Dr. John Petkau, for his input. This research was partially supported by the National Science and Engineering Research Council of Canada.

REFERENCES

G. Aneiros-Pérez & A. Quintela-del-Río (2001a). Asymptotic properties in partial linear models under dependence. *Test*, 10, 333–355.

G. Aneiros-Pérez & A. Quintela-del-Río (2001b). Modified cross-validation in semiparametric regression models with dependent errors. *Communications in Statistics: Theory and Methods*, 30, 289–307.

G. Aneiros-Pérez & A. Quintela-del-Río (2002). Plug-in bandwidth choice in partial linear models with autoregressive errors. *Journal of Statistical Planning and Inference*, 100, 23–48.

R. Bos, S. de Waele & P. M. T. Broersen (2002). Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. *IEEE Transactions on Instrumentation and Measurement*, 51, 1289–1294.

P. J. Brockwell & R. A. Davis (1991). *Time Series: Theory and Methods*. Second Edition. Springer-Verlag, New York.

P. M. T. Broersen (2000). Finite sample criteria for autoregressive order selection. *IEEE Transactions on Signal Processing*, 48, 3550–3558.

A. Buja, T. Hastie & R. Tibshirani (1989). Linear smoothers and additive models (with discussion). *The Annals of Statistics*, 17, 453–555.

B. David & G. Bastin (2001). An estimator of the inverse covariance matrix and its application to ML parameter estimation in dynamical systems. *Automatica*, 156, 99–106.

F. Dominici, A. McDermott, S. L. Zeger & J. M. Samet (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156, 193–203.

F. Dominici, A. McDermott & T. Hastie (2004). Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, 99, 938–948.

J. Fan (1993). Local linear regression smoothers and their minimax efficiency. *The Annals of Statistics*, 21, 196–216.

J. Fan & I. Gijbels (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20, 2008–2036.

M. Francisco-Fernández & J. M. Vilar-Fernández (2001). Local polynomial regression with correlated errors. *Communications in Statistics: Theory and Methods*, 30, 1271–1293.

I. R. Ghement, N. E. Heckman & J. A. Petkau (2006). Seasonal confounding and residual correlation in analyses of health effects of air pollution. *Technical Report #217*, Department of Statistics, University of British Columbia.

W. Härdle, H. Liang & J. Gao (2000). *Partially Linear Models.* Physica–Verlag, Heidelberg.

J. D. Opsomer & D. Ruppert (1999). A root-n consistent estimator for semi-parametric additive modelling. *Journal of Computational and Graphical Statistics*, 8, 715–732.

T. Ramsay, R. Burnett & D. Krewski (2003). The effect of concurvity in generalized additive models linking mortality and ambient air pollution. *Epidemiology*, 14, 18–23.

P. E. Speckman (1988). Regression analysis for partially linear models. *Journal of the Royal Statistical Association*, *Series B*, 50, 413–436.

J. You & G. Chen (2004). Block external bootstrap in partially linear models with nonstationary strong mixing error terms. *The Canadian Journal of Statistics*, 32, 335–346.

J. You, X. Zhou & G. Chen (2005). Jackknifing in partially linear regression models with serially correlated errors. *Journal of Multivariate Analysis*, 92, 386–404.

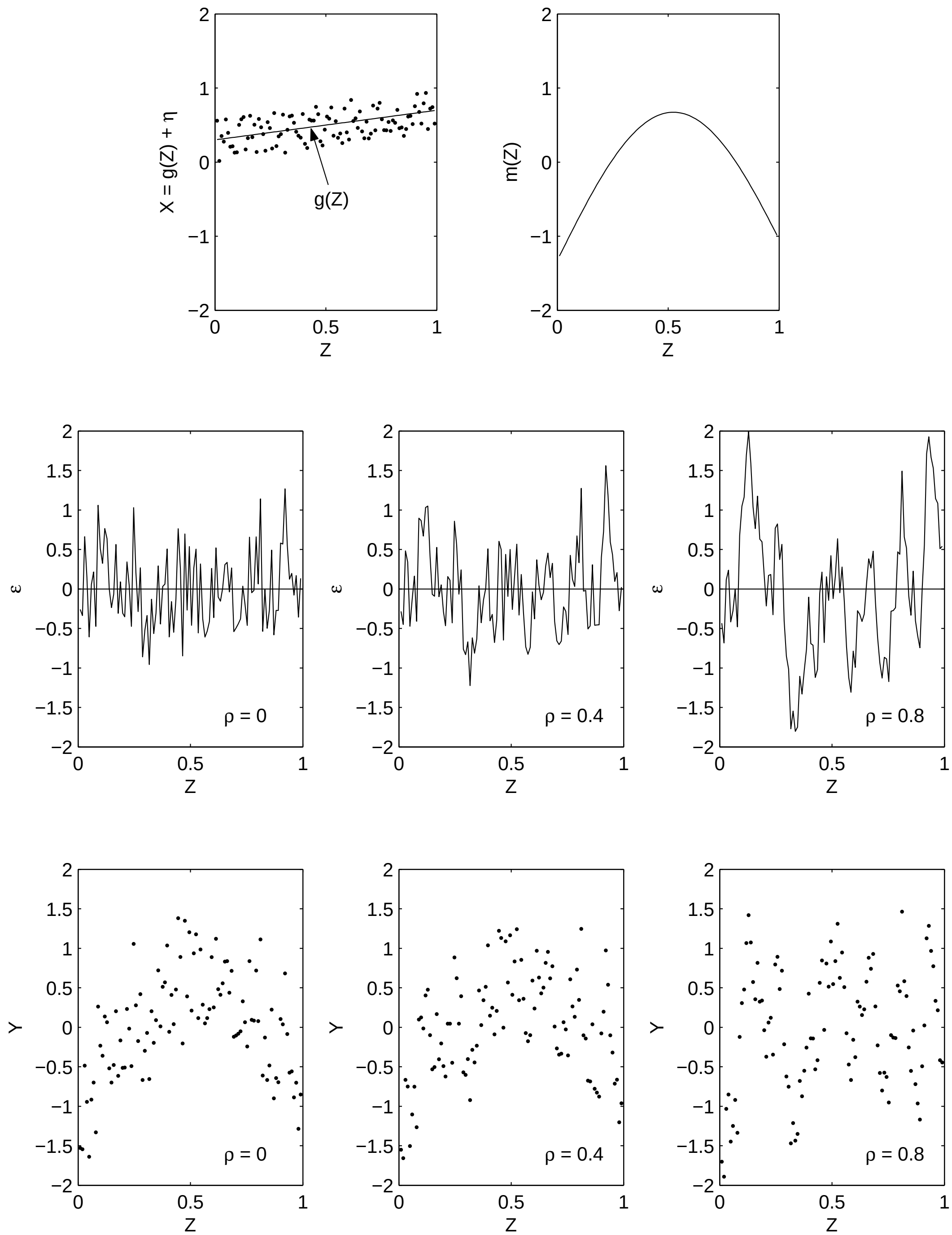

Figure 1: Data simulated from model (1) for $\rho = 0, 0.4, 0.8$ and $m(z) = m_1(z)$. The first row shows plots that do not depend on ρ. The second and third rows each show plots for $\rho = 0, 0.4$ and 0.8.

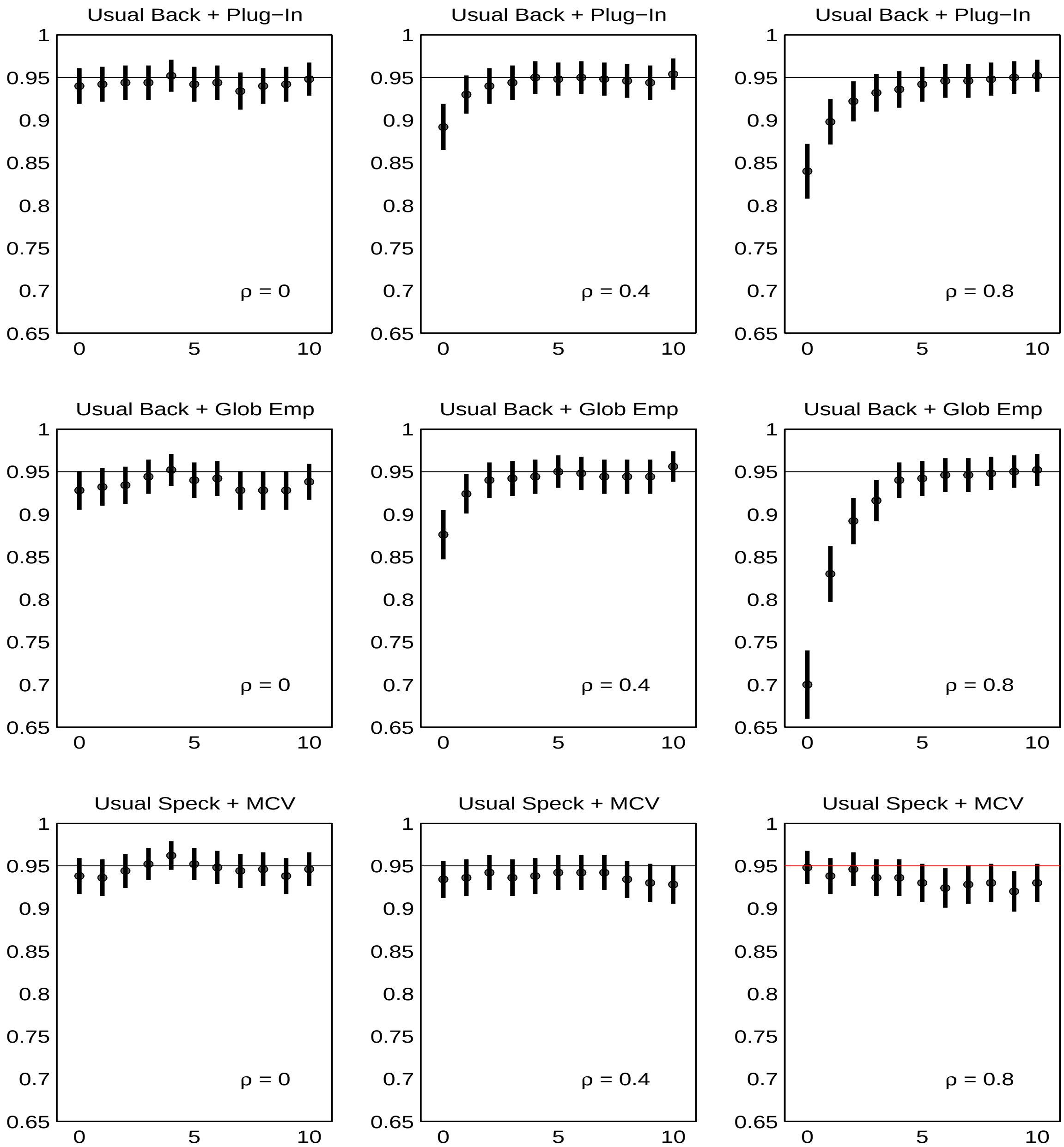

Figure 2: Point estimates (circles) and 95% confidence interval estimates (segments) for the true coverage achieved by three different methods for constructing standard 95% confidence intervals for the linear effect β_1 in model (1) with $p = 1$. The horizontal axis gives the tuning parameter $l = 0,1,...,10$. Estimates were obtained with $m(z) = m_2(z)$ and $\rho = 0$ (left column), $\rho = 0.4$ (middle column) or $\rho = 0.8$ (right column).

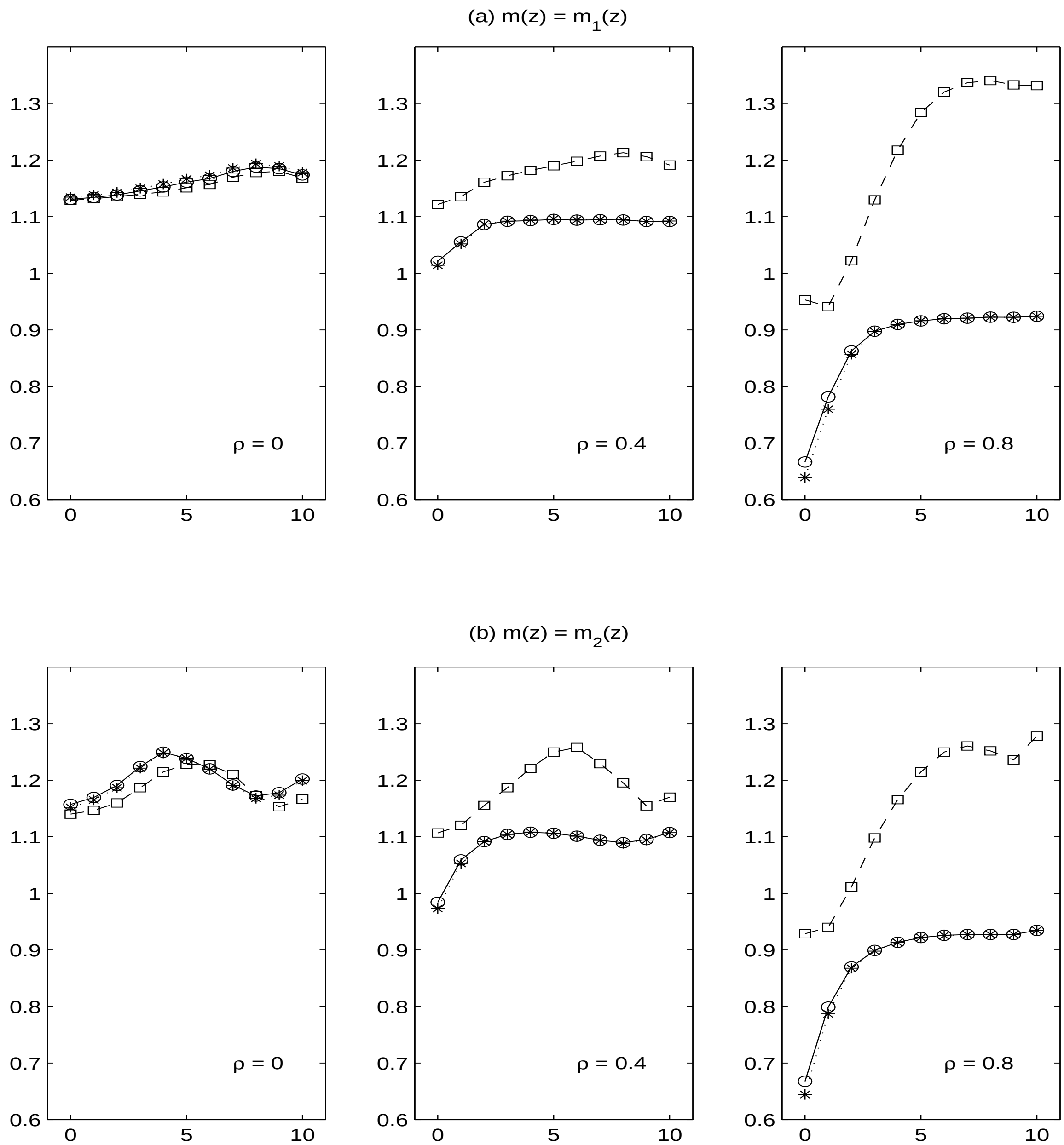

Figure 3: Average lengths of three types of standard 95% confidence intervals for the linear effect $\beta_1 = \boldsymbol{c}^\top \boldsymbol{\beta}$ in model (1) with $p = 1$ as a function of $l = 0, 1, \ldots, 10$. Circles and stars denote average lengths of intervals based on estimators of the form $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, with h chosen by the non-asymptotic plug-in method and the global empirical method, respectively. Squares denote average lengths of intervals from Speckman-based estimators. Lengths were computed with $\rho = 0, 0.4, 0.8$ and (a) $m(z) = m_1(z)$ or (b) $m(z) = m_2(z)$.

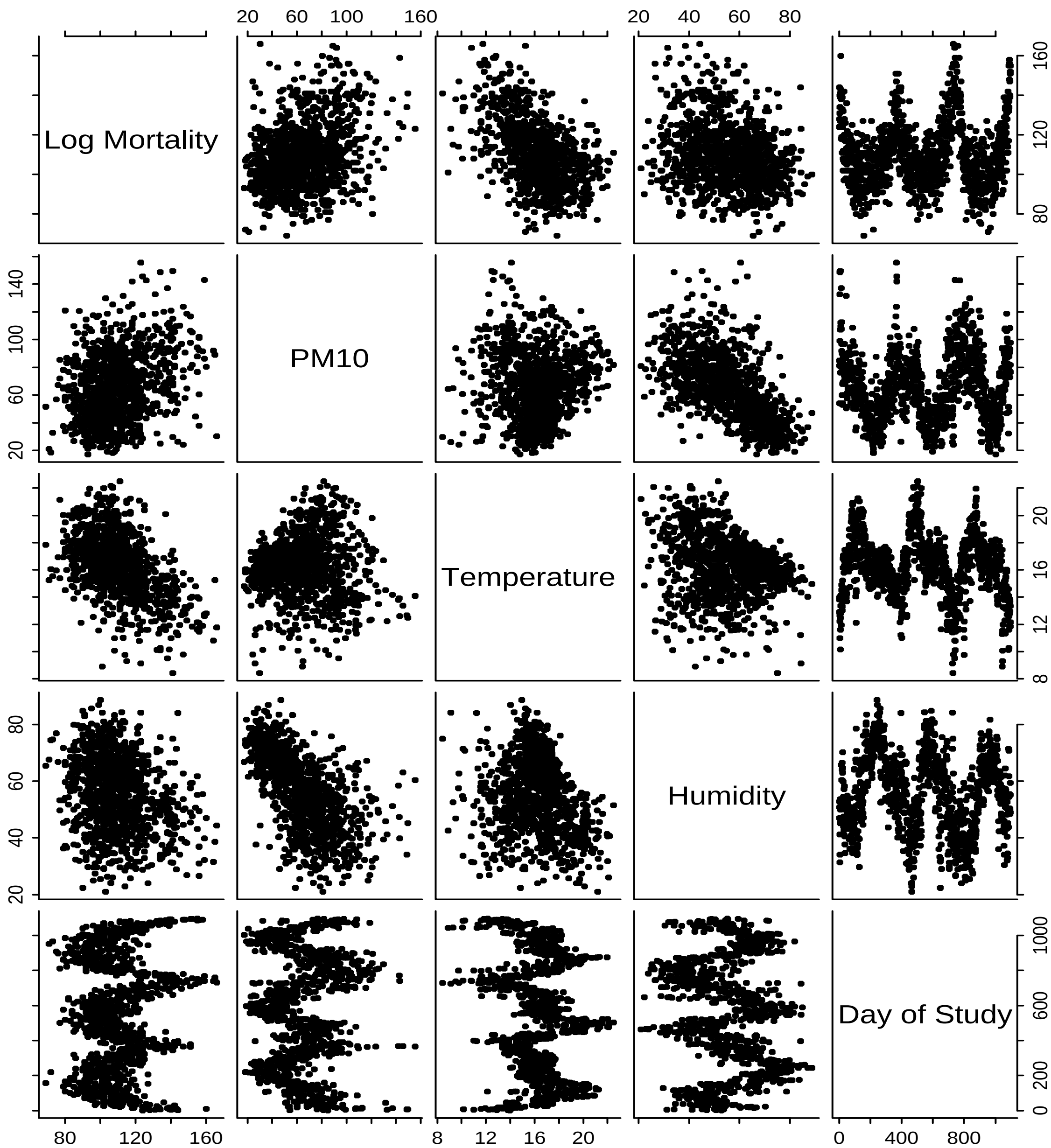

Figure 4: Pairwise scatter plots of log mortality, PM10, temperature, relative humidity and day of study.

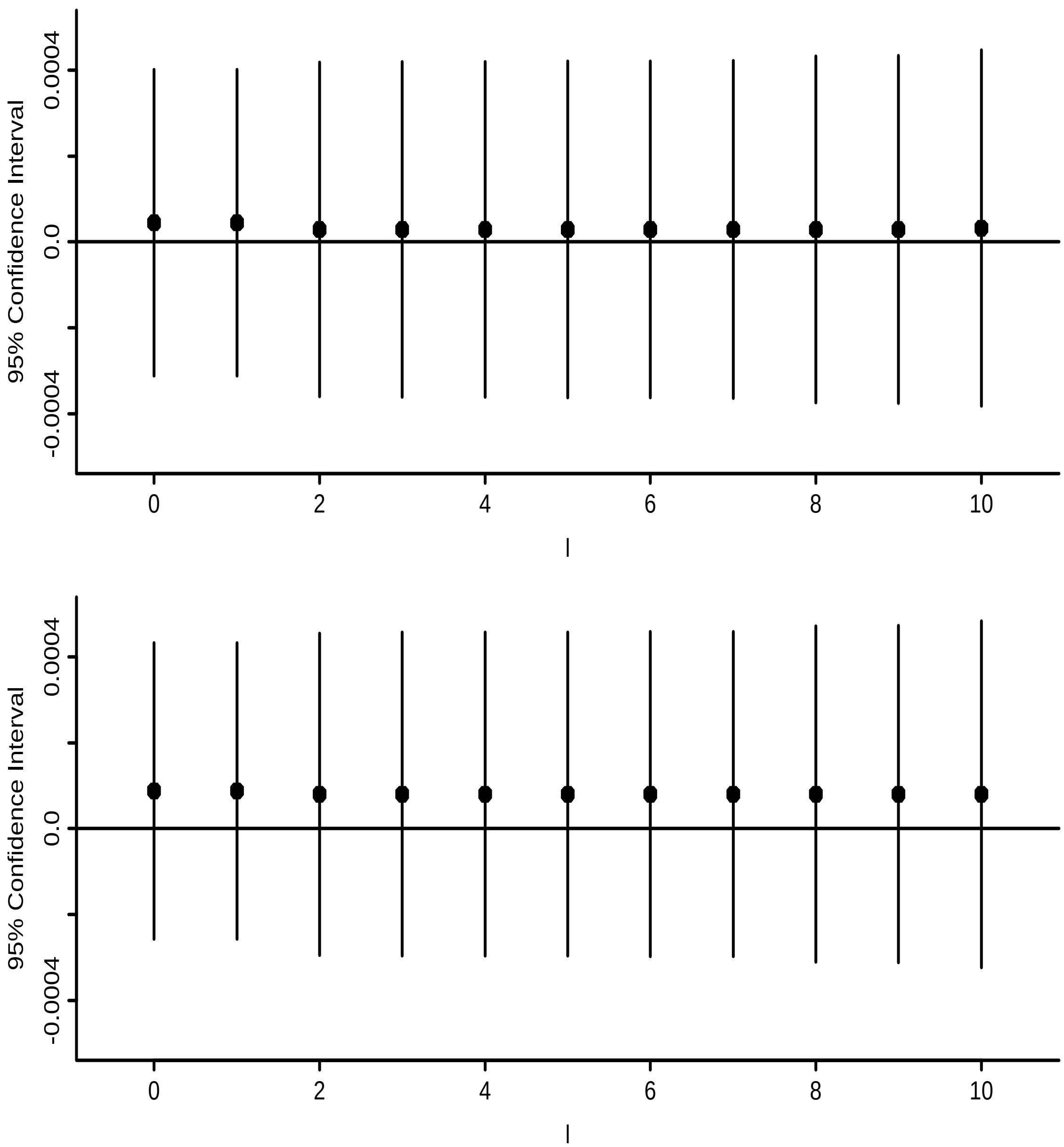

Figure 5: Standard 95% confidence intervals for β_1, the linear PM10 effect in model (15), as a function of the tuning parameter l, where $l = 0, 1, \ldots, 10$. *Top:* The non–asymptotic plug–in method for choosing h. *Bottom:* The global empirical method for choosing h.