Inference in Partially Linear Models with Correlated Errors

by

ISABELLA RODICA GHEMENT

B.Sc., The University of Bucharest, Romania, 1996

M.Sc., The University of Bucharest, Romania, 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Statistics)

The University of British Columbia

August 2005

Abstract

We study the problem of performing statistical inference on the linear effects in partially linear models with correlated errors. To estimate these effects, we introduce usual, modified and estimated modified backfitting estimators, relying on locally linear regression. We obtain explicit expressions for the conditional asymptotic bias and variance of the usual backfitting estimators under the assumption that the model errors follow a mean zero, covariance-stationary process. We derive similar results for the modified backfitting estimators under the more restrictive assumption that the model errors follow a mean zero, stationary autoregressive process of finite order. Our results assume that the width of the smoothing window used in locally linear regression decreases at a specified rate, and the number of data points in this window increases. These results indicate that the squared bias of the considered estimators can dominate their variance in the presence of correlation between the linear and non-linear variables in the model, therefore compromising their $\sqrt{n}$-consistency. We suggest that this problem can be remedied by selecting an appropriate rate of convergence for the smoothing parameter of the estimators. We argue that this rate is slower than the rate that is optimal for estimating the non-linear effect, and as such it 'undersmooths' the estimated non-linear effect. For this reason, data-driven methods devised for accurate estimation of the non-linear effect may fail to yield a satisfactory choice of smoothing for estimating the linear effects. We introduce three data-driven methods for accurate estimation of the linear effects. 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. We use the data-driven choices of smoothing supplied by these methods as a basis for constructing approximate confidence intervals and tests of hypotheses for the linear effects. Our inferential procedures do not account for the uncertainty associated with the fact that the choices of smoothing are data-dependent and the error correlation structure is estimated from the data. We investigate the finite sample properties of our procedures via a simulation study. We also apply these procedures to the analysis of data collected in a time-series air pollution study.

Contents

List of Tables

List of Figures

Acknowledgements

A huge thank you to my thesis supervisor, Dr. Nancy Heckman, for being such an inspirational mentor to me - amazingly generous with her time, ideas, advice and NSERC funding, immensely passionate about research and teaching, wonderfully encouraging and supportive.

A sincere thank you to Dr. John Petkau, Department of Statistics, University of British Columbia and to Professor Sverre Vedal and Dr. Eduardo Hernández-Garduño, formerly of the Respiratory Division, Department of Medicine, Faculty of Medicine, University of British Columbia, for kindly providing me with the Mexico City air pollution data.

A heartfelt thank you to Dr. John Petkau for generously funding me to analyze these data, for providing me with valuable feedback upon reading the manuscript of this thesis and for his excellent advice over the years.

A sincere thank you to Dr. Lang Wu, Dr. Jim Zidek, Dr. Michael Brauer and Dr. Jean Opsomer for their careful reading of the thesis manuscript and valuable comments and suggestions.

Thank you to the Department of Statistics and the University of British Columbia for providing me with funding that enabled me to pursue my degree.

I would like to thank all faculty, staff and graduate students in the Department of Statistics, University of British Columbia, for making my stay there such an enriching experience.

I would like to thank my family in Romania for believing in me and for loving me unconditionally. I would also like to thank my dear friends in Canada and Romania, whose affection and humour helped me stay grounded. Special thanks to Viviane Diaz-Lima, Lisa Kuramoto and Raluca Balan for their unwavering support and for being my friends.

Finally, thank you to Jeffie, my partner in mischief and adventure, for loving me and our family in Romania beyond measure, and for making magical things happen all the time.

ISABELLA RODICA GHEMENT

The University of British Columbia
August 2005

To Jeffie, my ever-loving, ever-caring, ever-there knight in shining armour,

and our loving family in Romania.

Chapter 1

Introduction

Semiparametric regression models combine the ease of interpretation of parametric regression models with the modelling flexibility of nonparametric regression models. They generalize parametric regression models by allowing one or more covariate effects to be non-linear. Just as in nonparametric regression models, the non-linear covariate effects are assumed to change gradually and are captured via smooth, unknown functions whose particular shapes will be revealed by the data.

In this thesis, we are interested in semiparametric regression models for which (i) the response variable is univariate, continuous, (ii) one of the covariate effects is allowed to be smooth, non-linear, and (iii) the remaining covariate effects are assumed to be linear. Given the data $(Y_i, \boldsymbol{X}_i^T, Z_i)$, $i = 1, \ldots, n$, such models can be specified as:

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + m(Z_i) + \epsilon_i, \; i = 1, \ldots, n, \tag{1.1}$$

where $\boldsymbol{\beta}$ is a vector of linear effects, m is a smooth, non-linear effect and the ϵ_i's are unobservable random errors with zero mean. Model (1.1) is typically referred to as a partially linear regression model.

In many applications, the smooth, non-linear effect m in model (1.1) is not of interest in itself but is included in the model because of its potential for confounding the linear effects $\boldsymbol{\beta}$, which are of main interest. The nature of this confounding is often too complex to specify parametrically. A non-parametric specification of this confounding effect is therefore preferred to avoid modelling biases. The practical choice of the degree of smoothness of the non-linear confounder effect is a delicate issue in these types of applications. This choice should yield accurate point estimators of the linear effects of interest. The choice may be highly sensitive to the correlations between the linear and non-linear variables in the model.

The potential correlation amongst model errors is a qualitatively different source of confounding on the linear effects of interest in a partially linear model. In practice, we need to decide carefully whether we should account for this correlation when assessing the significance and magnitude of the linear effects of interest. If one decides to ignore the error correlation, one should try to understand the impact of this decision on the validity of the ensuing inferences.

The issues of error correlation, non-linear confounding, and correlation between the linear and non-linear terms in a partially linear regression model are intimately connected. Their interplay needs to be judiciously considered when selecting the degree of smoothness of the estimated non-linear effect. Even when this selection yields accurate estimators of the linear effects of interest in the model, one needs to assess whether it also yields valid confidence intervals and testing procedures for assessing the magnitude and significance of these effects.

1.1 Literature Review

In this section, we provide a survey of some of the most important results in the literature of partially linear regression models of the form (1.1). We treat separately the case when the model errors, $\epsilon_i, i = 1, \ldots, n$, are uncorrelated and when they are correlated.

Note that, in (1.1), we observe only one sequence $Y_1, \ldots, Y_n$. In classical longitudinal studies we would observe multiple sequences. Even though in this thesis we are not interested in partially linear models for analyzing data collected in longitudinal studies, we do mention some results which are significant in the literature of these models.

1.1.1 Partially Linear Models with Uncorrelated Errors

The partially linear regression model (1.1) has been investigated extensively under the assumption of independent, identically distributed errors. In this section, we provide a brief overview of some of the most relevant results concerning inferences on $\boldsymbol{\beta}$, the parametric component of the model, that are available in the literature. These results have a common theme: seeing if $\boldsymbol{\beta}$ is estimated at the 'usual' parametric rate of $1/n$ - the rate that would be achieved if m were known. As Robinson (1988) points out, consistent estimators of $\boldsymbol{\beta}$ that do not have the 'usual' parametric rate of convergence have zero efficiency relative to estimators that have this rate.

Engle et al. (1983) and Wahba (1984) proposed estimating $\boldsymbol{\beta}$ and m simultaneously by minimizing a penalized least squares criterion with penalty based on the s^{th} derivative of m, with $s \geq 2$. The performance of the penalized least squares estimator of $\boldsymbol{\beta}$ depends on the correlation between the linear and non-linear variables in the model. Heckman (1986) established the $\sqrt{n}$-consistency of this estimator assuming that the linear and non-linear variables are uncorrelated. Rice (1986) showed that, if the linear and non-linear variables are correlated, the estimator becomes $\sqrt{n}$-inconsistent, unless one 'undersmooths' the estimated m. 'Undersmoothing' refers to the phenomenon of estimating m at a slower rate than the 'usual' nonparametric rate of $n^{-4/5}$ - the rate that would be achieved if $\boldsymbol{\beta}$ were known. Rice showed that if one didn't 'undersmooth', the squared bias of the estimated linear effects would dominate their variance. The author remarked that this would have disastrous consequences on the inferences carried out on the linear effects. For instance, conventional confidence intervals for these effects would be misleading. Rice called into question the utility of traditional methods such as cross-validation for choosing the degree of smoothness of the estimated non-linear effect when $\sqrt{n}$-consistency of the estimated linear effects is desired, and rightly so. These methods are devised for 'smoothing', not 'undersmoothing', the estimated non-linear effect.

Green, Jennison and Seheult (1985) proposed estimating $\boldsymbol{\beta}$ and m by minimizing a penalized least squares criterion with penalty based on a discretization of the second derivative of m. They termed their estimation method least squares smoothing and showed that it yields estimators that solve a system of backfitting equations. These equations combine a smoothing step for estimating m, carried out using a discretized version of smoothing splines, with a least squares regression step for estimating $\boldsymbol{\beta}$. Green, Jennison and Seheult generalized their least squares smoothing estimators by allowing the smoothing step in the backfitting equations to be carried out using any smoothing method. These generalized least squares smoothing estimators are referred to in the literature as the Green, Jennison and Seheult estimators. Speckman (1988) derived the asymptotic bias and variance of the Green, Jennison and Seheult estimator of $\boldsymbol{\beta}$, using locally constant regression with general kernel weights in the smoothing step. Speckman's findings paralleled those of Rice: in the presence of correlation between the linear and non-linear variables in the model, the Green, Jennison and Seheult estimator of $\boldsymbol{\beta}$ is $\sqrt{n}$-consistent only if one 'undersmooths' the estimated m. Speckman provided a heuristic argument for why the generalized cross-validation method cannot be used to choose the degree of smoothness of the estimated m in practice when $\sqrt{n}$-consistency of the Green, Jennison and Seheult estimator of $\boldsymbol{\beta}$ is desired.

Neither Rice nor Speckman proposed methods for 'undersmoothing' the estimated m. However, Speckman (1988) introduced a partial-residual flavoured estimator of $\boldsymbol{\beta}$ that does not require 'undersmoothing'. He argued that traditional methods such as generalized cross-validation could be used to select the degree of smoothness of the estimated m. Speckman did not address the important issue of whether such data-driven methods would produce amounts of smoothing that yield $\sqrt{n}$-consistent estimators of the linear effects of interest. Sy (1999) established that data-driven methods such as cross-validation and generalized cross-validation do indeed yield $\sqrt{n}$-consistent estimators of these effects, thus paving the way for carrying out valid inferences on these effects, at least for large sample sizes.

Opsomer and Ruppert (1999) proposed estimating $\boldsymbol{\beta}$ and m via the Green, Jennison and Seheult estimators, using locally linear regression with general kernel weights in the smoothing step. They showed that, unless one 'undersmooths' the estimated m, their estimator of $\boldsymbol{\beta}$ may not achieve $\sqrt{n}$-consistency. They then suggest how to use the data to choose the appropriate degree of smoothness for accurate estimation of $\boldsymbol{c}^T\boldsymbol{\beta}$, with $\boldsymbol{c}$ known. Opsomer and Ruppert's approach for choosing the right degree of smoothness, referred to as the Empirical Bias Bandwidth Selection (EBBS) method, will be discussed in more detail in Chapter 6. The authors conjectured that EBBS would produce a $\sqrt{n}$-consistent estimator of $\boldsymbol{c}^T\boldsymbol{\beta}$.

1.1.2 Partially Linear Models with Correlated Errors

The independence assumption for the errors associated with a partially linear regression model is not always appropriate in applications. For instance, when the data have been collected sequentially over time, it is likely that present response values will be correlated with past response values. Even in the presence of error correlation, it is desirable to obtain $\sqrt{n}$-consistent estimators for the linear effects in the model.

Engle et al. (1986) were amongst the first authors to consider a partially linear regression model with AR(1) errors. They noted that the correct error correlation structure can be used to transform this model into a model with serially uncorrelated errors, by quasi-differencing all of the data. They proposed estimating the linear effects $\boldsymbol{\beta}$ and the non-linear effect m in the original model by applying the penalized least squares method proposed by Engle et al. (1983) and Wahba (1984) to the quasi-differenced data. Engle et al. (1986) prove that their estimator of $\boldsymbol{\beta}$ is consistent when one estimates m at the 'usual' nonparametric rate of $n^{-4/5}$, but do not show it is $\sqrt{n}-$ consistent. They recommend choosing both the 'right' degree of smoothness of the estimated m and the autoregressive parameter by minimizing a generalized cross-validation criterion constructed from the quasi-differenced data. This data-driven choice of smoothing may not however yield an accurate estimator of $\boldsymbol{\beta}$, as it is geared at accurate estimation of m.

Schick (1996, 1999) considered partially linear regression models with AR(1) errors and ARMA(p,q) errors, respectively, where $p, q \geq 1$. He characterized and constructed efficient estimators for the parametric component $\boldsymbol{\beta}$ of these models, assuming appropriate theoretical choice of degree of smoothness for the estimated m. He did not however indicate how one might make this choice in practice.

Several authors investigated partially linear models with α-mixing errors. Before reviewing their respective contributions, we provide a definition for the α-mixing concept. For reference, see Ibragimov and Linnik (1971).

Definition 1.1.1 *A sequence of random variables $\{\epsilon_t, t = 0, \pm 1, \cdots\}$ is said to be α-mixing if*

$$\alpha(k) \equiv \underset{n}{sup} \underset{\{A \in \mathcal{F}_{-\infty}^{n},\, B \in \mathcal{F}_{n+k}^{\infty}\}}{sup} |P(A \cap B) - P(A)P(B)| \rightarrow 0 \tag{1.2}$$

as $k \rightarrow \infty$, where $\mathcal{F}_{-\infty}^{n}$ and $\mathcal{F}_{n+k}^{\infty}$ are two σ-fields generated by $\{\epsilon_t, t \leq n\}$ and $\{\epsilon_t, t \geq n + k\}$, respectively.

The mixing coefficient $\alpha(k)$ in (1.2) measures the amount of dependence between events involving variables separated by at least k lags. Note that for stationary sequences the supremum over n in (1.2) goes away.

Aneiros Pérez and Quintela del Río (2001a) considered a partially linear model with α-mixing, stationary errors. They proposed estimating $\boldsymbol{\beta}$ and m via modifications of the Speckman estimators. Their modifications account for the error correlation structure, assumed to be fully known. The smoothing step involved in estimating $\boldsymbol{\beta}$ and m is based on locally constant regression with Gasser-Müller weights (Gasser and Müller, 1984), adjusted for boundary effects. The authors derived the order of the conditional asymptotic bias and variance of the modified Speckman estimator of $\boldsymbol{\beta}$. They found that the conditional asymptotic bias of their estimator of $\boldsymbol{\beta}$ is negligible with respect to its conditional asymptotic variance, shown to have the 'usual' parametric rate of convergence of $1/n$. They concluded they do not need to 'undersmooth' their estimator for m in order to obtain a $\sqrt{n}$-consistent estimator for $\boldsymbol{\beta}$. The fact that the modified Speckman estimator of $\boldsymbol{\beta}$ does not require 'undersmoothing' in the presence of error correlation is not surprising. The estimator inherits this property from the usual Speckman estimator. Aneiros Pérez and Quintela del Río (2001b) proposed a data-driven modified cross-validation method for choosing the degree of smoothness required for accurate estimation of the regression function $r(\boldsymbol{X}_i, Z_i) = \boldsymbol{X}_i^T\boldsymbol{\beta} + m(Z_i)$ via modified Speckman estimators. It is not clear whether such a method would be suitable for accurate estimation of $\boldsymbol{\beta}$ itself. To address the problem of choosing the degree of smoothness for accurate estimation of $\boldsymbol{\beta}$ via the modified Speckman estimator, Aneiros Pérez and Quintela del Río (2002) developed an asymptotic plug-in method. Their method relies on the more restrictive assumption that the model errors are realizations of an autoregressive process of finite, known order.

You and Chen (2004) considered a partially linear model with α-mixing, possibly nonstationary errors. They estimated $\boldsymbol{\beta}$ and m using the usual Speckman estimators, which do not account for error correlation. They then applied a block external bootstrap approach to approximate the distribution of the usual Speckman estimator of $\boldsymbol{\beta}$ and provide a consistent estimator of its covariance matrix. Using this information, they constructed a large-sample confidence interval procedure for estimating $\boldsymbol{\beta}$. Based on a simulation study, the authors note that the block size seems to have a strong influence on the finite-sample performance of their procedure. However, they do not indicate how one might choose the block size in practice. In the simulation study, the smoothing parameter of the usual Speckman estimator of $\boldsymbol{\beta}$ was selected via cross-validation, modified for correlated errors. This method is appropriate for accurate estimation of m but may not be suitable for accurate estimation of $\boldsymbol{\beta}$.

You, Zhou and Chen (2005) considered a partially linear model with errors assumed to follow a moving average process of infinite order. They proposed a jackknife estimator for $\boldsymbol{\beta}$, which they obtained from a usual Speckman estimator. They showed their estimator to be asymptotically equivalent to the usual Speckman estimator, and proposed a method for estimating its asymptotic variance. They also constructed confidence intervals and tests of hypotheses for $\boldsymbol{\beta}$ based on the jackknife estimator and its estimated variance. In their simulation study, these authors find that confidence interval estimation based on their jackknife estimator has better finite-sample coverage properties than that based on the usual Speckman estimator, even though the latter uses the information on the error structure, while the former does not. In this study, the smoothing was performed with different nearest neighbor smoothing parameter values and the results were shown to be insensitive to the choice of this parameter. This may not always be the case for contexts that are different from that considered by these authors.

As we already mentioned, partially linear regression models with correlated errors can be used for analyzing longitudinal data, that is, data obtained by measuring each of several study units on multiple occasions over time. Longitudinal data are naturally correlated, as the measurements taken on the same study unit are correlated. In order to estimate the linear effects $\boldsymbol{\beta}$ and the non-linear effect m in such models, Moyeed and Diggle (1994) modified the Green, Jennison and Seheult and the Speckman estimators to account for the longitudinal data structure and for the error correlation, assumed to be known. Their smoothing step used local constant Nadaraya-Watson weights (Nadaraya, 1964 and Watson, 1964). They derived the order of the conditional asymptotic bias and variance of their estimators of $\boldsymbol{\beta}$, obtaining asymptotic constants only for the variance of these estimators. Their results are valid under the assumption that the number of study units goes to infinity and the number of occasions on which each study unit is being measured is kept constant. Note that Moyeed and Diggle did not treat m as a nuisance. To choose the degree of smoothness of the estimated m, these authors used a leave-one-subject-out cross-validation method. This method is geared towards accurate estimation of m and may not be appropriate for accurate estimation of $\boldsymbol{\beta}$.

None of the authors considered in this section looked simultaneously at how to choose the right degree of smoothing for accurate estimation of the linear effects and how to construct valid standard errors for the estimated linear effects. To do both requires accounting for the correlation structure of the model errors.

1.2 Thesis Objectives

Throughout this thesis, we will consider only partially linear models of the form (1.1) in which the non-linear effect m is treated as a nuisance. In contrast to the 'usual' view in regression models, we will think of the linear covariates as being random but consider the Z_i's to be fixed. The reason for this is that we are mainly interested in applications for which the Z_i's are consecutive time points (e.g. days, weeks, years). The results in this thesis can be easily modified to account for the case when the Z_i's are random instead of fixed. However, some expressions need to be re-defined to account for the randomness of the Z_i's. For instance, see the end of Sections 4.1 and 4.2. In this thesis, we will allow the linear covariates to be mutually correlated and assume they are related to the non-linear covariates via a non-parametric regression relationship. Most importantly, we will assume that the model errors are serially correlated. Within this framework, we will concentrate on developing formal methods for carrying out valid inferences on those linear effects in the model which are of main interest. This entails the following:

1. defining sensible estimators for the linear effects in the model, as well as for the nuisance non-linear effect;

2. deriving the asymptotic bias and variance of the proposed estimators of the linear effects;

3. developing methods for choosing the right degree of smoothness of the estimated non-linear effect in order to accurately estimate the linear effects of interest;

4. developing methods for estimating the correlation structure of the model errors for inference and smoothing;

5. developing methods for assessing the magnitude and statistical significance of the linear effects of interest;

6. investigating the performance of the proposed inferential methods via Monte Carlo simulation studies;

7. using the inferential methods developed in this thesis to answer specific questions related to the impact of air pollution on mortality in Mexico City during 1994-1996, after adjusting for weather patterns and temporal trends.

We conclude this chapter with an overview of the thesis which indicates where and how the above objectives are addressed.

In Chapter 2, we provide a formal definition of the partially linear model with correlated errors of interest in this thesis. We also introduce the notation and assumptions required for establishing the theoretical results in subsequent chapters.

In Chapter 3, we define the following types of estimators for $\boldsymbol{\beta}$ and m: (i) local linear backfitting estimators, (ii) modified local linear backfitting estimators, and (iii) estimated modified local linear backfitting estimators.

In Chapter 4, we derive asymptotic approximations for the exact conditional bias and variance of the local linear backfitting estimator of $\boldsymbol{\beta}$. Based on these results we conclude that, in general, the local linear backfitting estimator of $\boldsymbol{\beta}$ is not $\sqrt{n}$-consistent. We argue that the estimator can achieve $\sqrt{n}$-consistency provided we 'undersmooth' the corresponding local linear backfitting estimator of m.

In Chapter 5, we replicate the results in Chapter 4 for the modified local linear backfitting estimator of $\boldsymbol{\beta}$. We also provide sufficient conditions under which the estimated modified local linear backfitting estimator of $\boldsymbol{\beta}$ is asymptotically 'close' to its modified counterpart.

In Chapter 6, we develop three data-driven methods for choosing the degree of smoothness of the backfitting estimators of m defined in this thesis in order to accurately estimate $\boldsymbol{\beta}$. Two of these methods are modifications of the Empirical Bias Bandwidth Selection (EBBS) method of Opsomer and Ruppert (1999). The third method is a non-asymptotic plug-in method. All methods account for error correlation. We suspect that these methods 'undersmooth' the estimated m because they attempt to estimate the amount of smoothing that is optimal for estimating $\boldsymbol{\beta}$, not for estimating m. Our theoretical results suggest that, in general, the optimal amount of smoothing for estimating $\boldsymbol{\beta}$ is smaller than the optimal amount of smoothing for estimating m. In Chapter 6, we also introduce methods for estimating the correlation structure of the model errors needed to choose the amount of smoothing of the backfitting estimators of $\boldsymbol{\beta}$ and to carry out inferences on $\boldsymbol{\beta}$. These methods rely on a modified cross-validation criterion similar to that proposed by Aneiros Pérez and Quintela del Río (2001b).

In Chapter 7, we develop three kinds of confidence intervals and tests of hypotheses for assessing the magnitude and significance of a linear combination $\boldsymbol{c}^T\boldsymbol{\beta}$ of the linear effects in the model: standard, bias-adjusted and standard-error adjusted. To our knowledge, adjusting for bias in confidence intervals and tests of hypotheses has not been attempted in the literature of partially linear models.

In Chapter 8, we report the results of a Monte Carlo simulation study. In this study, we investigated the finite sample properties of the usual and estimated modified local linear backfitting estimators of $\boldsymbol{c}^T\boldsymbol{\beta}$ against those of the usual Speckman estimator. We chose the smoothing parameter of the backfitting estimators using the data-driven methods developed in Chapter 6. By contrast, we chose the smoothing parameter of the usual Speckman estimator using cross-validation, modified for correlated errors and for boundary effects. The main goals of our simulation study were (1) to compare the expected log mean squared error of the estimators and (2) to compare the performance of the confidence intervals built from these estimators and their associated standard errors. Our study suggested that quality of the inferences based on the usual local linear backfitting estimator was superior, and that this estimator should be computed with one of our modifications of EBBS or a non-asymptotic plug-in choice of smoothing. Even though the quality of the inferences based on the usual Speckman estimator was reasonable for most simulation settings, it was not as good as that of the inferences based on the usual local linear backfitting estimator. The quality of the inferences based on the estimated modified local linear estimator was poor for many simulation settings.

In Chapter 9, we use the inferential methods developed in this thesis to assess whether the pollutant PM10 had a significant short-term effect on log mortality in Mexico City during 1994-1996, after adjusting for temporal trends and weather patterns. Our data analysis suggests that there is no conclusive proof that PM10 had a significant short-term effect on log mortality.

In Chapter 10, we summarize the main contributions of this thesis and suggest possible extensions to our work.

Chapter 2

A Partially Linear Model with Correlated Errors

In Section 2.1 of this chapter, we provide a formal definition of the partially linear model of interest in this thesis. In Section 2.2, we introduce assumptions that we use to study the asymptotic behavior of our proposed estimators. In Section 2.3, we introduce some useful notation. In Section 2.4, we give several linear algebra definitions and results which will be utilized throughout this thesis. The chapter concludes with an Appendix which contains a useful theoretical result.

2.1 The Model

Given the data (Y_i, X_{ij}, Z_i), $i = 1, \ldots, n, j = 1, \ldots, p$, the specific form of the partially linear model considered in this thesis is:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{m} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ is the vector of responses, $\boldsymbol{X}$ is the design matrix for the parametric part of the model (to be defined shortly), $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the vector of unknown linear effects, $\boldsymbol{m} = (m(Z_1), \ldots, m(Z_n))^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ is the vector of model errors. Here,

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}, \tag{2.2}$$

where $X_{i1}, \ldots, X_{ip}$ are measurements on p variables $X_1, \ldots, X_p$, the Z_i's are fixed design points on $[0, 1]$ following a design density $f(\cdot)$ (see condition (A3) in Section 2.2 for the exact definition), and $m(\cdot)$ is a real-valued, unknown, smooth function defined on $[0, 1]$. Note that, unless we impose a restriction on $m(\cdot)$, model (2.1) is unidentifiable due to the presence of the intercept β_0 in the model. For instance, $\beta_0 + m(\cdot) = 0 + (m(\cdot) + \beta_0)$. To ensure identifiability, we assume that $m(\cdot)$ satisfies the integral restriction:

$$\int_0^1 m(z)f(z)dz = 0. \tag{2.3}$$

In practice, we replace (2.3) by the summation restriction:

$$\mathbf{1}^T\boldsymbol{m} = 0, \tag{2.4}$$

where the symbol $\mathbf{1}$ denotes an $n \times 1$ vector of 1's. One could think of the smooth function $m(\cdot)$ as being a transformation of the fixed design points $Z_i, i = 1, \ldots, n$, that ensures that the partially linear model (2.1) is an adequate description of the variability in the Y_i's. Alternatively, one could think of the function $m(\cdot)$ as representing the confounding effect of a random variable having density $f(\cdot)$ on the linear effects $\beta_1, \ldots, \beta_p$.

We assume that the errors ϵ_i in model (2.1) are such that $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma_\epsilon^2$ and $Corr(\epsilon_i, \epsilon_j) = \Psi_{i,j}$ for $i \neq j$, where $\sigma_\epsilon > 0$ and $\boldsymbol{\Psi} = (\Psi_{i,j})$ is the $n \times n$ error correlation matrix. Note that $\boldsymbol{\Psi}$ is not necessarily equal to the $n \times n$ identity matrix $\boldsymbol{I}$. In practice, both the error variance σ_ϵ^2 and the error correlation matrix $\boldsymbol{\Psi}$ are typically unknown and need to be estimated from the data.

An alternative formulation for the partially linear model (2.1) can be obtained by removing the constraint (2.3), setting $\boldsymbol{m}^* \equiv \beta_0 \mathbf{1} + \boldsymbol{m}$ and re-writing the model as:

$$\boldsymbol{Y} = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{m}^* + \boldsymbol{\epsilon}, \tag{2.5}$$

where $\boldsymbol{X}^*$ is an $n \times p$ matrix defined as:

$$\boldsymbol{X}^* = \begin{pmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{pmatrix} \tag{2.6}$$

and $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_p)^T$. The model formulation in (2.5) is frequently encountered in the partially linear model literature and does not require that we impose any identifiability conditions on the function $m^*(z) \equiv \beta_0 + m(z), z \in [0,1]$. Indeed, the absence of an intercept in model (2.5) ensures that $m^*(\cdot)$ is identifiable. In this thesis, however, we prefer to use the formulation in (2.1), as it makes it easier to understand that model (2.1) is a generalization of a linear regression model and a particular case of an additive model, which typically do contain an intercept.

2.2 Assumptions

The asymptotic results derived in Chapters 4 and 5 allow the linear variables in model (2.1) to be correlated with the non-linear variable via the following condition.

(A0) *The covariate values X_{ij} and the non-random design points Z_i are related via the nonparametric regression model:*

$$X_{ij} = g_j(Z_i) + \eta_{ij}, \; i = 1, \ldots, n, \; j = 1, \ldots, p, \tag{2.7}$$

where

(i) *the $g_j(\cdot)$'s are smooth, unknown functions having three continuous derivatives;*

(ii) *the* $(\eta_{i1}, \ldots, \eta_{ip})^T, i = 1, \ldots, n$, *are independent, identically distributed unobserved random vectors with mean zero and variance-covariance matrix* $\mathbf{\Sigma} = (\Sigma_{ij})$.

We impose two different sets of assumptions on the errors associated with model (2.1) for studying the asymptotic behaviour of two different estimators of $\boldsymbol{\beta}$. In Section 3.1.1 of Chapter 3 we define the so-called local linear backfitting estimator of $\boldsymbol{\beta}$. The definition of this estimator does not account for the correlation structure of the model errors. In Chapter 4, we study the asymptotic behaviour of this estimator under the assumption that the model errors satisfy the following condition.

(A1) (i) *The model errors* $\epsilon_i, i = 1, \ldots, n$, *represent* n *consecutive realizations from a general covariance-stationary process* $\{\epsilon_t\}$, $t = 0, \pm 1, \pm 2, \ldots$ *having mean* 0, *finite, non-zero variance* σ_ϵ^2 *and correlation coefficients:*

$$\rho_k = \frac{E(\epsilon_t \epsilon_{t-k})}{\sigma_\epsilon^2} = \frac{E(\epsilon_s \epsilon_{s+k})}{\sigma_\epsilon^2}, \ k = 1, 2, 3, \ldots, \tag{2.8}$$

where $t, s = 0, \pm 1, \pm 2, \ldots$.

(ii) *The error correlation matrix* $\mathbf{\Psi}$ *is assumed to be symmetric, positive-definite and to have a bounded spectral norm, that is* $||\mathbf{\Psi}||_S = \mathcal{O}(1)$ *as* $n \to \infty$. *(For a definition of the spectral norm of a matrix see Section 2.4.)*

(iii) *Let* $(\eta_{i1}, \ldots, \eta_{ip})^T, i = 1, \ldots, n$, *be as in* (A0)-(ii) . *Then there exists a* $(p+1) \times (p+1)$ *matrix* $\mathbf{\Phi}^{(0)}$ *such that the error correlation matrix* $\mathbf{\Psi}$ *satisfies:*

$$\frac{1}{n+1} \boldsymbol{\eta}^T \mathbf{\Psi} \boldsymbol{\eta} = \mathbf{\Phi}^{(0)} + o_P(1) \tag{2.9}$$

as $n \to \infty$, *where*

$$\boldsymbol{\eta} = \begin{pmatrix} 0 & \eta_{11} & \cdots & \eta_{1p} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \eta_{n1} & \cdots & \eta_{np} \end{pmatrix}. \tag{2.10}$$

(iv) ϵ_j *is independent of* $(\eta_{i1}, \ldots, \eta_{ip})^T$ *for any* $i, j = 1, \ldots, n$.

In Section 3.1.2 of Chapter 3 we define the so-called modified local linear backfitting estimator of the vector of linear effects $\boldsymbol{\beta}$ in model (2.1). The definition of this estimator assumes full knowledge of the correlation matrix of the model errors. In Chapter 5, we study the asymptotic behaviour of this estimator under the assumption that the model errors satisfy the following condition:

(A2) (i) *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:*

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \cdots + \phi_R\epsilon_{t-R} + u_t, \;\; t = 0, \pm 1, \pm 2, \ldots \tag{2.11}$$

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.

(ii) *ϵ_j is independent of $(\eta_{i1}, \ldots, \eta_{ip})^T$ for any $i, j = 1, \ldots, n$, where $(\eta_{i1}, \ldots, \eta_{ip})^T, i = 1, \ldots, n$, are as in* (A0)-(ii).

According to Comments 2.2.1 - 2.2.3 below, if the errors ϵ_i satisfy condition (A2), they also satisfy condition (A1).

Comment 2.2.1 If the errors $\epsilon_i, i = 1, \ldots, n$, satisfy condition (A2), then one can easily see that they also satisfy condition (A1)-(i). Moreover, one can show that their correlation matrix $\boldsymbol{\Psi} = (\Psi_{i,j})$ is given by $\Psi_{i,i} = 1$, $\Psi_{i,j} = \rho(|i-j|) \equiv \rho_{|i-j|}$, $i \neq j$, where ρ is a correlation function and the ρ_i's satisfy the Yule-Walker equations:

$$\rho_k = \phi_1\rho_{k-1} + \cdots + \phi_R\rho_{k-R}, \text{ for } k > 0.$$

The general solution of these difference equations is:

$$\rho_k = \psi_1\lambda_1^k + \psi_2\lambda_2^k + \cdots + \psi_R\lambda_R^k, \text{ for } k > 0$$

where the $\lambda_i, i = 1, \ldots, R$, are the roots of the polynomial equation:

$$z^R - \phi_1 z^{R-1} - \cdots - \phi_R = 0.$$

Initial conditions for determining $\psi_1, \ldots, \psi_R$ can be obtained by using $\rho_0 = 1$ together with the first $R-1$ Yule-Walker equations. For more details, see Chatfield (1989, page 38).

Comment 2.2.2 If the errors $\epsilon_i, i = 1, \ldots, n$, satisfy condition (A2), then their correlation matrix $\boldsymbol{\Psi} = (\Psi_{i,j})$ satisfies condition (A1)-(ii) by Comment 2.2.1 and result (5.34) of Lemma 5.7.2 (Appendix, Chapter 5). In other words, $\boldsymbol{\Psi}$ is symmetric, positive-definite and has finite spectral norm.

Comment 2.2.3 If the errors $\epsilon_i, i = 1, \ldots, n$, associated with model (2.1) satisfy condition (A2) then, by Lemma 2.5.1 in the Appendix of this chapter, $\boldsymbol{\Psi}$ satisfies (2.9) of condition (A1)-(iii), with $\boldsymbol{\Phi}^{(0)} = \boldsymbol{\Sigma}^{(0)}$ and $\boldsymbol{\Sigma}^{(0)}$ defined as in (2.15) .

Comment 2.2.4 Due to its parametric nature, assumption (A2) allows us to find an explicit expression for the inverse of the error correlation matrix $\boldsymbol{\Psi}$, making the derivation of the asymptotic results concerning the modified local linear estimator of $\boldsymbol{\beta}$ easier. We have not been able to modify our proof of these results to handle the more general assumption (A1), since finding an explicit expression for $\boldsymbol{\Psi}^{-1}$ under (A1) may not be possible.

The asymptotic results derived in Chapters 4 and 5 assume h, the half-width of the window of smoothing involved in the definition of the local linear backfitting estimator and the modified local linear estimator of $\boldsymbol{\beta}$, to be deterministic and to satisfy

$$h \rightarrow 0 \tag{2.12}$$

and

$$nh^3 \to \infty \tag{2.13}$$

as $n \to \infty$. These asymptotic results also rely on the conditions below.

(A3) *The Z_i's are non-random and follow a regular design, i.e. there exists a continuous strictly positive density $f(\cdot)$ on $[0,1]$ with:*

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

Moreover, $f(\cdot)$ admits two continuous derivatives.

(A4) *$m(\cdot)$ is a smooth function with 3 continuous derivatives.*

(A5) *$K(\cdot)$, the kernel function used in (3.7) and (3.8), is a probability density function symmetric about 0 and Lipschitz continuous, with compact support $[-1,1]$.*

2.3 Notation

Let $Z_i, i = 1, \ldots, n$, be design points satisfying the design condition (A3) and let $g_1(\cdot), \ldots,$ $g_p(\cdot)$ be functions satisfying the smoothness assumptions in condition (A0)-(i). We define the $n \times (p+1)$ matrix $\boldsymbol{G}$ as:

$$\boldsymbol{G} = \begin{pmatrix} 1 & g_1(Z_1) & \cdots & g_p(Z_1) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & g_1(Z_n) & \cdots & g_p(Z_n) \end{pmatrix} \equiv \begin{pmatrix} g_0(Z_1) & g_1(Z_1) & \cdots & g_p(Z_1) \\ \vdots & \vdots & \cdots & \vdots \\ g_0(Z_n) & g_1(Z_n) & \cdots & g_p(Z_n) \end{pmatrix}. \tag{2.14}$$

Furthermore, let the $n \times (p+1)$ matrix $\boldsymbol{\eta}$ be defined as in (2.10) (condition (A1)-(iii)). In light of condition (A0)-(ii), the transposed rows of $\boldsymbol{\eta}$ are independent, identically distributed degenerate random vectors with mean zero and variance-covariance matrix

$\boldsymbol{\Sigma}^{(0)}$, where:

$$\boldsymbol{\Sigma}^{(0)} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \Sigma_{11} & \cdots & \Sigma_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \Sigma_{p1} & \cdots & \Sigma_{pp} \end{pmatrix}. \tag{2.15}$$

Using equation (2.7) of condition (A0) ($X_{ij} = g_j(Z_i) + \eta_{ij}$) together with the definitions of $\boldsymbol{G}$ and $\boldsymbol{\eta}$ in equations (2.14) and (2.10), we can express the design matrix $\boldsymbol{X}$ in (2.2) as:

$$\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}. \tag{2.16}$$

Let $K(\cdot)$ be a kernel function satisfying condition (A5); if $z \in [0,1]$ and $h \in [0, 1/2]$, define the following quantity:

$$\nu_l(K, z, h) = \int_{-z/h}^{(1-z)/h} s^l K(s) ds, \; l = 0, 1, 2, 3. \tag{2.17}$$

Note that, if $z \in [h, 1-h]$, i.e. z is an 'interior' point of the interval $[0,1]$, then $\nu_l(K, z, h) = \int_{-1}^{1} s^l K(s) ds \equiv \nu_l(K)$ as $[-z/h, (1-z)/h] \supseteq [-1, 1]$ and $K(\cdot)$ has compact support on $[-1, 1]$ by condition (A5).

Now, for $g_0(\cdot), \ldots, g_p(\cdot)$ as above and $f(\cdot)$ a design density, we let:

$$\int_0^1 \boldsymbol{g}(z) f(z) dz = \left(\int_0^1 g_0(z) f(z) dz, \ldots, \int_0^1 g_p(z) f(z) dz \right)^T, \tag{2.18}$$

and

$$\int_0^1 \boldsymbol{g}(z) m''(z) f(z) dz = \left(\int_0^1 g_0(z) m''(z) f(z) dz, \ldots, \int_0^1 g_p(z) m''(z) f(z) dz \right)^T. \tag{2.19}$$

We also let $\int_0^1 \boldsymbol{g}(z)^T f(z) dz = \left[\int_0^1 \boldsymbol{g}(z) f(z) dz \right]^T$ and define the $(p+1) \times (p+1)$ matrix $\boldsymbol{V}$ as:

$$\boldsymbol{V} = \boldsymbol{\Sigma}^{(0)} + \int_0^1 \boldsymbol{g}(z) f(z) dz \cdot \int_0^1 \boldsymbol{g}(z)^T f(z) dz, \tag{2.20}$$

with $\boldsymbol{\Sigma}^{(0)}$ as in (2.15). We also define the $(p+1)$ vector $\boldsymbol{W}$ as:

$$\boldsymbol{W} = \frac{\nu_2(K)}{2}\int_0^1 \boldsymbol{g}(z)m''(z)f(z)dz - \frac{\nu_2(K)}{2}\int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 m''(z)f(z)dz. \quad (2.21)$$

Finally, define the $(p+1)\times(p+1)$ matrix $\boldsymbol{V}_{\boldsymbol{\Psi}}$ as:

$$\boldsymbol{V}_{\boldsymbol{\Psi}} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1+\sum_{k=1}^R \phi_k^2\right)\boldsymbol{\Sigma}^{(0)} + \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1-\sum_{k=1}^R \phi_k\right)^2 \int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 \boldsymbol{g}(z)^T f(z)dz. \quad (2.22)$$

2.4 Linear Algebra - Useful Definitions and Results

In this section, we first provide an overview of the vector and matrix norm definitions and properties used throughout the remainder of this thesis.

Let $\boldsymbol{A} = (A_{ij})$ be an arbitrary $m \times n$ matrix and $\boldsymbol{B} = (B_{kl})$ be an $n \times q$ matrix, both having real elements. Also, let $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ be an arbitrary $n \times 1$ vector with real elements. The spectral norm of the matrix $\boldsymbol{A}$ is defined as:

$$||\boldsymbol{A}||_S = \max_{||\boldsymbol{v}||_2 \neq 0} \frac{||\boldsymbol{A}\boldsymbol{v}||_2}{||\boldsymbol{v}||_2}$$

with $||\cdot||_2$ being the Euclidean norm of a vector, that is $||\boldsymbol{v}||_2^2 = \sum_{i=1}^n v_i^2$. Furthermore, the Frobenius norm of $\boldsymbol{A}$ is defined as:

$$||\boldsymbol{A}||_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}.$$

It is well-known that $||\boldsymbol{A}||_S \leq ||\boldsymbol{A}||_F$. Clearly, if $\boldsymbol{A}$ is a column vector (that is, $n = 1$), then $||\boldsymbol{A}||_S = ||\boldsymbol{A}||_2$. In particular, if $\boldsymbol{A}$ is a scalar (i.e., $m = n = 1$), then $||\boldsymbol{A}||_S$ equals the absolute value of this scalar. It is also known that $||\boldsymbol{A}\cdot\boldsymbol{B}||_F \leq ||\boldsymbol{A}||_F \cdot ||\boldsymbol{B}||_F$.

We conclude this section by reviewing the definitions of random bilinear and quadratic forms and providing formulas for computing the expected value of such forms.

Suppose $\boldsymbol{A} = (A_{ij})$ is an $n \times n$ matrix with real-valued elements, not necessarily symmetric. Similarly, suppose that $\boldsymbol{B} = (B_{ij})$ is an $n \times m$ matrix with real-valued elements. Let $\boldsymbol{u}$ be an arbitrary $n \times 1$ random vector having real-valued elements. Also, let $\boldsymbol{v}$ be an arbitrary $m \times 1$ random vector with real-valued elements.

A bilinear form in $\boldsymbol{u}$ and $\boldsymbol{v}$ with regulator matrix $\boldsymbol{B}$ is defined as:

$$B(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{B} \boldsymbol{v} = \sum_{i=1}^{n} \sum_{j=1}^{m} B_{ij} u_i v_j.$$

Note that $B(\boldsymbol{u}, \boldsymbol{v})$ is random, and its expected value can be computed using the following formula:

$$E(B(\boldsymbol{u}, \boldsymbol{v})) = trace(\boldsymbol{B} Cov(\boldsymbol{u}, \boldsymbol{v})^T) + E(\boldsymbol{u})^T \boldsymbol{B} E(\boldsymbol{v}). \tag{2.23}$$

In particular, a quadratic form in $\boldsymbol{u}$ with regulator matrix $\boldsymbol{A}$ is defined as:

$$Q(\boldsymbol{u}) = \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} u_i u_j,$$

with (2.23) reducing to:

$$E(Q(\boldsymbol{u})) = trace(\boldsymbol{A} Var(\boldsymbol{u})) + E(\boldsymbol{u})^T \boldsymbol{A} E(\boldsymbol{u}). \tag{2.24}$$

2.5 Appendix

The following result helps establish that condition (A2) is a special case of condition (A1).

Lemma 2.5.1 *Let $\boldsymbol{\eta}$ be defined as in equation* (2.10) *of condition (A1) and let $\boldsymbol{\Psi}$ be defined as in Comment 2.2.1. Then, as $n \to \infty$,*

$$\frac{1}{n+1} \boldsymbol{\eta} \boldsymbol{\Psi} \boldsymbol{\eta} = \boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1), \tag{2.25}$$

where $\boldsymbol{\Sigma}^{(0)}$ is defined as in (2.15).

Proof:

Let $\boldsymbol{\eta}_l$ denote the l^{th} column of η and consider $\boldsymbol{\eta}_l^T \boldsymbol{\Psi} \boldsymbol{\eta}_t$, where $l, t = 1, \ldots, p+1$. When $l = 1$ or $t = 1$, this is 0. For $l, t = 2, \ldots, p+1$, we have:

$$
\begin{aligned}
\frac{1}{n}\boldsymbol{\eta}_l^T \boldsymbol{\Psi} \boldsymbol{\eta}_{t+1} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \eta_{i,l}\Psi_{i,j}\eta_{j,t} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \rho(|i-j|)\eta_{i,l}\eta_{j,t} \\
&= \frac{1}{n}\sum_{i=1}^{n} \eta_{i,l}\eta_{i,t} + \sum_{k=1}^{[\frac{n}{2}]} \rho(|k|)\left(\frac{1}{n}\sum_{i=1}^{n-k}\eta_{i,l}\eta_{i+k,t} + \frac{1}{n}\sum_{i=k+1}^{n}\eta_{i,l}\eta_{i-k,t}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \eta_{i,l}\eta_{i,t} + \sum_{k=1}^{k_0} \rho(|k|)\left(\frac{1}{n}\sum_{i=1}^{n-k}\eta_{i,l}\eta_{i+k,t} + \frac{1}{n}\sum_{i=k+1}^{n}\eta_{i,l}\eta_{i-k,t}\right) \\
&\quad + \sum_{k=k_0+1}^{[\frac{n}{2}]} \rho(|k|)\left(\frac{1}{n}\sum_{i=1}^{n-k}\eta_{i,l}\eta_{i+k,t} + \frac{1}{n}\sum_{i=k+1}^{n}\eta_{i,l}\eta_{i-k,t}\right) \qquad (2.26)
\end{aligned}
$$

where $[n/2]$ denotes the integer part of $n/2$ and k_0 is chosen independently of n in the following fashion. Since $\sum_{k=1}^{\infty} |\rho(|k|)| < \infty$ (see Lemma 5.7.2 for a justification of this result), for any given $\epsilon > 0$ we can choose k_0 such that:

$$\sum_{k=k_0+1}^{\infty} |\rho(|k|)| < \frac{\epsilon^2}{\mathcal{C}}$$

for some large constant $\mathcal{C}$.

In light of condition (A0)-(ii), the first term in (2.26) converges to $\Sigma_{l,t}$ by the Weak Law of Large Numbers applied to the independent random variables $\eta_{i,l}\eta_{i,t}, i = 1, \ldots, n$.

The second term in (2.26) converges to zero in probability as $n \to \infty$ by the following argument. The random variables $\eta_{i,l}\eta_{i+k,t}, i = 1, \ldots, n-k$, are k-dependent and identically distributed by condition (A0)-(ii). The Weak Law of Large Numbers for k-dependent random variables implies that $\sum_{i=1}^{n-k} \eta_{i,l}\eta_{i+k,t}/(n-k)$ converges to $E(\eta_{1,l}\eta_{2,t}) = E(\eta_{1,l})E(\eta_{2,t}) = 0$ in probability as $n \to \infty$. A similar argument yields that the quantity $\sum_{i=k+1}^{n} \eta_{i,l}\eta_{i-k,t}/n$ converges to 0 in probability as $n \to \infty$.

Now, consider the third term in (2.26). By Markov's Inequality and condition (A0)-(ii), for n large enough, we have:

$$
\begin{aligned}
P&\left(\left|\sum_{k=k_0+1}^{[\frac{n}{2}]}\rho(|k|)\left(\frac{1}{n}\sum_{i=1}^{n-k}\eta_{i,l}\eta_{i+k,t}+\frac{1}{n}\sum_{i=k+1}^{n}\eta_{i,l}\eta_{i-k,t}\right)\right|>\epsilon\right)\\
&<\frac{1}{\epsilon}E\left|\sum_{k=k_0+1}^{[\frac{n}{2}]}\rho(|k|)\left(\frac{1}{n}\sum_{i=1}^{n-k}\eta_{i,l}\eta_{i+k,t}+\frac{1}{n}\sum_{i=k+1}^{n}\eta_{i,l}\eta_{i-k,t}\right)\right|\\
&<\frac{1}{\epsilon}\sum_{k=k_0+1}^{[\frac{n}{2}]}|\rho(|k|)|\left(\frac{1}{n}\sum_{i=1}^{n-k}E|\eta_{i,l}\eta_{i+k,t}|+\frac{1}{n}\sum_{i=k+1}^{n}E|\eta_{i,l}\eta_{i-k,t}|\right)\\
&=\frac{1}{\epsilon}\sum_{k=k_0+1}^{[\frac{n}{2}]}|\rho(|k|)|\left(2\frac{n-k}{n}E|\eta_{1,l}\eta_{1+k,t}|\right)\\
&<\frac{\mathcal{C}}{\epsilon}\sum_{k=k_0+1}^{[\frac{n}{2}]}|\rho(|k|)|<\frac{\mathcal{C}}{\epsilon}\sum_{k=k_0+1}^{\infty}|\rho(|k|)|<\frac{\mathcal{C}}{\epsilon}\cdot\frac{\epsilon^2}{\mathcal{C}}<\epsilon
\end{aligned}
$$

In conclusion, the third term in (2.26) converges to zero in probability as $n \to \infty$. Combining the previous results yields (2.25).

Chapter 3

Estimation in a Partially Linear Model with Correlated Errors

Obtaining sensible point estimators for the linear effects in a partially linear model with correlated errors is the first important step towards carrying out valid inferences on these effects. Such inferences include conducting hypotheses tests for assessing the statistical significance of the linear effects of interest, and constructing confidence intervals for these effects.

As we have seen in Sections 1.1.1-1.1.2, several methods for estimating the linear and non-linear effects in a partially linear model have been proposed in the literature, both in the presence and absence of correlation amongst model errors. In principle, any of these methods could be used to obtain point estimators for the linear effects in a partially linear model with correlated errors. However, those methods which ignore the correlation structure of the model errors might produce less efficient estimators than the methods which account explicitly for this correlation structure. It is still of interest to consider methods which do not account for the presence of correlation amongst the model errors when estimating the linear effects in the model. Indeed, these methods could yield valid testing procedures based on the inefficient point estimators they produce and the standard errors associated with these estimators.

In the present chapter we show that many of the estimation methods used in the literature for a partially linear model with known correlation structure can be conveniently viewed as particular cases of a generic Backfitting Algorithm. We also show how this generic Backfitting Algorithm can be modified for those instances when the error correlation structure is unknown and must be estimated from the data.

This chapter is organized as follows. In Section 3.1, we discuss the generic Backfitting Algorithm for estimating the linear and non-linear effects in model (2.1) when the error correlation structure is known. In particular, in Sections 3.1.1 and 3.1.2 we discuss the usual and modified generic backfitting estimators of these effects. In Section 3.1.3, we talk about appropriate modifications of these estimators that can be used when the error correlation structure is unknown. In Section 3.1.4, we discuss several generic backfitting estimators which are versions of the estimators introduced by Speckman (1988).

3.1 Generic Backfitting Estimators

In this section, we provide a formal definition for the *generic backfitting estimators* of the unknowns $\boldsymbol{\beta}$ and $\boldsymbol{m}$ in model (2.1). We also define and discuss various particular types of these estimators, clearly indicating which of these types we consider in this thesis.

We start by introducing some notation. Let $\boldsymbol{\Omega}$ be an $n \times n$ matrix of weights such that the $(p+1) \times (p+1)$ matrix $\boldsymbol{X}^T \boldsymbol{\Omega} \boldsymbol{X}$ is invertible. Also, let $\mathbb{S}_h$ be a smoother matrix depending on a smoothing parameter h which controls the width of the smoothing window. For example, the local linear smoother matrix is given in (3.6)-(3.8). Next, let $\mathbb{S}_h^c$ be the centered version of $\mathbb{S}_h$, obtained as:

$$\mathbb{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^T/n)\mathbb{S}_h. \tag{3.1}$$

Formal definitions for $\boldsymbol{\Omega}$ and $\mathbb{S}_h^c$ will be provided shortly. For now, we note that the matrix of weights $\boldsymbol{\Omega}$ may possibly depend on the known error correlation matrix $\boldsymbol{\Psi}$ and on the smoother matrix $\mathbb{S}_h^c$.

The constrained generic backfitting estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ of $\boldsymbol{\beta}$ and $\boldsymbol{m}$ are defined as the fixed points to the following generic backfitting equations:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = (\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{Y} - \widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}) \tag{3.2}$$

$$\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = \mathbb{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}). \tag{3.3}$$

Use of the matrix $\mathbb{S}_h^c$ instead of $\mathbb{S}_h$ in equation (3.3) ensures that $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ satisfies the identifiability condition $\mathbf{1}^T\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = 0$.

The motivation behind the generic backfitting equations introduced above is as follows. Given an estimator $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ of the unknown $\boldsymbol{m}$ in model (2.1), one can construct the vector of partial residuals $\boldsymbol{Y} - \widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$. Regressing these partial residuals on $\boldsymbol{X}$ via weighted least squares yields the generic backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ in equation (3.2). On the other hand, given an estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ of the unknown $\boldsymbol{\beta}$ in model (2.1), one can construct the vector of partial residuals $\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$. Smoothing these partial residuals on $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^T$ via the smoother matrix $\mathbb{S}_h^c$ yields the generic backfitting estimator $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ in equation (3.3).

In practice, one could solve the generic backfitting equations (3.2)-(3.3) for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ iteratively by employing a modification of the Backfitting Algorithm of Buja, Hastie and Tibshirani (1989), as follows.

The Generic Backfitting Algorithm

(i) Let $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{m}^{(0)}$ be initial estimators for $\boldsymbol{\beta}$ and $\boldsymbol{m}$ calculated as follows. We regress $\boldsymbol{Y}$ on the parametric and nonparametric covariates in the model via weighted least squares regression, obtaining:

$$\mathcal{P}(x_1, \ldots, x_p, z) = \widehat{\gamma}_0 + \widehat{\gamma}_1 \cdot x_1 + \cdots + \widehat{\gamma}_p \cdot x_p + \widehat{\gamma}_{p+1} \cdot (z - \overline{Z}).$$

Here, $\overline{Z} = (Z_1 + \cdots + Z_n)/n$. Note that, if $\overline{\boldsymbol{Z}} = (\overline{Z}, \ldots, \overline{Z})^T$ is an $n \times 1$ vector, the weighted least squares estimators $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_0, \widehat{\gamma}_1, \ldots, \widehat{\gamma}_p)^T$ and $\widehat{\gamma}_{p+1}$ above are obtained by minimizing the following criterion with respect to $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_p)^T$ and γ_{p+1}:

$$\left[\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\gamma} - \gamma_{p+1}(\boldsymbol{Z} - \overline{\boldsymbol{Z}})\right]^T \boldsymbol{\Omega} \left[\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\gamma} - \gamma_{p+1}(\boldsymbol{Z} - \overline{\boldsymbol{Z}})\right].$$

We let

$$m^{(0)}(z) = \widehat{\gamma}_{p+1} \cdot (z - \overline{Z})$$

and $\boldsymbol{m}^{(0)} = (m^{(0)}(Z_1), \ldots, m^{(0)}(Z_n))^T$. Also, we let $\boldsymbol{\beta}^{(0)} = \widehat{\boldsymbol{\gamma}}$. Note that $\boldsymbol{m}^{(0)}$ satisfies the identifiability condition (2.4).

(ii) Given the estimators $\boldsymbol{\beta}^{(I)}$ and $\boldsymbol{m}^{(I)}$, we construct $\boldsymbol{\beta}^{(I+1)}$ and $\boldsymbol{m}^{(I+1)}$ as follows:

$$\boldsymbol{\beta}^{(I+1)} = (\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{Y} - \boldsymbol{m}^{(I)})$$
$$\boldsymbol{m}^{(I+1)} = \mathbb{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}^{(I)}).$$

Note that $\boldsymbol{m}^{(I+1)}$ satisfies the identifiability condition (2.4), since $\mathbb{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^T/n)\mathbb{S}_h$, for some smoother matrix $\mathbb{S}_h$.

(iii) Repeat (ii) until $\boldsymbol{\beta}^{(I)}$ and $\boldsymbol{m}^{(I)}$ do not change much.

If the Generic Backfitting Algorithm converges at the iteration labeled as $I + 1$, say, we set:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = \boldsymbol{\beta}^{(I)}$$
$$\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = \boldsymbol{m}^{(I)}.$$

However, we need not iterate to find the generic backfitting estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$. Using the generic backfitting equations (3.2) and (3.3), we can easily derive an explicit expression for the generic backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$. Simply substitute the expression of $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ given in equation (3.3) into equation (3.2) and solve for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} &= (\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}\Big[\boldsymbol{Y} - \mathbb{S}_h^c(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c})\Big] \\ &= (\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}\Big[(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{Y} + \mathbb{S}_h^c\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}\Big].\end{aligned}$$

Pre-multiplying both sides of the above equation by $\boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X}$ and rearranging yields

$$\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = \boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{Y}.$$

Thus, provided the matrix $\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{X}$ is invertible,

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = (\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{Y}. \tag{3.4}$$

To obtain the generic backfitting estimator $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ without iterating, substitute the explicit expression of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},\mathbb{S}_h^c}$ obtained above in (3.3) to get:

$$\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},\mathbb{S}_h^c} = \Big[\mathbb{S}_h^c - \mathbb{S}_h^c\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\Big]\boldsymbol{Y}. \tag{3.5}$$

Results (3.4) and (3.5) above show that the generic backfitting equations (3.2)-(3.3) have a unique solution as long as the $(p+1)\times(p+1)$ matrix $\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \mathbb{S}_h^c)\boldsymbol{X}$ is invertible.

Various specifications for the smoother matrix $\mathbb{S}_h^c$ and the matrix of weights $\boldsymbol{\Omega}$ appearing in the generic backfitting equations (3.2) and (3.3) (or, equivalently, in the explicit equations (3.4) and (3.5)) lead to different types of generic backfitting estimators. In the rest of this section, we discuss several such specifications, together with the particular types of generic backfitting estimators they yield. Note that, if one wishes to estimate the unknowns $\boldsymbol{\beta}^*$ and $\boldsymbol{m}^*$ in the intercept-free model (2.5) one should carry out an unconstrained backfitting algorithm, using $\boldsymbol{X}^*$ instead of $\boldsymbol{X}$, and $\boldsymbol{S}_h$ instead of $\boldsymbol{S}_h^c$ in (3.2)-(3.3).

3.1.1 Usual Generic Backfitting Estimators

The usual generic backfitting estimators are obtained from (3.2)-(3.3) by taking $\boldsymbol{\Omega} = \boldsymbol{I}$. Clearly, these estimators are defined by ignoring the correlation structure of the model errors.

In this thesis, we consider a particular type of usual backfitting estimators, obtained by taking $\mathbb{S}_h$ to be a local linear smoother matrix $\boldsymbol{S}_h$, whose formal definition will be provided shortly. We refer to these estimators as local linear backfitting estimators and denote them by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. These estimators were introduced by Opsomer and Ruppert (1999) in the context of partially linear models with uncorrelated errors and discussed in Section 1.1.1.

Taking $\mathbb{S}_h$ to be $\boldsymbol{S}_h$ is motivated by the fact that local linear smoothing has been shown by Fan and Gijbels (1992) and Fan (1993) to be an effective smoothing method in nonparametric regression. It has the advantage of achieving full asymptotic minimax efficiency and automatically correcting for boundary bias. For more information on local linear smoothing, the reader is referred to Fan and Gijbels (1996).

We define the $(i,j)^{\text{th}}$ element of $\boldsymbol{S}_h$ as:

$$S_{ij} = \frac{w_j^{(i)}}{\sum_{j=1}^n w_j^{(i)}}, \tag{3.6}$$

with local weights $w_k^{(i)}$, $k = 1, \ldots, n$, given by:

$$w_j^{(i)} = K\Big(\frac{Z_i - Z_j}{h}\Big)\Big[\mathcal{S}_{n,2}(Z_i) - (Z_i - Z_j)\mathcal{S}_{n,1}(Z_i)\Big]. \tag{3.7}$$

Here:

$$\mathcal{S}_{n,l}(Z) = \sum_{j=1}^n K\Big(\frac{Z - Z_j}{h}\Big)(Z - Z_j)^l, \; l = 1, 2, \tag{3.8}$$

where $Z \in [0, 1]$, h is the half-width of the smoothing window and K is a kernel function specified by the user. One possible choice of K, which will be used later in this thesis, is the so-called Epanechnikov kernel:

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2), & \text{if } |u| < 1; \\ 0, & \text{else.} \end{cases} \tag{3.9}$$

3.1.2 Modified Generic Backfitting Estimators

The modified generic backfitting estimators are feasible when the error correlation matrix $\boldsymbol{\Psi}$ is fully known. These estimators are obtained from (3.2)-(3.3) by taking $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$. Unlike the usual generic backfitting estimators, which ignore the error correlation structure of the model errors, the modified generic backfitting estimators estimators account for this correlation structure and thus would be expected to be more efficient.

In this thesis, we consider a particular case of modified generic backfitting estimators, obtained by taking $\mathbb{S}_h$ to be the local linear smoother matrix $\boldsymbol{S}_h$, whose $(i,j)^{\text{th}}$ element is defined in (3.6)-(3.8). We refer to these estimators as modified local linear backfitting estimators and denote them by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1}, \boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\boldsymbol{\Psi}^{-1}, \boldsymbol{S}_h^c}$.

3.1.3 Estimated Modified Generic Backfitting Estimators

In practice, the error correlation matrix $\boldsymbol{\Psi}$ is never fully known. More commonly, $\boldsymbol{\Psi}$ is assumed to be known only up to a finite number of parameters, or assumed to be stationary, but otherwise left completely unspecified. In these situations, the modified generic backfitting estimators are no longer feasible. However, these estimators can be adjusted to become feasible by simply replacing $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ with $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$, where $\widehat{\boldsymbol{\Psi}}$ is an estimator of $\boldsymbol{\Psi}$. We refer to these adjusted estimators as being estimated modified generic backfitting estimators.

In this thesis, we consider a particular case of estimated modified generic backfitting estimators, obtained by taking $\mathbb{S}_h$ to be the local linear smoother matrix $\boldsymbol{S}_h$, whose $(i,j)^{\text{th}}$

element is defined in (3.6)-(3.8). We refer to these estimators as estimated modified local linear backfitting estimators and denote them by $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, \boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{\widehat{\boldsymbol{\Psi}}^{-1}, \boldsymbol{S}_h^c}$.

Surprisingly, not much information is available in the partially linear regression model literature on estimating the correlation structure of the model errors when it is known only up to a finite number of parameters, or assumed to be stationary, but otherwise left completely unspecified. Later in this thesis we discuss how one might obtain estimators for the error variance σ_ϵ^2 and the error correlation matrix $\boldsymbol{\Psi}$ in practice.

3.1.4 Usual, Modified and Estimated Modified Speckman Estimators

As we have seen earlier, the usual, modified and estimated modified backfitting estimators are obtained from (3.2)-(3.3) by taking $\boldsymbol{\Omega}$ to be $\boldsymbol{I}$, $\boldsymbol{\Psi}^{-1}$ and $\widehat{\boldsymbol{\Psi}}^{-1}$, respectively, with $\mathbb{S}_h^c$ determined by the smoothing method chosen. Other estimators are the usual, modified and estimated modified Speckman estimators, which are obtained from (3.2)-(3.3) by taking $\boldsymbol{\Omega}$ to be $(\boldsymbol{I}-\mathbb{S}_h^c)^T$, $(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1}$ and $(\boldsymbol{I}-\mathbb{S}_h^c)^T\widehat{\boldsymbol{\Psi}}^{-1}$, respectively. Here, $\widehat{\boldsymbol{\Psi}}$ is an estimator of $\boldsymbol{\Psi}$, while $\mathbb{S}_h^c$ depends on the smoothing method of our choice. We discuss these estimators below.

The usual Speckman estimators ignore the correlation structure of the model errors. In what follows, we denote these estimators by $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T, \mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T, \mathbb{S}_h^c}$. An explicit expression for $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T, \mathbb{S}_h^c}$ can be found by taking $\boldsymbol{\Omega} = (\boldsymbol{I}-\mathbb{S}_h^c)^T$ in (3.4):

$$\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T, \mathbb{S}_h^c} = (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{Y}}, \tag{3.10}$$

where $\tilde{\boldsymbol{X}} \equiv (\boldsymbol{I}-\mathbb{S}_h^c)\boldsymbol{X}$ and $\tilde{\boldsymbol{Y}} \equiv (\boldsymbol{I}-\mathbb{S}_h^c)\boldsymbol{Y}$ are partial residuals formed by smoothing $\boldsymbol{X}$ and $\boldsymbol{Y}$ as functions of $\boldsymbol{Z}$. The usual Speckman estimator $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T, \mathbb{S}_h^c}$ can thus be thought of as being the least squares estimator of $\boldsymbol{\beta}$ obtained by regressing the partial residuals $\tilde{\boldsymbol{Y}}$ on the partial residuals $\tilde{\boldsymbol{X}}$. Later in this thesis, we compare the finite sample behaviour of the usual Speckman estimator $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T,\, \mathbb{S}_h^c}$, with $\mathbb{S}_h^c$ being a local constant matrix with Nadaraya-Watson weights, against that of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\, \boldsymbol{S}_h^c}$, the local linear backfitting estimator, and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\, \boldsymbol{S}_h^c}$, the estimated modified local linear backfitting estimator.

The modified Speckman estimators are defined by taking into account the correlation structure of the errors associated with model (2.1) and are feasible when the correlation matrix of these errors is fully known. We denote these estimators by $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1},\, \mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1},\, \mathbb{S}_h^c}$ and note that an explicit expression for $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1},\, \mathbb{S}_h^c}$ can be found by taking $\boldsymbol{\Omega} = (\boldsymbol{I} - \mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1}$ in (3.4):

$$\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1},\, \mathbb{S}_h^c} = (\tilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{Y}}. \tag{3.11}$$

One can see that $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\boldsymbol{\Psi}^{-1},\, \mathbb{S}_h^c}$ is a weighted least squares estimator, obtained by regressing the partial residuals $\tilde{\boldsymbol{Y}}$ on the partial residuals $\tilde{\boldsymbol{X}}$. The large-sample properties of an unconstrained version of this estimator have been studied by Aneiros Pérez and Quintela del Río (2001a) under the assumption of α-mixing errors. Their estimator is given by:

$$\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{K}_h)^T\boldsymbol{\Psi}^{-1},\, \boldsymbol{K}_h} = (\tilde{\boldsymbol{X}}^{*T}\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{X}}^{*})^{-1}\tilde{\boldsymbol{X}}^{*T}\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{Y}}, \tag{3.12}$$

where $\tilde{\boldsymbol{X}}^* \equiv (\boldsymbol{I}-\boldsymbol{K}_h^c)\boldsymbol{X}^*$, $\boldsymbol{X}^*$ is defined as in (2.6) and $\boldsymbol{K}_h$ is an uncentered local constant smoother matrix with Gasser-Müller weights. Later in this thesis, we compare their asymptotic properties of $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{K}_h)^T\boldsymbol{\Psi}^{-1},\, \boldsymbol{K}_h}$ against those of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\, \boldsymbol{S}_h^c}$, the modified local linear backfitting estimator. We do not, however, compare the finite sample properties of these estimators, as neither estimator can be computed in practice. Indeed, both estimators depend on the true error correlation matrix, which is typically unknown in applications.

The estimated modified Speckman estimators are feasible in those situations where the error correlation matrix is unknown but estimable. We denote these estimators by $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\widehat{\boldsymbol{\Psi}}^{-1},\, \mathbb{S}_h^c}$ and $\widehat{\boldsymbol{m}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\widehat{\boldsymbol{\Psi}}^{-1},\, \mathbb{S}_h^c}$. An explicit expression for $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\mathbb{S}_h^c)^T\widehat{\boldsymbol{\Psi}}^{-1},\, \mathbb{S}_h^c}$ can be obtained by substituting $\widehat{\boldsymbol{\Psi}}$ instead of $\boldsymbol{\Psi}$ into (3.11).

In the remainder of this thesis, we concentrate on the following estimators of $\boldsymbol{\beta}$, the parametric component in model (2.1):

(i) $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, the local linear backfitting estimator;

(ii) $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$, the modified local linear backfitting estimator;

(iii) $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$, the estimated modified local linear backfitting estimator.

Opsomer and Ruppert (1999) studied the asymptotic behaviour of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ under the assumption that the model errors are uncorrelated. However, the asymptotic behaviour of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ has not been studied under the assumption of error correlation. In Chapter 4 of this thesis, we investigate the asymptotic behaviour of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ and discuss conditions under which this estimator is $\sqrt{n}$-consistent. In Chapter 5, we obtain similar results for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ for correctly specified $\boldsymbol{\Psi}$. Rather than assuming $\boldsymbol{\Psi}$ to have a general form as in Chapter 4, we restrict it to have a parametric (autoregressive) structure in order to simplify the proofs of all results in Chapter 5. We also give conditions under which $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent.

Chapter 4

Asymptotic Properties of the Local Linear Backfitting Estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$

In this chapter, we investigate the large-sample behaviour of the local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ as the number of data points in the local linear smoothing window increases and the window size decreases at a specified rate. Recall that an explicit expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ can be obtained from (3.4) by taking $\boldsymbol{\Omega} = \boldsymbol{I}$ and replacing $\mathbb{S}_h^c$ with the centered local linear smoother $\boldsymbol{S}_h^c$:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} = \left(\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{Y}. \tag{4.1}$$

Throughout this chapter, we assume that the errors associated with model (2.1) are a realization from a zero mean, covariance-stationary stochastic process satisfying condition (A1) of Section 2.2. We also assume that the non-linear variable in the model is a fixed design variable following a smooth design density $f(\cdot)$ (condition(A3), Section 2.2) and having a smooth effect $m(\cdot)$ on the mean response (condition (A4), Section 2.2). Finally, we allow the linear variables in the model to be mutually correlated and assume they are related with the non-linear variable via a non-parametric regression relationship (condition (A0), Section 2.2).

In Sections 4.1 and 4.2, we provide asymptotic expressions for the exact conditional bias and variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, given $\boldsymbol{X}, \boldsymbol{Z}$. In Section 4.3, we provide an asymptotic expression for an exact conditional quadratic loss criterion that measures the accuracy of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ as an estimator of $\boldsymbol{\beta}$. In Section 4.4, we discuss the circumstances under which the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ can be achieved given $\boldsymbol{X}$ and $\boldsymbol{Z}$. In particular, we show that one must 'undersmooth' $\widehat{\boldsymbol{m}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, the estimated non-parametric component, to ensure that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$. The results in Sections 4.1-4.4 focus on the local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. In Section 4.5, we indicate how these results can be generalized to local polynomials of higher degree. The chapter concludes with an Appendix containing several auxiliary results.

Throughout this chapter, we let $\boldsymbol{G}_i$ denote the i^{th} column of the matrix $\boldsymbol{G}$ defined in (2.14), and $\boldsymbol{\eta}_i$ denote the i^{th} column of the matrix $\boldsymbol{\eta}$ defined in (2.10). We also let $\widehat{\beta}_{i,\boldsymbol{I},\boldsymbol{S}_h^c}$ denote the i^{th} component of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$.

4.1 Exact Conditional Bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$

The modelling flexibility of the partially linear model (2.1) comes at a price. On one hand, the presence of the nonparametric term $\boldsymbol{m}$ in this model safeguards against model mis-specification bias in the estimated relationships between the linear variables $X_1, \ldots, X_p$ and the response. On the other hand, allowing $\boldsymbol{m}$ to enter the model causes the usual backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ to suffer from finite sample bias. Indeed, using the explicit expression of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ in (4.1), together with the model formulation in (2.1), we easily see the conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, given $\boldsymbol{X}, \boldsymbol{Z}$, to be:

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X}, \boldsymbol{Z}) - \boldsymbol{\beta} = \left(\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}, \tag{4.2}$$

an expression which generally does not equal zero.

Theorem 4.1.1 below provides an asymptotic expression for the exact conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$. As we already mentioned, this expression is obtained by assuming that the amount of smoothing h required for computing the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is deterministic and satisfies conditions (2.12) and (2.13).

Theorem 4.1.1 *Let* $\boldsymbol{V}$ *and* $\boldsymbol{W}$ *be defined as in equations* (2.20) - (2.21). *Under assumptions* (A0), (A1) *and* (A3) - (A5), *if* $n \to \infty$, $h \to 0$ *and* $nh^3 \to \infty$, *the conditional bias of the usual backfitting estimator* $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ *of* $\boldsymbol{\beta}$, *given* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$, *is:*

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} = -h^2 \cdot \boldsymbol{V}^{-1}\boldsymbol{W} + \boldsymbol{o}_P(h^2). \tag{4.3}$$

Comment 4.1.1 From equation (4.2) above, one can see that the exact conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, given $\boldsymbol{X}$ and $\boldsymbol{Z}$, does not depend upon the error correlation matrix $\boldsymbol{\Psi}$. Hence, it is not surprising that the leading term in (4.3) is unaffected by the possible correlation of the model errors.

Proof of Theorem 4.1.1:

Let:

$$\boldsymbol{B}_{n,\boldsymbol{I}} = \frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X},$$

where the dependence of $\boldsymbol{B}_{n,\boldsymbol{I}}$ upon h is omitted for convenience. We will see below that when $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, $\boldsymbol{B}_{n,\boldsymbol{I}}$ converges in probability to the quantity $\boldsymbol{V}$ defined in equation (2.20) . Since $\boldsymbol{V}$ is non-singular by Lemma 4.6.11, the explicit expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ in (4.1) holds on a set whose measure goes to 1 as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. We can use this expression to write:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} = \boldsymbol{B}_{n,\boldsymbol{I}}^{-1} \cdot \left\{\frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{Y}\right\}, \tag{4.4}$$

which holds on a set whose measure goes to 1 as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. Taking conditional expectation in both sides of (4.4) and subtracting $\boldsymbol{\beta}$ yields:

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} = \boldsymbol{B}_{n,\boldsymbol{I}}^{-1} \cdot \left\{\frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}\right\}. \tag{4.5}$$

We now show that $\boldsymbol{B}_{n,\boldsymbol{I}}$ converges in probability to $\boldsymbol{V}$ as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, that is:

$$\boldsymbol{B}_{n,\boldsymbol{I}} = \boldsymbol{V} + \boldsymbol{o}_P(1). \tag{4.6}$$

By equation (2.16), $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$, so $\boldsymbol{B}_{n,\boldsymbol{I}}$ can be decomposed as:

$$\begin{aligned}\boldsymbol{B}_{n,\boldsymbol{I}} &= \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta} \\ &+ \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta}.\end{aligned}$$

Using $\boldsymbol{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^T/n)\boldsymbol{S}_h$ (equation (3.1) with $\mathbb{S}_h = \boldsymbol{S}_h$), we re-write the first term, expand the last term and re-arrange to obtain:

$$\begin{aligned}\boldsymbol{B}_{n,\boldsymbol{I}} = &\frac{1}{n(n+1)}\boldsymbol{G}^T\mathbf{1}\mathbf{1}^T\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\eta} + \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G} \\ &+ \frac{1}{n(n+1)}\boldsymbol{G}^T\mathbf{1}\mathbf{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta} + \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G} \\ &- \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{S}_h^c\boldsymbol{\eta}.\end{aligned} \tag{4.7}$$

To establish (4.6), it suffices to show that

$$\frac{1}{n(n+1)}\boldsymbol{G}^T\mathbf{1}\mathbf{1}^T\boldsymbol{G} = \int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 \boldsymbol{g}(z)^T f(z)dz + \boldsymbol{o}(1), \tag{4.8}$$

$$\frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\eta} = \boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1), \tag{4.9}$$

whereas the remaining terms are $\boldsymbol{o}_P(1)$.

First consider $\boldsymbol{G}^T\mathbf{1}\mathbf{1}^T\boldsymbol{G}/n(n+1)$. Set $Z_0 \equiv 0$, $Z_{n+1} \equiv 1$ and use (A3), the design condition on the Z_i's, to get:

$$\begin{aligned}
\int_0^1 g_j(z)f(z)dz &= \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i} g_j(z)f(z)dz \\
&= \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i} [g_j(z)-g_j(Z_i)]f(z)dz + \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i} g_j(Z_i)f(z)dz \\
&= \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i} [g_j(z)-g_j(Z_i)]f(z)dz + \frac{1}{n+1}\sum_{i=1}^{n+1} g_j(Z_i) \\
&= \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i} [g_j(z)-g_j(Z_i)]f(z)dz + \left[\frac{1}{n+1}\boldsymbol{G}^T\mathbf{1}\right]_{j+1}
\end{aligned}$$

for $j = 0, \ldots, p$ fixed. Re-arranging and using the design condition (A3) and the Lipschitz-continuity of $g_j(\cdot)$ (consequence of (A0)-(i)) yields:

$$\begin{aligned}
&\left|\left[\frac{1}{n+1}\boldsymbol{G}^T\mathbf{1}\right]_{j+1} - \int_0^1 g_j(z)f(z)dz\right| = \left|\frac{g_j(1)}{n+1} + \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i}[g_j(z)-g_j(Z_i)]f(z)dz\right| \\
&\leq \frac{|g_j(1)|}{n+1} + \sum_{i=1}^{n+1}\int_{Z_{i-1}}^{Z_i}|g_j(z)-g_j(Z_i)|f(z)dz = \mathcal{O}\left(\frac{1}{n+1}\right)
\end{aligned}$$

for any $j = 0, \ldots, p$, so:

$$\frac{1}{n+1}\boldsymbol{G}^T\mathbf{1} = \int_0^1 \boldsymbol{g}(z)f(z)dz + o(1) \tag{4.10}$$

and (4.8) follows.

Next consider $\boldsymbol{\eta}^T\boldsymbol{\eta}/(n+1)$. Fix $i, j = 1, \ldots, p$, and use (A0)-(ii), which specifies the distributional assumptions on the rows of $\boldsymbol{\eta}$, to get:

$$\left[\frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\eta}\right]_{i+1,j+1} = \frac{1}{n+1}\sum_{k=1}^{n}\eta_{k,i}\eta_{k,j} \to E(\eta_{1i}\eta_{1j}) = \Sigma_{ij}$$

in probability. Since $[\boldsymbol{\eta}^T\boldsymbol{\eta}/(n+1)]_{i+1,j+1} = 0$ whenever $i = 0$ and $j = 0, \ldots, p$ or $i = 1, \ldots, p$ and $j = 0$, (4.9) follows.

It remains to show that all the other terms in (4.7) are $\boldsymbol{o}_P(1)$. It suffices to show that $\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n+1)$, $\boldsymbol{G}_{i+1}^T\mathbf{1}\mathbf{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{G}_{j+1}/n(n+1)$, $\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta}_{j+1}/(n+1)$,

$\boldsymbol{\eta}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{j+1}/(n+1)$ and $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\eta}_{j+1}/(n+1)$ are $o_P(1)$ for any $i, j = 0, 1, \ldots, p$. These facts follow from lemmas appearing in the Appendix of this chapter.

Let $i, j = 0, 1, \ldots, p$ be fixed and consider $\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n+1)$. By result (4.58) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} \equiv \boldsymbol{I}$ and $\boldsymbol{r} \equiv \boldsymbol{G}_{j+1}$, this quantity is $\mathcal{O}(h^2)$, so $\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n+1)$ is $o(1)$. Similarly, by result (4.59) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} \equiv \boldsymbol{I}$ and $\boldsymbol{r} \equiv \boldsymbol{G}_{j+1}$, $\boldsymbol{G}_{i+1}^T\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n(n+1))$ is $\mathcal{O}(h^2)$. Thus, $\boldsymbol{G}_{i+1}^T\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n(n+1))$ is $o(1)$.

Next consider $\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta}_{j+1}/(n+1)$. When $j = 0$, this is 0. For $j = 1, \ldots, p$, by result (4.60) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} \equiv \boldsymbol{I}$ and $\boldsymbol{\xi} \equiv \boldsymbol{\eta}_{j+1}$, this quantity is $\mathcal{O}_P(n^{-1/2}h^{-1/2}) = o_P(1)$. Similarly, when $i = 0$, $\boldsymbol{\eta}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{j+1}/(n+1) = 0$. For $i = 1, \ldots, 0$, result (4.61) of Lemma 4.6.9 establishes that $\boldsymbol{\eta}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{j+1}/(n+1) = o_P(1)$.

Finally, consider $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\eta}_{j+1}/(n+1)$. When $i = 0$ or $j = 0$, this is 0. By result (4.62) of Lemma 4.6.9 with $\boldsymbol{\xi}^* \equiv \boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega} \equiv \boldsymbol{I}$ and $\boldsymbol{\xi} \equiv \boldsymbol{\eta}_{j+1}$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\eta}_{j+1}/(n+1)$ is $\mathcal{O}_P(n^{-1/2}h^{-1/2}) = o_P(1)$ for $i, j = 1, \ldots, p$.

Combining these results, we conclude that

$$\boldsymbol{B}_{n,\boldsymbol{I}} = \Sigma^{(0)} + \int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 \boldsymbol{g}(z)^T f(z)dz + \boldsymbol{o}_P(1) = \boldsymbol{V} + \boldsymbol{o}_P(1).$$

But $\boldsymbol{V}$ is non-singular by Lemma 4.6.11, so

$$\boldsymbol{B}_{n,\boldsymbol{I}}^{-1} = \boldsymbol{V}^{-1} + \boldsymbol{o}_P(1). \tag{4.11}$$

To establish (4.3), by (4.5) it now suffices to show that:

$$\frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} = -h^2\boldsymbol{W} + \boldsymbol{o}_P(h^2). \tag{4.12}$$

This equality is established below with the help of lemmas stated in the Appendix of this chapter.

By equation (2.16), $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$, so $\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}/(n+1)$ can be decomposed as:

$$\frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} = \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} + \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}.$$

Using the identifiability condition on $m(\cdot)$ in (2.4) and the fact that $\boldsymbol{S}_h^c = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{S}_h$ we obtain:

$$\begin{aligned}\frac{1}{n+1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} &= \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} + \frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} \\ &+ \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}. \end{aligned} \tag{4.13}$$

By results (4.66) and (4.67) of Lemma 4.6.10, we obtain $\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}/(n+1) = -h^2(\nu_2(K)/2)\int_0^1 \boldsymbol{g}(z)m''(z)f(z)dz + \boldsymbol{o}_P(h^2)$ as well as $\boldsymbol{G}^T\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}/n(n+1) = h^2(\nu_2(K)/2)\int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 m''(z)f(z)dz + \boldsymbol{o}_P(h^2)$. Result (4.61) of Lemma 4.6.9 with $\boldsymbol{\xi}^* \equiv \boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega} \equiv \boldsymbol{I}$ and $\boldsymbol{r} \equiv \boldsymbol{m}$ establishes that $\boldsymbol{\eta}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}/(n+1) = \mathcal{O}_P(n^{-1/2}h^2) = o_P(h^2)$. Note that result (4.61) of Lemma 4.6.9 holds trivially when $\boldsymbol{\xi}^* = \boldsymbol{\eta}_1$, as $\boldsymbol{\eta}_1 \equiv \boldsymbol{0}$ by definition.

Thus, (4.12) holds. This, combined with (4.5) and (4.11) completes the proof of Theorem 4.1.1.

To better understand the effect of the correlation between the linear and non-linear variables in the model on the asymptotic conditional bias of $\widehat{\boldsymbol{\beta}}_{I,\boldsymbol{S}_h^c}$, we provide an alternative expression for this bias.

Corollary 4.1.1 *Let Z be a random variable with density function $f(\cdot)$ as in assumption* (A3). *Let $X_1, \ldots, X_p$ be random variables related to Z as:*

$$X_j = g_j(Z) + \eta_j,\ j = 1, \ldots, p,$$

where the $g_j(\cdot)$'s are smooth functions as in assumption (A0)-(1) *and the η_i's are random variables satisfying $E(\eta_j|Z) = 0$, $Var(\eta_j|Z) = \Sigma_{jj}$, $Cov(\eta_j, \eta_{j'}) = \Sigma_{jj'}$, $j \neq j'$, with $\boldsymbol{\Sigma} =$*

$(\Sigma_{jj'})$ *as in assumption* (A0)-(ii). *Also, let* $m(\cdot)$ *be a smooth satisfying assumption (A4) and denote its second derivative* $m''(\cdot)$. *Set* $\tilde{\boldsymbol{X}} = (X_1, \ldots, X_p)^T$. *Under the assumptions in Theorem 4.1.1, our previous bias expression can be re-written in terms of* $\tilde{\boldsymbol{X}}$ *and* Z *as:*

$$E(\widehat{\beta}_{0,\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \beta_0 = \frac{h^2\nu_2(K)}{2} E(\tilde{\boldsymbol{X}}|Z)^T Var(\tilde{\boldsymbol{X}}|Z)^{-1} Cov(\tilde{\boldsymbol{X}}, m''(Z)) + o_P(h^2) \tag{4.14}$$

and

$$E\left(\left.\begin{pmatrix} \widehat{\beta}_{1,\boldsymbol{I},\boldsymbol{S}_h^c} \\ \vdots \\ \widehat{\beta}_{p,\boldsymbol{I},\boldsymbol{S}_h^c} \end{pmatrix}\right| \boldsymbol{X},\boldsymbol{Z}\right) - \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = -\frac{h^2\nu_2(K)}{2} Var(\tilde{\boldsymbol{X}}|Z)^{-1} Cov(\tilde{\boldsymbol{X}}, m''(Z)) + \boldsymbol{o}_P(h^2). \tag{4.15}$$

Proof:

Let $\boldsymbol{a} = (\int_0^1 g_1(z)f(z)dz, \ldots, \int_0^1 g_p(z)f(z)dz)^T$ and let $\boldsymbol{W}$ be defined as in (2.21). Set $\boldsymbol{W} = (0, \boldsymbol{W}_2^T)^T$, with:

$$[\boldsymbol{W}_2]_j = \frac{\mu_2(K)}{2}\int_0^1 g_j(z)m''(z)f(z)dz - \frac{\mu_2(K)}{2}\int_0^1 g_j(z)f(z)dz \cdot \int_0^1 m''(z)f(z)dz,$$

for $j = 1, \ldots, p$. Substitute the explicit expression for $\boldsymbol{V}^{-1}$ (result (4.68), Lemma 4.6.11) into (4.3) to obtain:

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} &= -h^2 \left(\begin{array}{c:c} 1 + \boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & -\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1} \\ \hdashline -\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & \boldsymbol{\Sigma}^{-1} \end{array}\right) \begin{pmatrix} 0 \\ \boldsymbol{W}_2 \end{pmatrix} + \boldsymbol{o}_P(h^2) \\ &= -h^2 \begin{pmatrix} -\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{W}_2 \\ \boldsymbol{\Sigma}^{-1}\boldsymbol{W}_2 \end{pmatrix} + \boldsymbol{o}_P(h^2), \end{aligned}$$

with $\boldsymbol{\Sigma}$ as in assumption (A0)-(ii). Results (4.14) and (4.15) follow easily from the above by noting that $\boldsymbol{a} = E(\tilde{\boldsymbol{X}}|Z)$, $\boldsymbol{\Sigma} = Var(\tilde{\boldsymbol{X}}|Z)$ and $\boldsymbol{W}_2 = Cov(\tilde{\boldsymbol{X}}, Z)$.

Result (4.15) in Corollary 4.1.1 shows that the effect of the correlation between the linear variables and the non-linear variable in the model on the asymptotic bias of the local linear backfitting estimator of the linear effects $\beta_1, \ldots, \beta_p$ is through the variance-covariance matrix $Var(\tilde{\boldsymbol{X}}|Z)$ and the covariances $Cov(\tilde{\boldsymbol{X}}, m''(Z))$. Note that the latter depends on the curvature of the smooth non-linear effect $m(\cdot)$ through its second derivative $m''(\cdot)$. Therefore, the leading term in the bias of $\widehat{\beta}_{i,I,\boldsymbol{S}_h^c}$ disappears when there is no correlation between the corresponding linear and non-linear terms in the model, that is when the correlation between $g_i(Z)$ and $m''(Z)$ is zero. In particular, the leading term disappears if $m(\cdot)$ is a line, or if $g_i(\cdot) \equiv c_i$ for some constant c_i.

Opsomer and Ruppert (1999, Theorem 1) obtained a related bias result for the local linear backfitting estimator of the linear effects $\beta_1, \ldots, \beta_p$ in a partially linear model with independent, identically distributed errors. These authors derived their result under a different set of assumptions than ours. Specifically, they assumed the design points $Z_i, i = 1, \ldots, n$, to be random instead of fixed. Furthermore, they did not require that the covariate values X_{ij} and the design points Z_i be related via the nonparametric regression model (2.7). However, they assumed the linear covariates to have mean zero. Finally, they allowed h to converge to zero at a rate slower than ours by assuming $nh \to \infty$ instead of condition (2.13) $(nh^3 \to \infty)$.

The asymptotic bias expression derived by Opsomer and Ruppert is

$$-(h^2\nu_2(K)/2)\{E(Var(\tilde{\boldsymbol{X}}|Z))\}^{-1}Cov(\tilde{\boldsymbol{X}}, m''(Z)) + \boldsymbol{o}_P(h^2).$$

The leading term in this expression is a slight modification of our first term in (4.15), which accounts for the randomness of the Z_i's. The rate of the error associated with Opsomer and Ruppert's asymptotic bias approximation is $\boldsymbol{o}_P(h^2)$ and is of the same order as that associated with the bias approximation in (4.15).

4.2 Exact Conditional Variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ Given $\boldsymbol{X}$ and $\boldsymbol{Z}$

In this section, we derive an asymptotic expression for the exact conditional variance $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ of the usual backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ of $\boldsymbol{\beta}$, given $\boldsymbol{X}$ and $\boldsymbol{Z}$. But first, we obtain an explicit expression for the exact conditional variance $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$. Using the expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ in (4.1) together with the fact that

$$Var(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z}) = \sigma_\epsilon^2 \boldsymbol{\Psi} \tag{4.16}$$

from condition (A1), we get:

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) = \sigma_\epsilon^2 &\left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} \cdot \boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{X} \cdot \\ &\left(\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{X}\right)^{-1}. \end{aligned} \tag{4.17}$$

The next result provides an asymptotic expression for this variance.

Theorem 4.2.1 *Let $\boldsymbol{G}$, $\boldsymbol{V}$ and $\boldsymbol{S}_h^c$ be defined as in equations* (2.14), (2.20) *and* (3.1) *and let $\boldsymbol{I}$ be the $n \times n$ identity matrix. Under conditions* (A0) *and* (A3) - (A5), *if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$,*

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) = &\frac{\sigma_\epsilon^2}{n+1}\boldsymbol{V}^{-1}\boldsymbol{\Phi}^{(0)}\boldsymbol{V}^{-1} + \frac{\sigma_\epsilon^2}{(n+1)^2}\boldsymbol{V}^{-1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}\boldsymbol{V}^{-1} \\ &+ \boldsymbol{o_P}\left(\frac{1}{n}\right), \end{aligned} \tag{4.18}$$

where $\boldsymbol{\Phi}^{(0)}$ is defined in equation (2.9) *and $\boldsymbol{\Psi}$ is the error correlation matrix.*

Comment 4.2.1 From equation (4.17), $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ depends upon the error correlation matrix $\boldsymbol{\Psi}$, so we expect the asymptotic approximation of $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ to also depend upon the correlation structure of the model errors. Indeed, result (4.18) of

Theorem 4.2.1 shows that, for large samples, the first term in the asymptotic expression of $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ depends on $\boldsymbol{\Psi}$ indirectly via the limiting value $\boldsymbol{\Phi}^{(0)}$ of $\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{\eta}/(n+1)$, while the second term depends on $\boldsymbol{\Psi}$ directly.

Comment 4.2.2 By Lemma 4.6.12, the second term in (4.18) is at most $\mathcal{O}(1/n)$. Therefore, $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ has a rate of convergence of $1/n$.

Proof of Theorem 4.2.1:

From (4.6), $B_{n,\boldsymbol{I}} = \boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}/(n+1) = \boldsymbol{V} + \boldsymbol{o}_P(1)$, so $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ in (4.17) can be written as:

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) &= \sigma_\epsilon^2 B_{n,\boldsymbol{I}}^{-1}\cdot\frac{1}{(n+1)^2}\boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{X}\cdot\left(B_{n,\boldsymbol{I}}^T\right)^{-1} \\ &\equiv \frac{\sigma_\epsilon^2}{n+1}B_{n,\boldsymbol{I}}^{-1}\cdot\boldsymbol{C}_{n,\boldsymbol{I}}\cdot\left(B_{n,\boldsymbol{I}}^T\right)^{-1}, \end{aligned} \tag{4.19}$$

where $\boldsymbol{C}_{n,\boldsymbol{I}} \equiv \boldsymbol{X}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{X}/(n+1)$. The dependence of $\boldsymbol{C}_{n,\boldsymbol{I}}$ upon h is omitted for convenience.

To establish (4.18), it suffices to show that $\boldsymbol{C}_{n,\boldsymbol{I}}$ satisfies:

$$\boldsymbol{C}_{n,\boldsymbol{I}} = \boldsymbol{\Phi}^{(0)} + \boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}/(n+1) + \boldsymbol{o}_P(1). \tag{4.20}$$

Using $\boldsymbol{X} = \boldsymbol{G}+\boldsymbol{\eta}$ (equation (2.16)), $\boldsymbol{C}_{n,\boldsymbol{I}}$ can be decomposed as:

$$\begin{aligned} \boldsymbol{C}_{n,\boldsymbol{I}} &= \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta} \\ &+ \left[\frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}\right]^T + \frac{1}{n+1}\boldsymbol{\eta}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}. \end{aligned}$$

Expanding the last term and re-arranging yields:

$$\begin{aligned} \boldsymbol{C}_{n,\boldsymbol{I}} &= \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{\eta} + \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G} \\ &+ \frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta} + \left[\frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}\right]^T \\ &- \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta} - \left[\frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}\right]^T + \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{S}_h^c\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}. \end{aligned} \tag{4.21}$$

The first term, $\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{\eta}/(n+1)$, converges in probability to $\boldsymbol{\Phi}^{(0)}$ by condition (A1)-(iii). We now show that all the other terms, except for the second, are $\boldsymbol{o}_P(1)$. It suffices to show that $\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}_{j+1}/(n+1)$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$, and $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$ are $o_P(1)$; these facts follow from lemmas appearing in the Appendix of this chapter.

First consider $\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}_{j+1}/(n+1)$. Using Lemma 4.6.4 with $\boldsymbol{\xi}\equiv\boldsymbol{\eta}_{j+1}$ and $\boldsymbol{c}\equiv(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}$, as well as properties of vector and matrix norms from Section 2.4 of Chapter 2, we obtain:

$$\begin{aligned}\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}_{j+1} &= \frac{1}{n+1}\mathcal{O}_P(||(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}||_2)\\ &= \frac{1}{n+1}\mathcal{O}_P\left((1+||\boldsymbol{S}_h^c||_F)\cdot||\boldsymbol{\Psi}||_S\cdot||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}||_2\right)=\mathcal{O}_P(n^{-1/2}h^{-1/2}).\end{aligned}$$

The last equality was derived by using that $||\boldsymbol{S}_h^c||_F$ is $\mathcal{O}(h^{-1/2})$ by result (4.54) of Lemma 4.6.7, $||\boldsymbol{\Psi}||_S$ is $\mathcal{O}(1)$ by assumption (A1)-(ii), and $||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}||_2$ is $\mathcal{O}(n^{1/2})$ by result (4.53) of Lemma 4.6.7 with $\boldsymbol{r}\equiv\boldsymbol{G}_{i+1}$. We conclude that $\boldsymbol{G}_{i+1}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\eta}_{j+1}/(n+1)$ is $o_P(1)$. Note that Lemma 4.6.4 invoked earlier holds trivially for $\boldsymbol{\xi}\equiv\boldsymbol{\eta}_1$, as $\boldsymbol{\eta}_1\equiv\boldsymbol{0}$ by definition.

Next consider $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$ and $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$. When $i=0$ or $j=0$, these quantities are 0, so consider $i,j=1,\ldots,p$. By result (4.63) of Lemma 4.6.9 with $\boldsymbol{\xi}^*=\boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega}=\boldsymbol{\Psi}$ and $\boldsymbol{\xi}=\boldsymbol{\eta}_{j+1}$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$ is $\mathcal{O}_P(n^{-1/2}h^{-1/2})=o_P(1)$. By result (4.64) of the same lemma with $\boldsymbol{\xi}^*=\boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega}=\boldsymbol{I}$, $\boldsymbol{\Omega}^*=\boldsymbol{\Psi}$ and $\boldsymbol{\xi}=\boldsymbol{\eta}_{j+1}$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^c\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\eta}_{j+1}/(n+1)$ is $\mathcal{O}_P(n^{-1}h^{-1})=o_P(1)$.

Combining these results in (4.21) yields (4.20). This concludes our proof of Theorem 4.2.1.

We now provide an alternative expression for the asymptotic conditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ which will shed more light on the effect of the correlation between the linear and non-linear variables in model (2.1) on this variance.

Corollary 4.2.1 *Let $\boldsymbol{G}$ as in (2.14) and $\boldsymbol{\Phi}^{(0)}$ be as in (2.9). Set*

$$\left(\begin{array}{c|c} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \hline \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{array}\right) \equiv \boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}, \tag{4.22}$$

where $\boldsymbol{G}_{11}$ is a scalar, $\boldsymbol{G}_{12} = \boldsymbol{G}_{21}^T$ is a $1 \times p$ vector and $\boldsymbol{G}_{22}$ is a $p \times p$ matrix. Also, set:

$$\left(\begin{array}{c|c} \boldsymbol{\Phi}_{11}^{(0)} & \boldsymbol{\Phi}_{12}^{(0)} \\ \hline \boldsymbol{\Phi}_{21}^{(0)} & \boldsymbol{\Phi}_{22}^{(0)} \end{array}\right) \equiv \boldsymbol{\Phi}^{(0)}, \tag{4.23}$$

where $\boldsymbol{\Phi}_{11}^{(0)} = 0$, $\boldsymbol{\Phi}_{12}^{(0)} = (\boldsymbol{\Phi}_{21}^{(0)})^T = \boldsymbol{0}$ is a $1 \times p$ vector, and $\boldsymbol{\Phi}_{22}^{(0)}$ is a $p \times p$ matrix. If $\tilde{\boldsymbol{X}}$ and Z are as in Corollary 4.1.1 and the assumptions in Theorem 4.2.1 hold, then our previous variance expression can be re-written in terms of $\tilde{\boldsymbol{X}}$ and Z:

$$\begin{aligned} Var(\widehat{\beta}_{1,\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) &= \frac{\sigma_\epsilon^2}{n+1}E(\tilde{\boldsymbol{X}}|Z)^T Var(\tilde{\boldsymbol{X}}|Z)^{-1}\boldsymbol{\Phi}_{22}^{(0)}Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z) \\ &+ \frac{\sigma_\epsilon^2}{(n+1)^2}\Big\{\boldsymbol{G}_{11}(1+E(\tilde{\boldsymbol{X}}|Z)^T Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z))^2 - 2\boldsymbol{G}_{12}Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z) \\ &- 2E(\tilde{\boldsymbol{X}}|Z)^T Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z)\boldsymbol{G}_{12}Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z) \\ &+ E(\tilde{\boldsymbol{X}}|Z)^T Var(\tilde{\boldsymbol{X}}|Z)^{-1}\boldsymbol{G}_{22}Var(\tilde{\boldsymbol{X}}|Z)^{-1}E(\tilde{\boldsymbol{X}}|Z)\Big\} \end{aligned} \tag{4.24}$$

and

$$\begin{aligned} Var\left(\left(\begin{array}{c} \widehat{\beta}_{1,\boldsymbol{I},\boldsymbol{S}_h^c} \\ \vdots \\ \widehat{\beta}_{p,\boldsymbol{I},\boldsymbol{S}_h^c} \end{array}\right) \middle| \boldsymbol{X},\boldsymbol{Z}\right) &= \frac{\sigma_\epsilon^2}{n+1}Var(\tilde{\boldsymbol{X}}|Z)^{-1}\boldsymbol{\Phi}_{22}^{(0)}Var(\tilde{\boldsymbol{X}}|Z)^{-1} \\ &+ \frac{\sigma_\epsilon^2}{(n+1)^2}Var(\tilde{\boldsymbol{X}}|Z)^{-1}\Big\{\boldsymbol{G}_{22} - 2E(\tilde{\boldsymbol{X}}|Z)\boldsymbol{G}_{12} + \boldsymbol{G}_{11}E(\tilde{\boldsymbol{X}}|Z)E(\tilde{\boldsymbol{X}}|Z)^T\Big\}Var(\tilde{\boldsymbol{X}}|Z)^{-1} \\ &+ \boldsymbol{o}_P\left(\frac{1}{n}\right). \end{aligned} \tag{4.25}$$

Proof:

Let $\boldsymbol{a} = (\int_0^1 g_1(z)f(z)dz, \ldots, \int_0^1 g_p(z)f(z)dz)^T$ be as in Lemma 4.6.11 and $\boldsymbol{\Sigma} = (\Sigma_{ij})$ be the variance-covariance matrix introduced in condition (A0)-(ii). Substituting the explicit expression for V^{-1} (result (4.68), Lemma 4.6.11) into (4.18) yields:

$$Var(\widehat{\boldsymbol{\beta}}_{I,S_h^c}|\boldsymbol{X},\boldsymbol{Z}) = \left(\begin{array}{c:c} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \hdashline \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{array}\right) + \boldsymbol{o}_P\left(\frac{1}{n}\right)$$

where $\boldsymbol{V}_{11}$ is a scalar, $\boldsymbol{V}_{12} = \boldsymbol{V}_{21}^T$ is a $1 \times p$ vector and $\boldsymbol{V}_{22}$ is a $p \times p$ matrix given by:

$$\begin{aligned} \boldsymbol{V}_{11} = {} & \frac{\sigma_\epsilon^2}{n+1}\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{22}^{(0)}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} + \frac{\sigma_\epsilon^2}{(n+1)^2}\{\boldsymbol{G}_{11}(1+\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a})^2 - 2\boldsymbol{G}_{12}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} \\ & - 2\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a}\boldsymbol{G}_{12}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} + \boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{G}_{22}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}\}, \end{aligned} \tag{4.26}$$

$$\begin{aligned} \boldsymbol{V}_{12}^T = {} & -\frac{\sigma_\epsilon^2}{n+1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{22}^{(0)}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} + \frac{\sigma_\epsilon^2}{(n+1)^2}\{-\boldsymbol{G}_{11}(1+\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a})\boldsymbol{\Sigma}^{-1}\boldsymbol{a} \\ & + \boldsymbol{\Sigma}^{-1}\boldsymbol{a}\boldsymbol{G}_{12}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} + (1+\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a})\boldsymbol{\Sigma}^{-1}\boldsymbol{G}_{12}^T - \boldsymbol{\Sigma}^{-1}\boldsymbol{G}_{22}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}\} \end{aligned} \tag{4.27}$$

and

$$\boldsymbol{V}_{22} = \frac{\sigma_\epsilon^2}{n+1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}_{22}^{(0)}\boldsymbol{\Sigma}^{-1} + \frac{\sigma_\epsilon^2}{(n+1)^2}\boldsymbol{\Sigma}^{-1}\{\boldsymbol{G}_{22} - 2\boldsymbol{a}\boldsymbol{G}_{12} + \boldsymbol{G}_{11}\boldsymbol{a}\boldsymbol{a}^T\}\boldsymbol{\Sigma}^{-1}. \tag{4.28}$$

Results (4.24) and (4.25) follow from (4.26) and (4.28), respectively, since $\boldsymbol{a} = E(\tilde{\boldsymbol{X}}|Z)$ and $\boldsymbol{\Sigma} = Var(\tilde{\boldsymbol{X}}|Z)$.

Result (4.25) of Corollary 4.2.1 shows that the effect of the correlation between the linear variables and the non-linear variable in model (2.1) on the asymptotic variances of the local linear backfitting estimator of the linear effects $\beta_1, \ldots, \beta_p$ is through the conditional variance-covariance matrix $Var(\tilde{\boldsymbol{X}}|Z)$, the conditional mean vector $E(\tilde{\boldsymbol{X}}|Z)$ and the matrices $\boldsymbol{G}_{11}, \boldsymbol{G}_{12}, \boldsymbol{G}_{22}$ in (4.22).

Comment 4.2.3 In the case $\boldsymbol{\Psi} = \boldsymbol{I}$, $\boldsymbol{\eta}^T\boldsymbol{\Psi}\boldsymbol{\eta}/(n+1) = \boldsymbol{\eta}^T\boldsymbol{\eta}/(n+1) = \boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1)$ by result (4.9), with $\boldsymbol{\Sigma}^{(0)}$ as in (2.15). Therefore, $\boldsymbol{\Phi}_{22}^{(0)} = \boldsymbol{\Sigma} = Var(\tilde{\boldsymbol{X}}|Z)$. If we also assume, as Opsomer and Ruppert (1999) do, that $E(\tilde{\boldsymbol{X}}|Z) = \mathbf{0}$, then (4.25) becomes:

$$Var\left(\left(\begin{array}{c}\widehat{\beta}_{1,\boldsymbol{I},\boldsymbol{S}_h^c}\\ \vdots \\ \widehat{\beta}_{p,\boldsymbol{I},\boldsymbol{S}_h^c}\end{array}\right)\middle| \boldsymbol{X}, \boldsymbol{Z}\right) = \frac{\sigma_\epsilon^2}{n+1} Var(\tilde{\boldsymbol{X}}|Z)^{-1}$$

$$+ \frac{\sigma_\epsilon^2}{(n+1)^2} Var(\tilde{\boldsymbol{X}}|Z)^{-1} \boldsymbol{G}_{22} Var(\tilde{\boldsymbol{X}}|Z)^{-1} + \boldsymbol{o}_P\left(\frac{1}{n}\right), \tag{4.29}$$

Recall that these authors also used different conditions on the rate of convergence of the smoothing parameter h and the design points $Z_i, i = 1, \ldots, n$. Namely, they allowed h to converge to zero at a rate slower than ours by assuming $nh \to \infty$ instead of $nh^3 \to \infty$, and they assumed the design points $Z_i, i = 1, \ldots, n$, to be random instead of fixed.

The asymptotic variance expression derived by Opsomer and Ruppert (1999, Theorem 1) is $(\sigma_\epsilon^2/n) \cdot \{E(Var(\tilde{\boldsymbol{X}}|Z))\}^{-1} + \boldsymbol{\mathcal{O}}_P(h^2/n + 1/(n^2h))$. The leading term in this variance expression is $(\sigma_\epsilon^2/n) \cdot \{E(Var(\tilde{\boldsymbol{X}}|Z))\}^{-1}$, a slight modification of our first term in (4.29) which accounts for the randomness of the Z_i's. The rate of the error associated with their asymptotic variance approximation is $\boldsymbol{o}_P(h^2/n + 1/(n^2h))$ and is possibly of smaller order than the second term in (4.29), known to be at most $\boldsymbol{\mathcal{O}}_P(1/n)$ by result (4.69) of Lemma 4.6.12 (Appendix, Chapter 4) with $\boldsymbol{\Psi} = \boldsymbol{I}$.

4.3 Exact Conditional Measure of Accuracy of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$

Because $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is generally a biased estimator of $\boldsymbol{\beta}$ for finite samples, any suitable criterion for measuring the accuracy of this estimator should take into account both bias and variance. A natural way to take both effects into account is to consider

$$E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z}\right) = \left\{E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta}\right\}^T \left\{E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta}\right\} + trace\left\{Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})\right\}. \tag{4.30}$$

Using the above equality, which follows from (2.24), and the asymptotic expressions for $E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta}$ and $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ in Theorems 4.1.1 and 4.2.1, we obtain:

Corollary 4.3.1 *Assume that the conditions in Theorem 4.1.1 and Theorem 4.2.1 hold. Then:*

$$E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z}\right) = h^4 \cdot \boldsymbol{W}^T\boldsymbol{V}^{-2}\boldsymbol{W} + \frac{\sigma_\epsilon^2}{n+1} trace\left\{\boldsymbol{V}^{-1}\boldsymbol{\Phi}^{(0)}\boldsymbol{V}^{-1}\right\} + \frac{\sigma_\epsilon^2}{(n+1)^2} trace\left\{\boldsymbol{V}^{-1}\boldsymbol{G}^T(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{G}\boldsymbol{V}^{-1}\right\} + o_P(h^4) + o_P\left(\frac{1}{n}\right). \tag{4.31}$$

4.4 The $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$

For obvious reasons, we would like the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ to have the 'usual' parametric rate of convergence of $1/n$ - the rate that would be achieved if $\boldsymbol{m}$ were known, given $\boldsymbol{X}$ and $\boldsymbol{Z}$. If $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ has this rate of convergence, we say that it is $\sqrt{n}$-consistent. A sufficient condition for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ to be $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$ is for $E(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z})$ to be $\mathcal{O}_P(n^{-1})$.

By result (4.31) in Corollary 4.3.1, $E(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z})$ is $\mathcal{O}_P(h^4) + \mathcal{O}_P(n^{-1})$. This result is due to the fact that the conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is $\boldsymbol{\mathcal{O}}_P(h^2)$, while its conditional variance is $\boldsymbol{\mathcal{O}}_P(n^{-1})$. For $E(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z})$ to be $\mathcal{O}_P(n^{-1})$, we require $h^4 = \mathcal{O}(n^{-1})$, as well as $h \to 0$ and $nh^3 \to \infty$.

To understand the meaning of the above conditions, let us consider that $h = n^{-\alpha}$. For $h \to 0$, we require $\alpha > 0$. Also, for $nh^3 \to \infty$, we require $1 - 3\alpha > 0$. Finally, we want

$h^4 = n^{-4\alpha} = \mathcal{O}(n^{-1})$, so $\alpha \geq 1/4$. Thus, we require $\alpha \in [1/4, 1/3)$. In summary, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ achieves $\sqrt{n}$-consistency for $h = n^{-\alpha}$, with $\alpha \in [1/4, 1/3)$.

We argue that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ computed with an h optimal for estimating m is consistent, but not $\sqrt{n}$-consistent, given $\boldsymbol{X}$ and $\boldsymbol{Z}$. We argue this by finding the amount of smoothing h that is optimal for estimating $m(Z)$ via the local linear backfitting estimator $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ where, for $Z \in [0, 1]$ fixed,

$$\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z) \equiv \frac{\sum_{j=1}^{n} w_j^{(Z)} \left[\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}\right]_j}{\sum_{j=1}^{n} w_j^{(Z)}} \tag{4.32}$$

and

$$w_j^{(Z)} = K\left(\frac{Z - Z_j}{h}\right)[\mathcal{S}_{n,2}(Z) - (Z - Z_j)\mathcal{S}_{n,1}(Z)]. \tag{4.33}$$

Here, $\mathcal{S}_{n,l}(Z)$, $l = 1, 2$, is as in (3.8), K is a kernel function satisfying condition (A5) and the Z_j's are design points satisfying condition (A3).

We define the optimal h for estimating $m(Z)$ via $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ as:

$$h_{AMSE} = \underset{h}{argmin}\ AMSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right),$$

with $AMSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right)$ being an asymptotic approximation to the exact conditional mean squared error of $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$:

$$MSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right) = E\left\{\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z) - m(Z)\right)^2 \middle| \boldsymbol{X}, \boldsymbol{Z}\right\}.$$

To find the order of $AMSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right)$, and hence h_{AMSE}, note that:

$$MSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right) = \left\{E\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right) - m(Z)\right\}^2 + Var\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right).$$

By results (4.73) and (4.74) of Lemma 4.6.13, the first term is $\mathcal{O}_P(h^4)$ and the second term is $\mathcal{O}_P(1/(nh))$, so $MSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right)$ is $\mathcal{O}_P(h^4 + 1/(nh))$. Therefore, $AMSE\left(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X}, \boldsymbol{Z}\right)$ is $\mathcal{O}_P(h^4 + 1/(nh))$, and the h that minimizes it satisfies $h_{AMSE} = \mathcal{O}(n^{-1/5})$.

For $h = h_{AMSE}$, the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ has conditional bias of order $\mathcal{O}_P(n^{-2/5})$ and conditional variance of order $\mathcal{O}_P(n^{-1})$. Thus, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is consistent but not $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$, as its squared conditional bias asymptotically dominates its conditional variance. However, for $h = n^{-\alpha}, \alpha \in [1/4, 1/3)$, the squared conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ will no longer dominate its conditional variance asymptotically, ensuring that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ achieves $\sqrt{n}$-consistency given $\boldsymbol{X}$ and $\boldsymbol{Z}$. Note that the estimator $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ of $m(Z)$ computed with $h = n^{-\alpha}, \alpha \in [1/4, 1/3)$, is 'undersmoothed' relative to that computed with $h = h_{AMSE}$, since $n^{-\alpha} < n^{-1/5}$.

4.5 Generalization to Local Polynomials of Higher Degree

The asymptotic results in this chapter focus on the local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. A natural question that arises is whether these results generalize to the local polynomial backfitting estimator of $\boldsymbol{\beta}$. The latter estimator is obtained from (4.1) by replacing $\boldsymbol{S}_h^c$, the smoother matrix for locally linear regression, with the smoother matrix for locally polynomial regression of degree $D > 1$. See Chapter 3 in Fan and Gijbels (1996) for a definition of locally polynomial regression.

Recall that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ has conditional bias of order $\mathcal{O}_P(h^2)$ and conditional variance of order $\mathcal{O}_P(n^{-1})$ by Theorems 4.1.1 and 4.2.1. In keeping with the locally polynomial regression literature, we conjecture that the local polynomial backfitting estimator of $\boldsymbol{\beta}$ has conditional bias of order $\mathcal{O}_P(h^{D+1})$ and conditional variance of order $\mathcal{O}_P(n^{-1})$. Note that we may need boundary corrections if D is even. If our conjecture holds, we see that the conditional variance of the local polynomial backfitting estimator of $\boldsymbol{\beta}$ is of the same order as that of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. However, the conditional bias of the local polynomial backfitting estimator of $\boldsymbol{\beta}$ is of smaller order than that of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$.

In Section 4.4 we established that $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$ provided h

converges to zero at rate $n^{-\alpha}$, $\alpha \in [1/4, 1/3)$. To ensure that the local polynomial backfitting estimator of $\boldsymbol{\beta}$ is $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$, we conjecture that h should converge to zero at rate $n^{-\alpha}$, $\alpha \in [1/(2D+2), 1/3)$.

4.6 Appendix

Throughout this Appendix, the assumptions and notation introduced in Chapter 2 of this thesis hold, unless otherwise specified. The first result provides an asymptotic bias expression that will be useful for proving subsequent results.

Lemma 4.6.1 *Let $\boldsymbol{S}_h = (S_{ij})$ be the uncentered smoother matrix defined by equations* (3.6)-(3.8) *and $\boldsymbol{S}_h^c = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{S}_h$. Let $\boldsymbol{r} = (r(Z_1), \ldots, r(Z_n))^T$, where $r(\cdot) : [0,1] \to \mathbb{R}$ is a smooth function having three continuous derivatives and the Z_i's are fixed design points satisfying condition* (A3). *Furthermore, let K be a kernel function satisfying condition* (A5) *whose moments $\nu_l(K, z, h), z \in [0,1], l = 0,1,2,3,$ are defined as in* (2.17). *If $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, then the j^{th} element of the vector $(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}$ can be approximated as:*

$$[(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}]_j = B_r(K, Z_j, h) \cdot h^2 + o(h^2) \tag{4.34}$$

uniformly in $Z_j, j = 1, \ldots, n$, where

$$B_r(K, z, h) = \frac{r''(z)}{2} \cdot \frac{\nu_2(K,z,h)^2 - \nu_1(K,z,h)\nu_3(K,z,h)}{\nu_2(K,z,h)\nu_0(K,z,h) - \nu_1^2(K,z,h)}, \; z \in [0,1]. \tag{4.35}$$

Furthermore, if $\boldsymbol{r}^T\boldsymbol{1} = \boldsymbol{0}$, then the j^{th} element of the vector $(\boldsymbol{S}_h^c - \boldsymbol{I})\boldsymbol{r}$ can be approximated as:

$$[(\boldsymbol{S}_h^c - \boldsymbol{I})\boldsymbol{r}]_j = B_r(K, Z_j, h) \cdot h^2 - \left(\frac{1}{n}\sum_{j=1}^{n} B_r(K, Z_j, h)\right) \cdot h^2 + o(h^2). \tag{4.36}$$

Proof:

For $i = 1, \ldots, n$, let $y_i = r(Z_i) + e_i$, with the e_i's independent, identically distributed random variables with mean 0 and standard deviation $\sigma_e \in (0, \infty)$. Set $\boldsymbol{y} = (y_1, \ldots, y_n)^T$; if $\widehat{r}(Z_j) = [\boldsymbol{S}_h \boldsymbol{y}]_j$ is the local linear estimator of $r(Z_j)$ obtained by smoothing $\boldsymbol{y}$ on $Z_1, \ldots, Z_n$ via the local linear smoother matrix $\boldsymbol{S}_h$, then $Bias(\widehat{r}(Z_j)) = [(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}]_j$. Standard results on the asymptotic bias of a local linear estimator yield that $Bias(\widehat{r}(Z_j))$ is of order h^2, with asymptotic constant $B_r(K, Z_j, h)$, uniformly in $Z_j, j = 1, \ldots, n$ (Fan and Gijbels, 1993). So the proof of (4.34) is complete.

The definition of $\boldsymbol{S}_h^c$ and $\boldsymbol{r}^T\boldsymbol{1} = 0$ allow us to write:

$$
\begin{aligned}
[(\boldsymbol{S}_h^c - \boldsymbol{I})\boldsymbol{r}]_j &= \left\{ \left[\left(\boldsymbol{I} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n} \right) \boldsymbol{S}_h - \boldsymbol{I} \right] \boldsymbol{r} \right\}_j \\
&= [(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}]_j - \left[\frac{\boldsymbol{1}\boldsymbol{1}^T}{n} \boldsymbol{S}_h \boldsymbol{r} \right]_j \\
&= [(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}]_j - \left[\frac{\boldsymbol{1}\boldsymbol{1}^T}{n} (\boldsymbol{S}_h - \boldsymbol{I})\, \boldsymbol{r} \right]_j .
\end{aligned}
$$

Substituting (4.34) in the above result yields (4.36).

The next result establishes the boundedness of a function defined in terms of certain moments of a kernel function $K(\cdot)$. Subsequent results rely on this lemma.

Lemma 4.6.2 *Let $K(\cdot)$ be a kernel function satisfying condition* (A5) *and whose moments $\nu_l(K, z, h), z \in [0, 1], l = 0, 1, 2, 3$, are defined as in* (2.17). *Then, for $h_0 \in [0, 1/2]$ small enough and $l = 1, 2, 3$, we have:*

$$
\sup_{h \in [0, h_0]} \sup_{z \in [0,1]} \left| \frac{\nu_l(K, z, h)}{\nu_2(K, z, h)\nu_0(K, z, h) - \nu_1(K, z, h)^2} \right| < \infty. \tag{4.37}
$$

Proof:

For $z \in [0,1]$, we define the function:

$$z \longrightarrow \frac{\nu_l(K,z,h)}{\nu_2(K,z,h)\nu_0(K,z,h) - \nu_1(K,z,h)^2} \tag{4.38}$$

To establish the desired result, it suffices to show that, for any $l = 1,2,3$, this function is bounded when restricted to the intervals $[h, 1-h]$, $[0,h]$ and $[1-h, 1]$, where $h \leq h_0$ for some $h_0 \in [0, 1/2]$ small enough, and that the three bounds do not depend on h.

Let $l = 1,2,3$ be fixed and let $h \leq h_0$ for some $h_0 \in [0,1/2]$ small enough. The restriction of the function in (4.38) to the interval $[h, 1-h]$ is trivially bounded, as $\nu_l(K,z,h) \equiv \nu_l(K)$ for any $z \in [h, 1-h]$. Clearly, the bound of this restriction does not depend on h. To show that the restriction of this function to the interval $[0,h]$ is also bounded, let us note that, if $z \in [0,1]$, there exists $\alpha \in [0,1]$ such that $z = \alpha h$ and so

$$\begin{aligned}
\nu_l(K,z,h) &= \int_{-z/h}^{(1-z)/h} s^l K(s) ds \\
&= \int_{-\alpha}^{1/h-\alpha} s^l K(s) ds \\
&= \int_{-\alpha}^{1} s^l K(s) ds \\
&\equiv \phi_l(\alpha)
\end{aligned}$$

since $h \leq h_0$. Thus, when restricted to the interval $[0,h]$, the function in (4.38) is equivalent to:

$$\alpha \longrightarrow \frac{\phi_l(\alpha)}{\phi_0(\alpha)\phi_2(\alpha) - \phi_1(\alpha)^2} \equiv \frac{\phi_l(\alpha)}{D(\alpha)}$$

where $\alpha \in [0,1]$. To establish boundedness, it suffices to show that the nominator $\phi_l(\alpha)$ is bounded from above while the denominator $D(\alpha)$ is bounded from below for any $\alpha \in [0,1]$ and $l = 1,2,3$.

To bound $\phi_l(\alpha)$, note that:

$$|\phi_l(\alpha)| \leq \int_{-1}^{1} |s^l| K(s) ds$$

since $K(\cdot)$ is a continuous function with compact support. To bound $|D(\cdot)|$ from below, we show that $D(\cdot)$ is non-decreasing on $[0,1]$ and satisfies $D(0) > 0$. As

$$D'(\alpha) = \phi_0'(\alpha) \cdot \phi_2(\alpha) + \phi_0(\alpha) \cdot \phi_2'(\alpha) - 2\phi_1(\alpha) \cdot \phi_1'(\alpha),$$

and

$$\begin{aligned}\phi_l'(\alpha) &= \frac{d}{d\alpha}\left[\int_{-\alpha}^{1} s^l K(s) ds\right] \\ &= (-1)^l K(-\alpha) \\ &= (-1)^l K(\alpha)\end{aligned}$$

for any $l = 0, 1, 2$ (using Leibnitz's Rule and the symmetry of K), we obtain:

$$D'(\alpha) = K(\alpha)\left(\int_{-\alpha}^{1} s^2 K(s) ds + \alpha^2 \int_{-\alpha}^{1} K(s) ds + 2\alpha \int_{-\alpha}^{1} sK(s) ds\right).$$

Since K is non-negative and symmetric about 0, each term above is non-negative and so $D'(\alpha) \geq 0$, that is $D(\cdot)$ is non-decreasing on $[0,1]$. Further, with $K^*(s)$ the density $K(s)/\int_0^1 K(s)ds = 2K(s)$, we obtain:

$$\begin{aligned}D(0) &= \int_0^1 K(s)ds \cdot \int_0^1 s^2 K(s) ds - \left[\int_0^1 sK(s)\right]^2 \\ &= \frac{1}{4}\left[\int_0^1 s^2 K^*(s) ds - (sK^*(s)ds)^2\right].\end{aligned}$$

Thus, $D(0) = Var(D^*)/4 > 0$, with D^* a random variable with density K^*. Finally, note that the upper bound $\int_{-1}^{1} |s^l| K(s) ds / D(0)$ of the function $\phi_l(\alpha)/D(\alpha)$, $\alpha \in [0,1]$, does not depend on h.

A similar argument can be employed to establish that, when $h \leq h_0$, with $h_0 \in [0, 1/2]$, the restriction of the function defined in (4.38) to the interval $[1-h, 1]$ is bounded.

Now, we use Lemma 4.6.1 and Lemma 4.6.2 to derive asymptotic expressions for the Euclidean norms of the biases which can occur when using locally linear regression to estimate a smooth, unknown function $r(\cdot)$.

Lemma 4.6.3 *Let $\boldsymbol{r}$, $\boldsymbol{S}_h$ and $\boldsymbol{S}_h^c$ be as in Lemma 4.6.1. Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$:*

$$\frac{1}{n+1}||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{r}||_2^2 = \frac{\nu_2(K)^2}{4} \int_0^1 r''(z)^2 f(z) dz \cdot h^4 + o(h^4). \tag{4.39}$$

If $\boldsymbol{r}$ also satisfies $\boldsymbol{1}^T\boldsymbol{r} = 0$, then:

$$\frac{1}{n+1}||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}||_2^2 = \frac{\nu_2(K)^2}{4} \left[\int_0^1 r''(z)^2 f(z) dz - \left(\int_0^1 r''(z) f(z) dz \right)^2 \right] \cdot h^4 + o(h^4). \tag{4.40}$$

Proof:

To establish (4.39), use Lemma 4.6.1 to get:

$$\begin{aligned} \frac{1}{n+1}||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{r}||_2^2 &= \frac{1}{n+1} \sum_{j=1}^{n} [(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{r}]_j^2 \\ &= \frac{1}{n+1} \sum_{j=1}^{n} \left[B_r(K, Z_j, h) \cdot h^2 + o(h^2) \right]^2 \\ &= \left(\frac{1}{n+1} \sum_{j=1}^{n} B_r(K, Z_j, h)^2 \right) \cdot h^4 + o(h^4). \end{aligned} \tag{4.41}$$

The last equality using the boundedness of $B_r(K, z, h)$ for all $z \in [0, 1]$ and $h \leq h_0$, with $h_0 \in [0, 1/2]$ small enough, which is a consequence of Lemma 4.6.2 and the boundedness of $r''(\cdot)$. Now, we use $B_r(K, z, h) = r''(z)\nu_2(K)/2$ for $z \in [h, 1-h]$ to write:

$$\begin{aligned} \frac{1}{n+1} \sum_{j=1}^{n} B_r(K, Z_j, h)^2 = \frac{\nu_2(K)^2}{4(n+1)} \sum_{j=1}^{n} r''(Z_j)^2 - \frac{\nu_2(K)^2}{4(n+1)} \sum_{Z_j \notin [h, 1-h]} r''(Z_j)^2 \\ + \quad \frac{1}{n+1} \sum_{Z_j \notin [h, 1-h]} B_r(K, Z_j, h)^2. \end{aligned}$$

The first term can be shown to equal $(\nu_2(K)/2) \int_0^1 r''(z)^2 f(z) dz + o(1)$ by a Riemann integration argument. The second term is $o(1)$, as the sum contains $\mathcal{O}(nh)$ terms and $r''(z)$ is bounded for $z \notin [h, 1-h]$. The third term is also $o(1)$, as the sum contains

$\mathcal{O}(nh)$ terms that have been shown to be bounded for h small enough. Combining these results yields (4.39).

To establish (4.40), we use the fact that $\boldsymbol{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^T/n)\boldsymbol{S}_h$ (equation (3.1)) and $\mathbf{1}^T\boldsymbol{r} = 0$ to obtain:

$$\frac{1}{n+1}||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}||_2^2 = \sum_{j=1}^{n} [(\boldsymbol{S}_h^c - \boldsymbol{I})\boldsymbol{r}]_j^2 .$$

Substituting (4.36) in the above yields (4.40).

The following result provides a probability bound for a linear combination of independent, identically distributed random variables having zero mean and non-zero, finite variance.

Lemma 4.6.4 *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T$ be a vector whose components are independent and identically distributed real-valued random variables. If $E(\xi_1) = 0$ and $0 < Var(\xi_1) < \infty$, then:*

$$\boldsymbol{\xi}^T\mathbf{c} = \mathcal{O}_P(||\mathbf{c}||_2) \tag{4.42}$$

for any real-valued vector $\mathbf{c} = (c_1, \ldots, c_n)^T$.

Proof:

By Chebychev's Theorem, we have:

$$\begin{aligned}\boldsymbol{\xi}^T\mathbf{c} &= E\left\{\sum_{k=1}^{n} c_k\xi_k\right\} + \mathcal{O}_P\left(\sqrt{Var\left\{\sum_{k=1}^{n} c_k\xi_k\right\}}\right) \\ &= 0 + \mathcal{O}_P\left(\sqrt{\sum_{k=1}^{n} c_k^2}\right) = \mathcal{O}_P(||\mathbf{c}||_2).\end{aligned}$$

The next lemma provides asymptotic approximations for the elements $S_{ij},\ i,j=1,\dots,n$, of the local linear smoother matrix $\boldsymbol{S}_h$ defined in (3.6)-(3.8). These approximations are used to obtain uniform bounds for the elements of $\boldsymbol{S}_h$.

Lemma 4.6.5 *Let $S_{ij},\ i,j=1,\dots,n$, be local linear smoothing weights defined as in* (3.6)-(3.8). *Also, let $K(\cdot)$ and $\nu_l(K,z,h)$, $z\in[0,1]$, $l=0,1,2$, as in Lemma 4.6.2. Furthermore, let $Z_i, i=1,\dots,n$, be design points with density function $f(\cdot)$ satisfying condition* (A3). *Then, if $n\to\infty$, $h\to 0$ and $nh^3\to\infty$, we have:*

$$\begin{aligned} S_{ij} &= \frac{1}{f(Z_i)(n+1)h}\cdot\frac{\nu_2(K,Z_i,h)-\frac{Z_i-Z_j}{h}\nu_1(K,Z_i,h)}{\nu_2(K,Z_i,h)\nu_0(K,Z_i,h)-\nu_1(K,Z_i,h)^2}\cdot K\left(\frac{Z_i-Z_j}{h}\right)\\ &\quad+o\left(\frac{1}{nh}\right) \end{aligned} \tag{4.43}$$

uniformly in Z_i, $i=1,\dots,n$. Furthermore, for all $h\le h_0$, with $h_0\in[0,1/2]$ small enough, there exists a positive constant $\mathcal{C}$ so that:

$$|S_{ij}|\le\frac{\mathcal{C}}{(n+1)h}\cdot I(|Z_i-Z_j|\le h) \tag{4.44}$$

uniformly in Z_i and Z_j, $i,j=1,\dots,n$.

Proof:

Using the definition of S_{ij} in (3.6)-(3.8) and the fact that $\sum_{j=1}^n w_j^{(i)}=\mathcal{S}_{n,2}(Z_i)\mathcal{S}_{n,0}(Z_i)-\mathcal{S}_{n,1}(Z_i)^2$, we write:

$$\begin{aligned} (n+1)hS_{ij} &= \frac{(n+1)h\mathcal{S}_{n,2}(Z_i)}{\mathcal{S}_{n,2}(Z_i)\mathcal{S}_{n,0}(Z_i)-\mathcal{S}_{n,1}(Z_i)^2}K\left(\frac{Z_i-Z_j}{h}\right)\\ &\quad-\frac{(n+1)h^2\mathcal{S}_{n,1}(Z_i)}{\mathcal{S}_{n,2}(Z_i)\mathcal{S}_{n,0}(Z_i)-\mathcal{S}_{n,1}(Z_i)^2}K\left(\frac{Z_i-Z_j}{h}\right)\left(\frac{Z_i-Z_j}{h}\right). \end{aligned} \tag{4.45}$$

Let $l=0,1,2,3$ be fixed. By the definition of $\mathcal{S}_{n,l}(\cdot)$ in (3.8), the design condition (A3) on the Z_j's and a Riemann integration argument, we obtain that the following asymptotic expression for $\mathcal{S}_{n,l}(Z_i)/[(n+1)h^{l+1}]$:

$$\frac{1}{(n+1)h^{l+1}}\mathcal{S}_{n,l}(Z_i) = \frac{1}{(n+1)h}\sum_{j=1}^{n} K\left(\frac{Z_i - Z_j}{h}\right)\left(\frac{Z_i - Z_j}{h}\right)^l$$
$$= \frac{1}{h}\int_0^1 K\left(\frac{Z_i - z}{h}\right)\left(\frac{Z_i - z}{h}\right)^l f(z)dz + \mathcal{O}(n^{-1}h^{-2})$$

holds uniformly with respect to $Z_i, i = 1, \ldots, n$, as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. Making the change of variables $s = (Z_i - z)/h$ and using a Taylor series expansion of $f(\cdot)$, we express the leading term in the above asymptotic expression as:

$$\frac{1}{h}\int_0^1 K\left(\frac{Z_i - z}{h}\right)\left(\frac{Z_i - z}{h}\right)^l f(z)dz = \int_{-Z_i/h}^{(1-Z_i)/h} s^l K(s) f(Z_i + sh)ds$$
$$= \int_{-Z_i/h}^{(1-Z_i)/h} s^l K(s)\left[f(Z_i) + f'(Z_i)\cdot(sh) + \frac{f''(Z_i)}{2}\cdot(sh)^2 + o(h^2)\right]ds$$
$$= \int_{-Z_i/h}^{(1-Z_i)/h} s^l K(s)\left[f(Z_i) + \mathcal{O}(h)\right]ds \doteq f(Z_i)\int_{-Z_i/h}^{(1-Z_i)/h} s^l K(s)ds + \mathcal{O}(h)$$
$$= f(Z_i)\nu_l(K, Z_i, h) + \mathcal{O}(h)$$

Here, the $\mathcal{O}$ term holds uniformly with respect to $Z_i, i = 1, \ldots, n$ by the smoothness assumptions on $f(\cdot)$ given in condition (A3). Combining these results, we conclude that:

$$\frac{1}{(n+1)h^{l+1}}\mathcal{S}_{n,l}(Z_i) = f(Z_i)\nu_l(K, Z_i, h) + \mathcal{O}(h) + \mathcal{O}(n^{-1}h^{-2}) \tag{4.46}$$

uniformly in $Z_i, i = 1, \ldots, n$, as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$.

Now, for $l = 0, 1, 2, 3$, we substitute the asymptotic expression of $\mathcal{S}_{n,l}(Z_i)/[(n+1)h^{l+1}]$ in (4.46) in the right side of equation (4.45). Using that the quantities $f(z)$, $K(z)$ and $zK(z)$ are bounded for $z \in [0, 1]$ (conditions (A3) and (A5), respectively) and re-arranging, we easily obtain (4.43). The asymptotic bound for S_{ij} given in (4.44) follows immediately from Lemma 4.6.2 and (4.43).

The following result follows easily from Lemma 4.6.5. This result will be used to prove Lemma 4.6.7.

Lemma 4.6.6 *Let $\boldsymbol{S}_h$ be as in Lemma 4.6.1. Given $\tilde{C} > 0$, there exist $C_1^* > 0$ and $C_2^* > 0$ such that for any $n \geq 1$ and any $\boldsymbol{v} = (v_1, \dots, v_n)^T$ with $|v_j| \leq \tilde{C}$, we have:*

$$\left|[\boldsymbol{S}_h^T \boldsymbol{v}]_j\right| \leq C_1^* \tag{4.47}$$

and

$$|[\boldsymbol{S}_h \boldsymbol{v}]_j| \leq C_2^*. \tag{4.48}$$

Furthermore, we also have:

$$||\boldsymbol{S}_h^T \boldsymbol{v}||_2^2 \leq n(C_1^*)^2 \tag{4.49}$$

and

$$||\boldsymbol{S}_h \boldsymbol{v}||_2^2 \leq n(C_2^*)^2. \tag{4.50}$$

Proof:

Use result (4.44) of Lemma 4.6.5 to write:

$$\begin{aligned}\left|[\boldsymbol{S}_h^T \boldsymbol{v}]_j\right| &= \left|\sum_{k=1}^n S_{jk} v_j\right| \leq \sum_{k=1}^n |S_{jk}| \cdot |v_j| \leq \tilde{C} \sum_{k=1}^n |S_{jk}| \\ &= \tilde{C} \cdot \frac{\mathcal{C}}{(n+1)h} \sum_{k=1}^n I\left(|Z_k - Z_j| \leq h\right) = \frac{1}{(n+1)h} \mathcal{O}(nh) = \mathcal{O}(1).\end{aligned}$$

This proves (4.47). Result (4.48) can be derived using a similar reasoning.

By result (4.47), we have:

$$||\boldsymbol{S}_h^T \boldsymbol{v}||_2^2 = (\boldsymbol{S}_h^T \boldsymbol{v})^T (\boldsymbol{S}_h^T \boldsymbol{v}) = \sum_{j=1}^n \left[\boldsymbol{S}_h^T \boldsymbol{v}\right]_j^2 \leq n(C_1^*)^2,$$

so (4.49) is proven. Result (4.50) can be shown to hold in a similar manner.

Now, we use Lemmas 4.6.5 and 4.6.6 to establish the following asymptotic bounds.

Lemma 4.6.7 *Let $\boldsymbol{r}, \boldsymbol{S}_h^c$ and $\boldsymbol{I}$ be as in Lemma 4.6.1. Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$:*

$$||\boldsymbol{r}||_2 = \mathcal{O}(n^{1/2}), \tag{4.51}$$

$$||\boldsymbol{S}_h^{cT}\boldsymbol{r}||_2 = \mathcal{O}(n^{1/2}), \tag{4.52}$$

$$||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{r}||_2 = \mathcal{O}(n^{1/2}), \tag{4.53}$$

and

$$||\boldsymbol{S}_h^c||_F = \mathcal{O}(h^{-1/2}). \tag{4.54}$$

Proof:

Using the boundedness of $r(\cdot)$, we write:

$$||\boldsymbol{r}||_2^2 = \boldsymbol{r}^T\boldsymbol{r} = \sum_{i=1}^{n} r(Z_i)^2 = \mathcal{O}(n),$$

so (4.51) is proven.

Using $\boldsymbol{S}_h^c = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{S}_h$ and result (4.49) of Lemma 4.6.6 with $\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{r}$, we have:

$$||\boldsymbol{S}_h^{cT}\boldsymbol{r}||_2^2 = \left|\left|\boldsymbol{S}_h^T\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T\right)\boldsymbol{r}\right|\right|_2^2 \equiv ||\boldsymbol{S}_h^T\boldsymbol{v}||_2^2 \leq n \cdot (\mathcal{C}_1^*)^2$$

for some $\mathcal{C}_1^* > 0$ not depending on n. This proves (4.52).

Result (4.53) follows immediately from results (4.51) and (4.52). Finally, to show result (4.54), we use well-known properties of the Frobenius norm to get:

$$||\boldsymbol{S}_h^c||_F = \left|\left|\left(\boldsymbol{I} - \frac{\boldsymbol{1}\boldsymbol{1}^T}{n}\right)\boldsymbol{S}_h\right|\right|_F \leq ||\boldsymbol{S}_h||_F + \frac{1}{n}||\boldsymbol{1}\boldsymbol{1}^T||_F \cdot ||\boldsymbol{S}_h||_F \leq 2||\boldsymbol{S}_h||_F.$$

Thus, it suffices to show that $||\boldsymbol{S}_h||_F$ is of order $\mathcal{O}(h^{-1/2})$.

By result (4.44) of Lemma 4.6.5, we obtain:

$$||\boldsymbol{S}_h||_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}^2 \leq \frac{\mathcal{C}^2}{(n+1)^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(|Z_i - Z_j| \leq h)$$

for some positive constant $\mathcal{C}$. Since the number of non-zero terms in the double sum appearing on the right side of the above inequality is $n\mathcal{O}(nh)$, we conclude that $||\boldsymbol{S}_h||_F^2$ is $\mathcal{O}(h^{-1})$ or, equivalently, that $||\boldsymbol{S}_h||_F$ is $\mathcal{O}(h^{-1/2})$.

The next result provides a probability bound for the Euclidean norm of a vector of n independent, identically distributed random variables having zero mean and non-zero, finite variance. It also provides a probability bound for the Euclidean norm of a transformation of this vector, obtained by pre-multiplying the vector with the transpose of a centered local linear smoother matrix.

Lemma 4.6.8 *Let $\boldsymbol{\xi}$ be as in Lemma 4.6.4 and $\boldsymbol{S}_h^c$ be as in Lemma 4.6.1. Furthermore, let $\boldsymbol{\Omega}$ be an $n \times n$ symmetric, positive definite matrix with $||\boldsymbol{\Omega}||_S = \mathcal{O}(1)$. Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, we have:*

$$||\boldsymbol{\xi}||_2 = \mathcal{O}_P(n^{1/2}) \tag{4.55}$$

$$||\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}||_2 = \mathcal{O}_P(h^{-1/2}) \tag{4.56}$$

$$||\boldsymbol{S}_h^{c}\boldsymbol{\Omega}\boldsymbol{\xi}||_2 = \mathcal{O}_P(h^{-1/2}) \tag{4.57}$$

Proof:

By Markov's Theorem:

$$||\boldsymbol{\xi}||_2^2 = \mathcal{O}_P\left(E(||\boldsymbol{\xi}||_2^2)\right) = \mathcal{O}_P(nVar(\xi_1)) = \mathcal{O}_P(n),$$

so (4.55) is proven.

Next, consider (4.56). Set $\boldsymbol{B} \equiv \boldsymbol{\Omega} \boldsymbol{S}_h^c$. By Markov's Theorem, we have:

$$||\boldsymbol{S}_h^{cT} \boldsymbol{\Omega} \boldsymbol{\xi}||_2^2 = ||\boldsymbol{B}^T \boldsymbol{\xi}||_2^2 = \mathcal{O}_P(E(||\boldsymbol{B}^T \boldsymbol{\xi}||_2^2)) = E(\boldsymbol{\xi}^T \boldsymbol{B} \boldsymbol{B}^T \boldsymbol{\xi})$$

Thus, it suffices to show that $E(\boldsymbol{\xi}^T \boldsymbol{B}\boldsymbol{B}^T \boldsymbol{\xi})$ is $\mathcal{O}(h^{-1/2})$. Using result (2.24) with $\boldsymbol{u} \equiv \boldsymbol{\xi}$ and $\boldsymbol{A} \equiv \boldsymbol{B}\boldsymbol{B}^T$, together with the symmetry of $\boldsymbol{\Omega}$, we obtain:

$$\begin{aligned} E(\boldsymbol{\xi}^T \boldsymbol{B}\boldsymbol{B}^T \boldsymbol{\xi}) &= trace\left(\boldsymbol{B}\boldsymbol{B}^T \cdot Var(\boldsymbol{\xi})\right) + E(\boldsymbol{\xi})^T \cdot \boldsymbol{B}\boldsymbol{B}^T \cdot E(\boldsymbol{\xi}) \\ &= Var(\xi_1) \cdot trace\left(\boldsymbol{B}\boldsymbol{B}^T\right) + 0 = Var(\xi_1) \cdot ||\boldsymbol{B}||_F^2 \\ &\leq ||\boldsymbol{\Omega}||_S^2 \cdot ||\boldsymbol{S}_h^c||_F^2 = \mathcal{O}(1)\mathcal{O}(h^{-1}) = \mathcal{O}(h^{-1}), \end{aligned}$$

by result (4.54) of Lemma 4.6.7. This proves (4.56). Result (4.57) can be established using a similar argument.

The next lemma contains results concerning the asymptotic negligibility of various random or non-random terms. All of these terms depend on a matrix of weights $\boldsymbol{\Omega}$ and on centered or uncentered local linear smoother matrices. Some terms also depend on a matrix of weights $\boldsymbol{\Omega}^*$, possibly different than $\boldsymbol{\Omega}$ itself.

Lemma 4.6.9 *Let $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}^*$ be $n \times n$ symmetric, positive-definite matrices satisfying $||\boldsymbol{\Omega}||_S = \mathcal{O}(1) = ||\boldsymbol{\Omega}^*||_S$. Let $\boldsymbol{S}_h$ and $\boldsymbol{S}_h^c$ be as in Lemma 4.6.1. Set $\boldsymbol{r} = (r(Z_1), \ldots, r(Z_n))^T$ and $\boldsymbol{r}^* = (r^*(Z_1), \ldots, r^*(Z_n))^T$, where $r(\cdot) : [0,1] \to \mathbb{R}$ and $r^*(\cdot) : [0,1] \to \mathbb{R}$ are smooth functions having three continuous derivatives and the Z_i's are fixed design points satisfying condition* (A3). *Finally, let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)^T$ and $\boldsymbol{\xi}^* = (\xi_1^*, \ldots, \xi_n^*)^T$ be vectors whose components are independent, identically distributed random variables such that $E(\xi_1) = 0$, $Var(\xi_1) < \infty$ and $E(\xi_1^*) = 0$, $Var(\xi_1^*) < \infty$. Then, if $n \to \infty$, $h \to 0$ and*

$nh^3 \to \infty$, *we have:*

$$\frac{1}{n+1}\boldsymbol{r}^{*T}\boldsymbol{\Omega}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{r} = \mathcal{O}(h^2), \tag{4.58}$$

$$\frac{1}{n(n+1)}\boldsymbol{r}^{*T}\boldsymbol{\Omega}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{r} = \mathcal{O}(h^2), \tag{4.59}$$

$$\frac{1}{n+1}\boldsymbol{r}^{*T}\boldsymbol{\Omega}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\xi} = \mathcal{O}_P\left(n^{-1/2}h^{-1/2}\right), \tag{4.60}$$

$$\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{r} = \mathcal{O}_P(n^{-1/2}h^2), \tag{4.61}$$

$$\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^c\boldsymbol{\xi} = \mathcal{O}_P(n^{-1/2}h^{-1/2}) \tag{4.62}$$

$$\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^{cT}\boldsymbol{\xi} = \mathcal{O}_P(n^{-1/2}h^{-1/2}). \tag{4.63}$$

$$\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^c\boldsymbol{\Omega}^*\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi} = \mathcal{O}_P(n^{-1}h^{-1}) \tag{4.64}$$

$$\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}^*\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi} = \mathcal{O}_P(n^{-1}h^{-1}). \tag{4.65}$$

Proof:

Using properties of matrix and vector norms introduced in Section 2.4 of Chapter 2, we get:

$$\begin{aligned}\left|\tfrac{1}{n+1}\boldsymbol{r}^{*T}\boldsymbol{\Omega}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{r}\right| &\leq \tfrac{1}{n+1}||\boldsymbol{r}^*||_2 \cdot ||\boldsymbol{\Omega}||_S \cdot ||(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{r}||_2 \\ &= \tfrac{1}{n+1}\mathcal{O}(n^{1/2})\mathcal{O}(1)\mathcal{O}(n^{1/2}h^2) = \mathcal{O}(h^2)\end{aligned}$$

since $||\boldsymbol{r}^*||_2$ is $\mathcal{O}(n^{1/2})$ by result (4.51) with $\boldsymbol{r} = \boldsymbol{r}^*$ and $||(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{r}||_2^2/(n+1)$ is $\mathcal{O}(h^4)$ by result (4.39). Thus, (4.58) holds. Similarly, we obtain:

$$\begin{aligned}\left|\frac{1}{n(n+1)}\boldsymbol{r}^{*T}\boldsymbol{\Omega}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{r}\right| &\leq \frac{1}{n(n+1)}||\boldsymbol{r}^*||_2 \cdot ||\boldsymbol{\Omega}||_S \cdot ||\boldsymbol{1}\boldsymbol{1}^T||_F \cdot ||(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{r}||_2 \\ &\leq \frac{1}{n(n+1)}\mathcal{O}(n^{1/2})\mathcal{O}(1)\mathcal{O}(n)\mathcal{O}(n^{1/2}h^2) = \mathcal{O}(h^2),\end{aligned}$$

so (4.59) holds.

Using result (4.42) with $\boldsymbol{c} \equiv (\boldsymbol{I} - \boldsymbol{S}_h^c)^T \boldsymbol{\Omega} \boldsymbol{r}^*$, we have:

$$\begin{aligned}
\frac{1}{n+1}\boldsymbol{r}^{*T}\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\xi} &= \frac{1}{n+1}\mathcal{O}_P(||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{\Omega}\boldsymbol{r}^*||_2) \\
&\leq \frac{1}{n+1}\mathcal{O}_P\left((1 + ||\boldsymbol{S}_h^c||_F) \cdot ||\boldsymbol{\Omega}||_S \cdot ||\boldsymbol{r}^*||_2\right) = \frac{1}{n+1}\mathcal{O}_P(h^{-1/2})\mathcal{O}_P(1)\mathcal{O}_P(n^{1/2}) \\
&= \mathcal{O}_P(n^{-1/2}h^{-1/2}),
\end{aligned}$$

since $||\boldsymbol{S}_h^c||_F$ is $\mathcal{O}(h^{-1/2})$ by result (4.54) and $||\boldsymbol{r}^*||_2$ is $\mathcal{O}(n^{1/2})$ by result (4.51) with $\boldsymbol{r} = \boldsymbol{r}^*$. We conclude that (4.60) holds.

From result (4.42) with $\boldsymbol{c} \equiv \boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}$ and $\boldsymbol{\xi} = \boldsymbol{\xi}^*$, we get:

$$\begin{aligned}
\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r} &= \frac{1}{n+1}\mathcal{O}_P\left(||\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}||_2\right) \leq \frac{1}{n+1}\mathcal{O}_P(||\boldsymbol{\Omega}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}||_2) \\
&= \frac{1}{n+1}\mathcal{O}_P(1)\mathcal{O}_P(n^{1/2}h^2) = \mathcal{O}_P(n^{-1/2}h^2),
\end{aligned}$$

since $||(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{r}||_2^2/(n+1)$ is $\mathcal{O}(h^4)$ by result (4.40). Therefore, (4.61) holds.

To prove (4.62), write:

$$\begin{aligned}
\left|\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^c\boldsymbol{\xi}\right| &\leq \frac{1}{n+1}||\boldsymbol{\xi}^*||_2 \cdot ||\boldsymbol{\Omega}||_S \cdot ||\boldsymbol{S}_h^c\boldsymbol{\xi}||_2 \\
&= \frac{1}{n+1}\mathcal{O}_P(n^{1/2}) \cdot \mathcal{O}(1) \cdot \mathcal{O}_P(h^{-1/2}) = \mathcal{O}_P(n^{-1/2}h^{-1/2}),
\end{aligned}$$

since $||\boldsymbol{\xi}^*||_2$ is $\mathcal{O}_P(n^{1/2})$ by result (4.55) with $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ and $||\boldsymbol{S}_h^c\boldsymbol{\xi}||_2$ is $\mathcal{O}_P(h^{-1/2})$ by result (4.57) with $\boldsymbol{\Omega} = \boldsymbol{I}$. Result (4.63) follows via a similar argument, but with result (4.57) replaced by result (4.56).

Result (4.64) follows by noting that:

$$\begin{aligned}
\left|\frac{1}{n+1}\boldsymbol{\xi}^{*T}\boldsymbol{\Omega}\boldsymbol{S}_h^c\boldsymbol{\Omega}^*\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}\right| &\leq \frac{1}{n+1}||\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}^*||_2 \cdot ||\boldsymbol{\Omega}^*||_S \cdot ||\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}||_2 \\
&= \frac{1}{n+1}\mathcal{O}_P(h^{-1/2})\mathcal{O}(1)\mathcal{O}_P(h^{-1/2}) = \mathcal{O}_P(n^{-1}h^{-1}),
\end{aligned}$$

since both $||\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}^*||_2$ and $||\boldsymbol{S}_h^{cT}\boldsymbol{\Omega}\boldsymbol{\xi}||_2$ are $\mathcal{O}_P(h^{-1/2})$ by Lemma 4.6.8. A similar reasoning yields that (4.65) holds. This concludes our proof of the current lemma.

The next lemma provides asymptotic expressions for quantities involving the bias of a local linear estimator of an unknown, smooth regression function $m(\cdot)$.

Lemma 4.6.10 *Let $\boldsymbol{G}$ be as in* (2.14) *and $\boldsymbol{S}_h$ be as in Lemma 4.6.1. Furthermore, let $\boldsymbol{m} = (m(Z_1), \ldots, m(Z_n))^T$, where m satisfies the smoothness conditions in condition* (A4) *and $Z_1, \ldots, Z_n$ are fixed design points satisfying condition* (A3)*. Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, we have:*

$$\frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} = -h^2\frac{\nu_2(K)}{2}\int_0^1 \boldsymbol{g}(z)m''(z)f(z)dz + \boldsymbol{o}(h^2) \tag{4.66}$$

$$\frac{1}{(n+1)^2}\boldsymbol{G}^T\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} = h^2\frac{\nu_2(K)}{2}\int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 m''(z)f(z)dz + \boldsymbol{o}(h^2) \tag{4.67}$$

where $\int_0^1 \boldsymbol{g}(z)f(z)dz$ and $\int_0^1 \boldsymbol{g}(z)m''(z)f(z)dz$ are defined as in equations (2.18) *and* (2.19).

Proof:

Let $i = 0, 1, \ldots, p$, be fixed. By result (4.34) of Lemma 4.6.1 with $\boldsymbol{r} \equiv \boldsymbol{m}$, the $(i+1)^{\text{st}}$ element of $\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}/(n+1)$ is:

$$\begin{aligned}
\Big[\frac{1}{n+1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}\Big]_{i+1} &= -\frac{1}{n+1}\sum_{j=1}^n g_i(Z_j)\Big[(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}\Big]_j \\
&= -h^2\Big[\frac{1}{n+1}\sum_{j=1}^n g_i(Z_j)B_m(K, Z_j, h)\Big] + o(h^2).
\end{aligned}$$

Noting that $B_m(K, z, h) = m''(z)\nu_2(K)/2$ for $z \in [h, 1-h]$, we write:

$$\begin{aligned}
\frac{1}{n+1}\sum_{j=1}^n g_i(Z_j)B_m(K, Z_j, h) &= \frac{\nu_2(K)}{2(n+1)}\sum_{j=1}^n g_i(Z_j)m''(Z_j) \\
&- \frac{\nu_2(K)}{2(n+1)}\sum_{Z_j \notin [h, 1-h]} g_i(Z_j)m''(Z_j) + \frac{1}{n+1}\sum_{Z_j \notin [h, 1-h]} g_i(Z_j)B_m(K, Z_j, h).
\end{aligned}$$

The first term can be shown to equal $(\nu_2(K)/2) \cdot \int_0^1 g_i(z)m''(z)f(z)dz + o(1)$ by a Riemann integration argument. The second and third terms are $o(1)$, as both sums contain $\mathcal{O}(nh)$ terms and these terms are bounded for h small enough, by the following argument. The boundedness of $m''(z)$ for $z \notin [h, 1-h]$ is a consequence of condition (A4). Lemma 4.6.2 yields that the function $z \longrightarrow B_m(K, z, h)$ is bounded for all $z \in [0,1]$ and $h \le h_0$ with $h_0 \in [0, 1/2]$ small enough. Combining these results yields (4.66).

Now, consider (4.67). Since the first column of $\boldsymbol{G}$ is the vector $\mathbf{1}$, from (4.66):

$$\mathbf{1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}/(n+1) = -\frac{h^2\nu_2(K)}{2}\int_0^1 m''(z)f(z)dz + o(h^2).$$

Combining this with (4.10) proves (4.67).

The next result concerns the existence of an inverse for the $(p+1) \times (p+1)$ matrix $\boldsymbol{V}$ defined in (2.20). We do not provide a proof for this result, as one can easily verify that $\boldsymbol{V}\boldsymbol{V}^{-1} = \boldsymbol{V}^{-1}\boldsymbol{V} = \boldsymbol{I}$ using the expression for $\boldsymbol{V}^{-1}$ given below.

Lemma 4.6.11 *Let* $\boldsymbol{V} = \boldsymbol{\Sigma}^{(0)} + \int_0^1 \boldsymbol{g}(z)f(z)dz \cdot \int_0^1 \boldsymbol{g}(z)^T f(z)dz$ *be the* $(p+1) \times (p+1)$ *matrix introduced in* (2.20) *and set* $\boldsymbol{a} = (\int_0^1 g_1(z)f(z)dz, \ldots, \int_0^1 g_p(z)f(z)dz)^T$. *Also, let* $\boldsymbol{\Sigma} = (\Sigma_{i,j})$ *be the variance-covariance matrix introduced in condition* (A0)-(ii). *Then* $\boldsymbol{V}^{-1}$ *exists and is given by:*

$$\boldsymbol{V}^{-1} = \left(\begin{array}{c:c} 1 + \boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & -\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1} \\ \hdashline -\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & \boldsymbol{\Sigma}^{-1} \end{array}\right) \tag{4.68}$$

provided $\boldsymbol{\Sigma}^{-1}$ *exists.*

The last two lemmas in this Appendix provide several useful asymptotic bounds.

Lemma 4.6.12 *Suppose the assumptions in Theorem 4.2.1 hold. Then:*

$$\frac{1}{(n+1)^2}\boldsymbol{V}^{-1}\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}\boldsymbol{V}^{-1} = \boldsymbol{O}(n^{-1}). \tag{4.69}$$

Proof:

Since the elements of the $(p+1) \times (p+1)$ matrix $\boldsymbol{V}^{-1}$ do not depend upon n, it suffices to show that $\boldsymbol{G}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}/(n+1)^2$ is $\mathcal{O}(n^{-1})$. It is enough to show that $\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{j+1}/(n+1)^2$ is $\mathcal{O}(n^{-1})$ for any $i, j = 0, 1, \ldots, p$.

Let $i, j = 0, 1, \ldots, p$ be fixed. Using vector and matrix norm properties introduced in Section 2.4, we obtain:

$$\begin{aligned}&\left|\left|\frac{1}{(n+1)^2}\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{j+1}\right|\right|_2 \leq \\ &\frac{1}{(n+1)^2}||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}||_2 \cdot ||\boldsymbol{\Psi}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{j+1}||_2 \leq \\ &\frac{1}{(n+1)^2}\mathcal{O}(n^{1/2}) \cdot \mathcal{O}(1) \cdot \mathcal{O}(n^{1/2}) = \mathcal{O}(n^{-1})\end{aligned}$$

since $||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{i+1}||_2 = \mathcal{O}(n^{1/2}) = ||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{j+1}||_2$ by result (4.53) of Lemma 4.6.7 with $\boldsymbol{r} = \boldsymbol{G}_{i+1}$ and $\boldsymbol{r} = \boldsymbol{G}_{j+1}$, respectively, and $||\boldsymbol{\Psi}||_S = \mathcal{O}(1)$ by condition (A1)-(ii). Thus, $\boldsymbol{G}_{i+1}^T(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\boldsymbol{G}_{j+1}/(n+1)^2$ is $\mathcal{O}(n^{-1})$.

Lemma 4.6.13 *Suppose the assumptions in Theorems 4.1.1 and 4.2.1 hold. Let* $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ *be the local linear backfitting estimator of* $m(Z)$ *defined in* (4.32), *where* $Z \in [0,1]$ *is fixed. Also, let* $\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ *denote the local linear backfitting estimator of* $m(Z)$ *that would be obtained if* $\boldsymbol{\beta}$ *were known precisely:*

$$\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z) \equiv \frac{\sum_{j=1}^n w_j^{(Z)}\left[\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right]_j}{\sum_{j=1}^n w_j^{(Z)}}, \tag{4.70}$$

where the $w_j^{(Z)}$*'s are as in* (4.33). *Then, if* $n \to \infty$, $h \to 0$ *and* $nh^3 \to \infty$, *we have:*

$$E(\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) - m(Z) = \mathcal{O}(h^2), \tag{4.71}$$

$$Var(\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) = \mathcal{O}\left(\frac{1}{nh}\right), \tag{4.72}$$

and

$$E(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) - m(Z) = \mathcal{O}(h^2), \tag{4.73}$$

$$Var(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) = \mathcal{O}\left(\frac{1}{nh}\right). \tag{4.74}$$

Proof:

The proof of (4.71) and (4.72) can be found in Francisco-Fernández and Vilar-Fernández (2001), so we omit it.

To prove (4.73), use the definitions of $\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ and $\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)$ in (4.32) and (4.70) to write:

$$\begin{aligned}\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z) &= \frac{\sum_{j=1}^n w_j^{(Z)}\,[\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}]_j}{\sum_{j=1}^n w_j^{(Z)}} - \frac{\sum_{j=1}^n w_j^{(Z)}\left[\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}-\boldsymbol{X}\boldsymbol{\beta}\right]_j}{\sum_{j=1}^n w_j^{(Z)}}\\ &\equiv \widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z) - \boldsymbol{w}^T\boldsymbol{X}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}-\boldsymbol{\beta}).\end{aligned}$$

Thus:

$$\begin{aligned}E(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) - m(Z) &= \{E(\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) - m(Z)\}\\ &\quad - \boldsymbol{w}^T\boldsymbol{X}\{E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{\beta}\}\end{aligned} \tag{4.75}$$

and

$$\begin{aligned}Var(\widehat{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) &= Var(\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z}) - 2\boldsymbol{w}^T\boldsymbol{X}\cdot Cov(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c},\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z})\\ &\quad + \boldsymbol{w}^T\boldsymbol{X}\cdot Var(\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z})\cdot\boldsymbol{X}^T\boldsymbol{w}.\end{aligned} \tag{4.76}$$

Result (4.73) follows by combining (4.75) and (4.71) and using that $Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ is $\mathcal{O}_P(h^2)$ by Theorem 4.1.1 and $\boldsymbol{w}^T\boldsymbol{X}$ is $\mathcal{O}(1)$. The latter result is easy to establish using the fact that the $w_j^{(Z)}$'s are bounded by Lemma 4.6.5. Result (4.74) follows by combining (4.76) and (4.72) and using that $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ is $\mathcal{O}_P(1/n)$ by Theorem 4.2.1, $Cov(\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c},\tilde{m}_{\boldsymbol{I},\boldsymbol{S}_h^c}(Z)|\boldsymbol{X},\boldsymbol{Z})$ is $\mathcal{O}_P(1/(nh))$ by a Cauchy-Schwartz argument and $\boldsymbol{w}^T\boldsymbol{X}$ is $\mathcal{O}(1)$.

Chapter 5

Asymptotic Properties of the Modified and Estimated Modified Local Linear Backfitting Estimators, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$

In this chapter, we investigate the asymptotic behavior of the modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ of $\boldsymbol{\beta}$, with $\boldsymbol{\Psi}$ being the true correlation matrix of the model errors. Recall that an explicit expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ can be obtained from (3.4) by taking $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ and replacing $\mathbb{S}_h^c$ with the centered local linear smoother $\boldsymbol{S}_h^c$:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} = \left(\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{Y}. \tag{5.1}$$

To simplify the proofs of the asymptotic results derived in this chapter, we consider that the model errors satisfy assumption (A2), that is, they are consecutive realizations from a stationary AR process of finite order R. Assumption (A2) is a special case of the assumption (A1) considered in Chapter 4.

The structure of this chapter is similar to that of Chapter 4, where we studied the asymptotic behaviour of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. In the first part of the chapter, we study the asymptotic behaviour of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$. The proofs of the asymptotic results concerning $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ are however more complicated than those concerning $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ for the following reason: the exact conditional bias and variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$ depend on $\boldsymbol{\Psi}^{-1}$ whereas the exact conditional bias and variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$ do not depend on $\boldsymbol{\Psi}^{-1}$. Next, we mention how the asymptotic results concerning the modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ can be generalized to local polynomials of higher degree. We then provide sufficient conditions for the estimators $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ to be asymptotically 'close'. The chapter concludes with an Appendix containing several auxiliary results.

5.1 Exact Conditional Bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$

Just like the usual local linear backfitting estimate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, the modified local linear backfitting estimate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ suffers from finite sample bias. Indeed, using the explicit expression of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ given in equation (5.1), we obtain the exact conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$ as:

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{m}, \tag{5.2}$$

an expression which generally does not equal zero.

Theorem 5.1.1 below provides an asymptotic expression for the conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$, given $\boldsymbol{X}$ and $\boldsymbol{Z}$. These derivations assume that the value of h in $\boldsymbol{S}_h^c$ is deterministic and satisfies conditions (2.12)-(2.13).

Theorem 5.1.1 *Let $\boldsymbol{V}_{\boldsymbol{\Psi}}$ and $\boldsymbol{W}$ be defined as in equations* (2.21) - (2.22). *Under conditions* (A0) *and* (A2)-(A5), *if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, the conditional bias of the modified local linear backfitting estimate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ of $\boldsymbol{\beta}$, given $\boldsymbol{X}$ and $\boldsymbol{Z}$, is:*

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} = -h^2\frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1-\sum_{k=1}^{R}\phi_k\right)^2\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{W} + \boldsymbol{o}_P(h^2). \tag{5.3}$$

Comment 5.1.1 Aneiros Pérez and Quintela del Río (2001a) investigated the large sample properties of an estimator similar to $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$, namely $\widehat{\boldsymbol{\beta}}_{(\boldsymbol{I}-\boldsymbol{K}_h)^T\boldsymbol{\Psi}^{-1},\,\boldsymbol{K}_h}$, the unconstrained modified Speckman estimator in (3.12). Under similar assumptions as ours, Aneiros Pérez and Quintela del Río obtained a faster rate for the asymptotic conditional bias of their estimator, namely $\mathcal{O}_P(h^4)$. As seen in (5.3), the rate we obtained for the asymptotic conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ is $\mathcal{O}_P(h^2)$. However, they did not provide asymptotic constants for this bias, like we do in (5.3). They obtained the same rate of convergence for the asymptotic conditional variance of their estimator as we did for that of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$, namely $\mathcal{O}_P(1/n)$. Just like us, they do provide an asymptotic constant for this variance.

Proof of Theorem 5.1.1:

Let:

$$\boldsymbol{B}_{n,\boldsymbol{\Psi}} = \frac{1}{n+1}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}, \tag{5.4}$$

where the dependence of $\boldsymbol{B}_{n,\boldsymbol{\Psi}}$ upon h is omitted for convenience. We will see below that when $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, $\boldsymbol{B}_{n,\boldsymbol{\Psi}}$ converges in probability to the quantity $\boldsymbol{V}_{\boldsymbol{\Psi}}$ defined in equation (2.22). Since $\boldsymbol{V}_{\boldsymbol{\Psi}}$ is non-singular by Lemma 5.7.6, the explicit expression for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ in (5.1) holds on a set whose measure goes to 1 as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. We can use this expression to write:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} = \boldsymbol{B}_{n,\boldsymbol{\Psi}}^{-1} \cdot \left\{\frac{1}{n+1}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{Y}\right\}, \tag{5.5}$$

which holds on a set whose measure goes to 1 as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. Taking conditional expectation in both sides of (5.5) and subtracting $\boldsymbol{\beta}$ yields:

$$E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) - \boldsymbol{\beta} = \boldsymbol{B}_{n,\boldsymbol{\Psi}}^{-1} \cdot \left\{\frac{1}{n+1}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{m}\right\}. \tag{5.6}$$

We now show that $B_{n,\boldsymbol{\Psi}}$ converges in probability to $\boldsymbol{V}_{\boldsymbol{\Psi}}$ as $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, that is:

$$B_{n,\boldsymbol{\Psi}} = \boldsymbol{V}_{\boldsymbol{\Psi}} + \boldsymbol{o}_P(1). \tag{5.7}$$

Using the fact that $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$ (equation (2.16)), $\boldsymbol{B}_{n,\Psi}$ can be decomposed as:

$$\begin{aligned}\boldsymbol{B}_{n,\Psi} &= \frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta} \\ &+ \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta}. \end{aligned} \tag{5.8}$$

From equation (3.1) with $\mathbb{S}_h = \boldsymbol{S}_h$, $\boldsymbol{S}_h^c = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{S}_h$, so re-writing the first term, expanding the last term and re-arranging yields:

$$\begin{aligned}B_{n,\Psi} &= \frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{\eta}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta} + \frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{G} \\ &+ \frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{G} + \frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta} \\ &+ \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G} - \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c\boldsymbol{\eta}. \end{aligned} \tag{5.9}$$

To establish (5.7), it suffices to show that

$$\frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{G} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1-\sum_{k=1}^R\phi_k\right)^2\int_0^1 \boldsymbol{g}(z)f(z)dz\int_0^1 \boldsymbol{g}(z)^Tf(z)dz + \boldsymbol{o}(1), \tag{5.10}$$

$$\frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{\Psi}^{-1}\boldsymbol{\eta} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1+\sum_{k=1}^R\phi_k^2\right)\boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1), \tag{5.11}$$

while the remaining terms are $\boldsymbol{o}_P(1)$.

The proof of (5.10) is immediate by writing

$$\frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{G} = \left(\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\right)\cdot\left(\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{1}\right)^T\cdot\left(1+\frac{1}{n}\right).$$

and using Lemma 5.7.3 in the Appendix of this chapter and result (4.10).

Result (5.11) is proven in Lemma 5.7.4.

To prove the remaining terms in (5.9) are $\boldsymbol{o}_P(1)$, it suffices to show that the quantities $\boldsymbol{G}_{i+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n+1)$, $\boldsymbol{G}_{i+1}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{G}_{j+1}/n(n+1)$, $\boldsymbol{G}_{i+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\eta}_{j+1}/(n+1)$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{G}_{j+1}/(n+1)$ and $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c\boldsymbol{\eta}_{j+1}/(n+1)$ are $o_P(1)$.

These facts follow from lemmas appearing in the Appendices of this and the preceding chapter.

First consider $\boldsymbol{G}_{i+1}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{G}_{j+1}/(n+1)$. By result (4.58) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} \equiv \Psi^{-1}$ and $\boldsymbol{r} \equiv \boldsymbol{G}_{j+1}$, this quantity is $\mathcal{O}(h^2) = o(1)$. Similarly, from result (4.59) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} = \Psi^{-1}$ and $\boldsymbol{r} \equiv \boldsymbol{G}_{j+1}$, we have that $\boldsymbol{G}_{i+1}^T \Psi^{-1} \mathbf{1}\mathbf{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{G}_{j+1}/n(n+1)$ is $\mathcal{O}(h^2) = o(1)$.

By result (4.60) of Lemma 4.6.9 with $\boldsymbol{r}^* \equiv \boldsymbol{G}_{i+1}$, $\boldsymbol{\Omega} \equiv \Psi^{-1}$ and $\boldsymbol{\xi} \equiv \boldsymbol{\eta}_{j+1}$, we have that $\boldsymbol{G}_{i+1}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{\eta}_{j+1}/(n+1)$ is $O_p(n^{-1/2}h^{-1/2}) = o_P(1)$. Using a similar reasoning with (4.61) of Lemma 4.6.9, we obtain that $\boldsymbol{\eta}_{i+1}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{G}_{j+1}/(n+1)$ is also $o_P(1)$.

Finally, consider $\boldsymbol{\eta}_{i+1}^T \Psi^{-1} \boldsymbol{S}_h^c \boldsymbol{\eta}_{j+1}/(n+1)$. By result (4.62) of Lemma 4.6.9 with $\boldsymbol{\xi}^* \equiv \boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega} \equiv \Psi^{-1}$ and $\boldsymbol{\xi} \equiv \boldsymbol{\eta}_{j+1}$, this quantity is $\mathcal{O}_P(n^{-1/2}h^{-1/2}) = o_P(1)$. This concludes our proof of (5.7).

By Lemma 5.7.6 in the Appendix of this chapter, the matrix $\boldsymbol{V}_\Psi$ on the right side of (5.7) is non-singular and admits an inverse $\boldsymbol{V}_\Psi^{-1}$, so (5.7) leads to:

$$B_{n,\Psi}^{-1} = \boldsymbol{V}_\Psi^{-1} + \boldsymbol{o}_P(1). \tag{5.12}$$

To prove the theorem, by (5.6) and (5.12), it suffices to show that:

$$\frac{1}{n+1}\boldsymbol{X}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} = -h^2 \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1 - \sum_{k=1}^{R} \phi_k\right)^2 \boldsymbol{W} + \boldsymbol{o}_P(h^2). \tag{5.13}$$

From equation (2.16), $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$, so:

$$\frac{1}{n+1}\boldsymbol{X}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} = \frac{1}{n+1}\boldsymbol{G}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} + \frac{1}{n+1}\boldsymbol{\eta}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m}.$$

Using the identifiability condition on $\boldsymbol{m}$ in (2.4) and $\boldsymbol{S}_h^c = (\boldsymbol{I} - \mathbf{1}\mathbf{1}^T/n)\boldsymbol{S}_h$, we obtain:

$$\begin{aligned}\frac{1}{n+1}\boldsymbol{X}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{m} &= \frac{1}{n+1}\boldsymbol{G}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} + \frac{1}{n(n+1)}\boldsymbol{G}^T \Psi^{-1}\mathbf{1}\mathbf{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} \\ &+ \frac{1}{n+1}\boldsymbol{\eta}^T \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} + \frac{1}{n(n+1)}\boldsymbol{\eta}^T \Psi^{-1}\mathbf{1}\mathbf{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}\end{aligned} \tag{5.14}$$

By Lemma 5.7.5, the first two terms on the right side of (5.14) are equal to the right side of (5.13).

Now, consider $\boldsymbol{\eta}_{i+1}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}/(n+1)$, the $(i+1)^{\text{th}}$ element of the third term in (5.14). Using result (4.42) of Lemma 4.6.4 with $\boldsymbol{c} \equiv \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}$ and $\boldsymbol{\xi} = \boldsymbol{\eta}_{i+1}$, together with spectral norm properties introduced in Section 2.4, we obtain:

$$\begin{aligned}
&\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m} = \frac{1}{n+1}\mathcal{O}_P(||\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_2)\\
&= \frac{1}{n+1}\mathcal{O}_P(||\Psi^{-1}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_S) = \frac{1}{n+1}\mathcal{O}_P(||\Psi^{-1}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_2) = o_p(h^2).
\end{aligned}$$

The last equality was obtained by using that $||\Psi||^{-1}$ is bounded (result (5.35) of Lemma 5.7.2) and $||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_2 = \mathcal{O}(n^{1/2}h^2)$ by result (4.39) of Lemma 4.6.3 with $\boldsymbol{r} \equiv \boldsymbol{m}$.

Finally, consider $\boldsymbol{\eta}_{i+1}^T\Psi^{-1}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}/n(n+1)$, the $(i+1)^{\text{th}}$ element of the fourth term in (5.14). Using a similar reasoning as above, we obtain:

$$\begin{aligned}
\frac{1}{n(n+1)}\boldsymbol{\eta}_{i+1}^T\Psi^{-1}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} &= \frac{1}{n(n+1)}\mathcal{O}_P(||\Psi^{-1}\boldsymbol{1}\boldsymbol{1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}||_2)\\
&= \frac{1}{n(n+1)}\mathcal{O}_P(||\Psi^{-1}||_S \cdot ||\boldsymbol{1}\boldsymbol{1}^T||_F \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_S)\\
&= \frac{1}{n+1}\mathcal{O}_P(||\Psi^{-1}||_S \cdot ||(\boldsymbol{I} - \boldsymbol{S}_h)\boldsymbol{m}||_2) = o_p(h^2).
\end{aligned}$$

This proves (5.13) and completes our proof of Theorem 5.1.1.

5.2 Exact Conditional Variance of $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},S_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$

In this section, we derive an asymptotic expression for the exact conditional variance of $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},S_h^c}$, given $\boldsymbol{X}$,$\boldsymbol{Z}$:

$$Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},S_h^c}|\boldsymbol{X},\boldsymbol{Z}) = \sigma_\epsilon^2 B_{n,\Psi}^{-1} \cdot \boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{X} \cdot B_{n,\Psi}^{-1} \tag{5.15}$$

where $B_{n,\Psi}$ is defined as in (5.4). The above equality was obtained by using the explicit formula of $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}$ in (5.1), together with the fact that $Var(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z}) = \sigma_\epsilon^2\Psi$ by condition (A2).

Theorem 5.2.1 *Under conditions* (A0) *and* (A2)-(A5), *if* $n \to \infty$, $h \to 0$ *and* $nh^3 \to \infty$, *the conditional variance of the modified local linear backfitting estimator* $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}$ *of* $\boldsymbol{\beta}$, *given* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$, *is:*

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) = {} & \frac{1}{n+1}\cdot\frac{\sigma_\epsilon^4}{\sigma_u^2}\left(1+\sum_{k=1}^{R}\phi_k^2\right)\boldsymbol{V}_\Psi^{-1}\Sigma^{(0)}\boldsymbol{V}_\Psi^{-1} \\ & + \frac{\sigma_\epsilon^2}{(n+1)^2}\boldsymbol{V}_\Psi^{-1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\Psi(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{G}\boldsymbol{V}_\Psi^{-1} + o_P\left(\frac{1}{n}\right). \end{aligned} \tag{5.16}$$

Comment 5.2.1 By Lemma 5.7.7 in the Appendix of this chapter, the second term in the above asymptotic expression for $Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ is $\mathcal{O}_P(n^{-1})$ and hence it does not dominate the first term, which is $\mathcal{O}_P(n^{-1})$.

Proof of Theorem 5.2.1:

By (5.15), we have:

$$Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z}) \equiv \frac{\sigma_\epsilon^2}{n+1}B_{n,\Psi}^{-1}\cdot C_{n,\Psi}\cdot B_{n,\Psi}^{-1}, \tag{5.17}$$

where $C_{n,\Psi} = \boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\Psi(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{X}/(n+1)$. Since $B_{n,\Psi}^{-1} \xrightarrow[n\to\infty]{P} \boldsymbol{V}_\Psi^{-1}$ by result (5.12), to prove the theorem it suffices to show that:

$$C_{n,\Psi} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left(1+\sum_{k=1}^{R}\phi_k^2\right)\Sigma^{(0)} + \frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\Psi(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{G} + o_P(1). \tag{5.18}$$

This fact is shown below with the help of lemmas in the Appendix of this and the preceding chapter.

By (2.16), $\boldsymbol{X} = \boldsymbol{G} + \boldsymbol{\eta}$, so $C_{n,\Psi}$ can be decomposed as:

$$\begin{aligned} C_{n,\Psi} = & \frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{G} \\ & + \frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta} \\ & + \left[\frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta}\right]^T \\ & + \frac{1}{n+1}\boldsymbol{\eta}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta}. \end{aligned}$$

Expanding the last term and re-arranging yields:

$$\begin{aligned} C_{n,\Psi} = & \frac{1}{n+1}\boldsymbol{\eta}^T\Psi^{-1}\boldsymbol{\eta} + \frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{G} \\ & + \frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta} \\ & + \left[\frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta}\right]^T - \frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{S}_h^{cT}\Psi^{-1}\boldsymbol{\eta} \\ & - \left[\frac{1}{n+1}\boldsymbol{\eta}^T\boldsymbol{S}_h^{cT}\Psi^{-1}\boldsymbol{\eta}\right]^T + \frac{1}{n+1}\boldsymbol{\eta}^T\Psi^{-1}\boldsymbol{S}_h^c\Psi\boldsymbol{S}_h^{cT}\Psi^{-1}\boldsymbol{\eta}. \end{aligned} \tag{5.19}$$

The first term in the above converges to the first term on the right side of (5.18) by Lemma 5.7.4. The second term in the above is the same as the second term on the right side of (5.18). To show the remaining terms are $\boldsymbol{o}_P(1)$, it suffices to establish that $\boldsymbol{G}_{i+1}^T\Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{\eta}_{j+1}/(n+1)$, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^{cT}\Psi^{-1}\boldsymbol{\eta}_{j+1}/(n+1)$ and $\boldsymbol{\eta}_{i+1}^T\Psi^{-1}\boldsymbol{S}_h^c\Psi\boldsymbol{S}_h^{cT}\Psi^{-1}\boldsymbol{\eta}_{j+1}/(n+1)$ are $o_P(1)$ for all $i, j = 0, 1, \ldots, p$.

Let $i, j = 0, 1, \ldots, p$ be fixed. From result (4.42) of Lemma 4.6.4 with $\boldsymbol{c} \equiv \Psi^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\Psi(\boldsymbol{I} - \boldsymbol{S}_h^c)^T\Psi^{-1}\boldsymbol{G}_{i+1}$ and $\boldsymbol{\xi} = \boldsymbol{\eta}_{j+1}$ and from the spectral norm properties introduced in Section 2.4 of Chapter 2, we get:

$$\begin{aligned}
&\frac{1}{n+1}\boldsymbol{\eta}_{j+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{i+1} \\
&\quad = \frac{1}{n+1}\mathcal{O}_P(||\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S})\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{i+1}||_2) \\
&\quad \leq \frac{1}{n+1}\mathcal{O}_P(||\boldsymbol{\Psi}^{-1}||_S^2\cdot(1+||\boldsymbol{S}_h^c||_F)^2\cdot||\boldsymbol{\Psi}||_S\cdot||\boldsymbol{G}_{i+1}||_2) \\
&\quad = \mathcal{O}_P(n^{-1/2}h^{-1}) = o_P(1).
\end{aligned}$$

To derive the above result, we used Lemma 5.7.2 to obtain that $||\boldsymbol{\Psi}||_S$ and $||\boldsymbol{\Psi}^{-1}||_S$ are $\mathcal{O}(1)$. We also used the fact that $||\boldsymbol{S}_h^c||_F$ is $\mathcal{O}(h^{-1/2})$ by result (4.54) of Lemma 4.6.7, while $\boldsymbol{G}_{i+1}$ is $\mathcal{O}(n^{1/2})$ (take $\boldsymbol{r}\equiv\boldsymbol{G}_{i+1}$ in result (4.51) of Lemma 4.6.7).

Next, consider $\boldsymbol{\eta}_{i+1}^T\boldsymbol{S}_h^{cT}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_{j+1}/(n+1) = \boldsymbol{\eta}_{j+1}^T\boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c\boldsymbol{\eta}_{i+1}/(n+1)$. This quantity is $\mathcal{O}_P(n^{-1/2}h^{-1/2}) = o_P(1)$ by result (4.62) of Lemma 4.6.9 with $\boldsymbol{\xi}^*\equiv\boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega}=\boldsymbol{\Psi}^{-1}$ and $\boldsymbol{\xi}\equiv\boldsymbol{\eta}_{j+1}$. Finally, $\boldsymbol{\eta}_{i+1}^T\boldsymbol{\Psi}^{-1}\boldsymbol{S}_h^c\boldsymbol{\Psi}\boldsymbol{S}_h^{cT}\boldsymbol{\Psi}^{-1}\boldsymbol{\eta}_{j+1}/(n+1)$ is $O_P(n^{-1}h^{-1}) = o_P(1)$ by result (4.64) of Lemma 4.6.9 with $\boldsymbol{\xi}^*=\boldsymbol{\eta}_{i+1}$, $\boldsymbol{\Omega}=\boldsymbol{\Psi}^{-1}$, $\boldsymbol{\Omega}^*=\boldsymbol{\Psi}$ and $\boldsymbol{\xi}=\boldsymbol{\eta}_{j+1}$.

5.3 Exact Conditional Measure of Accuracy of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ Given $\boldsymbol{X}$ and $\boldsymbol{Z}$

Any suitable criterion for measuring the accuracy of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ should take into account both bias and variance effects. We use the following measure of accuracy for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$, which combines in a natural fashion these effects:

$$\begin{aligned}
E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}-\boldsymbol{\beta}||_2^2|\boldsymbol{X},\boldsymbol{Z}\right) &= \left\{E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{\beta}\right\}^T\left\{E(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{\beta}\right\} \\
&\quad + trace\left\{Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})\right\}. \qquad (5.20)
\end{aligned}$$

Using equation (5.20) above together with Theorem 5.1.1 and Theorem 5.2.1 we obtain the following result:

Corollary 5.3.1 *Assume that the conditions in Theorem 5.1.1 and Theorem 5.2.1 hold. Then, when $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, we have:*

$$
\begin{aligned}
E\left(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2 | \boldsymbol{X}, \boldsymbol{Z}\right) &= h^4 \cdot \frac{\sigma_\epsilon^4}{\sigma_u^4}\left(1 - \sum_{k=1}^R \phi_k\right)^4 \boldsymbol{W}^T \boldsymbol{V}_{\boldsymbol{\Psi}}^{-2} \boldsymbol{W} \\
&+ \frac{1}{n+1} \cdot \frac{\sigma_\epsilon^4}{\sigma_u^2} \cdot \left(1 + \sum_{k=1}^R \phi_k^2\right) trace\left\{\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} \boldsymbol{\Sigma}^{(0)} \boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\right\} \\
&+ \frac{\sigma_\epsilon^2}{(n+1)^2} trace\left\{\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} \boldsymbol{G}^T \boldsymbol{\Psi}^{-1} (\boldsymbol{I} - \boldsymbol{S}_h^c) \boldsymbol{\Psi} (\boldsymbol{I} - \boldsymbol{S}_h^c)^T \boldsymbol{\Psi}^{-1} \boldsymbol{G} \boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\right\} \\
&+ o_P(h^4) + o_P\left(\frac{1}{n}\right). \qquad (5.21)
\end{aligned}
$$

5.4 The $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$

Just as with the usual backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, we would like the modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ to be $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$, that is, we would like $E(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X}, \boldsymbol{Z})$ to be $\mathcal{O}_P(n^{-1})$.

By result (5.21) of Lemma 5.3.1, $E(||\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}||_2^2|\boldsymbol{X}, \boldsymbol{Z})$ is $\mathcal{O}_P(h^4) + \mathcal{O}_P(n^{-1})$. This result is due to the fact that the conditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ is $\boldsymbol{\mathcal{O}}_P(n^{-1})$ but its conditional bias is $\boldsymbol{\mathcal{O}}_P(h^2)$.

We are interested in assessing at what rate the smoothing parameter h should converge to zero so that the squared conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ tends to zero, but has the same order of magnitude as the conditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$. A similar argument as that employed in Section 4.4 yields that h should converge to zero at rate $n^{-\alpha}$, $\alpha \in [1/4, 1/3)$, to ensure that the modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$ - exactly as for the usual local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. Note that $n^{-\alpha} < n^{-1/5}$, so we must 'undersmooth' $\widehat{\boldsymbol{m}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ to achieve $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$. Here, $n^{-1/5}$ is the 'usual' rate of convergence for h, which we believe is optimal for estimating m via $\widehat{\boldsymbol{m}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$.

5.5 Generalization to Local Polynomials of Higher Degree

The asymptotic results in Sections 5.1-5.4 concern the modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$. We believe these results readily generalize to the modified local polynomial backfitting estimator of $\boldsymbol{\beta}$. The latter estimator is obtained from (5.1) by replacing $\boldsymbol{S}_h^c$, the smoother matrix for locally linear regression, with the smoother matrix for locally polynomial regression of degree $D > 1$.

In keeping with the locally polynomial regression literature, we conjecture that the modified local polynomial backfitting estimator of $\boldsymbol{\beta}$ has conditional bias of order $\boldsymbol{O}_P(h^{D+1})$ and conditional variance of order $\boldsymbol{O}_P(n^{-1})$. Note that we may need boundary corrections if D is even. We also conjecture that h should converge to zero at rate $n^{-\alpha}$, $\alpha \in [1/(2D+2), 1/3)$, for the modified local polynomial backfitting estimator of $\boldsymbol{\beta}$ to be $\sqrt{n}$-consistent given $\boldsymbol{X}$ and $\boldsymbol{Z}$.

5.6 The $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$

The estimated modified local linear backfitting estimator $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ can be obtained from (5.1) by replacing $\boldsymbol{\Psi}$ with an estimator $\widehat{\boldsymbol{\Psi}}$:

$$\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c} = \left(\boldsymbol{X}^T\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{Y}. \tag{5.22}$$

Deriving asymptotic approximations for the exact conditional bias and variance of $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$ is not possible, as these quantities are not tractable. The reason for this is that $\widehat{\boldsymbol{\Psi}}$ is random since it is computed from the data. In this section, we give sufficient conditions for $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ to be asymptotically 'close', in the sense that the difference between these estimators is $\boldsymbol{o}_P(n^{-1/2})$. Our conditions (5.23) and (5.24) are similar to those imposed by Aneiros Pérez and Quintela del Río (2001a) for establishing the asymptotic equivalence of their modified and estimated modified versions of the Speckman estimator.

Theorem 5.6.1 *Suppose that the conditions in Theorems 5.1.1 and 5.2.1 hold. In addition, suppose that:*

$$\frac{1}{n}\boldsymbol{X}^T(\widehat{\boldsymbol{\Psi}}^{-1} - \boldsymbol{\Psi}^{-1})(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X} = \boldsymbol{o}_P(1) \tag{5.23}$$

$$\frac{1}{\sqrt{n}}\boldsymbol{X}^T(\widehat{\boldsymbol{\Psi}}^{-1} - \boldsymbol{\Psi}^{-1})(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) = \boldsymbol{o}_P(1) \tag{5.24}$$

Then, if $h = n^{-\alpha}$, $\alpha \in [1/4, 1/3)$, we have:

$$\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c} = \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} + \boldsymbol{o}_P\left(\frac{1}{\sqrt{n}}\right). \tag{5.25}$$

Proof:

To establish (5.25), it suffices to show:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}) = \sqrt{n}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}) + \boldsymbol{o}_P(1). \tag{5.26}$$

Using the expression for $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ in (5.22) and $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{m} + \boldsymbol{\epsilon}$ (equation (2.1)), we write the left side of (5.26) as:

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c} - \boldsymbol{\beta}\right) &= \left(\frac{1}{n}\boldsymbol{X}^T\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} \cdot \frac{1}{\sqrt{n}}\boldsymbol{X}^T\widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) \\
&= \left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X} + \boldsymbol{o}_P(1)\right)^{-1} \cdot \left(\frac{1}{\sqrt{n}}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) + \boldsymbol{o}_P(1)\right) \\
&= \left[\left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} + \boldsymbol{o}_P(1)\right] \cdot \left(\frac{1}{\sqrt{n}}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) + \boldsymbol{o}_P(1)\right) \\
&= \left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} \cdot \frac{1}{\sqrt{n}}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) \\
&\quad + \left(\frac{1}{n}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1} \cdot \boldsymbol{o}_P(1) + \frac{1}{\sqrt{n}}\boldsymbol{X}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I} - \boldsymbol{S}_h^c)(\boldsymbol{m} + \boldsymbol{\epsilon}) \cdot \boldsymbol{o}_P(1) + \boldsymbol{o}_P(1).
\end{aligned}$$

By the definition of $\widehat{\boldsymbol{\beta}}_{\Psi,\boldsymbol{S}_h^c}$ in (5.1) we have:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\widehat{\Psi}^{-1},\boldsymbol{S}_h^c}-\boldsymbol{\beta}\right)=\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}-\boldsymbol{\beta}\right)+\left(\frac{1}{n}\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}\right)^{-1}\cdot \boldsymbol{o}_P(1)$$
$$+\frac{1}{\sqrt{n}}\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{m}+\boldsymbol{\epsilon})\cdot \boldsymbol{o}_P(1)+\boldsymbol{o}_P(1).$$

Therefore, to prove (5.26), it is enough to show that $\left(\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}/n\right)^{-1}$ and $\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{m}+\boldsymbol{\epsilon})/\sqrt{n}$ are $\boldsymbol{O}_P(1)$.

To prove the first fact, let $B_{n,\Psi}=\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}/(n+1)$. By (5.12), $B_{n,\Psi}^{-1}=\boldsymbol{V}_\Psi^{-1}+\boldsymbol{o}_P(1)$, with $\boldsymbol{V}_\Psi$ as in (2.22), so $\left(\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{X}/n\right)^{-1}=\boldsymbol{O}_P(1)$. To prove the second fact, use $B_{n,\Psi}=\boldsymbol{V}_\Psi+\boldsymbol{o}_P(1)$ (result (5.7)) and Chebychev's Theorem to write:

$$\frac{1}{\sqrt{n}}\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{m}+\boldsymbol{\epsilon})=\frac{1}{\sqrt{n}}\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})$$
$$=\frac{n+1}{\sqrt{n}}\cdot B_{n,\Psi}\cdot(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}-\boldsymbol{\beta})$$
$$=\frac{n+1}{\sqrt{n}}\cdot(\boldsymbol{V}_\Psi+\boldsymbol{o}_P(1))\cdot\left\{E(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{\beta}+\boldsymbol{O}_P\left(\sqrt{Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})}\right)\right\}.$$

By result (5.3) of Theorem 5.1.1, $E(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})-\boldsymbol{\beta}$ is $\boldsymbol{O}_P(h^2)=\boldsymbol{O}_P(n^{-2\alpha})$. Also, by result (5.16) of Theorem 5.2.1, $Var(\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}|\boldsymbol{X},\boldsymbol{Z})$ is $\boldsymbol{O}_P(n^{-1})$. Since $\alpha\geq 1/4$, we conclude:

$$\frac{1}{\sqrt{n}}\boldsymbol{X}^T\Psi^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)(\boldsymbol{m}+\boldsymbol{\epsilon})=\boldsymbol{O}_P(\sqrt{n})\cdot\left(\boldsymbol{O}_P(n^{-2\alpha})+\boldsymbol{O}_P\left(\frac{1}{\sqrt{n}}\right)\right)$$
$$=\boldsymbol{O}_P\left(n^{\frac{1-4\alpha}{2}}\right)+\boldsymbol{O}_P(1)=\boldsymbol{O}_P(1).$$

This completes our proof of Theorem 5.6.1.

Theorem 5.6.1 implies that $\widehat{\boldsymbol{\beta}}_{\widehat{\Psi}^{-1},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent since $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}$ is $\sqrt{n}$-consistent. One would expect the conditional bias and variance of $\widehat{\boldsymbol{\beta}}_{\widehat{\Psi}^{-1},\boldsymbol{S}_h^c}$ to be similar to those of $\widehat{\boldsymbol{\beta}}_{\Psi^{-1},\boldsymbol{S}_h^c}$.

5.7 Appendix

Throughout this Appendix, we assume that the assumptions and notation introduced in Sections 2.2 and 2.3 of this thesis hold, unless otherwise specified. We also let $I(\mathcal{S})$ denote the indicator function of an arbitrary set $\mathcal{S}$.

The first lemma in this Appendix shows that the correlation matrix of n consecutive observations arising from a stationary autoregressive process of finite order R is invertible. The lemma also provides an explicit formula for the inverse of this correlation matrix. A proof of this lemma can be found in David and Bastin (2001,Lemma 1).

Lemma 5.7.1 *Let $\epsilon_1, \ldots, \epsilon_n$ be successive observations from an AR process of finite order R satisfying condition* (A2)*. If $\mathbf{\Psi}$ is the correlation matrix of $\epsilon_1, \ldots, \epsilon_n$ defined in Comment 2.2.1, then its inverse exists and is given by:*

$$\mathbf{\Psi}^{-1} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\left[\mathcal{U}^T\mathcal{U} - \mathcal{V}^T\mathcal{V}\right], \tag{5.27}$$

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

$$\mathcal{U} = \begin{pmatrix} 1 & & & & & & \\ -\phi_1 & \cdot & & & & & \\ & \cdot & \cdot & & & & \\ -\phi_R & & \cdot & \cdot & & & \\ 0 & \cdot & & \cdot & \cdot & & \\ & \cdot & \cdot & & \cdot & \cdot & \\ 0 & & 0 & -\phi_R & & -\phi_1 & 1 \end{pmatrix} \quad \text{and } \mathcal{V} = \begin{pmatrix} 0 & & & & & & \\ & \cdot & & & & & \\ 0 & & \cdot & & & & \\ -\phi_R & \cdot & & \cdot & & & \\ & \cdot & \cdot & & \cdot & & \\ & & \cdot & \cdot & & \cdot & \\ -\phi_1 & & & -\phi_R & 0 & & 0 \end{pmatrix}. \tag{5.28}$$

Comment 5.7.1 Let $\mathcal{U}$ be as in (5.28) and define

$$[\boldsymbol{U}_{(k)}]_{i,j} = I(j = i - k, k + 1 \leq i \leq n) \tag{5.29}$$

for $k = 1, \ldots, R$. Then it can be easily seen that

$$\mathcal{U} = \boldsymbol{I} - \phi_1 \boldsymbol{U}_{(1)} - \cdots - -\phi_R \boldsymbol{U}_{(R)}.$$

Straightforward algebraic manipulations also yield

$$\mathcal{U}^T\mathcal{U} = -\sum_{k=1}^{R} \phi_k (\boldsymbol{U}_{(k)}^T + \boldsymbol{U}_{(k)}) + \sum_{\substack{p,\, q=1 \\ p<q}}^{R} \phi_p \phi_q (\boldsymbol{U}_{(p)}^T \boldsymbol{U}_{(q)} + \boldsymbol{U}_{(q)}^T \boldsymbol{U}_{(p)}) + \sum_{k=1}^{R} \phi_k^2 \boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)} + \boldsymbol{I}, \tag{5.30}$$

where

$$\left[\boldsymbol{U}_{(k)}^T\right]_{i,j} = I(j = i + k, 1 \leq i \leq n - k), \tag{5.31}$$

$$\left[\boldsymbol{U}_{(p)}^T \boldsymbol{U}_{(q)}\right]_{i,j} = I(j = i + p - q, 1 - p + q \leq i \leq n - p), \tag{5.32}$$

$$\left[\boldsymbol{U}_{(q)}^T \boldsymbol{U}_{(p)}\right]_{i,j} = I(j = i - p + q, 1 \leq i \leq n - q), \tag{5.33}$$

for $k, p, q = 1, \ldots, R$ and $p \leq q$.

The next lemma shows that, if $\boldsymbol{\Psi}$ is the correlation matrix of a sample of n consecutive observations arising from a stationary AR process of finite order R, then its spectral norm is bounded. Furthermore, the spectral norm of $\boldsymbol{\Psi}^{-1}$ is also bounded.

Lemma 5.7.2 *Let $\epsilon_1, \ldots, \epsilon_n$ be successive observations from an AR process of finite order R satisfying condition* (A2). *If $\boldsymbol{\Psi}$ is the correlation matrix of $\epsilon_1, \ldots, \epsilon_n$ defined in Comment 2.2.1, then:*

$$||\boldsymbol{\Psi}||_S = \mathcal{O}(1) \tag{5.34}$$

and

$$||\boldsymbol{\Psi}^{-1}||_S = \mathcal{O}(1). \tag{5.35}$$

Proof:

The boundedness of $||\Psi^{-1}||_S$ (result (5.35)) follows easily by using the explicit expression for Ψ^{-1} in equation (5.27).

To prove the boundedness of $||\Psi||_S$ (result (5.34)), use the symmetry of Ψ and a well-known result on spectral norms to get:

$$||\Psi||_S \le \left[\max_{1\le i\le n} \sum_{j=1}^{n} |[\Psi]_{i,j}|\right]^2 = \left[\max_{1\le i\le n} \sum_{h=1-i}^{n-i} |\rho_h|\right]^2.$$

According to Exercise 13 in Brockwell and Davis (1991), there exist constants $\mathcal{C} > 0$ and $s \in (0,1)$ so that:

$$|\rho_h| \le \mathcal{C}s^{|h|} \text{ for all } h.$$

Combining the previous results yields:

$$||\Psi||_S \le \left[\sum_{h=1-n}^{n-1} |\rho_h|\right]^2 \le \left(2\sum_{h=0}^{n-1} \mathcal{C}s^h\right)^2 = \left(2\mathcal{C}\frac{1}{1-s}\right)^2$$

and (5.35) follows.

The following lemma provides a useful asymptotic approximation.

Lemma 5.7.3 *Let $\epsilon_1, \ldots, \epsilon_n$ be successive observations from an AR process of finite order R satisfying condition* (A2). *Let Ψ be the correlation matrix of $\epsilon_1, \ldots, \epsilon_n$. Furthermore, let $\boldsymbol{G}$ be an $n \times (p+1)$ matrix defined as in* (2.14). *If $n \to \infty$, then:*

$$\frac{1}{n+1}\boldsymbol{G}^T\Psi^{-1}\mathbf{1} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\left(1 - \sum_{k=1}^{R}\phi_k\right)^2 \int_0^1 \boldsymbol{g}(z)f(z)dz + \boldsymbol{o}(1). \tag{5.36}$$

Proof:

By (5.27), the left side of (5.36) is

$$\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\boldsymbol{1} = \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}\boldsymbol{1} - \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}\boldsymbol{1},$$

so it suffices to show

$$\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}\boldsymbol{1} = \left(1-\sum_{k=1}^{R}\phi_k\right)^2\int_0^1 \boldsymbol{g}(z)f(z)dz + \boldsymbol{o}(1), \tag{5.37}$$

$$\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}\boldsymbol{1} = \boldsymbol{o}(1). \tag{5.38}$$

To establish (5.37), it is enough to show that, for any $i = 0, 1, \ldots, p$, we have:

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}\boldsymbol{1} = \left(1-\sum_{k=1}^{R}\phi_k\right)^2\int_0^1 g_i(z)f(z)dz + o(1).$$

Let $i = 0, 1, \ldots, p$, be fixed. Using the explicit expression for $\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}$ in result (5.30), we write:

$$\begin{aligned}\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}\boldsymbol{1} &= -\frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\sum_{k=1}^{R}\phi_k\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\left(\boldsymbol{U}_{(k)}^T+\boldsymbol{U}_{(k)}\right)\boldsymbol{1}\right] \\ &+ \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\sum_{\substack{p,\,q=1\\p<q}}^{R}\phi_p\phi_q\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\left(\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)}+\boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)}\right)\boldsymbol{1}\right] \\ &+ \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\sum_{k=1}^{R}\phi_k^2\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}^T\boldsymbol{U}_{(k)}\boldsymbol{1}\right] + \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{1}.\end{aligned} \tag{5.39}$$

Therefore, it suffices to prove that the following asymptotic approximations hold:

$$\sum_{k=1}^{R}\phi_k\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\left(\boldsymbol{U}_{(k)}^T+\boldsymbol{U}_{(k)}\right)\boldsymbol{1}\right] = 2\left(\sum_{k=1}^{R}\phi_k\right)\int_0^1 g_i(z)f(z)dz + o(1), \tag{5.40}$$

$$\sum_{\substack{p,\,q=1\\p<q}}^{R}\phi_p\phi_q\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\left(\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)}+\boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)}\right)\boldsymbol{1}\right] = 2\left(\sum_{\substack{p,\,q=1\\p<q}}^{R}\phi_p\phi_q\right)\int_0^1 g_i(z)f(z)dz + o(1), \tag{5.41}$$

$$\sum_{k=1}^{R} \phi_k^2 \left[\frac{1}{n+1} \boldsymbol{G}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)} \mathbf{1} \right] = \left(\sum_{k=1}^{R} \phi_k^2 \right) \int_0^1 g_i(z) f(z) dz + o(1), \tag{5.42}$$

$$\frac{1}{n+1} \boldsymbol{G}_{i+1}^T \mathbf{1} = \int_0^1 g_i(z) f(z) dz + o(1). \tag{5.43}$$

The last result follows from result (4.10).

To prove (5.40), it is enough to show that the equalities below hold for any $k = 1, \ldots, R$:

$$\frac{1}{n+1} \boldsymbol{G}_{i+1}^T \boldsymbol{U}_{(k)}^T \mathbf{1} = \int_0^1 g_i(z) f(z) dz + o(1) \tag{5.44}$$

$$\frac{1}{n+1} \boldsymbol{G}_{i+1}^T \boldsymbol{U}_{(k)} \mathbf{1} = \int_0^1 g_i(z) f(z) dz + o(1). \tag{5.45}$$

Using the expression of $\boldsymbol{U}_{(k)}^T$ in (5.31) and a Riemann integration argument, the left side of (5.44) can be written as:

$$\begin{aligned}
\frac{1}{n+1} \boldsymbol{G}_{i+1}^T \boldsymbol{U}_{(k)}^T \mathbf{1} &= \frac{1}{n+1} \sum_{t=1}^{n} \sum_{l=1}^{n} g_i(Z_t) \left[\boldsymbol{U}_{(k)}^T \right]_{t,l} \mathbf{1}_l \\
&= \frac{1}{n+1} \sum_{t=1}^{n} \sum_{l=1}^{n} g_i(Z_t) I \left(l = t + k, 1 \leq t \leq n - k \right) \\
&= \frac{1}{n+1} \sum_{t=1}^{n-k} g_i(Z_t) = \frac{1}{n+1} \sum_{t=1}^{n} g_i(Z_t) + o(1) \\
&= \int_0^1 g_i(z) f(z) dz + o(1).
\end{aligned}$$

Here, we have also used that k does not depend upon n, as R itself does not depend upon n. Similarly, using the expression for $\boldsymbol{U}_{(k)}$ in (5.29), we obtain that the left side of

(5.45) is:

$$\begin{aligned}\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}\mathbf{1} &= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} g_i(Z_t)\left[\boldsymbol{U}_{(k)}\right]_{t,l}\mathbf{1}_l \\ &= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} g_i(Z_t) I\left(l = t-k, k+1 \le t \le n\right) \\ &= \frac{1}{n+1}\sum_{t=k+1}^{n} g_i(Z_t) = \frac{1}{n+1}\sum_{t=1}^{n} g_i(Z_t) + o(1) \\ &= \int_0^1 g_i(z) f(z) dz + o(1).\end{aligned}$$

Thus, both (5.44) and (5.45) hold.

A similar argument can be used to derive (5.41) and (5.42). The only difference in the proofs is that the range of summation for t in $\sum_t g_i(Z_t)$ changes.

It remains to prove (5.38). To establish this result, it is enough to show that $\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}\mathbf{1}/(n+1)$ is $o_P(1)$ for all $i = 0, 1, \ldots, p$.

Let $i = 0, 1, \ldots, p$, be fixed. By the definition of $\boldsymbol{\mathcal{V}}$ in (5.28), we have:

$$\boldsymbol{\mathcal{V}}\boldsymbol{G}_{i+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\phi_R g_i(Z_1) \\ -\phi_{R-1} g_i(Z_1) - \phi_R g_i(Z_2) \\ \vdots \\ -\phi_1 g_i(Z_1) - \phi_2 g_i(Z_2) - \cdots - \phi_R g_i(Z_R) \end{pmatrix}.$$

Since $g_i(\cdot)$ is bounded by assumption (A0)-(i), $||\boldsymbol{\mathcal{V}}\boldsymbol{G}_{i+1}||_2 = \mathcal{O}(1)$. A similar argument yields $||\boldsymbol{\mathcal{V}}\mathbf{1}||_2 = \mathcal{O}(1)$. Combining these results, we obtain:

$$\left|\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}\mathbf{1}\right| \le \frac{1}{n+1}||\boldsymbol{\mathcal{V}}\boldsymbol{G}_{i+1}||_2 \cdot ||\boldsymbol{\mathcal{V}}\mathbf{1}||_2 = \frac{1}{n+1}\mathcal{O}(1)\cdot\mathcal{O}(1) = \mathcal{O}(1/n) = o(1),$$

so $\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}\mathbf{1}/(n+1)$ is $o(1)$.

The following lemma provides a result concerning the convergence in probability of a random matrix.

Lemma 5.7.4 *Suppose the assumptions in Lemma 5.7.1 hold. Let* $(\eta_{i1}, \ldots, \eta_{ip})^T, i = 1, \ldots, n$, *be as in condition* (A0)-(ii) *and let* $\boldsymbol{\eta}$ *be an* $n \times (p+1)$ *random matrix defined as in* (2.10). *Then, as* $n \to \infty$*:*

$$\frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\Psi}^{-1}\boldsymbol{\eta} = \frac{\sigma_\epsilon^2}{\sigma_u^2} \cdot \left(1 + \sum_{k=1}^{R} \phi_k^2\right) \boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1) \tag{5.46}$$

where $\boldsymbol{\Sigma}^{(0)}$ *is defined as in equation* (2.15).

Proof:

By (5.27), the left side of (5.46) can be written as:

$$\frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\Psi}^{-1}\boldsymbol{\eta} = \frac{\sigma_\epsilon^2}{\sigma_u^2} \cdot \frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}}\boldsymbol{\eta} - \frac{\sigma_\epsilon^2}{\sigma_u^2} \cdot \frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}}\boldsymbol{\eta},$$

so it suffices to show

$$\begin{aligned} \frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}}\boldsymbol{\eta} &= \left(1 + \sum_{k=1}^{R} \phi_k^2\right) \boldsymbol{\Sigma}^{(0)} + \boldsymbol{o}_P(1), \\ \frac{1}{n+1}\boldsymbol{\eta}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}}\boldsymbol{\eta} &= \boldsymbol{o}_P(1). \end{aligned}$$

In fact, if $\boldsymbol{\Sigma}^{(0)} = (\Sigma_{i,j})$, it is enough to show:

$$\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}}\boldsymbol{\eta}_{j+1} = \left(1 + \sum_{k=1}^{R} \phi_k^2\right) \Sigma_{i,j} + o_P(1), \tag{5.47}$$

$$\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{j+1} = o_P(1), \tag{5.48}$$

for any $i, j = 0, 1, \ldots, p$.

Let $i, j = 0, 1, \ldots, p$, be fixed. Using the explicit expression for $\boldsymbol{\Psi}^{-1}$ in (5.27), we write:

$$
\begin{aligned}
\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}} \boldsymbol{\eta}_{j+1} = & -\sum_{k=1}^{R} \phi_k \left[\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \left(\boldsymbol{U}_k^T + \boldsymbol{U}_{(k)}\right) \boldsymbol{\eta}_{j+1}\right] \\
& + \sum_{\substack{p,\, q=1 \\ p<q}}^{R} \phi_p \phi_q \left[\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \left(\boldsymbol{U}_{(p)}^T \boldsymbol{U}_{(q)} + \boldsymbol{U}_{(q)}^T \boldsymbol{U}_{(p)}\right) \boldsymbol{\eta}_{j+1}\right] \\
& + \sum_{k=1}^{R} \phi_k^2 \left[\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)} \boldsymbol{\eta}_{j+1}\right] + \frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\eta}_{j+1} \qquad (5.49)
\end{aligned}
$$

In order to establish (5.46), we will show that

$$
\sum_{k=1}^{R} \phi_k^2 \left[\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)} \boldsymbol{\eta}_{j+1}\right] = \left(\sum_{k=1}^{R} \phi_k^2\right) \Sigma_{i,j} + o_P(1), \qquad (5.50)
$$

$$
\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\eta}_{j+1} = \Sigma_{i,j} + o_P(1), \qquad (5.51)
$$

and the remaining terms in the right side of (5.49) are $o_P(1)$.

Result (5.51) holds by result (4.9). To prove (5.50), we use condition (A0)-(ii) and the Weak Law of Large Numbers for a sequence of independent variables to write:

$$
\begin{aligned}
\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)} \boldsymbol{\eta}_{j+1} &= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} \eta_{t,i} \left[\boldsymbol{U}_{(k)}^T \boldsymbol{U}_{(k)}\right]_{t,l} \eta_{l,j} \\
&= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} \eta_{t,i} I(l = t, 1 \le t \le n-k) \eta_{l,j} \\
&= \frac{1}{n+1}\sum_{t=1}^{n-k} \eta_{t,i}\eta_{t,j} \xrightarrow[n\to\infty]{P} E\left(\eta_{1,i}\eta_{1,j}\right) = \Sigma_{i,j},
\end{aligned}
$$

and (5.50) follows easily.

We now show that the first term in (5.49) is $o_P(1)$. We have:

$$
\begin{aligned}
\sum_{k=1}^{R} \phi_k \left[\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T (\boldsymbol{U}_{(k)}^T + \boldsymbol{U}_{(k)}) \boldsymbol{\eta}_{j+1}\right] &= \sum_{k=1}^{R} \phi_k \left(\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{\eta}_{j+1}\right) \\
&+ \sum_{k=1}^{R} \phi_k \left(\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)} \boldsymbol{\eta}_{j+1}\right), \qquad (5.52)
\end{aligned}
$$

so it suffices to analyze the term $\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{\eta}_{j+1}/(n+1)$, as $\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)} \boldsymbol{\eta}_{j+1}/(n+1)$ is its transpose, with i and j interchanged. Using the expression for $\boldsymbol{U}_{(k)}^T$ in (5.31), we obtain:

$$\begin{aligned}
\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{\eta}_{j+1} &= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} \eta_{t,i} \left[\boldsymbol{U}_{(k)}^T\right]_{t,l} \eta_{l,j} \\
&= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} \eta_{t,i} I(l = t+k, 1 \le t \le n-k) \eta_{l,j} \\
&= \frac{1}{n+1}\sum_{t=1}^{n-k} \eta_{t,i}\eta_{t+k,j} \xrightarrow[n\to\infty]{P} E\left(\eta_{1,i}\eta_{1+k,j}\right) = E(\eta_{1,i})E(\eta_{1+k,j}) = 0.
\end{aligned}$$

The above result was obtained by using the fact that $\{\eta_{t,i}\eta_{t+k,j}\}_{t=1}^{n-k}$ is a sequence of k-dependent, identically distributed random variables (condition (A0)-(ii)), so the quantity $\sum_{t=1}^{n-k} \eta_{t,i}\eta_{t+k,j}/(n+1)$ converges to $E\left(\eta_{1,i}\eta_{1+k,j}\right)$ by the Weak Law of Large Numbers for k-dependent sequences of random variables. We conclude that the term $\boldsymbol{\eta}_{i+1}^T \boldsymbol{U}_{(k)}^T \boldsymbol{\eta}_{j+1}/(n+1)$ is $o_P(1)$, so the first term in (5.49) is $o_P(1)$. Using a similar reasoning, we can show that the second term in (5.49) is also $o_P(1)$.

It remains to show (5.38). By the definition of $\boldsymbol{\mathcal{V}}$ in (5.28), we have:

$$\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{j+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\phi_R\eta_{1,j} \\ -\phi_{R-1}\eta_{1,j} - \phi_R\eta_{2,j} \\ \vdots \\ -\phi_1\eta_{1,j} - \phi_2\eta_{2,j} - \cdots - \phi_R\eta_{R,j} \end{pmatrix},$$

so $||\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{j+1}||_2 = \mathcal{O}(1)$ by assumption (A0)-(ii). A similar argument gives $||\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{i+1}||_2 = \mathcal{O}(1)$. Combining these results yields:

$$\begin{aligned}
\left|\frac{1}{n+1}\boldsymbol{\eta}_{i+1}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{j+1}\right| &\le \frac{1}{n+1}||\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{i+1}||_2 \cdot ||\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{j+1}||_2 \\
&= \frac{1}{n+1}\mathcal{O}_P(1)\cdot\mathcal{O}_P(1) = \mathcal{O}_P(1/n) = o_P(1),
\end{aligned}$$

so $\boldsymbol{\eta}_{i+1}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}} \boldsymbol{\eta}_{j+1}/(n+1)$ is $o_P(1)$. This completes our proof of Lemma 5.7.4.

The following lemma provides asymptotic approximations for non-random quantitities involving the bias associated with estimating a smooth function $m(\cdot)$ via locally linear regression.

Lemma 5.7.5 *Suppose the assumptions in Lemma 5.7.1 hold. Let $\boldsymbol{G}$ be an $n \times (p+1)$ matrix defined as in* (2.14) *such that condition* (A0)-(i) *is satisfied, and $\boldsymbol{S}_h$ be an $n \times n$ local linear smoother matrix defined as in* (3.6)- (3.8). *Set $\boldsymbol{m} = (m(Z_1), \ldots, m(Z_n))^T$, where $m(\cdot)$ satisfies condition* (A4) *and the Z_i's satisfy the design condition* (A3). *Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$:*

$$\frac{1}{n+1} \boldsymbol{G}^T \boldsymbol{\Psi}^{-1} (\boldsymbol{I} - \boldsymbol{S}_h) \boldsymbol{m} = -h^2 \frac{\sigma_\epsilon^2}{\sigma_u^2} \left(1 - \sum_{k=1}^{R} \phi_k\right)^2 \frac{\nu_2(K)}{2} \int_0^1 \boldsymbol{g}(z) m''(z) f(z) dz + \boldsymbol{o}(h^2) \tag{5.53}$$

and

$$\frac{1}{n(n+1)} \boldsymbol{G}^T \boldsymbol{\Psi}^{-1} \boldsymbol{1}\boldsymbol{1}^T (\boldsymbol{S}_h - \boldsymbol{I}) \boldsymbol{m} = h^2 \frac{\sigma_\epsilon^2}{\sigma_u^2} \left(1 - \sum_{k=1}^{R} \phi_k\right)^2 \frac{\nu_2(K)}{2} \int_0^1 \boldsymbol{g}(z) m''(z) f(z) dz \times \int_0^1 \boldsymbol{g}(z) f(z) dz + \boldsymbol{o}_P(h^2), \tag{5.54}$$

where $\int_0^1 \boldsymbol{g}(z) f(z) dz$ and $\int_0^1 \boldsymbol{g}(z) m''(z) f(z) dz$ are defined as in equations (2.18) *and* (2.19).

Proof:

We first prove (5.53). Using the explicit expression for $\boldsymbol{\Psi}^{-1}$ in equation (5.27), we write

$$\begin{aligned} \frac{1}{n+1} \boldsymbol{G}^T \boldsymbol{\Psi}^{-1} (\boldsymbol{I} - \boldsymbol{S}_h) \boldsymbol{m} &= \frac{\sigma \epsilon^2}{\sigma_u^2} \cdot \frac{1}{n+1} \boldsymbol{G}^T \boldsymbol{\mathcal{U}}^T \boldsymbol{\mathcal{U}} (\boldsymbol{I} - \boldsymbol{S}_h) \boldsymbol{m} \\ &- \frac{\sigma \epsilon^2}{\sigma_u^2} \cdot \frac{1}{n+1} \boldsymbol{G}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}} (\boldsymbol{I} - \boldsymbol{S}_h) \boldsymbol{m}, \end{aligned}$$

so it suffices to show that

$$\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} = -h^2\left(1-\sum_{k=1}^{R}\phi_k\right)^2\frac{\nu_2(K)}{2}\int_0^1 \boldsymbol{g}(z)m''(z)f(z)dz + \boldsymbol{o}(h^2)$$
$$\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} = \boldsymbol{o}(h^2).$$

These facts follow by proving that

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} = -h^2\left(1-\sum_{k=1}^{R}\phi_k\right)^2\frac{\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2) \tag{5.55}$$

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} = o(h^2), \tag{5.56}$$

for any $i = 0, 1, \ldots, p$.

First, consider (5.55). Using the expression for $\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}$ in (5.30), we have:

$$\begin{aligned}
\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} &= \sum_{k=1}^{R}\phi_k\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{U}_k^{(T)}+\boldsymbol{U}_{(k)})(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{m}\right] \\
&- \sum_{\substack{p,\,q=1\\ p<q}}^{R}\phi_p\phi_q\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)}+\boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)})(\boldsymbol{S}_h-\boldsymbol{I})m\right] \\
&- \sum_{k=1}^{R}\phi_k^2\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}^T\boldsymbol{U}_{(k)}(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{m}\right] \\
&- \frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{m}
\end{aligned} \tag{5.57}$$

Thus, to establish (5.53), it suffices to show that the last two terms are $o(h^2)$ and the remaining terms can be approximated as:

$$\begin{aligned}
&\sum_{k=1}^{R}\phi_k\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{U}_{(k)}^T+\boldsymbol{U}_{(k)})(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{m}\right] \\
&\qquad = h^2\left(2\sum_{k=1}^{R}\phi_k\right)\frac{\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2),
\end{aligned} \tag{5.58}$$

$$\sum_{\substack{p,\,q=1\\p<q}}^{R} \phi_p\phi_q \left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)} + \boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)})(\boldsymbol{S}_h - \boldsymbol{I})m\right]$$

$$= h^2\left(2\sum_{\substack{p,\,q=1\\p<q}}^{R}\phi_p\phi_q\right)\frac{\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2), \quad (5.59)$$

$$\sum_{k=1}^{R}\phi_k^2\left[\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}^T\boldsymbol{U}_{(k)}(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}\right]$$

$$= h^2\left(\sum_{k=1}^{R}\phi_k^2\right)\frac{\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2), \quad (5.60)$$

and

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} = \frac{h^2\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2). \quad (5.61)$$

The last result follows easily from result (4.66) of Lemma 4.6.10.

To prove (5.58), it suffices to show that the equalities below hold for any $k = 1, \ldots, R$:

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} = \frac{h^2\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2), \quad (5.62)$$

$$\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} = \frac{h^2\nu_2(K)}{2}\int_0^1 g_i(z)m''(z)f(z)dz + o(h^2). \quad (5.63)$$

Consider the left side in (5.62). Using the expression for $\boldsymbol{U}_{(k)}^T$ in (5.31), the boundedness of $g_i(\cdot)$ (condition (A0)-(i)) and result (4.34) of Lemma 4.6.1 with $\boldsymbol{r} \equiv \boldsymbol{m}$, we obtain:

$$\begin{aligned}
\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{U}_{(k)}^T(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m} &= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} g_i(Z_t)\left[\boldsymbol{U}_{(k)}^T\right]_{t,l}\cdot\left[(\boldsymbol{S}_h - \boldsymbol{I})\boldsymbol{m}\right]_l \\
&= \frac{1}{n+1}\sum_{t=1}^{n}\sum_{l=1}^{n} g_i(Z_t)I(l = t+k, 1 \le t \le n-k)\left[B_m(K, Z_l, h)\cdot h^2 + o(h^2)\right] \\
&= h^2\left[\frac{1}{n+1}\sum_{t=1}^{n-k} g_i(Z_t)B_m(K, Z_{t+k}, h)\right] + o(h^2) \\
&\equiv h^2\left(\mathcal{P}_n + \mathcal{Q}_n\right) + o(h^2), \quad (5.64)
\end{aligned}$$

where $\mathcal{P}_n = \sum_{t=1}^{n-k} g_i(Z_{t+k}) B_m(K, Z_{t+k}, h)/(n+1)$ and
$\mathcal{Q}_n = \sum_{t=1}^{n-k} \left(g_i(Z_t) - g_i(Z_{t+k})\right) B_m(K, Z_{t+k}, h)/(n+1)$.

A Riemann integration argument allows us to approximate $\mathcal{P}_n$ as:

$$\begin{aligned}
\mathcal{P}_n &= \frac{1}{n+1} \sum_{t=1}^{n} g_i(Z_t) B_m(K, Z_t, h) - \frac{1}{n+1} \sum_{t=1}^{k} g_i(Z_t) B_m(K, Z_t, h) \\
&= \frac{\nu_2(K)}{2} \int_0^1 g_i(z) m''(z) f(z) dz + o(1).
\end{aligned}$$

The last equality was obtained by using the fact that k does not depend on n and $g_i(\cdot)$ is bounded (condition (A0)-(i)). We also used the fact that $B_m(K, z, h)$ is bounded for all $z \in [0,1]$ and $h \leq h_0$, with $h_0 \in [0, 1/2]$ small enough, by result (4.35) of Lemma 4.6.1, Lemma 4.6.2 and condition (A4).

Using the fact that $g_i(\cdot)$ is Lipschitz continuous with Lipschitz constant $\mathcal{C}^*$ (condition (A0)-(i)) and that the Z_t's satisfy the design condition (A3), we bound $\mathcal{Q}_n$ as:

$$\begin{aligned}
|\mathcal{Q}_n| &\leq \frac{1}{n+1} \sum_{t=1}^{n-k} |g_i(Z_t) - g_i(Z_{t+k})| \cdot |B_m(K, Z_{t+k}, h)| \\
&\leq \mathcal{C} \frac{1}{n+1} \sum_{t=1}^{n-k} \sum_{l=0}^{k-1} |g_i(Z_{t+l}) - g_i(Z_{t+l+1})| \\
&\leq \mathcal{C} \frac{1}{n+1} \sum_{t=k+1}^{n} \left(k \frac{\mathcal{C}^*}{n+1} \right) = \mathcal{C}\mathcal{C}^* \frac{(n-k)k}{(n+1)^2} = o(1).
\end{aligned}$$

Substituting the results concerning $\mathcal{P}_n$ and $\mathcal{Q}_n$ in (5.64) yields (5.62). A similar argument can be used to prove (5.63). The only difference in the proofs is that the range of summation for $\sum_t g_i(Z_t) B_m(K, Z_{t+k}, h)$ changes.

Combining (5.62) and (5.63) yields (5.58). Similar arguments can be employed to obtain results (5.59) and (5.60).

It remains to prove (5.56). By the definition of $\boldsymbol{\mathcal{V}}$ in (5.28), we have:

$$\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\phi_R[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_1 \\ -\phi_{R-1}[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_1 - \phi_R[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_2 \\ \vdots \\ -\phi_1[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_1 - \phi_2[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_2 - \cdots - \phi_R[(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}]_R \end{pmatrix}.$$

But $||\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}||_2 = \mathcal{O}(h^2)$, since for $l = 1, \ldots, R$, $[(\boldsymbol{S}_h - \boldsymbol{I})m]_l = \mathcal{O}(h^2)$ by Lemma 4.6.2. We know that $||\boldsymbol{\mathcal{V}}\boldsymbol{\eta}_{i+1}||_2 = \mathcal{O}(1)$. Combining these results yields:

$$\begin{aligned} \left|\frac{1}{n+1}\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}\right| &\leq \frac{1}{n+1}||\boldsymbol{\mathcal{V}}\boldsymbol{G}_{i+1}||_2 \cdot ||\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}||_2 \\ &= \frac{1}{n+1}\mathcal{O}(1)\cdot\mathcal{O}(h^2) = \mathcal{O}(1/n) = o(h^2), \end{aligned}$$

so $\boldsymbol{G}_{i+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}}(\boldsymbol{I}-\boldsymbol{S}_h)\boldsymbol{m}/(n+1)$ is $o(h^2)$.

Result (5.54) follows easily, by writing:

$$\frac{1}{n(n+1)}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\mathbf{1}\mathbf{1}^T(\boldsymbol{S}-\boldsymbol{I})\boldsymbol{m} = \left(\frac{1}{n+1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}\mathbf{1}\right)\cdot\left[\frac{1}{n+1}\mathbf{1}^T(\boldsymbol{S}_h-\boldsymbol{I})\boldsymbol{m}\right]\cdot\left(1+\frac{1}{n}\right).$$

and using Lemma 5.7.3 and result (4.66) of Lemma 4.6.10 with $\boldsymbol{G}$ replaced with $\mathbf{1}$. The proof of Lemma 5.7.5 is now complete.

Let $\boldsymbol{V}_{\boldsymbol{\Psi}}$ be the matrix defined in (2.22). The next result concerns the existence of an inverse for $\boldsymbol{V}_{\boldsymbol{\Psi}}$ and provides an explicit expression for this inverse. We do not provide a proof for this result, as one can easily verify that $\boldsymbol{V}_{\boldsymbol{\Psi}}\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1} = \boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}\boldsymbol{V}_{\boldsymbol{\Psi}} = \boldsymbol{I}$ by using the expression of $\boldsymbol{V}_{\boldsymbol{\Psi}}^{-1}$ given in Lemma 5.7.6 below.

Lemma 5.7.6 *Let $\boldsymbol{V}_{\Psi}$ be the $(p+1)\times(p+1)$ matrix introduced in* (2.22) *and define the $p\times 1$ vector $\boldsymbol{a}$ as $\boldsymbol{a} = (\int_0^1 g_1(z)f(z)dz, \dots, \int_0^1 g_p(z)f(z)dz)^T$. Here, $f(\cdot)$ is a design density satisfying condition* (A3) *and $g_1(\cdot), \dots, g_p(\cdot)$ are smooth functions satisfying condition* (A0)-(i). *Furthermore, let $\boldsymbol{\Sigma} = (\Sigma_{i,j})$ be the variance-covariance matrix introduced in condition* (A0)-(ii). *Then $\boldsymbol{V}_{\Psi}^{-1}$ exists and is given by:*

$$\boldsymbol{V}_{\Psi}^{-1} = \frac{\sigma_u^2}{\sigma_\epsilon^2}\left(\begin{array}{c|c} \dfrac{1}{(1-\sum_{k=1}^R \phi_k)^2} + \dfrac{1}{1+\sum_{k=1}^R \phi_k^2}\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & -\dfrac{1}{1+\sum_{k=1}^R \phi_k^2}\boldsymbol{a}^T\boldsymbol{\Sigma}^{-1} \\ \hline -\dfrac{1}{1+\sum_{k=1}^R \phi_k^2}\boldsymbol{\Sigma}^{-1}\boldsymbol{a} & \dfrac{1}{1+\sum_{k=1}^R \phi_k^2}\boldsymbol{\Sigma}^{-1} \end{array}\right)$$

provided $\boldsymbol{\Sigma}^{-1}$ exists.

The last result in this Appendix provides a useful asymptotic bound.

Lemma 5.7.7 *Suppose the assumptions in Theorem 5.2.1 hold. Then, if $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$, we have:*

$$\frac{1}{(n+1)^2}\boldsymbol{V}_{\Psi}^{-1}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}\boldsymbol{V}_{\Psi}^{-1} = \mathcal{O}(n^{-1}). \tag{5.65}$$

Proof:

To prove (5.65) it suffices to show that

$$\frac{1}{(n+1)^2}\boldsymbol{G}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G} = \mathcal{O}(n^{-1}), \tag{5.66}$$

since the elements of the $(p+1)\times(p+1)$ matrix $\boldsymbol{V}_{\Psi}^{-1}$ do not depend upon n. Result (5.66) follows by showing that $\boldsymbol{G}_{i+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}/(n+1)^2$ is $\mathcal{O}(n^{-1})$ for any $i, j = 0, 1, \dots, p$.

Let $i, j = 0, 1, \ldots, p$ be fixed. Using vector and matrix norm properties introduced in Section 2.4, we obtain:

$$\begin{aligned}
&\left|\left|\frac{1}{(n+1)^2}\boldsymbol{G}_{i+1}^T\boldsymbol{\Psi}^{-1}(\boldsymbol{I}-\boldsymbol{S}_h^c)\boldsymbol{\Psi}(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}\right|\right|_2 \\
&\leq \frac{1}{(n+1)^2}||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{i+1}||_2 \cdot ||\boldsymbol{\Psi}||_S \cdot ||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2 \\
&\leq \mathcal{O}\left(n^{-2}\right) \cdot ||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{i+1}||_2 \cdot ||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2,
\end{aligned} \tag{5.67}$$

since $||\boldsymbol{\Psi}||_S$ is $\mathcal{O}(1)$ by result (5.34) of Lemma 5.7.2. Thus, it suffices to show that $||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{i+1}||_2$ and $||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2$ are $\mathcal{O}(n^{1/2})$.

Let $\boldsymbol{v} \equiv \boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}$; using $\boldsymbol{S}_h^c = (\boldsymbol{I} - \boldsymbol{1}\boldsymbol{1}^T/n)\boldsymbol{S}_h$ (equation (3.1) with $\mathbb{S}_h = \boldsymbol{S}_h$) we write:

$$\begin{aligned}
||(\boldsymbol{I}-\boldsymbol{S}_h^c)^T\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2 &\leq ||\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2 + ||\boldsymbol{S}_h^{cT}\boldsymbol{\Psi}^{-1}\boldsymbol{G}_{j+1}||_2 \\
&= ||\boldsymbol{v}||_2 + ||\boldsymbol{S}_h^{cT}\boldsymbol{v}||_2 \\
&= ||\boldsymbol{v}||_2 + \left|\left|\boldsymbol{S}_h^T(\boldsymbol{I} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T)\boldsymbol{v}\right|\right| \\
&\leq ||\boldsymbol{v}||_2 + ||\boldsymbol{S}_h^T\boldsymbol{v}||_2 + \left|\left|\boldsymbol{S}_h^T\left(\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T\boldsymbol{v}\right)\right|\right|.
\end{aligned} \tag{5.68}$$

If v_t denotes the t^{th} component of $\boldsymbol{v}$, we can show that there exists $\tilde{C} > 0$ such that $|v_t| < \tilde{C}$ for all $t = 1, \ldots, n$. Indeed, by the expression for $\boldsymbol{\Psi}^{-1}$ in (5.27) and Comment 5.7.1, we have:

$$\begin{aligned}
\boldsymbol{v}^T = \boldsymbol{G}_{j+1}^T\boldsymbol{\Psi}^{-1} &= \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\boldsymbol{G}_{j+1}^T\boldsymbol{\mathcal{U}}^T\boldsymbol{\mathcal{U}} - \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\boldsymbol{G}_{j+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}} \\
&= \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\boldsymbol{G}_{j+1}^T\left[-\sum_{k=1}^R\phi_k(\boldsymbol{U}_{(k)}^T + \boldsymbol{U}_{(k)}) + \sum_{\substack{p,\,q=1\\p<q}}^R\phi_p\phi_q(\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)} + \boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)})\right. \\
&\quad \left. + \sum_{k=1}^R\phi_k^2\boldsymbol{U}_{(k)}^T\boldsymbol{U}_{(k)} + \boldsymbol{I}\right] - \frac{\sigma_\epsilon^2}{\sigma_u^2}\cdot\boldsymbol{G}_{j+1}^T\boldsymbol{\mathcal{V}}^T\boldsymbol{\mathcal{V}},
\end{aligned}$$

so it suffices to show that the quantities $\boldsymbol{G}_{j+1}^T\boldsymbol{U}_{(k)}^T$, $\boldsymbol{G}_{j+1}^T\boldsymbol{U}_{(k)}$, $\boldsymbol{G}_{j+1}^T\boldsymbol{U}_{(p)}^T\boldsymbol{U}_{(q)}$, $\boldsymbol{G}_{j+1}^T\boldsymbol{U}_{(q)}^T\boldsymbol{U}_{(p)}$ and $\boldsymbol{G}_{j+1}^T\boldsymbol{U}_{(k)}^T\boldsymbol{U}_{(k)}$ have bounded components for all $k, p, q = 1, \ldots, R$, $p < q$, and

$\boldsymbol{G}_{j+1}^T \boldsymbol{\mathcal{V}}^T \boldsymbol{\mathcal{V}}$ also has bounded components. These facts follow easily from the sparseness of $\boldsymbol{U}_{(k)}$ and $\boldsymbol{\mathcal{V}}$ (see Comment 5.7.1) and the boundedness of $\boldsymbol{G}_{j+1}$'s components (condition (A0)-(i)).

The boundedness of $\boldsymbol{v}$'s components implies that $||\boldsymbol{v}||_2$ is $\mathcal{O}(n^{1/2})$, $||\boldsymbol{S}_h^T \boldsymbol{v}||_2$ is $\mathcal{O}(n^{1/2})$ and $||\boldsymbol{S}_h^T(\mathbf{1}\mathbf{1}^T \boldsymbol{v}/n)||_2$ is $\mathcal{O}(n^{1/2})$. The last two results follow by result (4.49) of Lemma 4.6.6. Using these asymptotic bounds in (5.68) yields that $||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T \boldsymbol{\Psi}^{-1} \boldsymbol{G}_{j+1}||_2$ is $\mathcal{O}(n^{1/2})$. A similar argument gives that $||(\boldsymbol{I} - \boldsymbol{S}_h^c)^T \boldsymbol{\Psi}^{-1} \boldsymbol{G}_{i+1}||_2$ is $\mathcal{O}(n^{1/2})$.

Chapter 6

Choosing the Correct Amount of Smoothing

The estimators of the linear effects in model (2.1) considered in this thesis depend on a smoothing parameter h. This parameter has a dual function. On one hand, it influences the statistical properties of the estimated linear effects. On the other hand, it controls the shape and smoothness of the estimated non-linear effect. Our focus in this chapter is on developing data-driven methods for choosing h so that we obtain accurate estimators for the *linear* effects of interest. These methods may not be the most appropriate for accurate estimation of the *non-linear* effect, as they may undersmooth its estimator.

This chapter is organized as follows. In Section 6.1, we introduce some useful notation. In Section 6.2, we introduce methods for choosing the correct smoothing parameter for the usual and modified local linear backfitting estimators of the linear effects of interest in model (2.1). These methods require the accurate estimation of the nonparametric component $\boldsymbol{m}$ and the error correlation structure, topics discussed in Section 6.3. Finally, in Section 6.4 we introduce methods for choosing the correct smoothing parameter for the estimated modified local linear backfitting estimators of the linear effects of interest in model (2.1).

6.1 Notation

In what follows we are interested in the accurate estimation of a linear combination $\boldsymbol{c}^T\boldsymbol{\beta}$ of the linear effects $\boldsymbol{\beta}$ in model (2.1), where $\boldsymbol{c} = (c_0, c_1, \ldots, c_p)^T$ is a known vector with real-valued components (e.g: $\boldsymbol{c} = (0, \ldots, 0, 1, 0, \ldots, 0)^T$).

Throughout this chapter, we denote $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ generically by $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ in order to emphasize their dependence upon the amount of smoothing h. We want to choose the amount of smoothing h to accurately estimate $\boldsymbol{c}^T\boldsymbol{\beta}$ via $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. Given that $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\boldsymbol{S}_h^c}$ is conceptually qualitatively different than the other estimators considered here, we defer its discussion to Section 6.4. In the remainder of this chapter, unless otherwise stated, we assume that $\boldsymbol{\Omega}$ stands for $\boldsymbol{I}$ or $\boldsymbol{\Psi}^{-1}$.

The correct choice of h depends on the conditional bias and variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$. We provide below explicit expressions for these quantities.

The exact conditional variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ equals $\boldsymbol{c}^T Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{c}$. Expressions for $Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z})$ are found in (4.17) when $\boldsymbol{\Omega} = \boldsymbol{I}$ and in (5.15) when $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$. Thus:

$$\boldsymbol{c}^T Var(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z})\boldsymbol{c} = \sigma_\epsilon^2 \boldsymbol{c}^T \boldsymbol{M}_{\boldsymbol{\Omega},h}\boldsymbol{\Psi}\boldsymbol{M}_{\boldsymbol{\Omega},h}^T\boldsymbol{c} \equiv Var(h;\boldsymbol{\Omega}), \tag{6.1}$$

where

$$\boldsymbol{M}_{\boldsymbol{\Omega},h} = (\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{S}_h^c). \tag{6.2}$$

The exact conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ equals $\boldsymbol{c}^T Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z})$ and can be obtained from (4.2) when $\boldsymbol{\Omega} = \boldsymbol{I}$ or (5.2) when $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$:

$$\boldsymbol{c}^T Bias(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z}) = \boldsymbol{c}^T\boldsymbol{M}_{\boldsymbol{\Omega},h}\boldsymbol{m} \equiv Bias(h;\boldsymbol{\Omega}). \tag{6.3}$$

6.2 Choosing h for $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ and $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},\boldsymbol{S}_h^c}$

The estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ depends on the smoothing parameter h. To obtain an accurate estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ of $\boldsymbol{c}^T\boldsymbol{\beta}$ we choose h so that it minimizes a measure of accuracy of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. Although the smoothing parameter h quantifies the degree of smoothness of $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$, a 'good' value for h should not necessarily be chosen to minimize a measure of accuracy of $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$ as, in the present context, $\boldsymbol{m}$ is merely a nuisance.

Since $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ is generally biased in finite samples, we assess its accuracy via its exact conditional mean squared error, given $\boldsymbol{X}$ and $\boldsymbol{Z}$:

$$MSE(h;\boldsymbol{\Omega}) = Bias(h;\boldsymbol{\Omega})^2 + Var(h;\boldsymbol{\Omega}). \tag{6.4}$$

We define the MSE-optimal amount of smoothing for estimating $\boldsymbol{c}^T\boldsymbol{\beta}$ via $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ as the minimizer of MSE:

$$h_{MSE}^{\boldsymbol{\Omega}} = \underset{h}{argmin}\ MSE(h;\boldsymbol{\Omega}). \tag{6.5}$$

From equations (6.1) and (6.3), one can see that $h_{MSE}^{\boldsymbol{\Omega}}$ depends upon the unknown nonparametric component $\boldsymbol{m}$ as well as the error variance σ_ϵ^2 and the error correlation matrix $\boldsymbol{\Psi}$, which are typically unknown. Thus, $h_{MSE}^{\boldsymbol{\Omega}}$ is not directly available.

To date, no methods have been proposed for estimating $h_{MSE}^{\boldsymbol{\Omega}}$ when the model errors are correlated. However, when the model errors are uncorrelated and $\boldsymbol{\Omega} = \boldsymbol{I}$, Opsomer and Ruppert (1999) proposed an empirical bias bandwidth selection (EBBS) method for estimating $h_{MSE}^{\boldsymbol{\Omega}}$. We describe this method in Section 6.2.1. We propose modifications of the EBBS method to estimate $h_{MSE}^{\boldsymbol{\Omega}}$ when the errors are correlated not only for $\boldsymbol{\Omega} = \boldsymbol{I}$, but also for $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ in Section 6.2.2. Finally, in Section 6.2.3 we propose a non-asymptotic plug-in method for estimating $h_{MSE}^{\boldsymbol{\Omega}}$ in the presence of error correlation when $\boldsymbol{\Omega}$ equals $\boldsymbol{I}$ or $\boldsymbol{\Psi}^{-1}$. Each method minimizes an estimator of $MSE(\cdot;\boldsymbol{\Omega})$ over h in some grid. Throughout, we let $\mathcal{H} = \{h(1), \ldots, h(N)\}$ denote the grid, for some integer N.

6.2.1 Review of Opsomer and Ruppert's EBBS method

In this section, we provide a detailed review of Opsomer and Ruppert's EBSS method. Throughout this section only, we assume that the errors associated with model (2.1) satisfy $\boldsymbol{\Psi} = \boldsymbol{I}$. Specifically, we assume that these errors satisfy the assumption:

(A6) The model errors $\epsilon_i, i = 1, \ldots, n$, are independent, identically distributed, having mean 0 and variance $\sigma_\epsilon^2 \in (0, \infty)$.

We also consider $\boldsymbol{\Omega} = \boldsymbol{I}$, so that the results in this section will apply exclusively to $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, the usual local linear backfitting estimator of $\boldsymbol{c}^T\boldsymbol{\beta}$.

Under the above conditions, the EBBS method attempts to estimate $h_{MSE}^{\boldsymbol{I}}$ by minimizing an estimator of $MSE(\cdot; \boldsymbol{I})$ over $\mathcal{H}$, a grid of possible values of h. For a given $h(j) \in \mathcal{H}$, Opsomer and Ruppert find an estimator for $MSE(h(j); \boldsymbol{I})$ by combining an empirical estimator of $Bias(h(j); \boldsymbol{I})$ with a residual-based estimator of $Var(h(j); \boldsymbol{I})$. We discuss the details related to computing these estimators below.

Opsomer and Ruppert use a higher order asymptotic expression for $E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}, \boldsymbol{Z}) - \boldsymbol{c}^T\boldsymbol{\beta}$, the exact conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$, to obtain:

$$Bias(h; \boldsymbol{I}) = \sum_{t=2}^{T} a_t h^t + o(h^{1+T}), \tag{6.6}$$

as $h \to 0$, where $a_t, t = 1, \ldots, T$, are unknown asymptotic constants referred to as bias coefficients. This expression can be obtained by a more delicate Taylor series analysis in (4.3). This yields the approximation:

$$E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{c}^T\boldsymbol{\beta} + \sum_{t=2}^{T+1} a_t h^t + o(h^{1+T}). \tag{6.7}$$

For fixed $h(j) \in \mathcal{H}$, Opsomer and Ruppert estimate $Bias(h(j); \boldsymbol{I})$, the exact conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)}$, as follows. They calculate $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(k)}$ for $k \in \{j - k_1, \dots, j + k_2\}$, for some k_1, k_2. Note that j must be between $k_1 + 1$ and $N - k_2$, inclusive. They then fit the model:

$$E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}, \boldsymbol{Z}) = a_0 + a_2 \cdot h^2 + \cdots + a_{T+1} \cdot h^{T+1} \tag{6.8}$$

to the 'data' $\left\{\left(h(k), \boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(k)}\right) : k = j - k_1, \dots, j + k_2\right\}$ using ordinary least squares. This results in the fit:

$$\widehat{E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}|\boldsymbol{X}}, \boldsymbol{Z}) = \widehat{a}_0 + \widehat{a}_2 \cdot h^2 + \cdots + \widehat{a}_{T+1} \cdot h^{T+1}. \tag{6.9}$$

An estimator for $Bias(h(j); \boldsymbol{I})$ is then:

$$\begin{aligned} \widehat{Bias}(h(j); \boldsymbol{I}) &= E(\boldsymbol{c}^T\widehat{\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)}}|\boldsymbol{X}, \boldsymbol{Z}) - \widehat{a}_0 \\ &= \widehat{a}_2 \cdot h(j)^2 + \cdots + \widehat{a}_{T+1} \cdot h(j)^{T+1}. \end{aligned} \tag{6.10}$$

Here, k_1, k_2 and T are tuning parameters that must be chosen by the user. We must have $k_1 + k_2 > T$ since the $T + 1$ parameters $a_0, a_1, \dots, a_T$ will be estimated using $k_1 + k_2 + 1$ 'data' points.

Opsomer and Ruppert estimate $Var(h(j); \boldsymbol{I})$, the exact conditional variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)}$, by using (6.1) with $\boldsymbol{\Psi} = \boldsymbol{I}$ but with σ_ϵ^2 replaced by the following residual-based estimator:

$$\widehat{\sigma}^2_{\epsilon,\boldsymbol{I}} = \frac{||\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)} - \widehat{\boldsymbol{m}}_{\boldsymbol{I},h(j)}||^2}{n}.$$

This yields:

$$\widehat{Var}(h(j); \boldsymbol{I}) = \widehat{\sigma}^2_{\epsilon,\boldsymbol{I}}\boldsymbol{c}^T\boldsymbol{M}_{\boldsymbol{I},h(j)}\boldsymbol{M}^T_{\boldsymbol{I},h(j)}\boldsymbol{c}. \tag{6.11}$$

Finally, Opsomer and Ruppert combine (6.10) and (6.11) to obtain the following estimator of $MSE(h(j); \boldsymbol{I})$, $k_1 + 1 \le j \le N - k_2$:

$$\widehat{MSE}(h(j); \boldsymbol{I}) = \widehat{Bias}(h(j); \boldsymbol{I})^2 + \widehat{Var}(h(j); \boldsymbol{I}).$$

They then estimate $h^{\boldsymbol{I}}_{MSE}$, the minimizer of $MSE(\cdot;\boldsymbol{I})$, as follows:

$$\widehat{h}^{\boldsymbol{I}}_{MSE} = \underset{k_1+1\leq j\leq N-k_2}{argmin} \widehat{MSE}(h(j);\boldsymbol{I}).$$

We see that $\widehat{h}^{\boldsymbol{I}}_{MSE}$ attempts to estimate $h^{\boldsymbol{I}}_{MSE}$, the smoothing parameter which is MSE-optimal for estimation of $\boldsymbol{c}^T\boldsymbol{\beta}$. It is not clear however whether using $\widehat{h}^{\boldsymbol{I}}_{MSE}$ yields a $\sqrt{n}$-consistent estimator of $\boldsymbol{c}^T\boldsymbol{\beta}$.

The variance estimator $\widehat{Var}(h(j);\boldsymbol{I})$ in (6.11) depends on the matrix $\boldsymbol{M}_{\boldsymbol{I},h(j)}$. To speed computation of $\boldsymbol{M}_{\boldsymbol{I},h(j)}$, and hence $\widehat{Var}(h(j);\boldsymbol{I})$, Opsomer and Ruppert suggest the following. First, take $\boldsymbol{\Omega} = \boldsymbol{I}$ and $h = h(j)$ in (6.2) and re-arrange to obtain an alternative expression for $\boldsymbol{M}_{\boldsymbol{I},h(j)}$:

$$\begin{aligned}\boldsymbol{M}_{\boldsymbol{I},h(j)} &= (\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}^c_{h(j)})\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{I} - \boldsymbol{S}^c_{h(j)}) \\ &= (\boldsymbol{X}^T(\boldsymbol{X} - \boldsymbol{S}^c_{h(j)}\boldsymbol{X}))^{-1}(\boldsymbol{X}^T - \boldsymbol{X}^T\boldsymbol{S}^c_{h(j)}).\end{aligned} \tag{6.12}$$

Then, compute the product $\boldsymbol{S}^c_{h(j)}\boldsymbol{X}$ in (6.12) by smoothing each column of $\boldsymbol{X}$ against the design points $Z_1,\ldots,Z_n$. Finally, compute the product $\boldsymbol{X}^T\boldsymbol{S}^c_{h(j)}$ in (6.12) by using the approximation $\boldsymbol{X}^T\boldsymbol{S}^c_{h(j)} \approx (\boldsymbol{S}^c_{h(j)}\boldsymbol{X})^T$. This approximation is justified by the near symmetry of $\boldsymbol{S}^c_{h(j)}$. These computational tricks can also be used to ease the burden involved in calculating $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)}, h(j) \in \mathcal{H}$, as $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h(j)}$ can be easily seen to depend upon $\boldsymbol{M}_{\boldsymbol{I},h(j)}$.

A peculiar feature of the estimator $\widehat{\sigma}^2_{\epsilon,\boldsymbol{I}}$ of σ^2_ϵ is that it uses residuals based on the 'working' bandwidth $h(j) \in \mathcal{H}$, instead of a bandwidth optimized for estimation of σ^2_ϵ. As an alternative to estimating σ^2_ϵ with the 'working' bandwidth h, Opsomer and Ruppert suggest that one could use residuals based on a bandwidth optimized for estimation of σ^2_ϵ as in Opsomer and Ruppert (1998).

For implementing the EBBS method in practice, Opsomer and Ruppert (1999) suggest using a grid size $N = 18$ and grid values equally spaced on the log scale. They recommend using the following values for the tuning parameters involved in this method: $k_1 = 1, k_2 = 2$ and $T = 1$. For situations where $\widehat{MSE}(\cdot; \boldsymbol{I})$ is found to have more than one minimum as a function of h, they suggest that one could take $\widehat{h}^{\boldsymbol{I}}_{MSE}$ to be either the h value where the global minimum occurs, or the h value where the first local minimum occurs. The authors advise that they found the former approach to be superior to the latter in their simulation studies.

6.2.2 Modifications to the EBBS method

Here we adjust the EBBS method to deal with estimating $h^{\boldsymbol{\Omega}}_{MSE}$ when the model errors are correlated and $\boldsymbol{\Omega} = \boldsymbol{I}$ or $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$. The modified EBBS method attempts to estimate $h^{\boldsymbol{\Omega}}_{MSE}$ by minimizing an estimator of $MSE(\cdot; \boldsymbol{\Omega})$ over the grid $\mathcal{H}$. For a given $h(j) \in \mathcal{H}$, this estimator is obtained by combining an empirical estimator of $Bias(h(j); \boldsymbol{\Omega})$, the exact conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, with a residual-based estimator of $Var(h(j); \boldsymbol{\Omega})$, the exact conditional variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. Specifics are provided below.

The modified EBBS method uses a similar bias-estimation scheme to that employed in the EBBS method in order to estimate $Bias(h(j); \boldsymbol{\Omega})$. This scheme relies on the following asymptotic bias approximation:

$$E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{c}^T\boldsymbol{\beta} + \sum_{t=2}^{T+1} a_t h^t + o(h^{1+T}), \tag{6.13}$$

which parallels (6.7), and yields the estimator $\widehat{Bias}(h; \boldsymbol{\Omega})$.

However, the modified EBBS can no longer rely on the residual estimation scheme utilized in the EBBS method for estimating $Var(h(j); \boldsymbol{\Omega})$. The reason for this is that $Var(h(j); \boldsymbol{\Omega})$ depends not only on the error variance σ^2_ϵ, but also on the error correlation matrix $\boldsymbol{\Psi}$. For $\boldsymbol{\Omega} = \boldsymbol{I}$, we propose to estimate $Var(h; \boldsymbol{\Omega})$ via:

$$\widehat{Var}(h;\boldsymbol{\Omega}) = \begin{cases} \widehat{\sigma}_\epsilon^2 \boldsymbol{c}^T \boldsymbol{M}_{\boldsymbol{\Omega},h} \boldsymbol{\Psi} \boldsymbol{M}_{\boldsymbol{\Omega},h}^T \boldsymbol{c}, & \text{if } \boldsymbol{\Psi} \text{ is known and } \sigma_\epsilon^2 \text{ is unknown;} \\ \sigma_\epsilon^2 \boldsymbol{c}^T \boldsymbol{M}_{\boldsymbol{\Omega},h} \widehat{\boldsymbol{\Psi}} \boldsymbol{M}_{\boldsymbol{\Omega},h}^T \boldsymbol{c}, & \text{if } \boldsymbol{\Psi} \text{ is unknown and } \sigma_\epsilon^2 \text{ is known;} \\ \widehat{\sigma}_\epsilon^2 \boldsymbol{c}^T \boldsymbol{M}_{\boldsymbol{\Omega},h} \widehat{\boldsymbol{\Psi}} \boldsymbol{M}_{\boldsymbol{\Omega},h}^T \boldsymbol{c}, & \text{if } \boldsymbol{\Psi} \text{ is unknown and } \sigma_\epsilon^2 \text{ is unknown.} \end{cases} \tag{6.14}$$

For $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$, if σ_ϵ^2 is unknown, we propose to estimate $Var(h; \boldsymbol{\Omega})$ via:

$$\widehat{Var}(h;\boldsymbol{\Omega}) = \widehat{\sigma}_\epsilon^2 \boldsymbol{c}^T \boldsymbol{M}_{\boldsymbol{\Omega},h} \boldsymbol{\Psi} \boldsymbol{M}_{\boldsymbol{\Omega},h}^T \boldsymbol{c}. \tag{6.15}$$

The estimators in (6.14)-(6.15) have been obtained from (6.1) by substituting $\widehat{\sigma}_\epsilon^2$ for σ_ϵ^2 and $\widehat{\boldsymbol{\Psi}}$ for $\boldsymbol{\Psi}$ whenever appropriate. Details on how to obtain reasonable estimators $\widehat{\sigma}_\epsilon^2$ and $\widehat{\boldsymbol{\Psi}}$ are provided in Section 6.3.2.

In summary, the modified EBBS method finds the minimizer:

$$\widehat{h}_{MSE}^{\boldsymbol{\Omega}} = \underset{h}{argmin} \left\{ \widehat{Bias}(h;\boldsymbol{\Omega})^2 + \widehat{Var}(h;\boldsymbol{\Omega}) \right\} \equiv \widehat{h}_{EBBS-L}^{\boldsymbol{\Omega}}, \tag{6.16}$$

with $h \in \mathcal{H} = \{h(1), \ldots, h(N)\}$, $\widehat{Bias}(h; \boldsymbol{\Omega})$ obtained via the bias-estimation scheme described earlier, and $\widehat{Var}(h; \boldsymbol{\Omega})$ as in (6.14) if $\boldsymbol{\Omega} = \boldsymbol{I}$ or as in (6.15) if $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ and σ_ϵ^2 is unknown. Here, the label '$EBBS - L$' denotes the fact that the modified EBBS method estimates $Bias(h; \boldsymbol{\Omega})$ by *local* ordinary least squares regression.

It is possible to estimate $Bias(h; \boldsymbol{\Omega})$ by performing global, rather than local, ordinary least squares fitting. Specifically, we can perform just one least squares regression, using the 'data' $\left\{ \left(h(k), \boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h(k)} \right) : k = 1, \ldots, N \right\}$. We refer to the method that finds the minimizer of (6.16), with $\widehat{Bias}(h; \boldsymbol{\Omega})$ obtained by global ordinary least squares fitting, as the global modified EBBS method. We denote the amount of smoothing this method yields by $\widehat{h}_{EBBS-G}^{\boldsymbol{\Omega}}$.

Before concluding this section, we indicate how the modified EBBS methods can be generalized if one is interested in smoothing parameter selection for accurate estimation of $\boldsymbol{c}^T\boldsymbol{\beta}$ via the usual or modified local polynomial backfitting estimators. For simplicity, in this section only, we denote both of these estimators by $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$.

The variance-estimation scheme to be used in the generalized modified EBBS methods should be the same as that employed in (6.14)-(6.15). Obviously, the quantities $\widehat{\sigma}_\epsilon^2$, $\boldsymbol{M}_{\boldsymbol{\Omega},h}$ and $\widehat{\boldsymbol{\Psi}}$ involved in these equations should be computed based on locally polynomial regression of degree $D > 1$, instead of locally linear regression. We conjecture that the bias-estimation scheme would have to rely on the asymptotic approximation

$$E(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}|\boldsymbol{X},\boldsymbol{Z}) = \boldsymbol{c}^T\boldsymbol{\beta} + \sum_{t=D+1}^{T+1} a_t h^T + o(h^{1+T})$$

instead of (6.13). Note that we must have $T > D$.

6.2.3 Plug-in method

In this section we introduce yet another method for estimating the optimal amount of smoothing $h_{MSE}^{\boldsymbol{\Omega}}$ in the presence of error correlation whenever $\boldsymbol{\Omega} = \boldsymbol{I}$ or $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$, namely the non-asymptotic plug-in method. Recall that $h_{MSE}^{\boldsymbol{\Omega}}$ was defined as the minimizer of $MSE(\cdot;\boldsymbol{\Omega})$ in (6.4). Thus, we might find a reasonable estimator for $h_{MSE}^{\boldsymbol{\Omega}}$ by minimizing an estimator of $MSE(\cdot;\boldsymbol{\Omega})$ over a grid of possible values for the smoothing parameter h. We propose estimating $MSE(h(j);\boldsymbol{\Omega})$ by assembling plug-in estimators of its exact bias and variance components, $Bias(h(j);\boldsymbol{\Omega})$ and $Var(h(j);\boldsymbol{\Omega})$.

More specifically, we propose to estimate $Var(h(j);\boldsymbol{\Omega})$ using (6.14) if $\boldsymbol{\Omega} = \boldsymbol{I}$ and (6.15) if $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ and σ_ϵ^2 is unknown. Furthermore, we propose to estimate $Bias(h(j);\boldsymbol{\Omega})$ using (6.3), but with $\boldsymbol{m}$ replaced by an accurate estimator $\widehat{\boldsymbol{m}}$:

$$\widehat{Bias}(h;\boldsymbol{\Omega}) = \boldsymbol{c}^T\boldsymbol{M}_{\boldsymbol{\Omega},h}\widehat{\boldsymbol{m}}. \tag{6.17}$$

Details on how to obtain an accurate estimator $\widehat{\boldsymbol{m}}$ of $\boldsymbol{m}$ are provided in Section 6.3.1. As remarked before, when $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$, $\boldsymbol{M}_{\boldsymbol{\Omega},h}$ depends upon the error correlation matrix $\boldsymbol{\Psi}$. Thus, if $\boldsymbol{\Psi}$ is unknown, we must substitute $\widehat{\boldsymbol{\Psi}}$ for $\boldsymbol{\Psi}$ in the expression for $\boldsymbol{M}_{\boldsymbol{\Omega},h}$, where $\widehat{\boldsymbol{\Psi}}$ is obtained as in Section 6.3.2.

Finally, minimizing the estimator of $MSE(\cdot;\boldsymbol{\Omega})$ obtained by combining (6.17) with (6.14) for $\boldsymbol{\Omega} = \boldsymbol{I}$, or (6.15) for $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ and σ_ϵ^2 unknown, over a grid of possible values for h yields the desired plug-in estimator of $h_{MSE}^{\boldsymbol{\Omega}}$:

$$\widehat{h}_{MSE}^{\boldsymbol{\Omega}} = \underset{h\in\mathcal{H}}{argmin} \left\{ \widehat{Bias}(h(j);\boldsymbol{\Omega})^2 + \widehat{Var}(h(j);\boldsymbol{\Omega}) \right\} \equiv \widehat{h}_{PLUG-IN}^{\boldsymbol{\Omega}}. \tag{6.18}$$

6.3 Estimating $\boldsymbol{m}$, σ_ϵ^2 and $\boldsymbol{\Psi}$

Here we introduce methods for (1) accurately estimating the nonparametric component $\boldsymbol{m}$ in model (2.1) in the presence of error correlation and (2) estimating the variance σ_ϵ^2 and the correlation matrix $\boldsymbol{\Psi}$ of the errors associated with model (2.1). Estimating $\boldsymbol{m}$, σ_ϵ^2 and $\boldsymbol{\Psi}$ is difficult because of the confounding between the linear, non-linear and correlation effects. We hope that the combined way of estimating $\boldsymbol{m}$, σ_ϵ^2 and $\boldsymbol{\Psi}$ proposed in this thesis will enable us to do well when estimating $\boldsymbol{\beta}$.

6.3.1 Estimating $\boldsymbol{m}$

In this section, we consider the issue of accurately estimating the nonparametric component $\boldsymbol{m}$ in model (2.1) when the model errors are correlated. Recall that we need an accurate estimation of $\boldsymbol{m}$ for estimating the exact conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ in the plug-in method in (6.17). We propose estimating $\boldsymbol{m}$ via $\widehat{\boldsymbol{m}}_{\boldsymbol{\Omega},h}$, with $\boldsymbol{\Omega} = \boldsymbol{I}$ and with h chosen by cross-validation, modified for correlated errors.

Throughout this section, we thus consider that $\boldsymbol{\Omega} = \boldsymbol{I}$. We also let $\boldsymbol{X}_i^T = (1, X_{i1}, \ldots, X_{ip})$ denote the i^{th} row of the matrix $\boldsymbol{X}$ in (2.2).

To assess the accuracy of $\widehat{\boldsymbol{m}}_{\boldsymbol{I},h}$ as an estimator of $\boldsymbol{m}$ for a given amount of smoothing h, we use the mean average squared error of $\widehat{\boldsymbol{m}}_{\boldsymbol{I},h}$:

$$
\begin{aligned}
MASE(h;\boldsymbol{I}) &= E\left[\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{m}_{\boldsymbol{I},h}(Z_i) - m(Z_i)\Big)^2\Big|\boldsymbol{X},\boldsymbol{Z}\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{m}_{\boldsymbol{I},h}(Z_i) + \boldsymbol{X}_i^T\boldsymbol{\beta} - m(Z_i) - \boldsymbol{X}_i^T\boldsymbol{\beta}\Big)^2\Big|\boldsymbol{X},\boldsymbol{Z}\right] \\
&\equiv E\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Y}_i - E(Y_i|\boldsymbol{X}_i, Z_i)\right)^2\Big|\boldsymbol{X},\boldsymbol{Z}\right]
\end{aligned}
\tag{6.19}
$$

where $\widehat{Y}_i = \widehat{m}_{\boldsymbol{I},h}(Z_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}$ and $E(Y_i|\boldsymbol{X}_i, Z_i) = m(Z_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}$. We define the $MASE$-optimal amount of smoothing for accurate estimation of $\boldsymbol{m}$ via $\widehat{\boldsymbol{m}}_{\boldsymbol{I},h}$ as:

$$
h_{MASE}^{\boldsymbol{I}} = \underset{h}{argmin}\; MASE(h;\boldsymbol{I}). \tag{6.20}
$$

From (6.19) we can see that $h_{MASE}^{\boldsymbol{I}}$ depends on the bias and variance associated with estimating the non-parametric component $\boldsymbol{m}$, which in turn depend on $\boldsymbol{m}$ itself. Since in practice $\boldsymbol{m}$ is unknown, $h_{MASE}^{\boldsymbol{I}}$ has to be estimated.

To estimate $h_{MASE}^{\boldsymbol{I}}$, we propose using the modified (or leave-$(2l+1)$-out) cross-validation method originally formulated by Hart and View (1990) in the context of density estimation and studied by Chu and Marron (1991) and Härdle and Vieu (1992) in the context of nonparametric regression with correlated errors. Aneiros Pérez and Quintela del Río (2001b) recommend modified cross-validation in the context of partially linear models with α-mixing errors. These authors used a version of the Speckman estimator with boundary-adjusted Gasser-Müller weights to estimate $\boldsymbol{m}$.

The modified cross-validation method estimates $h_{MASE}^{\boldsymbol{I}}$ by minimizing an estimator of $MASE(h;\boldsymbol{I})$:

$$
\widehat{MASE}(h;\boldsymbol{I}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Y}_{i,h}^{(-i;\,l)} - Y_i\right)^2 \tag{6.21}
$$

This estimator is obtained from (6.19) by dropping the outer expectation sign, substituting $E(Y_i|\boldsymbol{X}_i, Z_i)$ with Y_i, and replacing $\widehat{Y}_i$ with $\widehat{Y}_{i,h}^{(-i;\,l)}$, a prediction of $Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} +$

$m(Z_i) + \epsilon_i$ based on data points $(Y_j, X_{j1}, \ldots, X_{jp}, Z_j)$ which are believed to be uncorrelated with Y_i. More specifically,

$$\widehat{Y}_{i,h}^{(-i;\,l)} = \boldsymbol{X}_i^T \widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}^{(-i;\,l)} + \widehat{m}_{\boldsymbol{I},h}^{(-i;\,l)}(Z_i), \tag{6.22}$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}^{(-i;\,l)}$ and $\widehat{m}_{\boldsymbol{I},h}^{(-i;\,l)}(Z_i)$ are estimators of $\boldsymbol{\beta}$ and $m(Z_i)$ obtained from the data points $(Y_j, X_{j1}, \ldots, X_{jp}, Z_j)$ with j such that $|i-j| > l$. The estimation procedure used for obtaining $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}^{(-i;\,l)}$ and $\widehat{m}_{\boldsymbol{I},h}^{(-i;\,l)}(Z_i)$ is the same as that utilized for obtaining $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ and $\widehat{m}_{\boldsymbol{I},h}(Z_i)$.

Recall that the estimation procedure utilized for obtaining the estimators $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h}$ and $\widehat{m}_{\boldsymbol{I},h}(Z_i)$ of $\boldsymbol{\beta}$ and $\boldsymbol{m}$ is the usual backfitting algorithm, with a (centered) local linear smoother matrix in the smoothing step. However, the backfitting algorithm allows us to evaluate $\widehat{m}_{\boldsymbol{I},h}^{(-i;\,l)}(\cdot)$ only at Z_j's with j such that $|i-j| > l$. We cannot evaluate $\widehat{m}_{\boldsymbol{I},h}^{(-i;\,l)}(\cdot)$ at Z_i. To overcome this problem, we propose to estimate $\boldsymbol{\beta}$ and $m(Z_i)$ as indicated below.

We first carry out the usual backfitting algorithm on all data to obtain the estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ of $\boldsymbol{\beta}$ using all n data points. We then define the partial residuals:

$$r_{j,h} = Y_j - \boldsymbol{X}_j^T \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h},\ j = 1, \ldots, n. \tag{6.23}$$

From now on, these residuals will become working responses for the modified cross-validation and our 'data set' is $(r_{j,h}, Z_j), j = 1, \ldots, n$. Fix i, $1 \leq i \leq n$. We temporarily remove from the 'data' the $(2l+1)$ 'data points' $(r_{j,h}, Z_j)$ with $|i-j| \leq l$. We use the remaining $n-(2l+1)$ data points in a usual local linear regression to obtain the $n-(2l+1)$ estimators $\widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_i)$ and $\widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_j)$, with j such that $|i-j| > l$. These estimators are not centered. Subtracting the average of $\widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_j)$, with j such that $|i-j| > l$ from $\widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_i)$ yields a centered estimator for $m(Z_i)$:

$$\widehat{m}_{\boldsymbol{\Omega},h}^{(-i;\,l)}(Z_i) = \widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_i) - \frac{1}{\#\{j:\ |i-j| > l\}} \sum_{j:\,|i-j|>l} \widehat{m}_{\boldsymbol{\Omega},h}^{*(-i;\,l)}(Z_j). \tag{6.24}$$

The centering approach used above is admittedly ad-hoc, but nevertheless attempts to address the need of subjecting $m(\cdot)$ to an appropriate identifiability restriction.

Next, we use the estimators in (6.24) in a computationally feasible modified cross-validation criterion:

$$MCV_l(h) = \frac{1}{n}\sum_{i=1}^{n}\left(r_{i,h} - \widehat{m}_{\boldsymbol{\Omega},h}^{(-i;\,l)}(Z_i)\right)^2. \tag{6.25}$$

Minimizing this criterion yields the desired cross-validation amount of smoothing for accurate estimation of $\boldsymbol{m}$ via $\widehat{\boldsymbol{m}}_{\boldsymbol{I},h}$ when the model errors are correlated.

Note that it is possible to compute a full scale modified cross-validation criterion, by calculating a different estimator of $\boldsymbol{\beta}$ for each i. Specifically, we could replace $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ in the right side of (6.23) with $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}^{(-i;\,l)}$, the estimator obtained from all data less those data points $(Y_j, X_{j1}, \ldots, X_{jp}, Z_j)$ with j such that $|i-j| \leq l$. However, computing the full scale modified cross-validation criterion would be more involved than computing the computationally convenient criterion in (6.25). Given that $\boldsymbol{\beta}$ is easier to estimate than $\boldsymbol{m}$, we believe that the computational simplification used to estimate $\boldsymbol{\beta}$ 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.

Although we do not have theoretical results that establish the properties of the modified cross-validation method, our simulation study suggests that it has reasonable finite sample performance and that it produces a reliable estimator of $\boldsymbol{m}$, provided l is taken to be large enough. It is not clear how to best choose l in practice. Recall that l should be specified such that the correlation between Y_i and $(Y_j, X_{j1}, \ldots, X_{jp}, Z_j)$, with $|i-j| \leq l$, is effectively removed when predicting Y_i by the value $\widehat{Y}_{i,h}^{-i;l}$ in (6.22). Choosing an l value that is too small may not succeed in removing the correlation between these data values, therefore producing an undersmoothed estimator of $\boldsymbol{m}$. Choosing an l value that is too large may remove too much of the underlying systematic structure in the data, therefore producing an estimator of $\boldsymbol{m}$ that is oversmoothed. Whenever possible, one should examining a whole range of values for l to gain more understanding about the sensitivity of the final results to the choice of l. Our simulation study suggests that small values of l should probably be avoided.

6.3.2 Estimating σ_ϵ^2 and Ψ

In this section, we propose a method for estimating the variance σ_ϵ^2 and correlation matrix Ψ of the errors associated with model (2.1). The method we propose relies on assumption (A2), that the model errors follow a stationary autoregressive process of unknown, but finite, order. To estimate the order and the corresponding parameters of this process, we apply standard time series techniques to suitable estimators of the model errors. Monte Carlo simulation studies conducted in Chapter 9 indicate that this method performs reasonably well.

Assumption (A2) will clearly not be appropriate for all applications. However, we expect it to cover those situations where the errors can be assumed to be realizations of a stationary stochastic process. Indeed, it can be shown that almost any stationary stochastic process can be modelled as an unique infinite order autoregressive process, independent of the origin of the process. In practice, finite order autoregressive processes are sufficiently accurate because higher order parameters tend to become small and not significant for estimation (Bos, de Waele, Broersen, 2002).

If the ϵ_i's were observed, we could estimate the order R of the autoregressive process they are assumed to follow by using the finite sample criterion for autoregressive order selection developed by Broersen (2000). This criterion selects the order of the process by achieving a trade-off between over-fitting (selecting an order that is too large) and under-fitting (selecting an order that is too small). Traditional autoregressive order selection criteria either fail to resolve these issues (i.e., the Akaike Information Criterion) or address just the issue of over-fitting (i.e., the corrected Akaike Information Criterion). In addition, Broersen's criterion performs well even when the order of the autoregressive error process is large.

After estimating R, we could estimate the error variance σ_ϵ^2 and the corresponding autoregressive parameters $\phi_1, \ldots, \phi_R$ by using Burg's algorithm. This algorithm is described, for instance, in Brockwell and Davis (1991). A comparison of various methods for autoregressive parameter estimation has shown that the Burg algorithm is the preferred method (Broersen, 2000). Finally, we could estimate the error correlation matrix $\boldsymbol{\Psi}$ by replacing $\phi_1, \ldots, \phi_R$ with their estimated values in the expression for $\boldsymbol{\Psi}$ provided in Comment 2.2.1. For instance, if R was estimated to be 1, we would estimate the $(i,j)^{th}$ element of $\boldsymbol{\Psi}$ as:

$$\widehat{\Psi}_{ij} = \widehat{\phi}_1^{\,|i-j|}$$

where $\widehat{\phi}_1$ is the estimator of the autoregressive coefficient ϕ_1.

However, the ϵ_i's are unobserved, so we must first estimate them via suitably defined model residuals and then apply the methodology described above to these residuals in order to obtain the desired estimators of σ_ϵ^2 and $\boldsymbol{\Psi}$.

We propose to estimate the vector of errors $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ by the model residuals $\widehat{\boldsymbol{\epsilon}}_{\boldsymbol{I},h} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},h} - \widehat{\boldsymbol{m}}_{\boldsymbol{I},h}$, where h is chosen by modified cross-validation, as described in Section 6.3.1. As argued in Section 6.3.1, this choice of h is expected to provide an accurate estimator for $\boldsymbol{X\beta}+\boldsymbol{m}$, and therefore a reasonable estimator for $\boldsymbol{\epsilon} = \boldsymbol{Y}-\boldsymbol{X\beta}-\boldsymbol{m}$.

For those applications where the reasonableness of assumption (A2) is questionable, we believe that one could still use the modified cross-validation residuals to estimate the model errors, since the modified cross-validation method does not rely on explicitly incorporating the error correlation structure. For instance, under the more general assumption

(A1), one could estimate σ_ϵ^2 and $\boldsymbol{\Psi} = (\Psi_{i,j})$ from $\widehat{\boldsymbol{\epsilon}}_{I,h} \equiv (\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n)^T$ 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 pursue this approach in this thesis.

6.4 Choosing h for $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, \boldsymbol{S}_h^c}$

We conclude this chapter by discussing the choice of smoothing parameter h for the estimated modified local linear backfitting estimator $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, \boldsymbol{S}_h^c}$. As indicated in Section 6.1, we denote $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, \boldsymbol{S}_h^c}$ by $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, h}$ to emphasize its dependence upon h. Our theoretical goal is to choose values of h which minimize measures of accuracy for $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, h}$ similar to those introduced for $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{I,h}$ and $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1}, h}$. Namely, if $Bias(h; \widehat{\boldsymbol{\Psi}}^{-1}) \equiv \boldsymbol{c}^T Bias(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, h} | \boldsymbol{X}, \boldsymbol{Z})$ and $Var(h; \widehat{\boldsymbol{\Psi}}^{-1}) \equiv \boldsymbol{c}^T Var(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, h} | \boldsymbol{X}, \boldsymbol{Z}) \boldsymbol{c}$, we wish to choose the value of h that minimizes the quantity $MSE(h; \widehat{\boldsymbol{\Psi}}^{-1})$, obtained by taking $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$ in (6.4). Denote this value by $h_{MSE}^{\widehat{\boldsymbol{\Psi}}^{-1}}$. In practice, we have to estimate this value from the data. The difficulty that we face is that, since $\widehat{\boldsymbol{\Psi}}$ is estimated and thus random, an expression for $MSE(h; \widehat{\boldsymbol{\Psi}}^{-1})$ is not tractable.

To avert this issue, we ignore the effect of estimating $\boldsymbol{\Psi}$ and simply replace $\boldsymbol{\Psi}$ by $\widehat{\boldsymbol{\Psi}}$ in the expression for $MSE(h; \boldsymbol{\Psi}^{-1})$. We have seen earlier in this thesis that, under certain conditions, $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1}, h}$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1}, h}$ are asymptotically 'close', so we expect our approach to be reasonable for large sample sizes.

We propose to choose h using suitable modifications of the EBBS and plug-in methods discussed in Sections 6.2.2 and 6.2.3. The global and local modified EBBS methods for choosing the smoothing parameter h of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ attempt to estimate $h_{MSE}^{\widehat{\boldsymbol{\Psi}}^{-1}}$:

$$\widehat{h}_{MSE}^{\widehat{\boldsymbol{\Psi}}^{-1}} = \underset{h}{argmin}\{\widehat{Bias}(h; \widehat{\boldsymbol{\Psi}}^{-1}) + \widehat{Var}(h; \widehat{\boldsymbol{\Psi}}^{-1})\}, \tag{6.26}$$

with $h \in \mathcal{H}$. For both methods, $\widehat{Var}(h; \widehat{\boldsymbol{\Psi}}^{-1})$ is computed by substituting $\boldsymbol{\Psi}$ with $\widehat{\boldsymbol{\Psi}}$ into the expression of $Var(h; \boldsymbol{\Psi}^{-1})$, the exact conditional variance of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$. This expression is obtained by taking $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ in (6.1). The global modified EBBS method estimates $Bias(h; \widehat{\boldsymbol{\Psi}}^{-1})$ empirically by fitting a global ordinary least squares regression model to the 'data' points $\left\{\left(h(k), \boldsymbol{c}^T\widehat{\beta}_{\widehat{\boldsymbol{\Psi}}^{-1},h(k)}\right) : k = 1, \ldots, N\right\}$. We denote the amount of smoothing this method yields by $\widehat{h}_{EBBS-G}^{\widehat{\boldsymbol{\Psi}}^{-1}}$. The local modified EBBS method uses only a fraction of these data to accomplish the same task. We denote the amount of smoothing supplied by this method by $\widehat{h}_{EBBS-L}^{\widehat{\boldsymbol{\Psi}}^{-1}}$.

The plug-in method for choosing the value of h in $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},h}$ tries to estimate (an approximation to) $h_{MSE}^{\widehat{\boldsymbol{\Psi}}^{-1}}$:

$$\widehat{h}_{MSE}^{\widehat{\boldsymbol{\Psi}}^{-1}} = \underset{h\in\mathcal{H}}{argmin}\{\widehat{Bias}(h; \widehat{\boldsymbol{\Psi}}^{-1}) + \widehat{Var}(h; \widehat{\boldsymbol{\Psi}}^{-1})\} \equiv \widehat{h}_{PLUG-IN}^{\boldsymbol{\Psi}^{-1}}. \tag{6.27}$$

Here, $\widehat{Var}(h; \widehat{\boldsymbol{\Psi}}^{-1})$ is as above, and $\widehat{Bias}(h; \widehat{\boldsymbol{\Psi}}^{-1})$ is constructed by substituting $\boldsymbol{\Psi}$ with $\widehat{\boldsymbol{\Psi}}$ into the expression of $Bias(h; \boldsymbol{\Psi}^{-1})$, the exact conditional bias of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Psi}^{-1},h}$. This expression is obtained by taking $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$ in (6.3).

Chapter 7

Confidence Interval Estimation and Hypothesis Testing

In this chapter, we develop statistical methods for assessing the magnitude and statistical significance of a linear combination of linear effects $\boldsymbol{c}^T\boldsymbol{\beta}$ in model (2.1), where $\boldsymbol{c} = (c_0, \ldots, c_p)^T$ is a known vector with real valued components. Specifically, we propose several confidence intervals for assessing the magnitude of $\boldsymbol{c}^T\boldsymbol{\beta}$, as well as several tests of hypotheses for testing whether $\boldsymbol{c}^T\boldsymbol{\beta}$ is significantly different than some known value of interest.

7.1 Confidence Interval Estimation

We propose to construct approximate $100(1-\alpha)\%$ confidence intervals for $\boldsymbol{c}^T\boldsymbol{\beta}$ from the usual, modified or estimated modified local linear backfitting estimators considered in this thesis, and their associated estimated standard errors. In what follows, we use the notation in Section 6.1 to denote these estimators generically by $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, where $\boldsymbol{\Omega}$ can be $\boldsymbol{I}$, $\boldsymbol{\Psi}^{-1}$ or $\widehat{\boldsymbol{\Psi}}^{-1}$, respectively, and h is an amount of smoothing that must be chosen from the data. Our confidence intervals use an estimated standard error $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ obtained as follows:

$$\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}) = \sqrt{\widehat{Var}(h;\boldsymbol{\Omega})}. \tag{7.1}$$

Here, $\widehat{Var}(h;\boldsymbol{\Omega})$ is an estimator of $Var(h;\boldsymbol{\Omega})$, the conditional variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$. Specifically, for $\boldsymbol{\Omega} = \boldsymbol{I}$, $\widehat{Var}(h;\boldsymbol{\Omega})$ is defined as in (6.14). For $\boldsymbol{\Omega} = \boldsymbol{\Psi}^{-1}$, if σ_ϵ^2 is unknown, $\widehat{Var}(h;\boldsymbol{\Omega})$ is defined as in (6.15). Finally, for $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$, $\widehat{Var}(h;\boldsymbol{\Omega})$ is obtained from (6.15) by replacing $\boldsymbol{\Psi}$ with $\widehat{\boldsymbol{\Psi}}^{-1}$. Note that the standard error expression in (7.1) does not account for the estimation of σ_ϵ^2 and $\boldsymbol{\Psi}$ when these quantities are unknown, nor does it account for the data-dependent choice of h. Rather, it is a purely 'plug-in' expression.

The performance of a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ depends to a great extent on how well we choose the smoothing parameter h of the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. A poor choice of h can affect the mean squared error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, resulting in a confidence interval with poor coverage and/or length properties. We want to choose an h for which (i) the bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ is small, so the interval is centered near the true $\boldsymbol{c}^T\boldsymbol{\beta}$, and (ii) the variance of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ is small, so the interval has small length. Choosing h to ensure that the confidence interval is valid (in the sense of achieving the nominal coverage) and has reasonable length is crucial to the quality of inferences about $\boldsymbol{c}^T\boldsymbol{\beta}$.

In this thesis, we choose the amount of smoothing h needed for constructing confidence intervals for $\boldsymbol{c}^T\boldsymbol{\beta}$ via the following data-driven choices of h, introduced in Chapter 6:

1. the (local) modified EBBS choice, $\widehat{h}^{\boldsymbol{\Omega}}_{EBBS-L}$;

2. the global modified EBBS choice, $\widehat{h}^{\boldsymbol{\Omega}}_{EBBS-G}$;

3. the (non-asymptotic) plug-in choice, $\widehat{h}^{\boldsymbol{\Omega}}_{PLUG-IN}$.

Recall that each of these choices is expected to yield an accurate estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ of $\boldsymbol{c}^T\boldsymbol{\beta}$. Throughout the rest of this chapter, unless otherwise specified, we assume that the smoothing parameter h of the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ refers to any of $\widehat{h}^{\boldsymbol{\Omega}}_{EBBS-L}$, $\widehat{h}^{\boldsymbol{\Omega}}_{EBBS-G}$ or $\widehat{h}^{\boldsymbol{\Omega}}_{PLUG-IN}$.

The performance of a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ also depends on how well we estimate $SE(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$, the true standard error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. As already mentioned, we will estimate $SE(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ by $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ as defined in (7.1). Recall that $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ depends on another smoothing parameter, needed for estimating $\boldsymbol{\Psi}$ via $\widehat{\boldsymbol{\Psi}}$, as described in Section 6.3.2. It is not clear whether the modified cross-validation choice of smoothing proposed in Section 6.3.2 yields a reasonable estimator of $SE(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. The Monte Carlo simulations presented in Chapter 8 will shed more light on this issue.

The standard $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ is given by

$$\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} \pm z_{\alpha/2}\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}), \tag{7.2}$$

where $z_{\alpha/2}$ is the $100(1-\alpha)\%$ quantile of the standard normal distribution. According to the asymptotic results in this thesis, the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ is biased in finite samples. Consequently, the standard confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ may not be correctly centered and may not provide $1-\alpha$ coverage. We propose two strategies for dealing with this problem. One strategy is to perform a bias adjustment to the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, to try to ensure that the confidence interval is better centered. This approach, referred to as bias-adjusted confidence interval construction, is discussed in Section 7.1.1. Another strategy is to perform an adjustment to the estimated standard error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. The purpose of this adjustment is to inflate the estimated standard error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ to reflect the bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$. This approach, referred to as standard error-adjusted confidence interval construction, is discussed in Section 7.1.2.

Throughout, we assume we can use standard normal probability tables to construct the confidence interval in (7.2) and those proposed in Sections 7.1.1 - 7.1.2. This assumption is justified provided the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ is asymptotically normal and our standard error estimators are consistent. Opsomer and Ruppert (1999) established the asymptotic normality of the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ for the case when the model errors are uncorrelated and $\boldsymbol{\Omega} = \boldsymbol{I}$. However, no asymptotic normality results are available as yet for the cases when the model errors are correlated, for either $\boldsymbol{\Omega} = \boldsymbol{I}$ or more general $\boldsymbol{\Omega}$. The simulations conducted in Chapter 8 support the use of normal tables when constructing 95% confidence intervals.

Note that, for small sample sizes, one might widen the confidence intervals by using t-tables instead of standard normal probability tables. The issue of how one might specify the degrees of freedom involved in these t-tables needs to be considered carefully and is beyond the scope of this thesis.

7.1.1 Bias-Adjusted Confidence Interval Construction

The idea underlying the bias-adjusted confidence interval estimation of $\boldsymbol{c}^T\boldsymbol{\beta}$ is to first adjust the estimator $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ for possible finite sample bias effects. Then a bias-adjusted $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ is given by:

$$\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \widehat{Bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}) \pm z_{\alpha/2}\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}), \tag{7.3}$$

where $\widehat{Bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ estimates the finite sample conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, given $\boldsymbol{X}$ and $\boldsymbol{Z}$, and is defined either as in (6.17) for $h = \widehat{h}^{\boldsymbol{\Omega}}_{PLUG-IN}$, or as in (6.10) for $h = \widehat{h}^{\boldsymbol{\Omega}}_{EBBS-G}$ and $h = \widehat{h}^{\boldsymbol{\Omega}}_{EBBS-L}$. Neither of these bias expressions takes into account the data-dependent choice of h. Furthermore, these bias expressions do not account for the estimation of $\boldsymbol{\Psi}$ when $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}^{-1}$.

The length of the bias-adjusted confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ in (7.3) is the same as that of the standard confidence interval in (7.2). The coverage properties of the bias-adjusted confidence interval may, however, be better than those of the standard confidence interval, because the bias-adjusted confidence interval may be better centered.

Note that the estimated standard error $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ in (7.3) reflects the variability of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, instead of the variability of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \widehat{Bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. One could, of course, replace $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ by an estimator of the true standard error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \widehat{Bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. But such an estimator may be difficult to obtain in practice, unless one resorts to computationally expensive bootstrapping methods, and may not necessarily yield a confidence interval with better coverage properties than those of the standard confidence interval.

7.1.2 Standard Error-Adjusted Confidence Interval Construction

We have suggested in Section 7.1.1 that the standard confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ in (7.2) can be improved upon by replacing $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ with its bias-adjusted version $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \widehat{Bias}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. Another possible way to improve upon the standard confidence interval in (7.2) is to replace $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$ with $\sqrt{\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})}$, the square root of the estimated conditional mean squared error of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$. The motivation for this latter adjustment is that, compared to $\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$, $\sqrt{\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})}$ is a better measure of the uncertainty associated with estimating $\boldsymbol{c}^T\boldsymbol{\beta}$ via $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$, as it tries to account for the finite sample bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$.

A standard error-adjusted $100(1-\alpha)\%$ confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ is given by:

$$\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} \pm z_{\alpha/2}\sqrt{\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})} \tag{7.4}$$

where

$$\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}) = \left[\widehat{Bias}(h;\boldsymbol{\Omega})\right]^2 + \left[\widehat{SE}(h;\boldsymbol{\Omega})\right]^2. \tag{7.5}$$

Here, $\widehat{Bias}(h;\boldsymbol{\Omega})$ estimates the conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$, and is defined either as in (6.17) for $h = \widehat{h}^{\boldsymbol{\Omega}}_{PLUG-IN}$, or as in (6.10) for $h = \widehat{h}^{\boldsymbol{\Omega}}_{EBBS-G}$ and $h = \widehat{h}^{\boldsymbol{\Omega}}_{EBBS-L}$.

Note that the length of the standard error-adjusted confidence interval for $\boldsymbol{c}^T\boldsymbol{\beta}$ in (7.4) is wider than that of the standard confidence interval in (7.2) due to the fact that

$\sqrt{\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})} \geq \widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})$. This may translate into improved coverage properties for the standard error-adjusted confidence interval.

7.2 Hypothesis Testing

In this section, we exploit the duality between confidence interval estimation and hypothesis testing to develop tests of hypotheses for $\boldsymbol{c}^T\boldsymbol{\beta}$.

Suppose we are interested in testing the null hypothesis

$$H_0 : \; \boldsymbol{c}^T\boldsymbol{\beta} = \delta \tag{7.6}$$

against the alternative hypothesis

$$H_a : \; \boldsymbol{c}^T\boldsymbol{\beta} \neq \delta, \tag{7.7}$$

where δ is a constant.

From the confidence intervals introduced in Section 7.1, we construct three test statistics for testing H_0 against H_a:

$$Z^{(1)}_{\boldsymbol{\Omega},h} = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \delta}{\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})}, \tag{7.8}$$

$$Z^{(2)}_{\boldsymbol{\Omega},h} = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \widehat{Bias}(h;\boldsymbol{\Omega}) - \delta}{\widehat{SE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})}, \tag{7.9}$$

$$Z^{(3)}_{\boldsymbol{\Omega},h} = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h} - \delta}{\sqrt{\widehat{MSE}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Omega},h})}}. \tag{7.10}$$

We will reject H_0 at significance level α if $|Z^{(k)}_{\boldsymbol{\Omega},h}| > z_{\alpha/2}$.

Chapter 8

Monte Carlo Simulations

In this chapter we report the results of a Monte Carlo study on the finite sample properties of estimators and confidence intervals for the linear effect β_1 in the model:

$$Y_i = \beta_0 + \beta_1 X_i + m(Z_i) + \epsilon_i, \ i = 1, \dots, n, \tag{8.1}$$

obtained by taking $p = 1$ in (2.1). Even though this model is not too complicated, we hope that it will allow us to understand how the properties of these estimators and confidence intervals will be affected by (1) dependency between the X_i's and the Z_i's, and (2) correlation amongst the ϵ_i's.

For our study, we have deliberately chosen to use a context similar to that considered by Opsomer and Ruppert (1999) for independent ϵ_i's, so that we can make direct comparisons. Given this context, the main goals of our simulation study were to:

1. Compare the expected log mean squared error (MSE) of the estimators for β_1.
2. Compare the performance of the confidence intervals for β_1 built from these estimators and their associated standard errors.

The rest of this chapter is organized as follows. In Section 8.1, we discuss how we generated the data in our simulation study. In Section 8.2, we provide an overview of the estimators for β_1 considered in this study. We also specify the methods used for choosing the smoothing parameters of these estimators. In Section 8.3, we compare the expected log mean squared errors (MSE) of the estimators for all simulation settings in our study. Finally, in Sections 8.4 and 8.5, we assess the coverage and length properties of various approximate 95% confidence intervals for β_1 constructed from these estimators and their associated approximate standard errors.

8.1 The Simulated Data

The data (Y_i, X_i, Z_i), $i = 1, \ldots, n$, in our simulation study were generated from model (8.1) using a modification of the simulation setup adopted by Opsomer and Ruppert (1999). Specifically, we took the sample size n to be 100 and set the values of the linear parameters β_0 and β_1 to zero. We considered two $m(\cdot)$ functions:

- $m_1(z) = 2sin(3z) - 2(cos(0) - cos(3))/3,\ z \in [0,1]$;
- $m_2(z) = 2sin(6z) - 2(cos(0) - cos(6))/6,\ z \in [0,1]$.

The Z_i's were equally spaced on $[0,1]$, being defined as $Z_i = i/(n+1)$. Furthermore, $X_i = g(Z_i) + \eta_i$, with $g(z) = 0.4z + 0.3, z \in [0,1]$, and $\eta_i = (1 - 0.4)U_i - 0.3$, where the U_i's were independent, identically distributed having a $Unif(0,1)$ distribution.

The ϵ_i's followed a stationary $AR(1)$ model with normal distribution:

$$\epsilon_i = \rho\epsilon_{i-1} + u_i, \tag{8.2}$$

where ρ is an autoregressive parameter quantifying the degree of correlation amongst the ϵ_i's. The u_i's were independent, identically distributed normal random variables having mean 0 and standard deviation $\sigma_u = 0.5$. The u_i's were independent of the ϵ_i's. In our simulation study, we used $\rho = 0$ to include the case of independence, as well as $\rho =$ 0.2, 0.4, 0.6 and 0.8 to model positive correlation ranging from weak to strong.

The simulation settings corresponding to $\rho = 0$ (the case of independent errors) are the same as those considered by Opsomer and Ruppert (1999), with the following exceptions: (i) we considered $n = 100$ instead of $n = 250$, (ii) we 'centered' the $m(\cdot)$ functions, that is, we subtracted a constant so that these functions integrate to 0 over the interval $[0, 1]$ and (iii) we scaled the errors η_i to have $E(\eta_i) = 0$ instead of $E(\eta_i) = 0.3$. Opsomer and Ruppert did not specify what value they used for β_1.

For each model configuration, we generated 500 data sets. Note that there are 10 model configurations altogether, one for each combination of autoregressive parameter ρ and non-linear effect $m(\cdot)$ considered.

Figure 8.1 displays data generated from model (8.1) for $\rho = 0, 0.4, 0.8$ and $m_1(z)$. Figure 8.2 provides the same display for $m_2(z)$. The responses Y_i are qualitatively different for different values of ρ. For $\rho = 0$, the responses vary randomly about the $m(\cdot)$ curve. As ρ increases from 0.4 to 0.8, the variation of the Y_i's about the curve $m(\cdot)$ makes it virtually impossible to distinguish the non-linear signal $m(\cdot)$ from the autoregressive noise that masks it.

8.2 The Estimators

In this section, we provide an overview of the estimators for the linear effect β_1 in model (8.1) considered in our simulation study. We also provide an overview of the methods used for choosing the smoothing parameter of these estimators.

Note that $\beta_1 = \boldsymbol{c}^T\boldsymbol{\beta}$, where $\boldsymbol{c} = (0, 1)^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$. The estimators of β_1 considered in our simulation study are of the form $\widehat{\beta}_1 = \boldsymbol{c}^T\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is:

(i) $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\mathbb{S}_h^c}$, the usual backfitting estimator defined in (3.4) with $\boldsymbol{\Omega} = \boldsymbol{I}$;

(ii) $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Psi}}^{-1},\mathbb{S}_h^c}$, the estimated modified backfitting estimator defined in (3.4) with $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Psi}}$;

(iii) $\widehat{\beta}_{(I-\mathbb{S}_h^c)^T,\mathbb{S}_h^c}$, the usual Speckman estimator defined in (3.4) with $\Omega = (I - \mathbb{S}_h^c)^T$.

In all three estimators, $\mathbb{S}_h^c$ is a centered smoother matrix, defined in terms of the Epanechnikov kernel in (3.9). For the two backfitting estimators, we take $\mathbb{S}_h^c$ to be a centered local linear smoother matrix. For the usual Speckman estimator, we take $\mathbb{S}_h^c$ to be a centered local constant smoother matrix with Nadaraya-Watson weights. The latter choice is motivated by the fact that the usual Speckman estimator is typically used with local constant smoother matrices with kernel weights. We are not sure to what extent the differences in performance between the usual Speckman estimator and the two backfitting estimators may be due to this difference in the method of local smoothing.

Note that $\widehat{\beta}_{\Psi^{-1},\mathbb{S}_h^c}$, the modified backfitting estimator obtained from (3.4) with $\Omega = \Psi$, was omitted from our simulation study. This estimator may have value as a benchmark, but has no practical value due to the fact that the error correlation matrix Ψ is never fully known in applications. For similar reasons, we also omitted $\widehat{\beta}_{(I-S_h^c)^T\Psi^{-1},\mathbb{S}_h^c}$, the modified Speckman estimator obtained from (3.4) with $\Omega = (I - S_h^c)^T\Psi^{-1}$. Another estimator not included in our study is $\widehat{\beta}_{(I-S_h^c)^T\widehat{\Psi}^{-1},\mathbb{S}_h^c}$, the estimated modified Speckman estimator obtained from (3.4) with $\Omega = (I - S_h^c)^T\widehat{\Psi}^{-1}$. Recall that Aneiros Pérez and Quintela del Río (2001a) investigated the large sample properties of a similar estimator, based on local constant smoothing with Gasser-Müller weights. These authors have a suggestion for estimating Ψ from the data, but they did not explore how well it works in practice. In our simulation study, the estimator $\widehat{\beta}_{\widehat{\Psi}^{-1},\mathbb{S}_h^c}$ - which is similar to $\widehat{\beta}_{(I-S_h^c)^T\widehat{\Psi}^{-1},\mathbb{S}_h^c}$ - does poorly in general. We believe this may be due to a combination of the following: (1) Ψ is hard to estimate in the presence of confounding between the linear, non-linear and correlation effects and (2) the additional variability introduced by estimating Ψ is not properly taken into account when selecting the smoothing parameter and when constructing standard errors for $\widehat{\beta}_{\widehat{\Psi}^{-1},\mathbb{S}_h^c}$ from small samples. We suspect that, if one were to use the methods proposed in this thesis to estimate Ψ for computing $\widehat{\beta}_{(I-S_h^c)^T\widehat{\Psi}^{-1},\mathbb{S}_h^c}$, one would also get an estimator with poor finite sample behaviour.

All three estimators in (i)-(iii) require a data driven choice of smoothing parameter. For the three backfitting estimators we consider EBBS-G and EBBS-L (see Section 6.2.2) and PLUG-IN (see Section 6.2.3). For the usual Speckman estimator, we use cross-validation, modified for correlated errors (MCV) and for boundary effects. The MCV criterion is similar to that in (6.21), namely:

$$MCV_l(h) = \frac{1}{n}\sum_{i=1}^{n} W(Z_i)\left(Y_i - \widehat{Y}_{i,h}^{(-i;l)}\right)^2.$$

Here, $\widehat{Y}_{i,h}^{(-i;l)}$ is obtained as in (6.22), but with $\boldsymbol{\Omega} = (\boldsymbol{I} - \mathbb{S}_h^c)^T$, where $\mathbb{S}_h^c$ is the centered local constant smoother matrix. Also, W is a weight function introduced to allow elimination (or at least significant reduction) of boundary effects that may affect the estimation of the non-linear effect m in model (8.1), and hence the prediction of Y_i. W is defined as in Chu and Marron (1991):

$$W(u) = \begin{cases} \frac{5}{3}, & \text{if } \frac{1}{5} \leq u \leq \frac{4}{5}; \\ 0, & \text{if } 0 \leq u < \frac{1}{5} \text{ or } \frac{4}{5} < u \leq 1. \end{cases}$$

Recall that EBBS-G depends on the tuning parameters l, N and T, whereas EBBS-L depends on the tuning parameters l, N, T, k_1 and k_2. Also, recall that PLUG-IN and MCV depend on the tuning parameter l. In our simulation study, we consider $N = 50$, $T = 2$, $k_1 = 5$, $k_2 = 5$, and $l = 0, 1, \ldots, 10$.

For convenience, throughout the remainder of this chapter, we use the notation $\widehat{\beta}_{U,PLUG-IN}^{(l)}$, $\widehat{\beta}_{U,EBBS-G}^{(l)}$ and $\widehat{\beta}_{U,EBBS-L}^{(l)}$ for the usual local linear backfitting estimators of β_1. We use the notation $\widehat{\beta}_{EM,PLUG-IN}^{(l)}$, $\widehat{\beta}_{EM,EBBS-G}^{(l)}$ and $\widehat{\beta}_{EM,EBBS-L}^{(l)}$ for the estimated modified local linear backfitting estimators. Finally, we use the notation $\widehat{\beta}_{S,MCV}^{(l)}$ to refer to the usual Speckman estimator of β_1. Wherever necessary, we refer to these estimators generically as $\widehat{\beta}_1^{(l)}$.

8.3 The MSE Comparisons

In this section, we identify the estimators $\widehat{\beta}_1^{(l)}$, including bandwidth selection methods, that appear to be best, in the sense of being most accurate for all simulation settings and for most values of l, the tuning parameter used in the modified cross-validation. Recall that the measure of accuracy of $\widehat{\beta}_1^{(l)}$ considered in this thesis is the conditional MSE of $\widehat{\beta}_1^{(l)}$, $MSE(\widehat{\beta}_1^{(l)})$, defined in (6.4). Specifics are provided below.

To compare the accuracy of two estimators for a given simulation setting, we look at the boxplot of differences in the log MSE's of these estimators. If the boxplot is symmetric about 0, then the two estimators have comparable accuracy. We also conduct a level 0.05 two-sided paired t-test to compare the expected log MSE's of the estimators. If the test is significant, we label the boxplot with an S. The log MSE's of the two estimators are evaluated from the 500 data sets generated for the given simulation setting.

For each backfitting estimation method (usual, estimated modified), we recommend a way to choose the smoothing parameter h. Then we compare the resulting backfitting estimators, including a comparison with the usual Speckman estimator to determine an estimator that is best, in the sense of being most accurate for all simulation settings and most values of l.

In Figures A.1-A.10 in Appendix A, we study the methods of bandwidth choice for the usual local linear backfitting estimator. We display boxplots of pairwise differences in the log MSE's of the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{U,EBBS-L}$, $l = 0, 1, \ldots, 10$. Each figure corresponds to a different simulation setting. From these figures, we see that $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ have comparable accuracy across all simulation settings, provided l is large enough, say $l \geq 4$. They also have better accuracy than $\widehat{\beta}^{(l)}_{U,EBBS-L}$, which performs poorly for several simulation settings (see, for instance, Figures A.6-A.7). Therefore, we recommend using PLUG-IN and EBBS-G to choose the smoothing parameter for the usual local linear backfitting estimator.

Figures A.11-A.20 display the corresponding plots for the estimated modified local linear backfitting estimator. We see that $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ is the most accurate across all simulation settings, provided l is large enough, say $l \geq 4$. We also see that $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ and $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$ perform very poorly relative to $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ for most simulation settings and most values of l. Therefore, we recommend using EBBS-G to choose the smoothing parameter for the estimated modified local linear backfitting estimator.

In Figures A.21-A.30 we compare estimators using our favourite bandwidth selection method. We display boxplots of pairwise differences in the log MSE's of the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$, $l = 0, 1, \ldots, 10$. Each figure corresponds to a different simulation setting. From these figures, we conclude that the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ have comparable accuracy for all simulation settings, provided l is large enough, say $l \geq 4$. The estimator $\widehat{\beta}^{(l)}_{S,MCV}$ is less accurate than these three estimators for most simulation settings and most values of l. In particular, plots such as those in Figures A.24, A.25, A.29 and A.30 strongly support the elimination of $\widehat{\beta}^{(l)}_{S,MCV}$. The poor performance of $\widehat{\beta}^{(l)}_{S,MCV}$ with respect to the log MSE criterion could be due to the fact that this estimator uses local constant smoothing, instead of local linear smoothing. But it could also be due to the fact that $\widehat{\beta}^{(l)}_{S,MCV}$ is computed with an MCV choice of smoothing. Recall that this choice attempts to estimate the amount of smoothing optimal for estimation of $\boldsymbol{X\beta} + \boldsymbol{m}$. It is not clear whether this choice will provide a reliable estimate of the amount of smoothing optimal for estimation of $\boldsymbol{c}^T\boldsymbol{\beta}$.

8.4 Confidence Interval Coverage Comparisons

In this section, we assess and compare the coverage properties of various confidence intervals for β_1 constructed from all estimators considered in our simulation study. Our goals are to:

1. Identify those estimators which yield standard confidence intervals for β_1 with good coverage properties across all simulation settings and most values of l.

2. Establish whether the coverage properties of standard confidence intervals for β_1 can be improved through bias or standard error adjustments.

To assess the coverage properties of a confidence interval C for a given simulation setting, we proceed as follows. We evaluate the confidence interval for each of the 500 simulated data sets. We calculate the proportion of these intervals which contain the true value of β_1 and denote it by $\widehat{p}$. If $\widehat{p} \pm 1.96\sqrt{\widehat{p}(1-\widehat{p})/500}$, the 95% confidence interval for the true coverage, contains the nominal level of C, we say that C is valid. If the upper (lower) confidence limit is smaller (bigger) than the nominal level of C, we say that C is anti-conservative (conservative).

The confidence intervals for β_1 considered in our simulation study fall into three categories: standard, bias-adjusted and standard-error adjusted, as defined in (7.2), (7.3) and (7.4).

8.4.1 Standard Confidence Intervals

We now assess the coverage properties of the standard 95% confidence intervals for β_1 obtained from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{U,EBBS-L}$, $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ and $\widehat{\beta}^{(l)}_{S,MCV}$, where $l = 0, 1, \ldots, 10$. Point estimates and 95% confidence interval estimates for the true coverage achieved by these intervals are displayed in Figures B.1-B.10 in Appendix B. Each figure corresponds to a different simulation setting.

Figures B.1-B.10 show that the standard confidence intervals constructed from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$ are valid for all simulation settings provided the value of l is large enough. However, the standard confidence intervals obtained from the estimators $\widehat{\beta}^{(l)}_{U,EBBS-L}$, $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ have extremely poor coverage for many simulation settings and for many values of l; see, for instance, Figures B.6 and B.7. In view of these findings, the preferred estimators for constructing standard confidence intervals for β_1 are $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$. The other estimators cannot be trusted to produce valid inferences on β_1. More details concerning our findings are provided below.

The standard confidence intervals constructed from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ are valid for all simulation settings, provided l is large enough, as shown in Table 8.1. From this table, we see that taking $l \geq 1$ when $\rho = 0.2$, $l \geq 2$ when $\rho = 0.4$, $l \geq 3$ when $\rho = 0.6$, and $l \geq 4$ when $\rho = 0.8$ yields valid intervals for the contexts considered. We recommend using these intervals to conduct inferences on β_1, with values of l that are large enough. Clearly, taking $l = 0, 1, 2, 3$ is not advised, unless one is certain that ρ is small.

What is not apparent from Table 8.1 is why the confidence intervals constructed from $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ are not valid for smaller values of l. Typically, for small l's, the estimates of β_1 constructed from the simulated data have a tendency to underestimate the true value of β_1 when $m(z) = m_2(z)$. Furthermore, the estimated standard errors associated with these estimates have a tendency to underestimate the true standard errors both when $m(z) = m_1(z)$ and when $m(z) = m_2(z)$. However, as l increases, the estimates of β_1 and their associated standard errors improve significantly for all simulation settings.

The standard confidence intervals constructed from the usual Speckman estimator $\widehat{\beta}^{(l)}_{S,MCV}$ are generally valid across all simulation settings even for smaller l values. However, $\widehat{\beta}^{(l)}_{S,MCV}$ does not yield valid confidence intervals when $m(z) = m_2(z)$ and (i) $\rho = 0.4$ and $l = 1$ or 4 and (ii) $\rho = 0.8$ and $l = 3, 4, 5, 6, 7, 8$ or 10. In these two cases, $\widehat{\beta}^{(l)}_{S,MCV}$ yields confidence intervals that are slightly anti-conservative. This lack of continuity in behaviour is of concern and might not be attributable to simulation variability. Indeed, Figures B.6-B.10 show that, for $m(z) = m_2(z)$, $\widehat{\beta}^{(l)}_{S,MCV}$ seems to exhibit an anti-conservative pattern for most l's.

When $\rho = 0$ and $m(z) = m_1(z)$, the standard confidence intervals obtained from the estimators $\widehat{\beta}^{(l)}_{U,EBBS-L}$, $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ provide the nominal coverage, regardless of how we choose l (see Figure B.1). However, when $\rho = 0$ and $m(z) = m_2(z)$, the intervals constructed from $\widehat{\beta}^{(l)}_{U,EBBS-L}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ are extremely anti-conservative for all values of l (see Figure B.6). In addition, the intervals constructed from $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ are mildly anti-conservative for many values of l (see Figure B.6).

As ρ increases, the coverage provided by some of the standard confidence intervals obtained from $\widehat{\beta}^{(l)}_{U,EBBS-L}$, $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ deteriorates for many small and/or large values of l, depending on the specification of $m(\cdot)$. For instance, when $m(z) = m_2(z)$, the coverage properties of the intervals constructed from $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ are extremely poor (see Figures B.7-B.10). The coverage properties of the intervals constructed from $\widehat{\beta}^{(l)}_{U,EBBS-L}$ are also poor for small ρ values (see Figures B.7-B.8). Finally, the coverage properties of the intervals constructed from $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ worsen as ρ increases, but not dramatically. We do not recommend using these intervals to carry out inferences on β_1.

8.4.2 Bias-Adjusted Confidence Intervals

In this section, we assess the coverage properties of the bias-adjusted 95% confidence intervals for β_1. We did not consider a bias-adjusted confidence interval for the usual Speckman estimator $\widehat{\beta}^{(l)}_{S,MCV}$, as this estimator is known to have good bias properties both when $\rho = 0$ (see Speckman, 1988) and when $\rho > 0$ (see Aneiros-Pérez and Quintela-del-Río, 2001a).

Plots (not shown) of the point estimates and 95% confidence interval estimates for the true coverage achieved by the bias-adjusted intervals yield some general conclusions.

Only the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ yield bias-adjusted confidence intervals that are valid for all simulation settings provided the value of l is large enough. These values of l are almost identical to those reported in Table 8.1. Again, we see that one should avoid using $l = 0, 1, 2, 3$ unless one is sure that ρ is small enough.

8.4.3 Standard Error-Adjusted Confidence Intervals

Here, we assess the coverage properties of the standard error-adjusted 95% confidence intervals for β_1. We did not consider a standard error-adjusted confidence interval for the usual Speckman $\widehat{\beta}^{(l)}_{S,MCV}$, due to its good bias properties. Plots (not shown) indicate that only the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ provide standard error-adjusted confidence intervals that are valid for all simulation settings provided the value of l is large enough. These values of l are nearly identical to those reported in Table 8.1. Yet again, we see that one should avoid using $l = 0, 1, 2, 3$ unless one is sure that ρ is small enough.

To sum up, we see no reason to recommend bias adjustments to the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ or to their associated standard errors. Indeed, such adjustments do not seem to improve the coverage properties of the confidence intervals obtained from these estimators.

8.5 Confidence Interval Length Comparisons

Recall from the previous section that we identified $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ as the only estimators of β_1 in our simulation study that yielded valid 95% standard confidence intervals for all simulation settings provided the value of l is large enough. The standard intervals based on $\widehat{\beta}^{(l)}_{S,MCV}$ were found to be competitive, but just not as good. Also recall that the coverage properties of the standard confidence intervals constructed from

$\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ could not be improved by performing bias-adjustments to these estimators or to their associated standard errors. Before recommending any of the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ for practical use, we must compare the lengths of the standard confidence intervals for β_1 constructed from these estimators. We choose to include standard intervals constructed from $\widehat{\beta}^{(l)}_{S,MCV}$ in our comparison to gain more understanding into their properties. When several confidence interval procedures are valid (in the sense of achieving the desired nominal level), we prefer the one with the shortest length.

In this section, we conduct visual and formal comparisons of the lengths of the standard 95% confidence intervals for β_1 constructed from these estimators. We only consider values of l that are large enough to guarantee the validity of the ensuing confidence intervals, as in Section 8.4. Specifically, we take $l \geq 1$ for $\rho = 0.2$, $l \geq 2$ for $\rho = 0.4$, $l \geq 3$ for $\rho = 0.6$ and $l \geq 4$ for $\rho = 0.8$.

To compare the lengths of two confidence intervals for a given simulation setting we look at the boxplot of differences in the log lengths of these intervals. The lengths are evaluated from the 500 data sets generated for the given simulation setting. If the boxplot is symmetric about 0, then the two confidence intervals have comparable length.

Figures C.1 - C.10 in Appendix C (bottom three rows) display boxplots of pairwise differences in the log length of the standard 95% confidence intervals constructed from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$. From these figures, we see that for all simulation settings with $\rho > 0$ and for values of l that are large enough (e.g., larger than 3), the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ yield shorter confidence intervals than those based on $\widehat{\beta}^{(l)}_{S,MCV}$. This was to be expected, as the log MSE behaviour of $\widehat{\beta}^{(l)}_{S,MCV}$ was seen to be inferior to that of $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$. Furthermore, we notice that the lengths of the confidence intervals constructed from $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ tend to be comparable for many of these l values.

Our previous findings are supported by the results of pairwise level 0.05 two-sided paired t-tests for comparing the expected log lengths of the confidence intervals under consideration for all simulation settings and for values of l that are large enough. We describe these tests below.

Given a simulation setting, for fixed l, conduct $\binom{3}{2}$ two-sided paired t-tests to compare the expected log lengths of the intervals obtained from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$. For each test, the null hypothesis is that the expected log lengths of the intervals being compared are the same. The test result is considered significant if the p-value associated with the test is smaller than 0.05. Use the results of the t-tests to identify which estimators yield the shortest confidence interval. If all tests give significant results, we claim that there is a clear winner; in other cases, we say that two estimators might be tied for best.

Figures C.1-C.10 (top row) show the average length of the confidence intervals obtained, with standard error bars superimposed. The figures indicate which of these estimators produces the shortest confidence interval for values of l of interest.

8.6 Conclusions

Based on the results of our simulation study, we recommend using the usual local linear backfitting estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and the usual Speckman estimator $\widehat{\beta}^{(l)}_{S,MCV}$ to carry out valid inferences about the linear effect β_1 in model (8.1). The value of l used when computing these estimators should be large enough, that is, at least 4.

Our findings indicate that $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ have comparable accuracy for large values of l, and that they are in general more accurate than $\widehat{\beta}^{(l)}_{S,MCV}$. All three estimators yield valid standard 95% confidence intervals for β_1 when l is large enough. However, the intervals based on $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ tend to have shorter length and are therefore preferred over the interval based on $\widehat{\beta}^{(l)}_{S,MCV}$.

We see no reason to recommend bias-adjustments to the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$ and $\widehat{\beta}^{(l)}_{U,EBBS-G}$ or to their associated estimated standard errors. Such adjustments do not seem to improve the coverage properties of the corresponding confidence intervals.

Finally, we do not recommend using the usual backfitting estimator $\widehat{\beta}^{(l)}_{U,EBBS-L}$ or the estimated modified backfitting estimators $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ to carry out inferences about β_1. These estimators yielded confidence intervals with poor coverage for many simulation settings and many values of l, owing to the difficulties associated with estimating their standard errors.

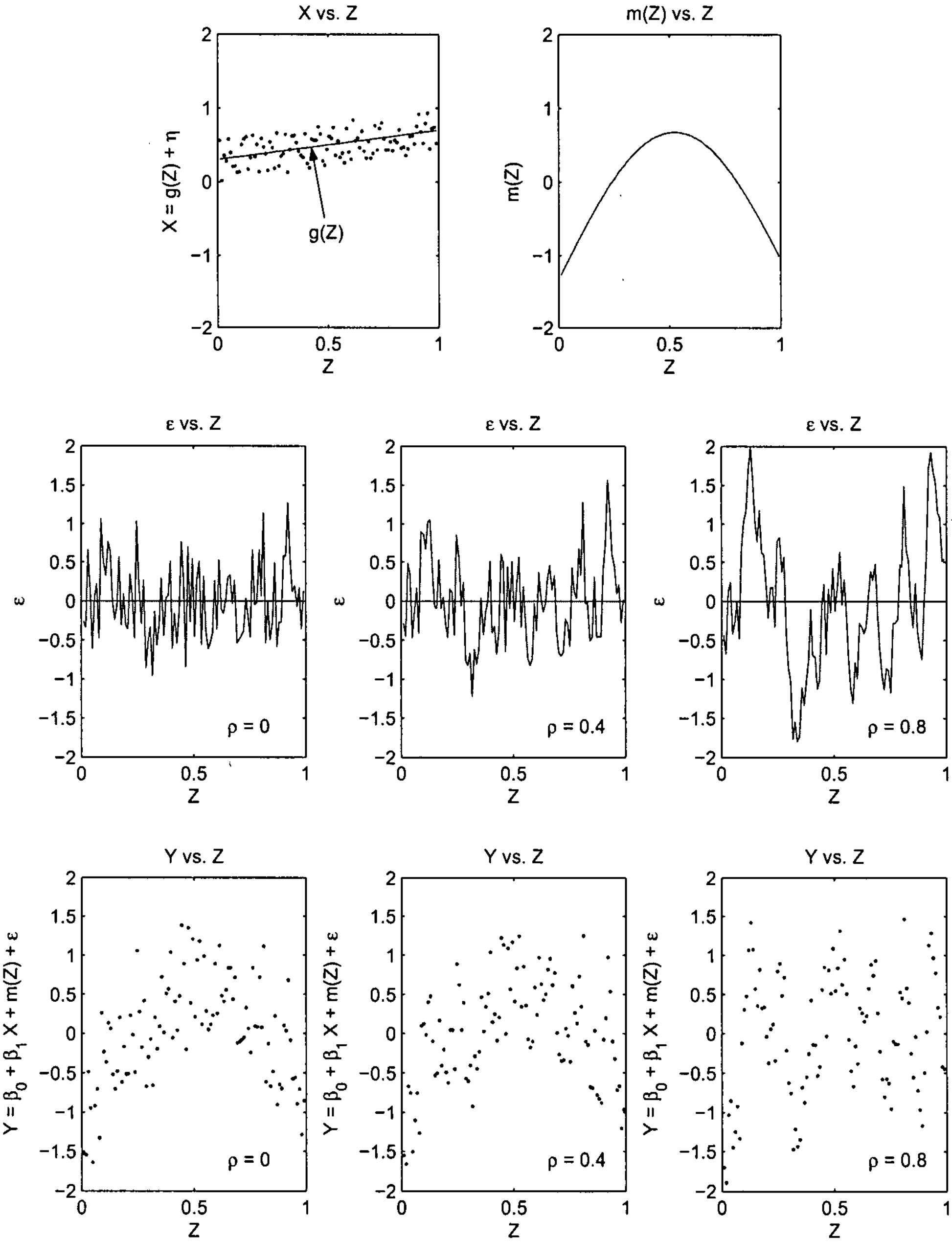

Figure 8.1: *Data simulated from model* (8.1) *for* $\rho = 0, 0.4, 0.8$ *and* $m(z) = m_1(z)$. *The first row shows plots that do not depend on* ρ. *The second and third rows each show plots for* $\rho = 0, 0.4, 0.8$.

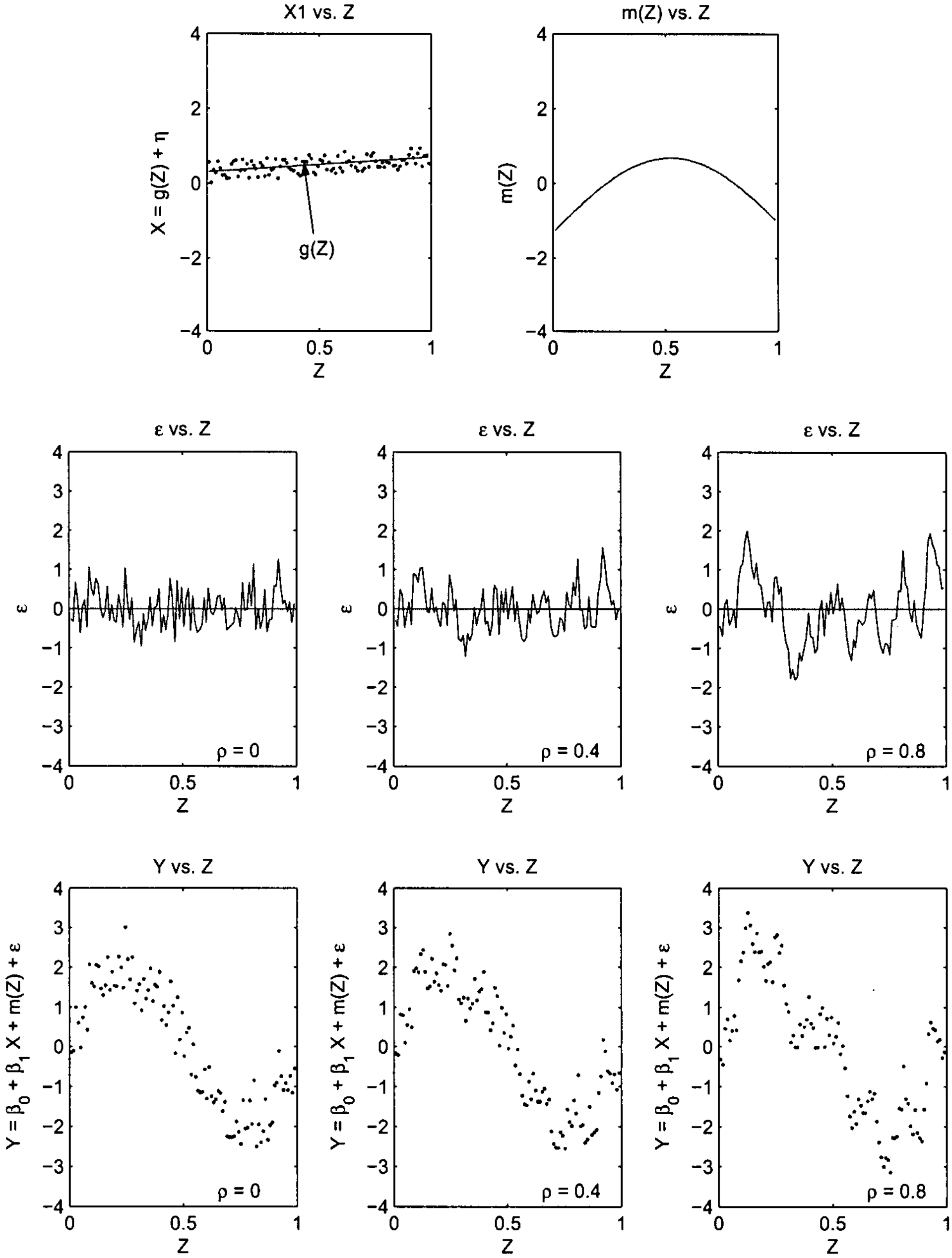

Figure 8.2: *Data simulated from model* (8.1) *for* $\rho = 0, 0.4, 0.8$ *and* $m(z) = m_2(z)$. *The first row shows plots that do not depend on* ρ. *The second and third rows each show plots for* $\rho = 0, 0.4, 0.8$.

Table 8.1: *Values of l for which the standard 95% confidence intervals for β_1 constructed from the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$ are valid (in the sense of achieving the nominal coverage) for each setting in our simulation study.*

	$m_1(z)$		
	$\widehat{\beta}^{(l)}_{U,PLUG-IN}$	$\widehat{\beta}^{(l)}_{U,EBBS-G}$	$\widehat{\beta}^{(l)}_{S,MCV}$
$\rho = 0$	$l \in \{0, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.2$	$l \in \{0, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$	$l \in \{1, \ldots, 10\}$
$\rho = 0.4$	$l \in \{1, \ldots, 10\}$	$l \in \{2, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.6$	$l \in \{2, \ldots, 10\}$	$l \in \{3, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.8$	$l \in \{3, \ldots, 10\}$	$l \in \{3, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
	$m_2(z)$		
	$\widehat{\beta}^{(l)}_{U,PLUG-IN}$	$\widehat{\beta}^{(l)}_{U,EBBS-G}$	$\widehat{\beta}^{(l)}_{S,MCV}$
$\rho = 0$	$l \in \{0, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.2$	$l \in \{0, \ldots, 10\}$	$l \in \{1, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.4$	$l \in \{1, \ldots, 10\}$	$l \in \{2, \ldots, 10\}$	$l \in \{0\} \cup \{2, 3\} \cup \{5, \ldots, 10\}$
$\rho = 0.6$	$l \in \{3, \ldots, 10\}$	$l \in \{3, \ldots, 10\}$	$l \in \{0, \ldots, 10\}$
$\rho = 0.8$	$l \in \{3, \ldots, 10\}$	$l \in \{4, \ldots, 10\}$	$l \in \{0, 1, 2\} \cup \{9\}$

Chapter 9

Application to Air Pollution Data

Many community-level studies have provided evidence that air pollution is associated with mortality. Statistical analyses of data collected in such studies face various methodological challenges: (1) controlling for observed and unobserved factors, such as season and temperature, that might confound the true association between air pollution and mortality, (2) accounting for serial correlation in the residuals that might underestimate statistical uncertainty of the estimated association, and (3) assessing and reporting uncertainty associated with the choice of statistical model.

Various statistical models can be used to describe the true association between air pollution and health outcomes of interest based on community-level data. However, the most widely used have been the generalized additive models (GAMs) introduced by Hastie and Tibshirani (1990). These models include a single 'time series' response (e.g. non-accidental mortality rates) and various covariates (e.g. pollutants of interest, time, temperature). The effects of the pollutants of interest on the response are typically presumed to be linear, whereas those of the remaining covariates are presumed to be smooth, non-linear. Schwartz (1994), Kelsall, Samet and Zeger (1997), Schwartz (1999), Samet, Dominici, Curriero et al. (2000), Katsouyani, Toulomi, Samoli et al. (2001), Moolgavkar (2000), Schwartz (2000) are just some of the authors who relied on GAMs in order to assess the acute effects of air pollution on health outcomes such as mortality or hospital admissions.

There are various problems that researchers must consider when using GAMs to analyze air pollution data arising from community-level studies. Some of these problems are purely computational, whereas others are more delicate and pertain to the theoretical underpinnings of these models.

Several computational issues associated with the S-Plus implementation of methodology developed by Hastie and Tibshirani (1990) for estimation of GAMs have been brought to light in recent years. We describe these problems here. The linear and non-linear effects in GAMs applied to air pollution data have been typically estimated using the S-Plus function *gam*. Dominici et al. (2002) showed that *gam* may provide incorrect estimates of the linear effects in GAMs and their standard errors if used with the original default parameters. Although the defaults have recently been revised (Dominici et al., 2002), an important problem that remains is that *gam* calculates the standard errors of the linear effects by assuming that the non-linear effects are effectively linear, resulting in an underestimation of uncertainty (Ramsay et al., 2003a). In air pollution studies, this assumption is likely inadequate, resulting in underestimation of the standard error of the linear pollutant effect (Ramsay et al., 2003a).

The practical choice of the degree of smoothness of the estimated non-linear confounding effects of time and meteorology variables is a delicate issue in air pollution studies which utilize GAMs. Given that the confounding effects are viewed as a nuisance in such studies, the appropriate choice should be informed by the objective of conducting valid inferences about the pollution effect. Most choices performed in the air pollution literature are based on exploratory analyses (see, for instance, Kelsall, Samet and Zeger, 1997) and seem to be justified by a different objective, namely doing well at estimating the non-linear confounding effects. This objective typically ignores the impact of residual correlation on the choice of degree of smoothness, as well as the dependencies between the various variables in the model.

In the present chapter we apply the methodology developed in this thesis to analyze air pollution data collected in Mexico City between January 1, 1994 and December 31, 1996. Our goal is to determine whether the pollutant PM10 has a significant short-term effect on the non-accidental death rate in Mexico City after adjusting for temporal and weather confounding. We give a description of the data in Section 9.1 and analyze the data in Section 9.2.

9.1 Data Description

PM10 - airborne particulate matter less than 10 microns in diameter - is a major component of air pollution, arising from natural sources (e.g. pollen), road transport, power generation, industrial processes, etc. When inhaled, PM10 particles tend to be deposited in the upper parts of the human respiratory system from which they can be eventually expelled back into the throat. Health problems begin as the body reacts to these foreign particles. PM10 is associated with mortality, exacerbation of airways disease and decrement in lung function. Although PM10 can cause health problems for everyone, certain people are especially vulnerable to its adverse health effects. These "sensitive populations" include children, the elderly, exercising adults, and those suffering from heart and lung disease.

The data to be analyzed in this chapter were collected in Mexico City over a period of three years, from January 1, 1994 to December 31, 1996, in order to determine if there is a significant short term effect of PM10 on mortality, after adjusting for potential temporal and weather confounders. The data consist of daily counts of non-accidental deaths, daily levels of ambient concentration of PM10 ($10\mu g/m^3$), and daily levels of temperature (^oC) and relative humidity (%). The ambient concentration of PM10 corresponding to a given day was obtained by averaging the PM10 measurements over all the stations in Mexico City.

Pairwise scatter plots of the data are shown in Figure 9.1. The most striking features in these plots are the strong annual cycles in the log mortality levels, the daily level of ambient concentration of PM10, and the daily levels of temperature and relative humidity. It is likely that the annual cycles in the log mortality levels are produced by unobserved seasonal factors such as influenza and respiratory infections. Note that log mortality and PM10 peak at the same time with respect to the annual cycles. Our analysis of the health effects of PM10 must account for the potential confounding effect of these temporal cycles on the association between PM10 and log mortality. We believe the strength of these cycles will make it difficult to detect whether this association is significant.

9.2 Data Analysis

The following is an overview of our data analysis. First, we introduce the four statistical models that we use to capture the relationship between PM10 and mortality, adjusted for seasonal and meteorological confounding. Three of these models contain smooth non-parametric terms which attempt to control for these confounding effects. Next, we illustrate the importance of choosing the amount of smoothing for estimating the nonparametric terms in these models when the main objective is accurate estimation of the true association between PM10 and mortality. We then focus on determining which of the four models is most relevant for the data. Finally, we use this model as a basis for carrying out inference about the true association between PM10 and mortality.

9.2.1 Models Entertained for the Data

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. The models that we entertain for our data are:

$$log(D_i) = \beta_0 + \beta_1 P_i + \epsilon_i \tag{9.1}$$

$$log(D_i) = \beta_0 + \beta_1 P_i + m_1(i) + \epsilon_i \tag{9.2}$$

$$log(D_i) = \beta_0 + \beta_1 P_i + m_1(i) + \beta_2 T_i + \beta_3 H_i + \beta_{23} T_i \cdot H_i + \epsilon_i \tag{9.3}$$

$$log(D_i) = \beta_0 + \beta_1 P_i + m_1(i) + m_2(T_i, H_i) + \epsilon_i. \tag{9.4}$$

Here, $i = 1, 2, \ldots, 1096$. Also, m_1 is a smooth univariate function, whereas m_2 is a smooth bivariate surface. The function m_1 serves as a linear filter on the log mortality and PM10 series and removes any seasonal or long-term trends in the data. For the time being, the error terms in all four models are assumed to be independent, identically distributed, with mean 0 and constant variance $\sigma_\epsilon^2 < \infty$. The independence assumption will be relaxed later.

Models (9.1)-(9.4) treat the log mortality counts as a continuous response. Furthermore, they assume the relationship between PM10 and log mortality to be linear, to allow for easily interpretable inferences about the effect of PM10 on log mortality. The models differ, however, in their specification of the potential seasonal and weather confounding on this relationship. Specifically, model (9.1) ignores the possible seasonal and weather confounding on the relationship between PM10 and log mortality. Models (9.2)-(9.4), however, allow us to adjust this relationship for potential seasonal and weather confounding.

Models (9.2) and (9.3) require that we specify the amount of smoothing needed for estimating m_1. Model (9.4) requires that we specify the amount of smoothing necessary for estimating both m_1 and m_2.

To fit models (9.2)-(9.4) to the data, we use the S-Plus function *gam* with the more stringent convergence parameters recommended by Dominici et al. (2002). We employ a univariate loess smoother to estimate m_1 and a bivariate loess smoother to estimate

m_2. The loess smoothers are local linear smoothers relying on spans corresponding to a fixed number of nearest neighbours instead of a bandwidth.

9.2.2 Importance of Choice of Amount of Smoothing

The inferences made on the linear PM10 effect β_1 in any of the models (9.2)-(9.4) may be severely affected by the choice of amount of smoothing for estimating the smooth confounding effects in these models. To illustrate the impact of this choice on the conclusions of such inferences, we restrict attention to model (9.3). Later, we will see that this model is the most appropriate for the data.

Figure 9.2 compares the impact of various choices of smoothing for the seasonal effect m_1 in model (9.3) on the following quantities:

(i) *gam* estimates of β_1,

(ii) *gam* standard errors for the estimates in (i),

(iii) 95% confidence intervals for β_1 constructed from the estimates in (i) and (ii),

(iv) *gam* p-values associated with standard t-tests of significance of β_1.

These quantities were obtained by fitting model (9.3) to the data using *gam* with loess as a basic smoother. The loess span used for smoothing m_1 was allowed to take on values in the range 0.01 to 0.50. The reference distribution for the 95% confidence intervals and the p-values depicted in Figure 9.2 is a t-distribution whose degrees of freedom are the residual (or error) degrees of freedom associated with model (9.3). Note that the estimated standard errors reported by *gam* do not account for error correlation.

Changing the span for smoothing m_1 greatly affects the estimates, standard errors, confidence intervals and p-values in Figure 9.2 and hence the conclusions of our inferences on β_1, the short-term PM10 effect on log mortality. In particular, using large spans for smoothing m_1 suggests that the data provide strong evidence in favour of a significant PM10 effect on log mortality, after adjusting for seasonal and weather confounding. Using small spans for smoothing m_1 suggests that the data do not provide enough evidence in support of a significant PM10 effect on log mortality in Mexico City.

Proper choice of amount of smoothing for estimating the seasonal effect m_1 in model (9.3) is crucial for making inferences on β_1, as seen in Figure 9.2. Given the sensitivity of our conclusions to the choice of smoothing, the natural question that arises is: how can we choose the amount of smoothing to be able to make valid inferences on β_1?

The correct choice of smoothing should be appropriate for accurate estimation of β_1, not for accurate estimation of m_1. This choice should account for the strong relationships between the linear and non-linear variables in the model seen in Figure 9.1, and for potential correlation amongst model errors.

It is important to note that the S-Plus function *gam* provides no data-driven method for choosing the amount of smoothing. Using *gam*'s default choice of smoothing is not advised when one is concerned with accurate estimation of β_1. The default choice of smoothing used by *gam* is 0.50, or 50% of the nearest neighbours. This choice is much larger than the choices that we recommend for estimating m_1 (shown in the next section). The theoretical results in this thesis suggest that the correct choice of smoothing for estimating β_1 should undersmooth the estimated m_1. Therefore, this choice of smoothing is most likely smaller than the one we recommend for estimating m_1, and certainly not larger.

9.2.3 Choosing an Appropriate Model for the Data

In this section, we focus on the issue of selecting an appropriate model for the data amongst models (9.1)-(9.4). Selecting such a model requires that we balance model complexity with model parsimony. In what follows, we show that model (9.3) is the most appropriate for describing the variability in the log mortality counts, as it is complex enough to capture the main features present in the data, yet relatively inexpensive to fit to these data in terms of degrees of freedom.

Model (9.1) is the simplest of models (9.1) -(9.4) and, not too surprisingly given the strong cycles apparent in Figure 9.1, it provides an inadequate description for the variability in the log mortality counts. In fact, the linear relationship between PM10 and log mortality postulated by model (9.1) explains only 9.25% of the total variability in the log mortality counts. Figure 9.3 (top panel) shows that the log mortality counts are widely scattered about the regression line obtained by fitting model (9.1) to the data. Figure 9.3 (bottom panel) shows that model (9.1) displays clear lack-of-fit, as it fails to account for the strong annual cycles present in the model residuals. We therefore drop model (9.1) from our pool of candidate models and concentrate instead on models (9.2)-(9.4).

Model (9.4) is the most complex of these models, and will consume significantly more degrees of freedom when fitted to the data than either model (9.2) or model (9.3). As we shall see shortly, comparing model (9.4) against model (9.2) via a series of approximate F-tests suggests that we can drop model (9.4) in favour of model (9.2).

We could therefore consider the simpler model (9.2) as being adequate for describing the variability in the log mortality counts. However, given that the weather variables are typically included in models for PM10 mortality data, we prefer to use model (9.3). This model is more flexible than model (9.2), as it includes linear marginal effects for the weather variables together with a linear interaction effect between these variables. Compared to model (9.2), this model can be fitted to the data at the expense of just three additional degrees of freedom. Given the large size of the data set, this is an insignificant price to pay for achieving more modelling flexibility.

We now provide more details concerning the choice of an appropriate model for our data amongst models (9.2)-(9.4). As a first step we need to identify spans that are reasonable for smoothing the seasonal effect m_1 in these models.

To identify a reasonable range of spans for smoothing m_1 in model (9.2), we fit model (9.2) to the data by smoothing m_1 with spans ranging from 0.01 to 0.50 in increments of 0.01 and examine plots of the fitted m_1 and corresponding model residuals. From Figures 9.4 and 9.5 we see that the data suggest spans in the range $0.09 - 0.12$. Using spans smaller than 0.09 for estimating m_1 leads to under-smoothed fits, that are visually noisy. On the other hand, using spans larger than 0.12 leads to over-smoothed fits, that fail to reflect important seasonal features of the data. In summary, the range $0.09 - 0.12$ is reasonable for smoothing the seasonal effect m_1 in model (9.2).

Plots of the fitted additive component m_1 in models (9.3) and (9.4) (not shown) corresponding to spans in the range 0.09 to 0.12 are similar to those in Figure 9.4 and suggest that this range is also reasonable for smoothing the seasonal effect m_1 in models (9.3) and (9.4).

We now show that we can reduce model (9.4) to model (9.2). We use a series of approximate F-tests to compare models (9.4) and (9.2). Each F-test compares a fit of model (9.4), obtained by smoothing m_1 with the span s_1, against a fit of model (9.2), obtained by smoothing m_1 with the span s_1 and m_2 with the span s_2. The test statistic for each F-test is obtained in the usual fashion from the residual sums of squares and the residual (or error) degrees of freedom associated with the two model fits. The residual degrees of freedom of these fits are obtained as the difference between the size of the data set $n = 1096$ and the trace of the hat matrix associated with the model fit. We allow the span s_1 to range between 0.09 and 0.12 in increments of 0.01, and the span s_2 to range between 0.01 and 0.50 in increments of 0.01.

The p-values associated with these F-tests are displayed in Figure 9.6. P-values corresponding to spans s_2 bigger than 0.04 are quite large, suggesting that the smooth weather surface m_2 need not be included in model (9.4). P-values corresponding to spans s_2 of $0.02, 0.03$ or 0.04 are a bit smaller, suggesting that perhaps the surface m_2 should be included in the model. However, Figures 9.7 and 9.8, for $s_1 = 0.09$, show that very small spans are not appropriate for estimating the surface m_2, as they yield visually rough surfaces that consume unacceptably high numbers of degrees of freedom. Using a span s_1 of $0.10, 0.11$ or 0.12 instead yielded plots (not shown) that were basically identical to those in Figures 9.7 and 9.8.

In conclusion, the smooth weather surface m_2 contributes little to model (9.4), so there is no real need to include either temperature or relative humidity in this model. In other words, we can reduce model (9.4) to model (9.2). Coplots (not shown) of the residuals associated with model (9.2) versus temperature, given relative humidity, and versus relative humidity, given temperature, support this conclusion.

Since there is no real need to include the weather variables, temperature and relative humidity, we could consider the simpler model (9.2) as being adequate for describing the variability in the log mortality counts. However, for reasons explained earlier, we prefer to use the more flexible model (9.3).

How well does model (9.3) fit the data? To answer this question, we examine a series of diagnostic plots. Figure 9.9 shows plots of the residuals associated with model (9.3) against PM10 and day of study. These residuals were obtained by smoothing the unknown m_1 with a span of 0.09; using spans of $0.10, 0.11$ or 0.12 yielded similar plots (not shown). The functional form of the relationship between PM10 and log mortality postulated by model (9.3) is not violated by the data, since no systematic structure is apparent in the plot of residuals versus PM10. The plot of residuals against day of study also shows no systematic structure, suggesting that the seasonal component m_1 of the model accounts for the long-term temporal variation in the data reasonably well. Figures 9.10-9.11 show that the functional specification of the weather portion of model (9.3) is not violated by the data. Indeed, these plots display no obvious systematic structure. The weather coplots corresponding to spans of $0.10, 0.11$ and 0.12 were similar, so we omitted them. Finally, Figure 9.12 presents autocorrelation and partial autocorrelation plots for the residuals associated with model (9.3). From these plots, it is apparent that the magnitude of the residual correlation is small. We believe this is due to the fact that most of the short-term temporal variation in log mortality counts has been accounted for by the seasonal component m_1 of the model. Comparing Figure 9.12 against Figure 9.13, which displays autocorrelation and partial autocorrelation plots for the raw log mortality counts, supports this belief.

In summary, the assumptions underlying the systematic part of model (9.3) seem reasonable. However, there is some modest suggestion that the independence assumption concerning the error terms in this model may not hold for these data. This assumption will be relaxed to account for the slight temporal correlation present in the data.

Model (9.3) can therefore be used as a basis for carrying out inferences on β_1, the linear PM10 effect on log mortality, adjusted for seasonal and weather confounding. Accounting for error correlation when conducting such inferences is perhaps not as important as accounting for the strong relationships between the linear and non-linear variables in the model evident in Figure 9.1.

9.2.4 Inference on the PM10 Effect on Log Mortality

In order to conduct valid inferences about the linear effect β_1 in model (9.3), we must not only estimate it accurately, but also calculate correct standard errors for this estimate.

For model (9.3), $\beta_1 = \boldsymbol{c}^T\boldsymbol{\beta}$, where $\boldsymbol{c} = (0, 1, 0, 0, 0)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_{23})^T$. We propose to estimate β_1 via $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ is the usual local linear backfitting estimate of $\boldsymbol{\beta}$. Figure 9.14 displays a plot of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ versus the smoothing parameter h, which controls the width of the smoothing window. The large variation in the values of these estimates re-iterates the importance of choosing h appropriately from the data so as to obtain accurate estimates of β_1.

To choose appropriate values of h from the data, we use the preferred PLUG-IN and EBBS-G methods developed in Chapter 6. Both methods use a grid $\mathcal{H} = \{2, 3, \ldots, 548\}$, where the values in the grid represent half-widths of local linear smoothing windows. Recall that both of these methods require that we estimate the underlying correlation structure of the model errors. In addition, PLUG-IN requires that we estimate the sesonal effect m_1 in the model. We discuss these topics below.

We estimate the seasonal effect m_1 and the error correlation structure using modified (or leave-(2l+1)-out) cross-validation, as outlined in Sections 6.3.1 and 6.3.2. We allow the tuning parameter l to take on the values $0, 1, \ldots, 26$. Recall that l quantifies our belief about the range and magnitude of the error correlation. For instance, $l = 0$ signifies that we believe the errors to be independent. When the model errors are truly correlated, we suspect that values of l that are too small may produce under-smoothed estimates of m_1, whereas values of l that are too large may produce over-smoothed estimates of m_1.

To ascertain what values of l are reasonable for the data, we examine plots of the estimated seasonal effect m_1 in model (9.3) corresponding to $l = 0, 1, \ldots, 26$; see Figure 9.15. These plots suggest that using $l = 0$ or $l = 1$ is probably not appropriate, as the corresponding estimates of m_1 are visually too rough. Using values of l in the range $2 - 17$ seems to yield reasonable estimates of m_1. Values of l in the range $18 - 26$ seem to yield over-smoothed estimates of m_1, so perhaps should be avoided.

Next, we estimate the error terms in model (9.3) via modified (or leave-(2l+1)-out) cross-validation residuals, defined as in Section 6.3.1. Figure 9.16 shows plots of these residuals for various values of l.

Now, we use the modified cross-validation residuals to estimate the correlation structure of the model errors. We will operate under the assumption that these errors follow a covariance-stationary autoregressive process of finite order R. To estimate R, we use the finite sample criterion for autoregressive order selection developed by Broersen (2000). Figure 9.17 shows that our estimate of R is influenced by how we choose the value of the tuning parameter l. Choosing $l = 0$ or 1 yields an R of 28. Choosing larger l's yields R's like $0, 2, 3$ or 4. Recall that values of l like $0, 1$ or $18, \ldots, 26$ are likely not appropriate for these data.

Finally, after determining the order $R = R(l), l = 0, 1, \ldots, 26$, of the autoregressive error process, we estimate the error variance σ_ϵ^2 and the autoregressive parameters $\phi_1, \ldots, \phi_R$ using Burg's method (Brockwell and Davis, 1991). Furthermore, we estimate the error correlation matrix $\boldsymbol{\Psi}$ by plugging in the estimated values of $\phi_1, \ldots, \phi_R$ in the expression for $\boldsymbol{\Psi}$ provided in Comment 2.2.1.

Having estimated the seasonal effect m_1 and the error correlation structure for model (9.3), we can now tackle the issue of data-driven choice of h for accurate estimation of β_1 via $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$. The estimated bias squared, variance and mean squared error curves used for determining the PLUG-IN choice of smoothing for $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ are shown in Figure 9.18. The different curves correspond to different values of l, where $l = 0, 1, \ldots, 26$. In general, the mean squared error curves corresponding to small values of l dominate those corresponding to large values of l. Figure 9.19 displays similar plots used for determining the EBBS-G choice of smoothing. Note that the bias curve in this figure does not depend on l. Also note that mean squared error curves in this figure that correspond to large values of l dominate, in general, the curves that correspond to small values of l.

Figures 9.20 and 9.21 display the PLUG-IN and EBBS-G choices of smoothing parameter obtained by minimizing the estimated mean squared error curves in Figures 9.18 and 9.19. Both choices are remarkably stable for values of l that seem appropriate for these data. However, the PLUG-IN choices are much smaller in magnitude than the EBBS-G choices. The PLUG-IN choices that seem appropriate for the data indicate that the seasonal effect m_1 should be smoothed using $h \approx 28$. On the other hand, the corresponding EBBS-G choices indicate that m_1 should be smoothed using $h \approx 69$.

Figures 9.22 and 9.23 show the 95% confidence intervals constructed for β_1 with PLUG-IN and EBBS-G choices of smoothing for values of l ranging from 0 to 26. These intervals were obtained from formula (7.2), with $\boldsymbol{\Omega} = \boldsymbol{I}$. Both figures suggest that the choice of l (among those that are reasonable for the data) is not that important. This finding is consistent with the Monte Carlo simulation study conducted in Chapter 8 that indicated these choices of smoothing were appropriate for conducting inferences on the linear effect β_1 in model (8.1) provided l was large enough.

From Figure 9.22, there is no conclusive proof that β_1, the short-term PM10 effect on log mortality, is significantly different from 0. Indeed, the standard confidence intervals for β_1 based on $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$, with h chosen via PLUG-IN, cross the zero line for all values of l that are appropriate for the data. The stability of these confidence intervals across various values of l is quite remarkable, but not entirely surprising given the stability of the corresponding PLUG-IN choices of smoothing shown in Figure 9.20.

Figure 9.23 supports the same conclusion for β_1, at least in part. However, for all values of l that are appropriate for the data, these intervals either narrowly miss zero or barely contain it, suggesting that perhaps PM10 does have a significant effect on log mortality.

What could explain the discrepancy between Figures 9.22 and 9.23? The standard errors of the estimated PM10 effects are comparable in both figures. However, the PM10 effect estimates obtained with a PLUG-IN choice of smoothing are much smaller than those obtained with EBBS-G. As seen in Figures 9.20 and 9.21, the PLUG-IN choices of smoothing parameter for these data are about 28 or so, and are much smaller than the EBBS-G choices, which are about 69 or so. Figure 9.14 shows that using choices of smoothing parameter h of 28 or so yields smaller PM10 estimates than using values of h of 69 or so. We favour smaller choices of smoothing parameter. We believe EBBS-G yielded large choices because it used a grid range that was too wide. Recall that EBBS-G attempts to estimate the conditional bias of $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}$ by assuming a specific form for the relationship between this bias and the smoothing parameter h. This relationship is motivated by asymptotic considerations as in (6.13), so it may break down for values of $h \in \mathcal{H}$ that are too large. Estimating this relationship based on all the "data"

$\left\{\left(h, \boldsymbol{c}^T\widehat{\boldsymbol{\beta}}_{\boldsymbol{I},\boldsymbol{S}_h^c}\right) : h \in \mathcal{H}\right\}$, may therefore not be appropriate. One should perhaps use only "data" for which h is reasonably small to ensure the asymptotic considerations underlying EBBS-G are valid. In other words, one should use a smaller grid range for EBBS-G. We used EBBS-G with a grid $\mathcal{H} = \{2, \ldots, 100\}$ instead of $\mathcal{H} = \{2, \ldots, 548\}$ and got a similar result to that obtained via PLUG-IN (see Figure 9.24): there is no conclusive proof that PM10 has a significant effect on log mortality. This finding is not surprising given the strength of the annual cycles present in Figure 9.1.

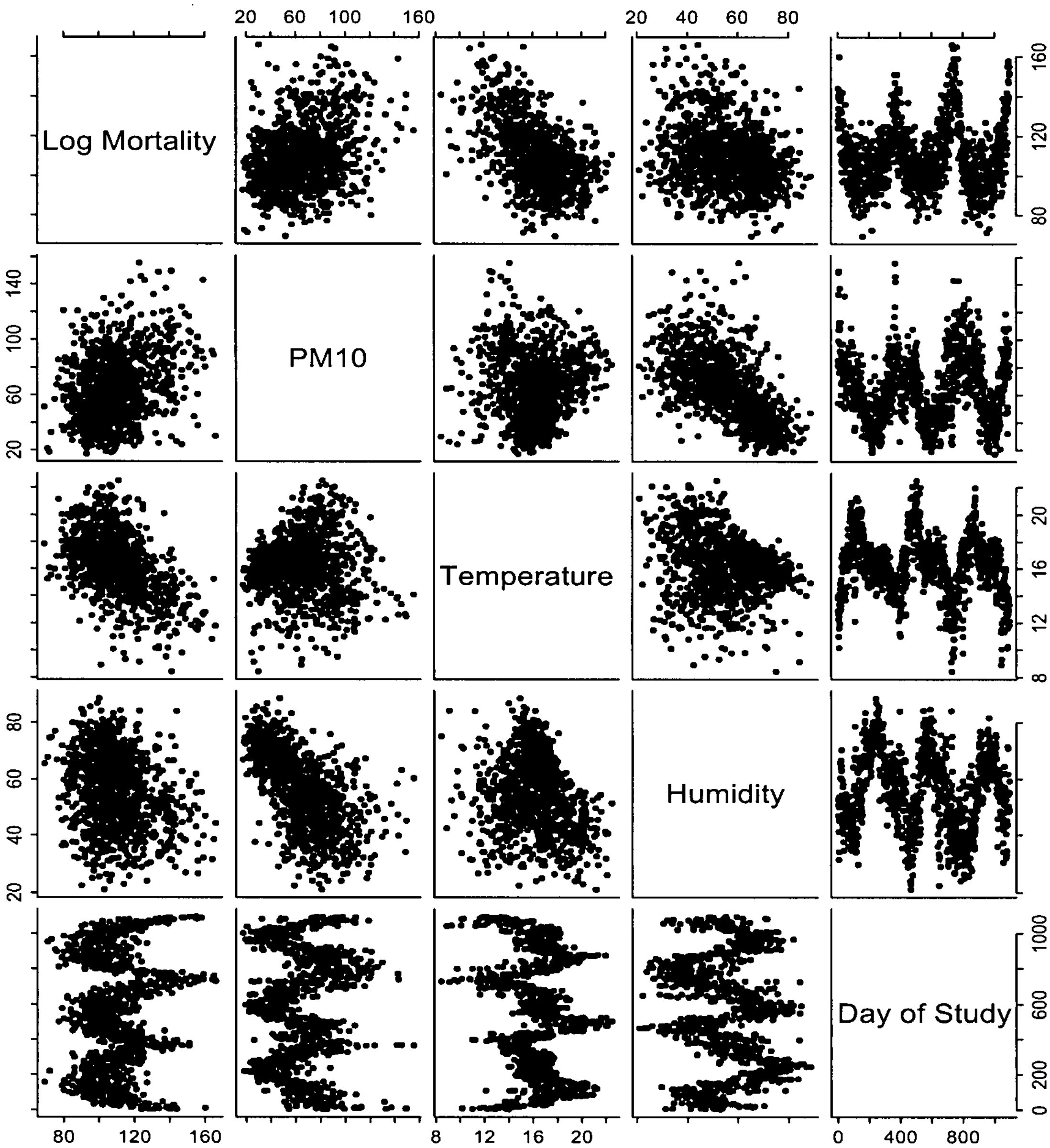

Figure 9.1: *Pairwise scatter plots of the Mexico City air pollution data.*

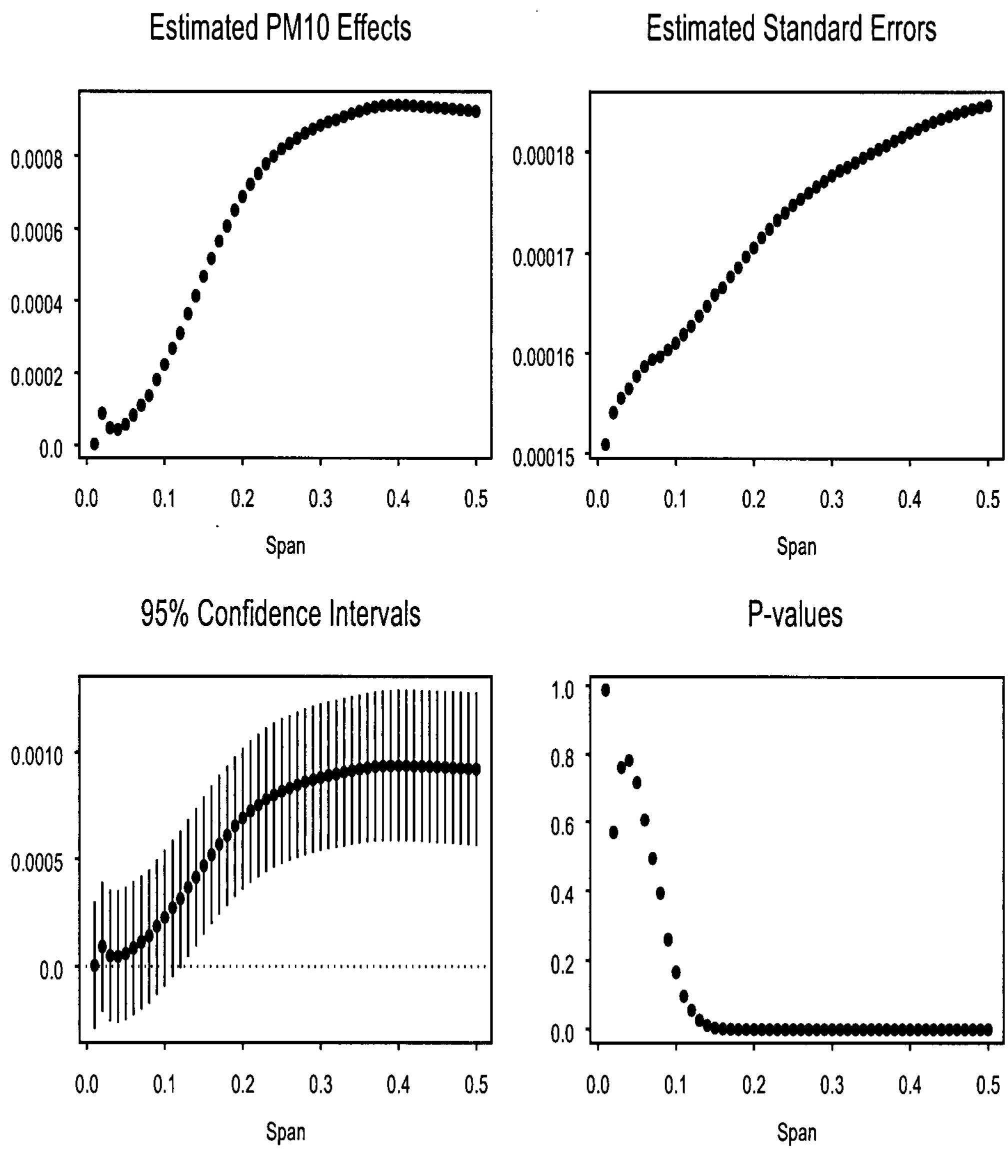

Figure 9.2: *Results of* gam *inferences on the linear PM10 effect β_1 in model* (9.3) *as a function of the span used for smoothing the seasonal effect m_1: estimated PM10 effects (top left), associated standard errors (top right),* 95% *confidence intervals for β_1 (bottom left) and p-values of t-tests for testing the statistical significance of β_1.*

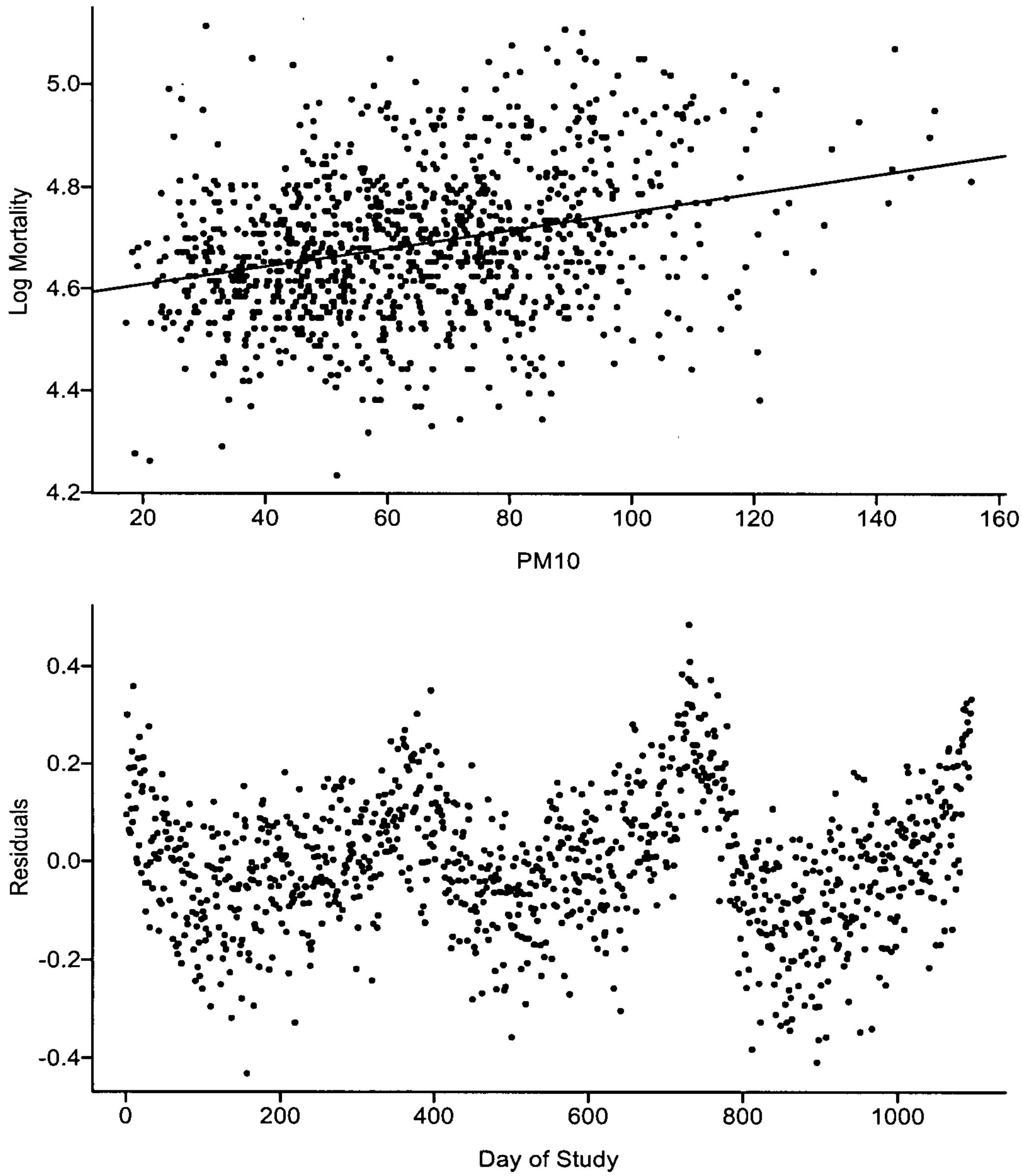

Figure 9.3: *The top panel displays a scatter plot of log mortality versus PM10. The ordinary least squares regression line of log mortality on PM10 is superimposed on this plot. The bottom panel displays a plot of the residuals associated with model* (9.1) *versus day of study.*

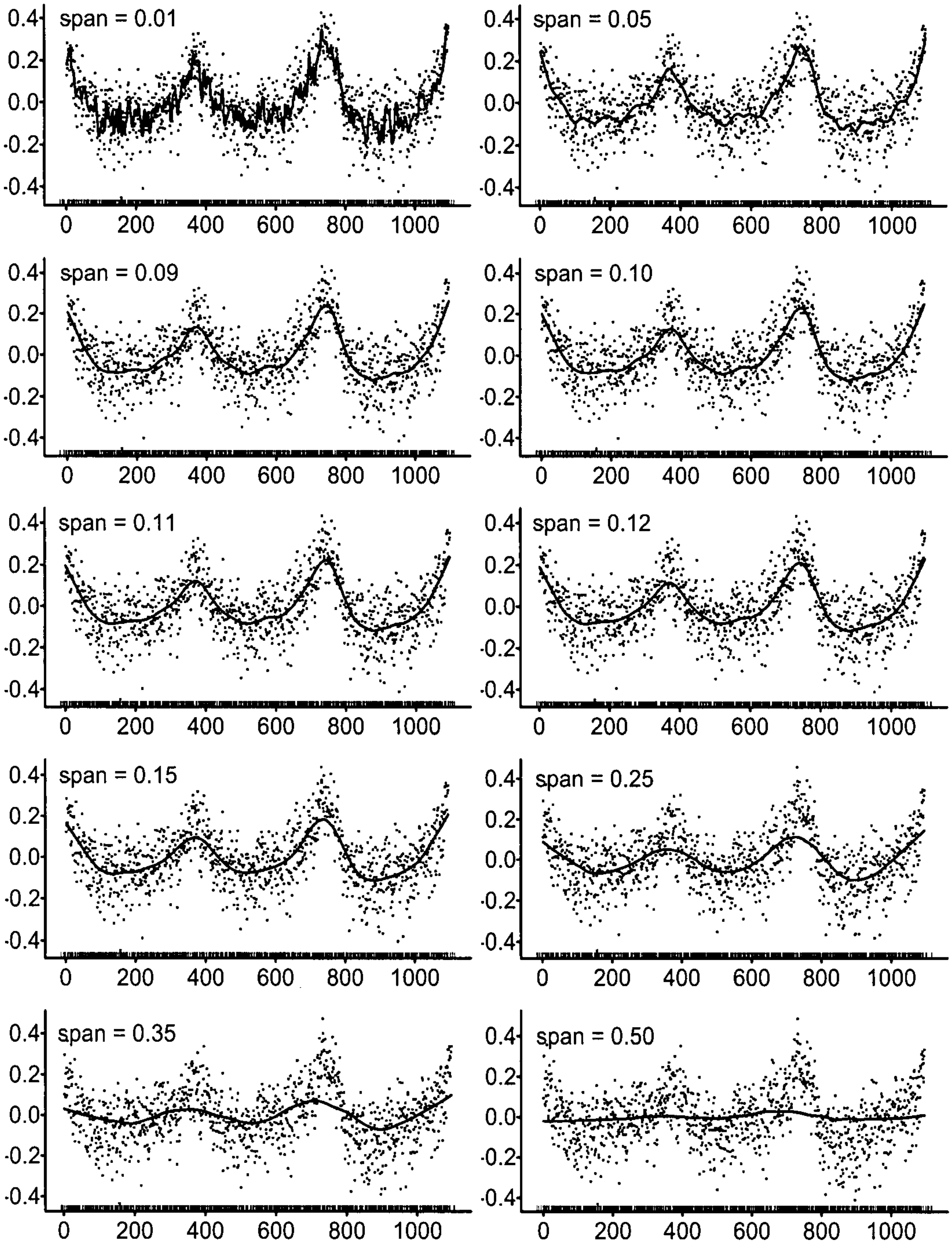

Figure 9.4: *Plots of the the fitted seasonal effect* m_1 *in model* (9.2) *for various spans. Partial residuals, obtained by subtracting the fitted parametric part of the model from the responses, are superimposed as dots.*

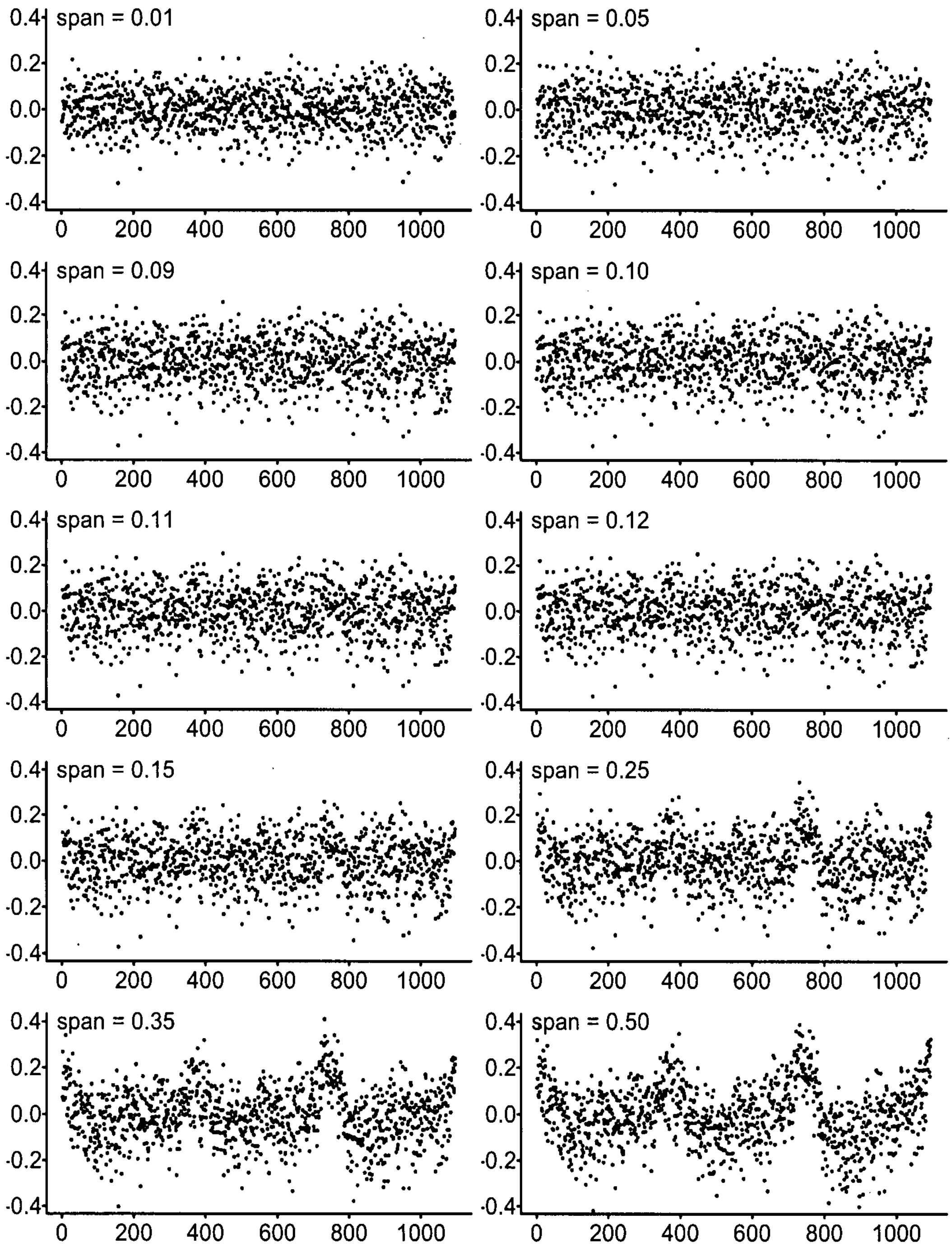

Figure 9.5: *Plots of the residuals associated with model* (9.2) *for various spans.*

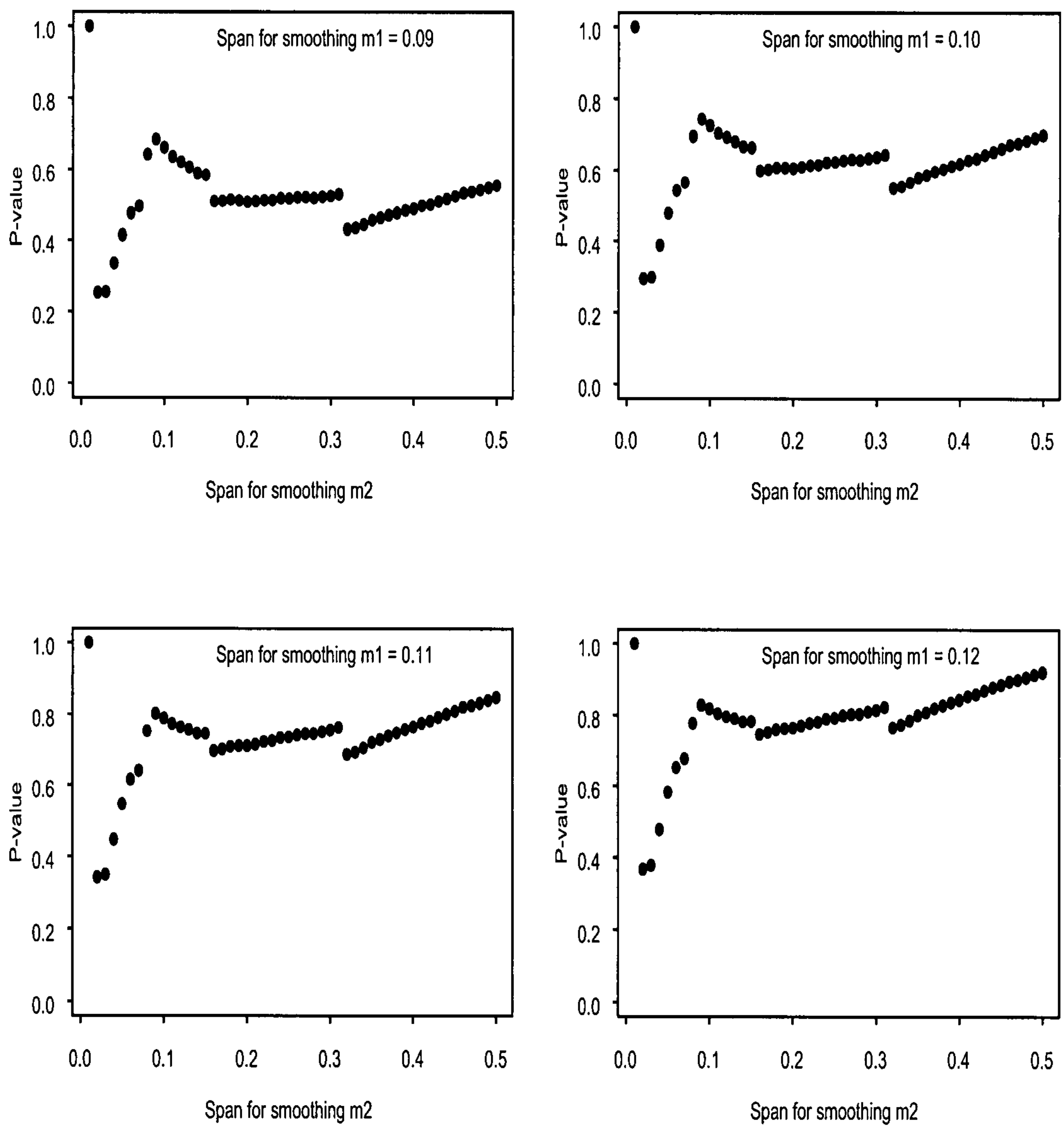

Figure 9.6: *P-values associated with a series of crude F-tests for testing model* (9.4) *against model* (9.2).

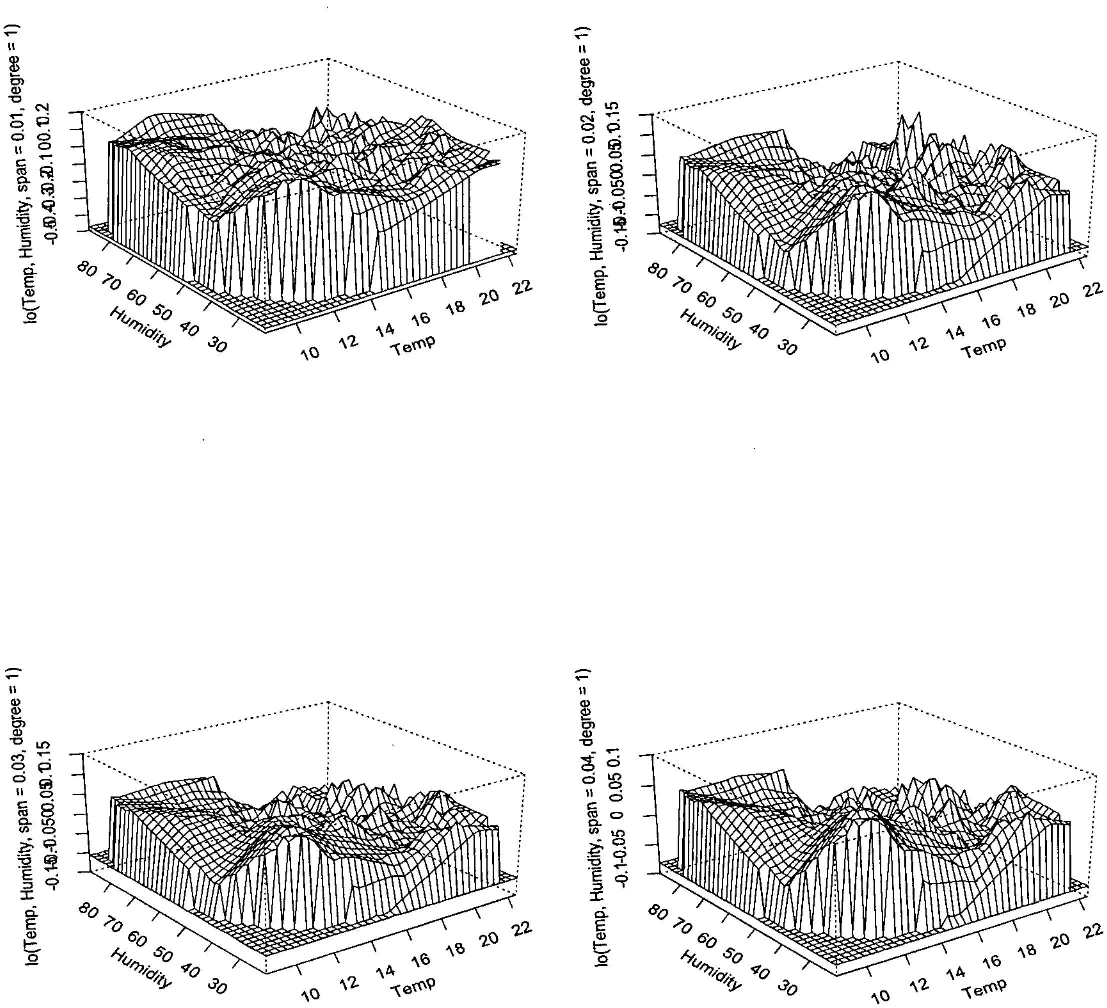

Figure 9.7: *Plots of the fitted weather surface* m_2 *in model* (9.4) *when the fitted seasonal effect* m_1 *(not shown) was obtained with a span of* 0.09. *The surface* m_2 *was smoothed with spans of* 0.01 *(top left),* 0.02 *(top right),* 0.03 *(bottom left) or* 0.04 *(bottom right).*

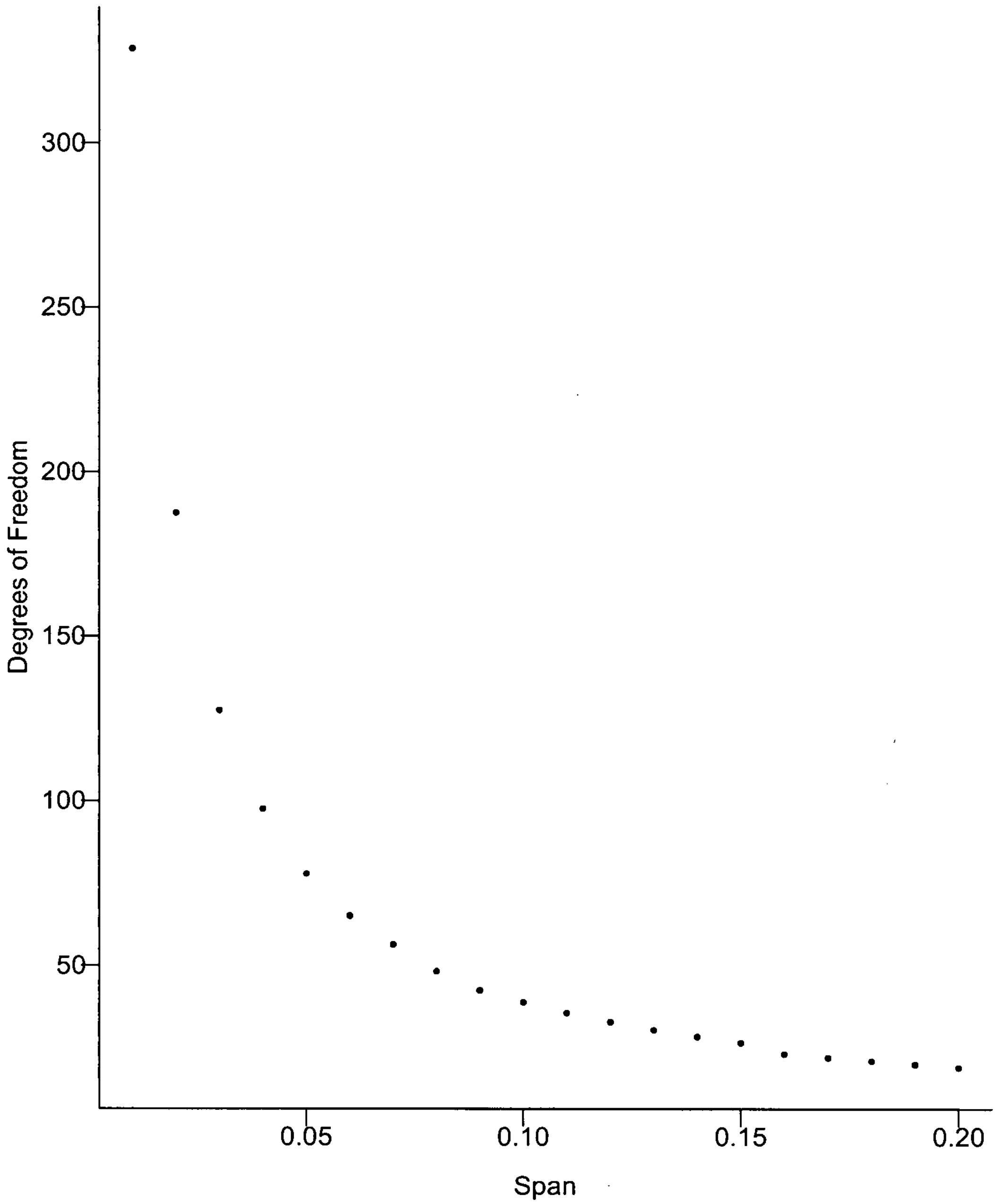

Figure 9.8: *Degrees of freedom consumed by the fitted weather surface m_2 in model* (9.4) *versus the span used for smoothing m_2 when the fitted seasonal effect m_1 (not shown) was obtained with a span of* 0.09.

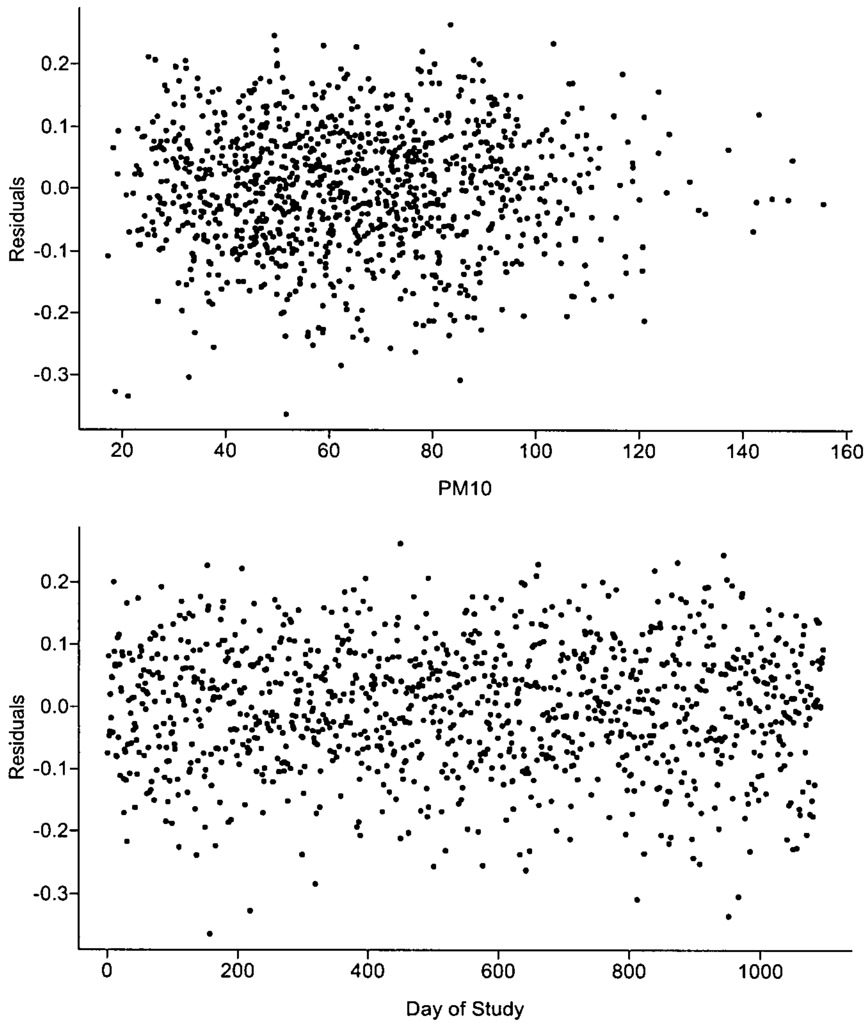

Figure 9.9: *Plot of residuals associated with model* (9.3) *versus PM10 (top row) and day of study (bottom row). The span used for smoothing the unknown* m_1 *in model* (9.3) *is* 0.09.

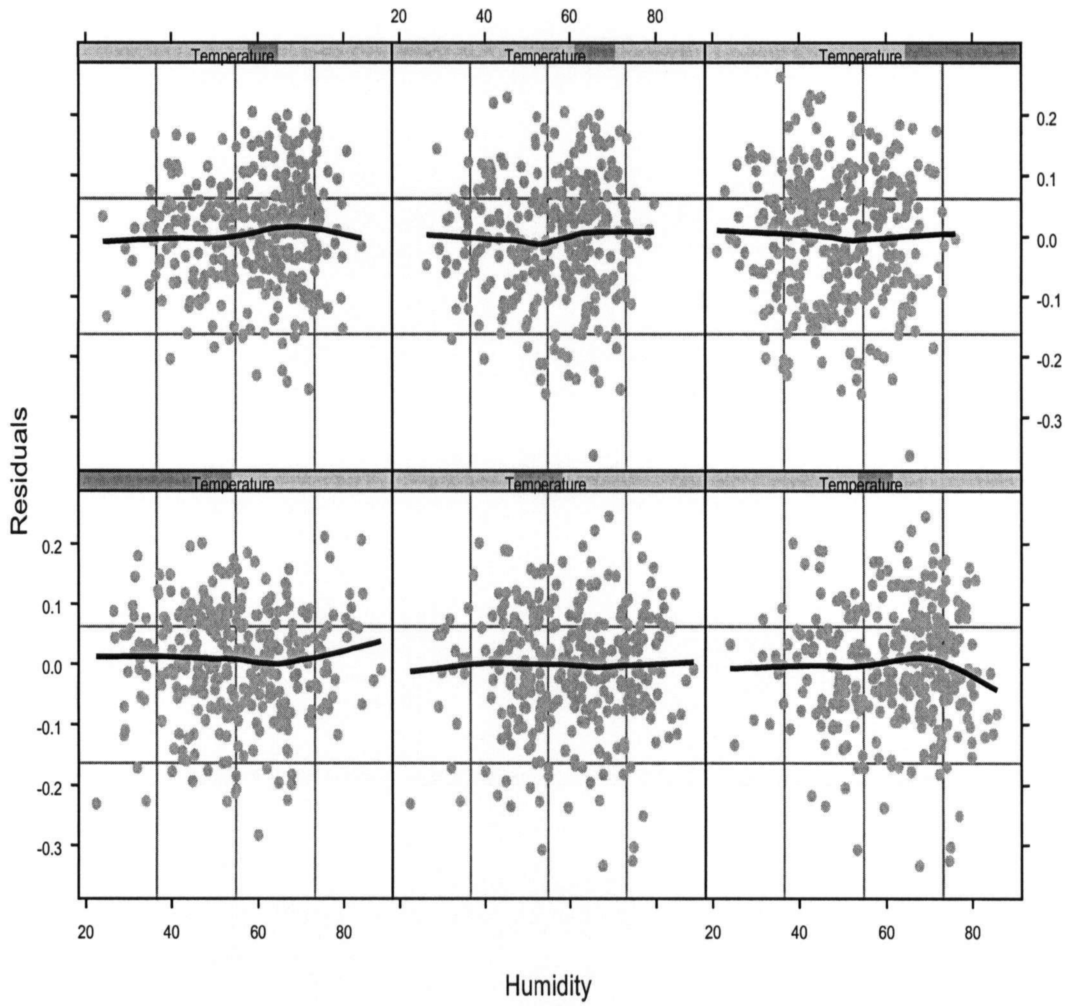

Figure 9.10: *Plot of residuals associated with model* (9.3) *versus relative humidity, given temperature. The span used for smoothing the unknown* m_1 *in model* (9.3) *is* 0.09.

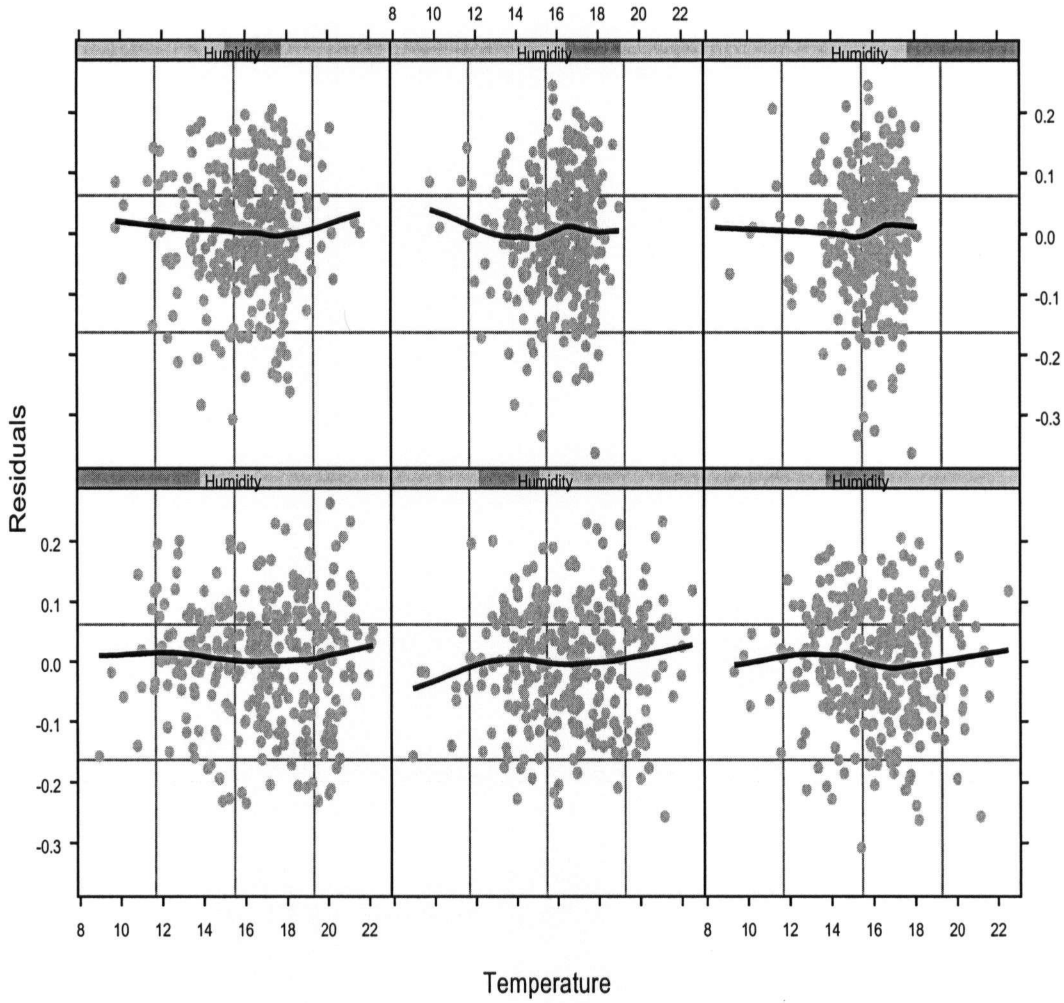

Figure 9.11: *Plot of residuals associated with model* (9.3) *versus temperature, given relative humidity. The span used for smoothing the unknown* m_1 *in model* (9.3) *is* 0.09.

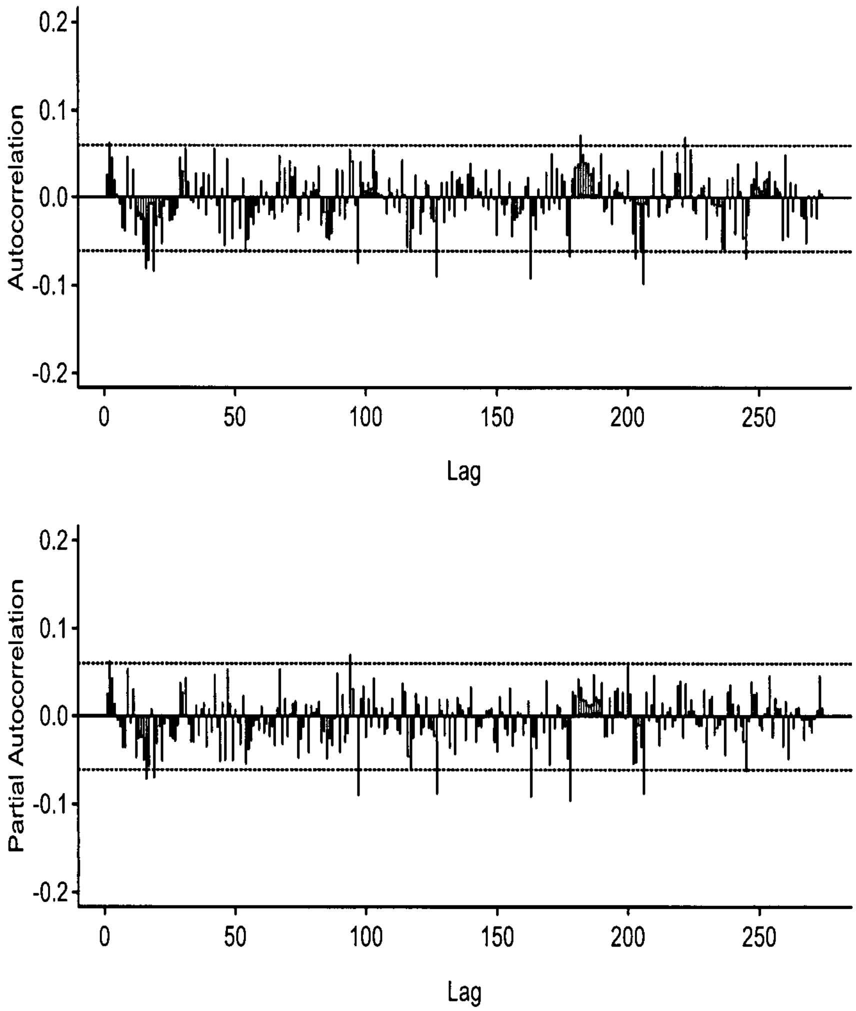

Figure 9.12: *Autocorrelation plot (top row) and partial autocorrelation plot (bottom row) of the residuals associated with model* (9.3). *The span used for smoothing the unknown* m_1 *in model* (9.3) *is* 0.09.

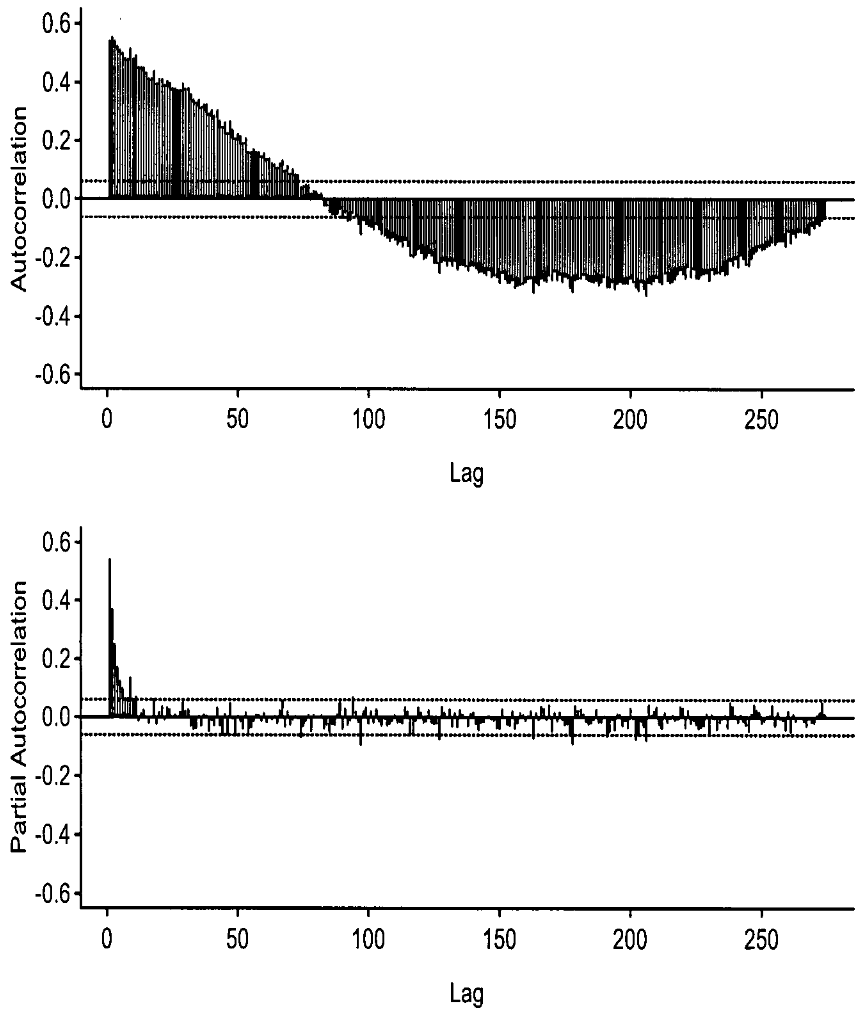

Figure 9.13: *Autocorrelation plot (top row) and partial autocorrelation plot (bottom row) of the responses in model* (9.3).

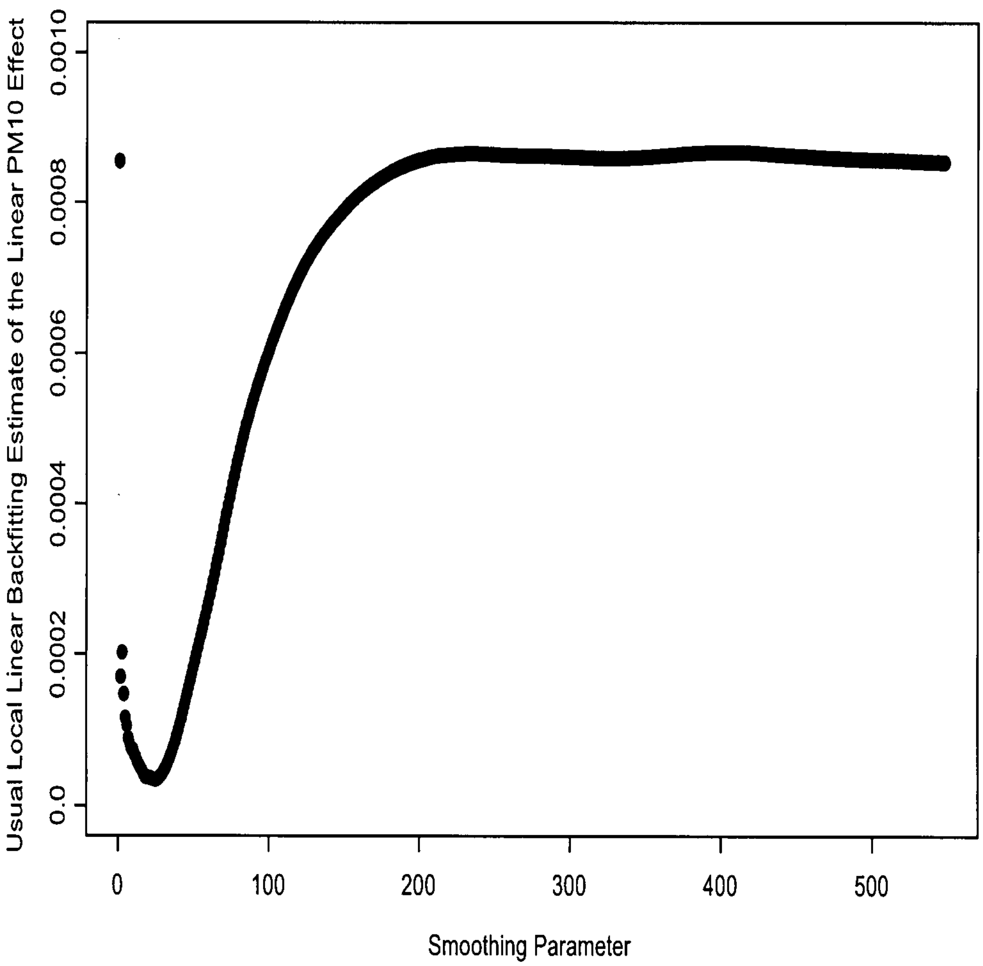

Figure 9.14: *Usual local linear backfitting estimate of the linear PM10 effect in model* (9.4) *versus the smoothing parameter.*

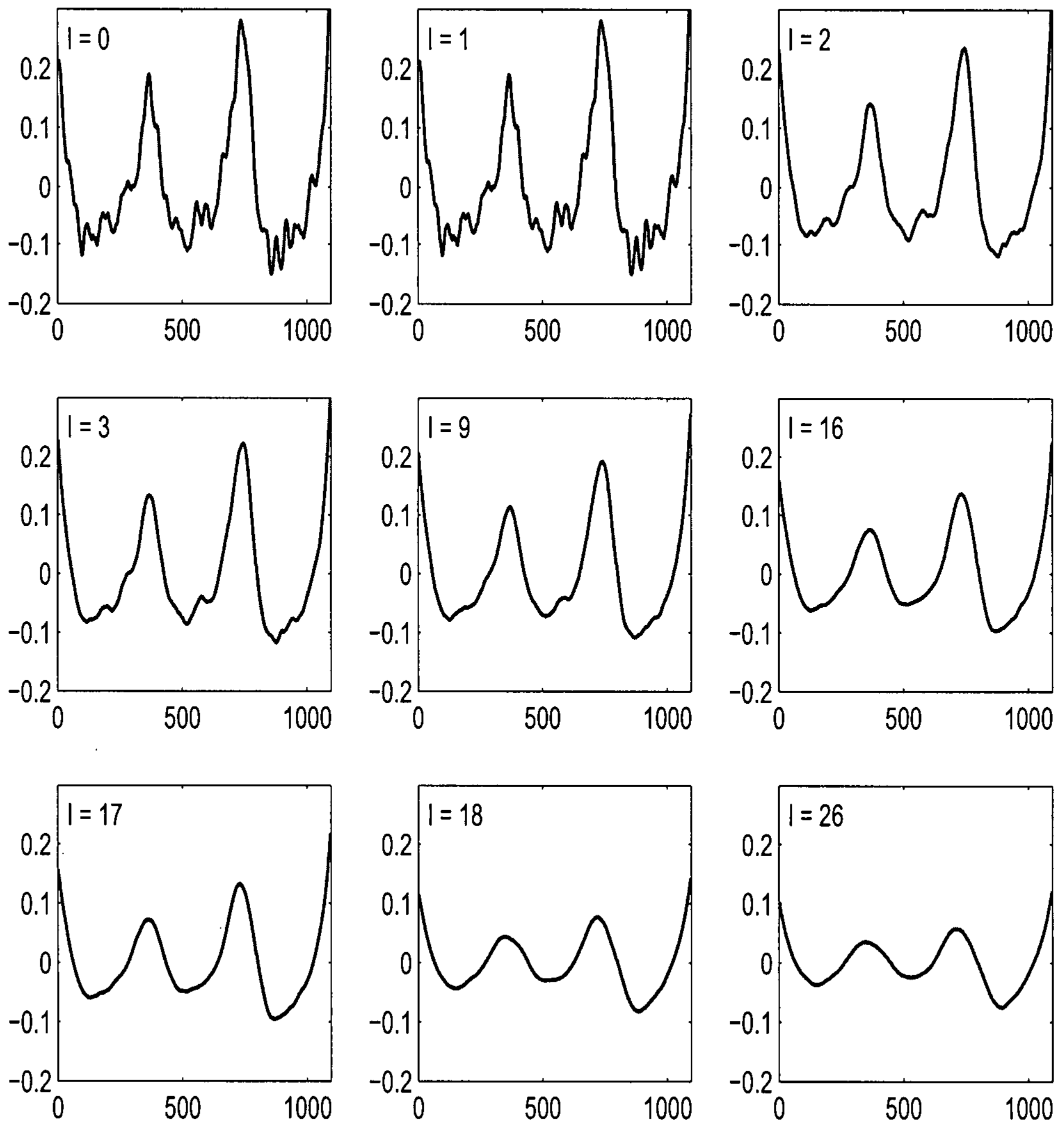

Figure 9.15: *Preliminary estimates of the seasonal effect m in model* (9.3), *obtained with a modified (or leave-$2l+1$-out) cross-validation choice of amount of smoothing.*

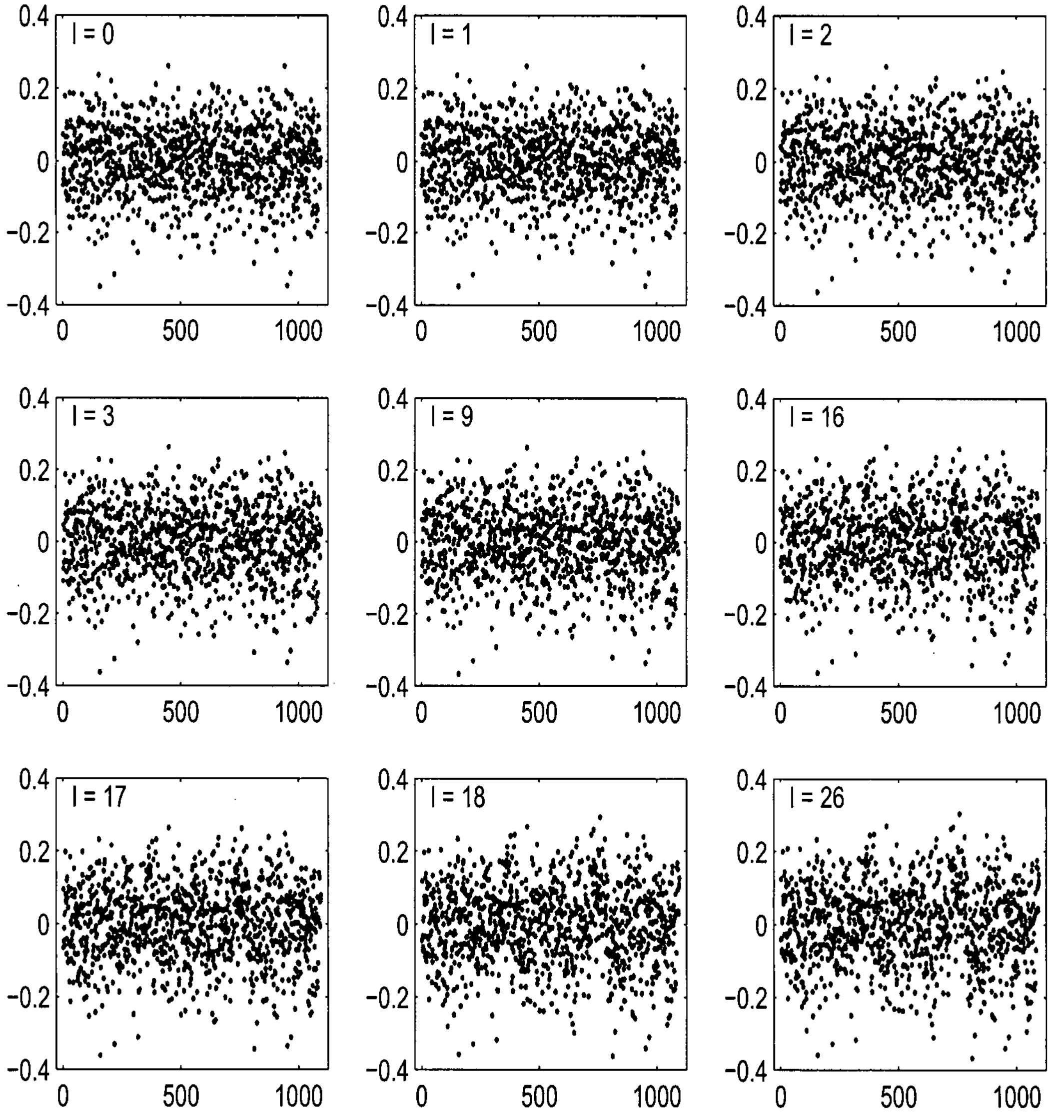

Figure 9.16: *Residuals associated with model* (9.3), *obtained by estimating* m_1 *with a modified (or leave-*$(2l+1)$*-out) cross-validation choice of amount of smoothing.*

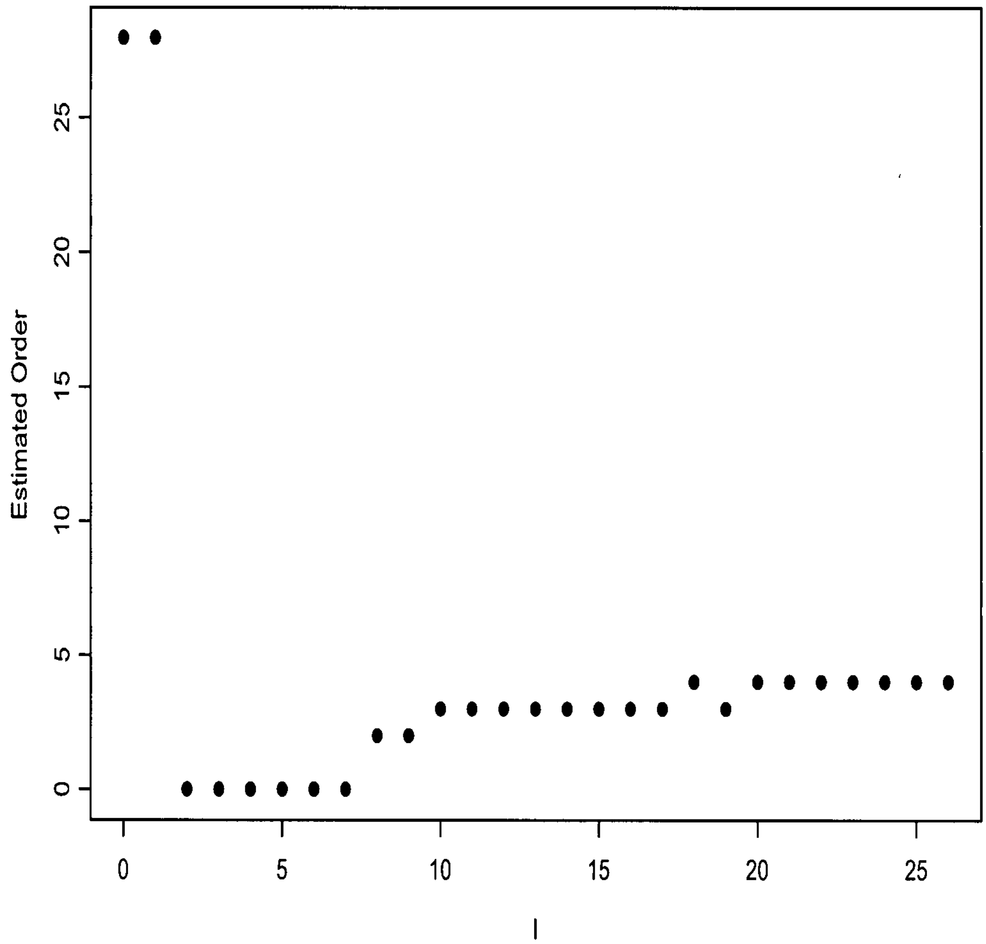

Figure 9.17: *Estimated order for AR process describing the serial correlation in the residuals associated with model* (9.3) *versus* l, *where* $l = 0, 1, \ldots, 26$. *Residuals were obtained by estimating* m_1 *with a modified (or leave-*$(2l+1)$*-out) cross-validation choice of amount of smoothing.*

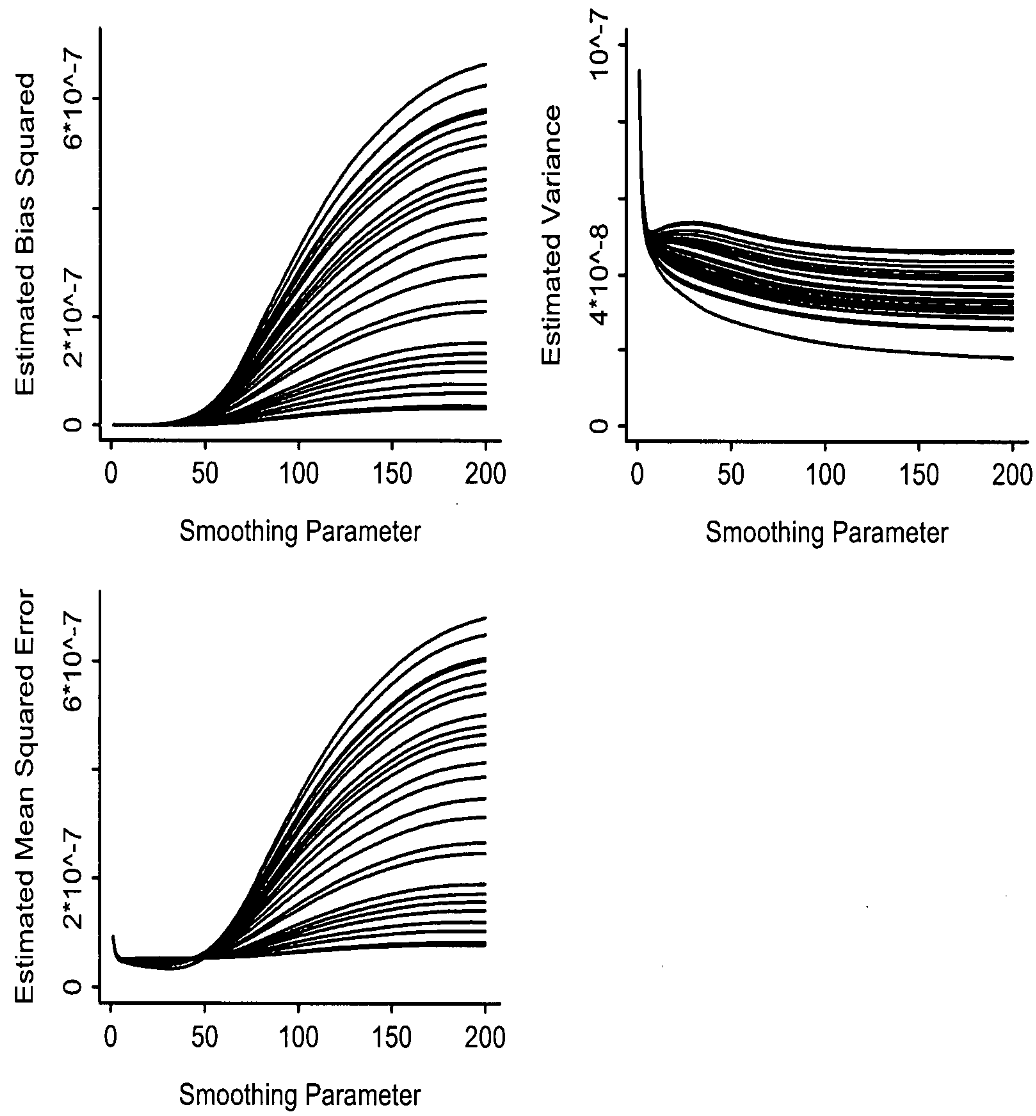

Figure 9.18: *Estimated bias squared, variance and mean squared error curves used for determining the plug-in choice of smoothing for the usual local linear backfitting estimate of β_1. The different curves correspond to different values of l, where $l = 0, 1, \ldots, 26$. The estimated variance curves corresponding to small values of l are dominated by those corresponding to large values of l when the smoothing parameter is large. In contrast, the estimated squared bias and mean squared error curves corresponding to small values of l dominate those corresponding to large values of l when the smoothing parameter is large.*

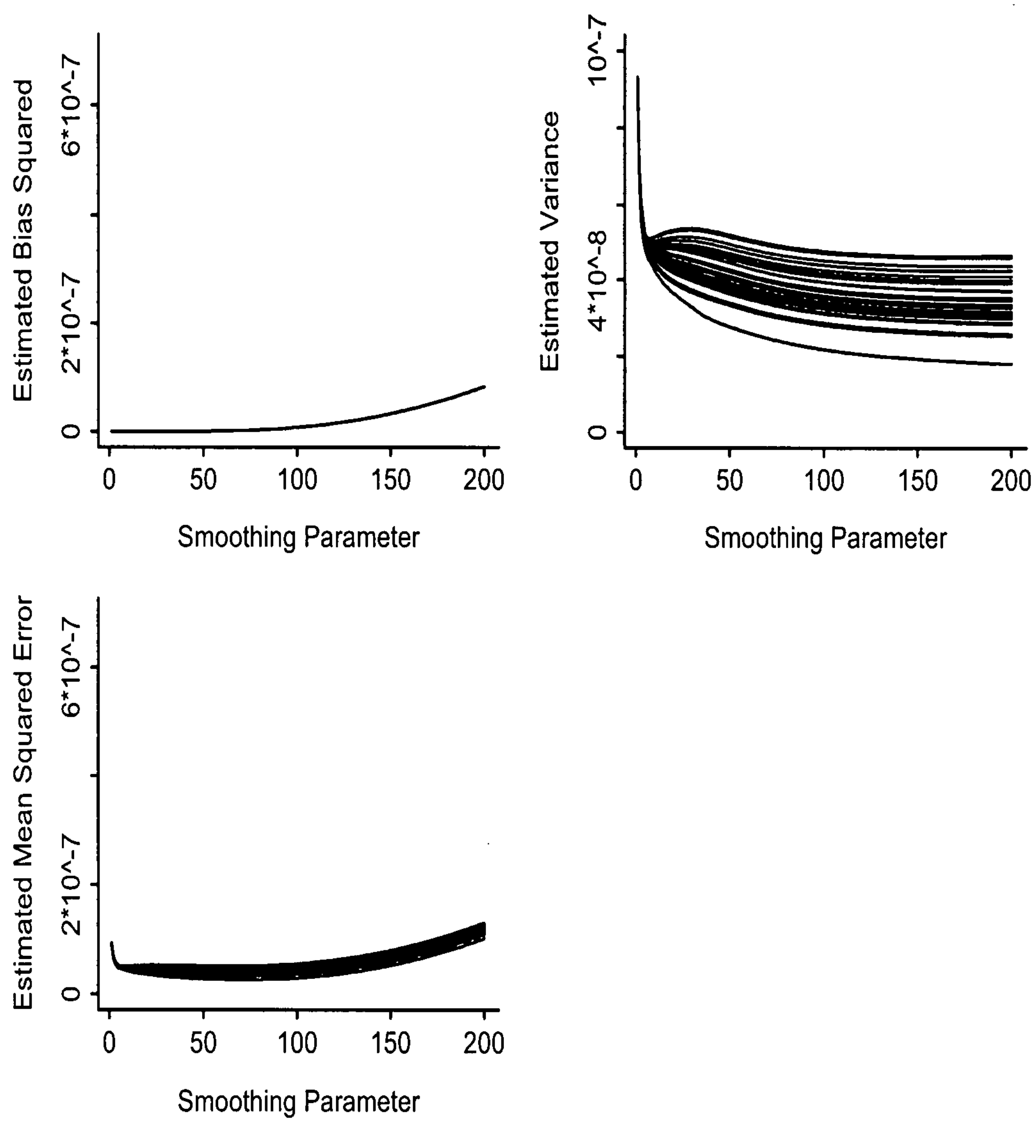

Figure 9.19: *Estimated bias squared, variance and mean squared error curves used for determining the global EBBS choice of smoothing for the usual local linear backfitting estimate of β_1. The different curves correspond to different values of l, where $l = 0, 1, \ldots, 26$. The curves corresponding to large values of l dominate those corresponding to small values of l.*

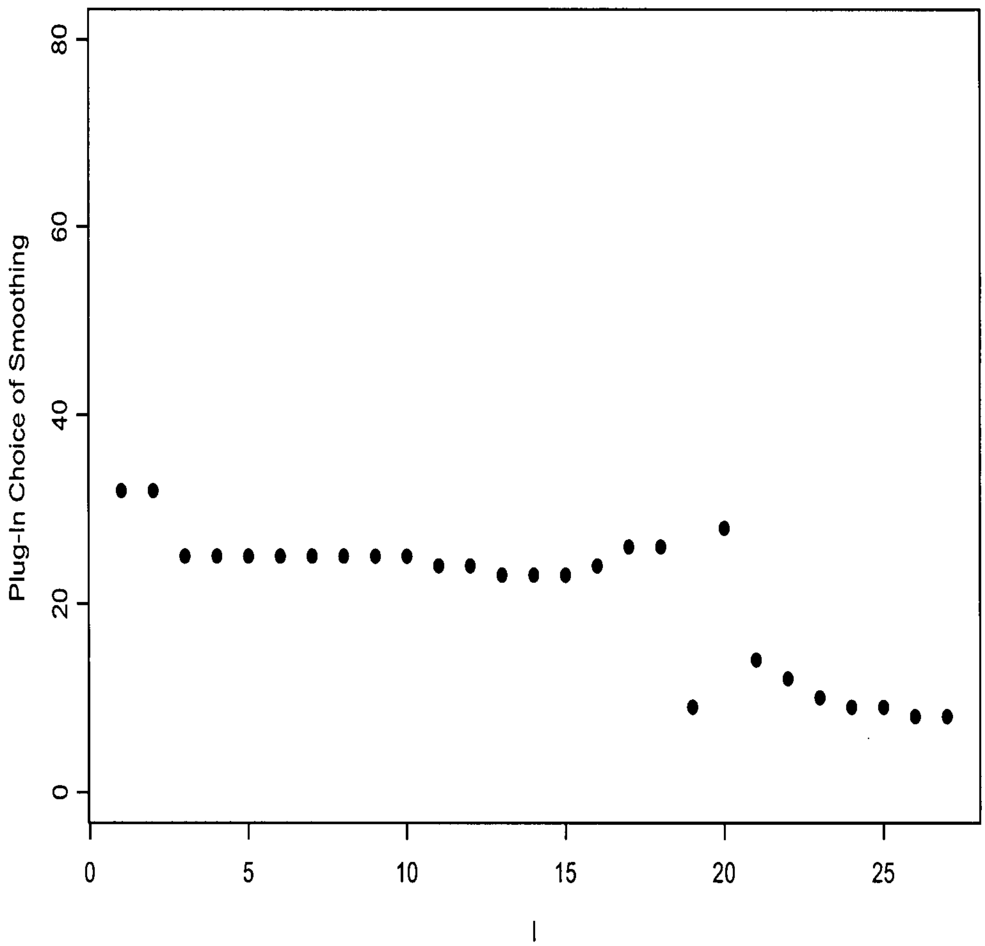

Figure 9.20: *Plug-in choice of smoothing for estimating β_1 versus l, where $l = 0, 1, \ldots, 26$.*

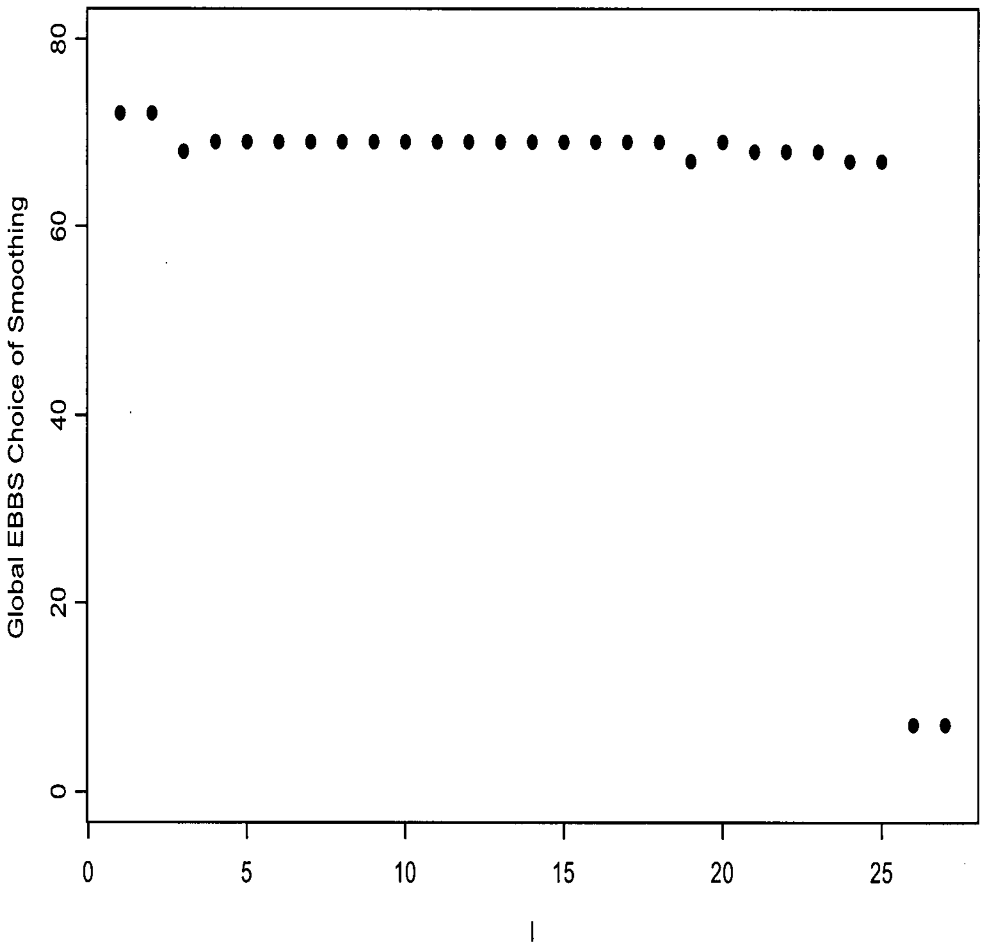

Figure 9.21: *Global EBBS choice of smoothing for estimating β_1 versus l, where* $l = 0, 1, \ldots, 26$.

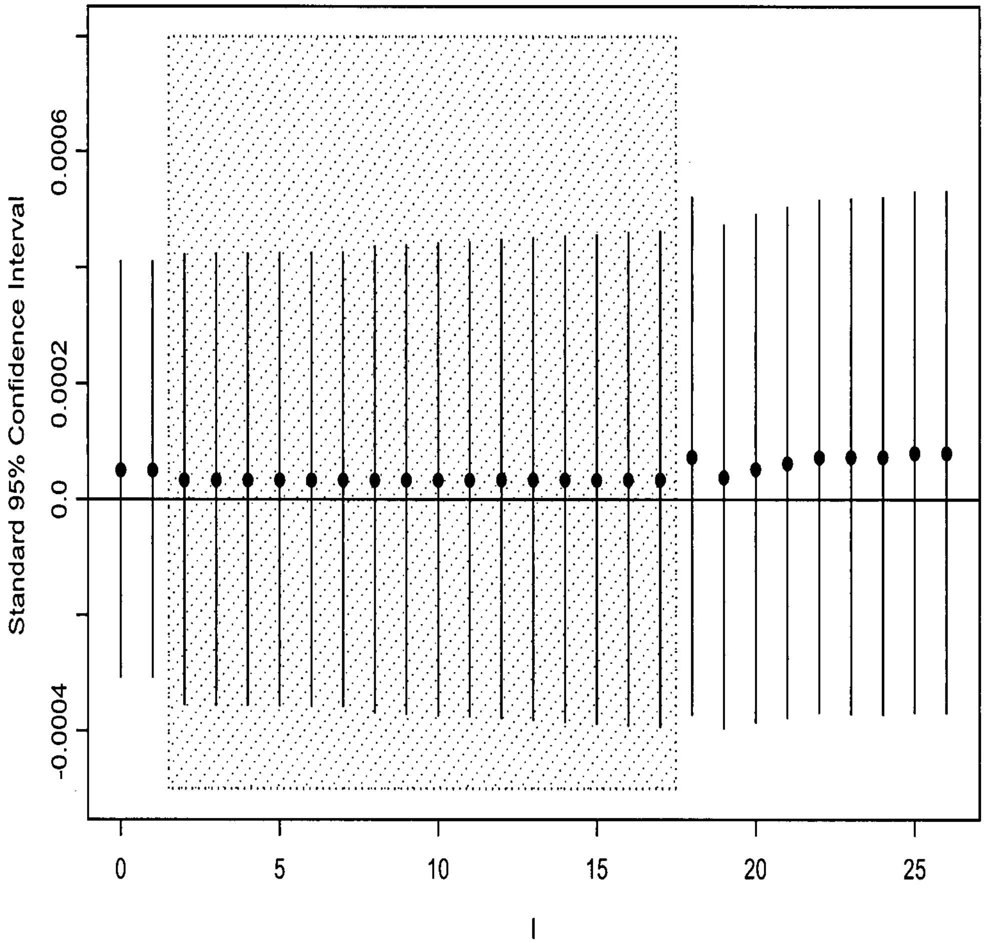

Figure 9.22: *Standard 95% confidence intervals for β_1 based on local linear back-fitting estimates of β_1 with plug-in choices of smoothing. The different intervals correspond to different values of l, where $l = 0, 1, \ldots, 26$. The shaded area represents confidence intervals corresponding to values of l that are reasonable for the data.*

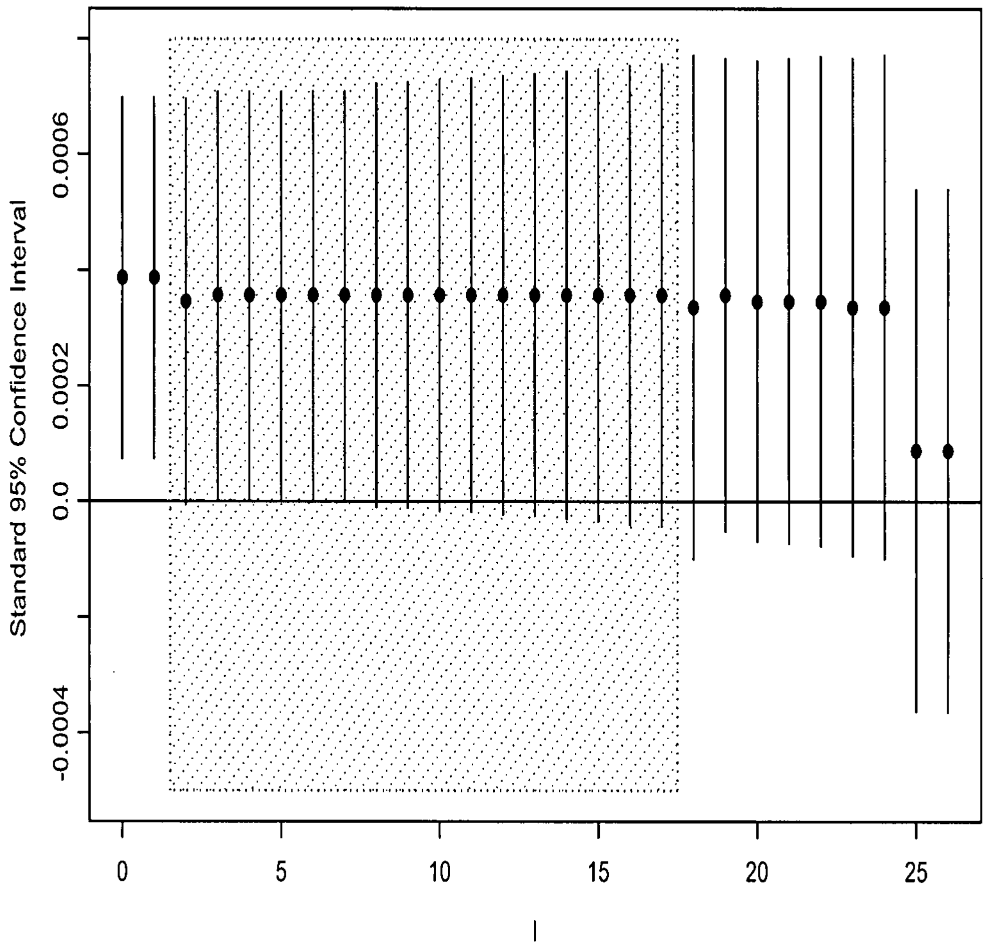

Figure 9.23: *Standard* 95% *confidence intervals for* β_1 *based on local linear backfitting estimates of* β_1 *with global EBBS choices of smoothing. The different intervals correspond to different values of* l*, where* $l = 0, 1, \ldots, 26$*. The shaded area represents intervals corresponding to values of* l *that are reasonable for the data; the intervals corresponding to* $l = 3, \ldots, 7$ *do not cross the horizontal line passing through zero.*

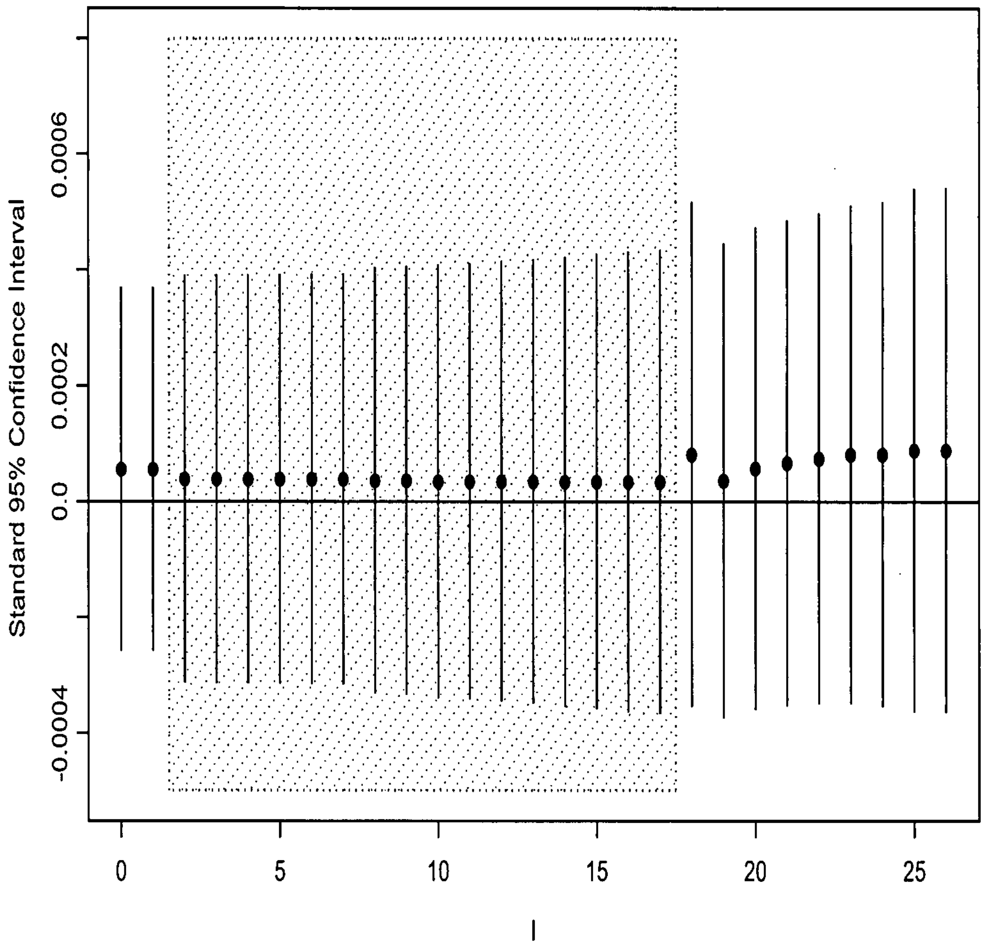

Figure 9.24: *Standard* 95% *confidence intervals for* β_1 *based on local linear backfitting estimates of* β_1 *with global EBBS choices of smoothing obtained by using a smaller grid range. The different intervals correspond to different values of* l*, where* $l = 0, 1, \ldots, 26$. *The shaded area represents confidence intervals corresponding to values of* l *that are reasonable for the data.*

Chapter 10

Conclusions

In this chapter, we provide an overview of the research problem considered in this thesis. We then outline the main contributions of this thesis and summarize the contents of each chapter. Finally, we suggest possible extensions to our work.

Partially Linear Models

Partially linear models are flexible tools for analyzing data from a variety of applications. They generalize linear regression models by allowing one of the variables in the model to have a non-linear effect on the response.

Inferences on the Linear Effects in Partially Linear Models

In many applications, the primary focus is on conducting inferences on the linear effects β in a partially linear model. In these applications, the non-linear effect m in the model is treated as a nuisance. This nuisance effect is a double-edged sword - while it affords greater modelling flexibility, it is also more difficult to estimate than the linear effects and, as such, it complicates the inferences on these effects.

Inferential Goals

Depending on the application, various goals could be relevant to the problem of conducting inferences on the linear effects in a partially linear models with correlated errors.

One goal would be to choose the correct amount of smoothing for accurately estimating the linear effects. One would hope that the methodology used for making this choice produces an amount of smoothing for which the linear effects are estimated at the 'usual' parametric rate of $1/n$ - the rate that would be achieved if the non-linear effect were known.

Another goal would be to construct valid standard errors for the estimated linear effects.

An additional goal would be to use the estimated linear effects and their associated standard errors to construct valid confidence intervals and tests of hypotheses for assessing the magnitude and statistical significance of the linear effects, possibly adjusting for smoothing bias. Little has been done in the literature to address this goal.

Research Questions Concerning the Inferential Goals

Various research questions emerge in connection with the inferential goals listed above:

1. How can we choose the correct amount of smoothing for accurate estimation of the linear effects?

2. How can we estimate the correlation structure of the model errors for conducting inferences on the linear effects?

3. How can we construct valid standard errors for the estimated linear effects?

4. How can we construct valid confidence intervals and tests of hypotheses for assessing the magnitude and statistical significance of the linear effects?

5. What is the impact of the choice of amount of smoothing on the validity of the confidence intervals and tests of hypotheses?

6. Could inefficient estimates of the linear effects provide valid inferences?

Thesis Contributions

The major contributions of this thesis to the research questions stated above are: (1) defining sensible estimators of the linear and non-linear effects in partially linear models with correlated errors, (2) deriving explicit expressions for the asymptotic conditional bias and variance of the proposed estimators of the linear effects, (3) developing data-driven methods for selecting the appropriate amount of smoothing for accurate estimation of the linear effects, (4) developing confidence interval and hypothesis testing procedures for assessing the magnitude and statistical significance of the linear effects of main interest, (5) studying the finite-sample properties of these procedures, and (6) applying these procedures to the analysis of an air pollution data set. These contributions are discussed in more detail below.

The estimators we proposed in this thesis are backfitting estimators, relying on locally linear regression, which is known to posses attractive theoretical and practical properties. Many of the backfitting estimators proposed in the literature of partially linear regression models with correlated errors rely on locally constant regression, a method that does not enjoy the good properties of locally linear regression.

In Chapters 4 and 5 of this thesis, we studied the large-sample behaviour of the estimators of linear effects introduced in this thesis as the width of the smoothing window used in locally linear regression decreases at a specified rate, and the number of data points in this window increases. Specifically, we obtained explicit expressions for the conditional asymptotic bias and variance of these estimators. Our asymptotic results are important as they show that, in the presence of correlation between the linear and non-linear variables in the model, the bias of the estimators of the linear effects can dominate their variance asymptotically, therefore compromising their $\sqrt{n}$-consistency. This problem can be remedied however by selecting an appropriate rate of convergence for the smoothing parameter of the estimators. This rate is slower than the rate that is optimal for estimation of the non-linear effect, and as such it 'undersmooths' the estimated non-linear effect.

Selecting the appropriate amount of smoothing for the estimators of the linear effects is a crucial problem, which is complicated by the presence of error correlation and dependencies between the linear and nonlinear components of the model. Our theoretical results indicate that the amount of smoothing that is 'optimal' for estimating the non-linear effect is not 'optimal' for estimating the linear effects. Data-driven methods devised for accurate estimation of the non-linear effect will likely fail to yield a satisfactory choice of smoothing for estimating the linear effects. In this thesis, we proposed three data-driven smoothing parameter selection methods. Two of these methods are modifications of the EBBS method of Opsomer and Ruppert (1999) and rely on the asymptotic bias results derived in this thesis. The third method is a non-asymptotic plug-in method. Our methods fill a gap in the literature of partially linear models with correlated errors, as they are designed specifically for accurate estimation of the linear effects. These methods 'undersmooth' the estimated non-linear effect because they attempt to estimate the amount of smoothing that is MSE-optimal for estimating the linear effects, not the amount of smoothing that is MSE-optimal for estimating the non-linear effect. Our theoretical results suggest that, in general, the amount of smoothing that is MSE-optimal for estimating the linear effects is smaller than the amount of smoothing that is MSE-optimal for estimating the non-linear effect.

The issue of conducting valid inferences on the linear effects in a partially linear model with correlated errors is inter-connected with the appropriate choice of smoothing for estimating these effects. Most literature results devoted to this issue use choices of smoothing that 'do well' for estimation of the non-linear effect and are deterministic. Such choices may not be satisfactory when one wishes to 'do well' for estimation of the linear effects and hence have little practical value in such contexts. The confidence interval and hypothesis testing procedures proposed in this thesis are constructed with data-driven choices of smoothing. They are either standard, bias-adjusted or standard-error adjusted. To our knowledge, adjusting for bias in confidence intervals and tests of hypotheses has not been attempted in the literature of partially linear models. The inferential procedures we introduced in this thesis do not account for the uncertainty associated with the fact that the choice of smoothing is data-dependent and the error correlation structure is estimated from the data. However, simulations indicate that several of these procedures perform reasonably well for finite samples.

In Chapter 8, we conducted a Monte Carlo simulation study to investigate the finite sample properties of the linear effects estimators proposed in this thesis, namely, the usual and estimated modified local linear backfitting estimators. We also compared the properties of these estimators against those of the usual Speckman estimator. In our simulation study, we chose the smoothing parameter of the backfitting estimators using the data-driven methods developed in Chapter 6. By contrast, we chose the smoothing parameter of the usual Speckman estimator using cross-validation, modified for correlated errors (MCV) and for boundary effects. The main goals of our simulation study were (1) to compare the expected log mean squared error of the estimators and (2) to compare the performance of the confidence intervals built from these estimators and their associated standard errors. Our study suggested that the usual local linear backfitting estimator should be used in practice, with either a global modified EBBS or a non-asymptotic plug-in choice of smoothing. To ensure the validity of the inferences based on this estimator and its associated standard error, one should never use small values of l in the modified (or leave-(2l+1)-out) cross-validation criterion utilized in estimating the error correlation structure. Adjusting these inferences for possible bias effects did not affect the quality of our results. The quality of the inferences based on the estimated modified local linear estimator was poor for many simulation settings, owing to the fact that the associated standard errors were too variable. The quality of the inferences based on the Speckman estimator was reasonable for most simulation settings, but not as good as that of the inferences based on the usual local linear backfitting estimator.

In Chapter 9, we used the inferential methods developed in this thesis to assess whether the pollutant PM10 had a significant short-term effect on log mortality in Mexico City during 1994-1996, after adjusting for temporal trends and weather patterns. Our data analysis suggested that there is no conclusive proof that PM10 had a significant short-term effect on log mortality. Our data analysis differs from standard analyses in that it relies on objective methods to adjust this effect for temporal confounding.

Further Work to be Done

As usual, there is further work to be done. The following are just a few of the issues that need additional investigation.

Proofs of the asymptotic normality of the linear effects estimators proposed in this thesis are still pending. These proofs will provide formal justification for using standard confidence intervals and tests of hypotheses based on these estimators and their associated standard errors.

Further investigation into the appropriate choice of l in the modified cross-validation criterion used in estimating the error correlation structure is needed. This choice should take into account the range and magnitude of the error correlation.

Possible Extensions to Our Work

The work in this thesis can be extended in various directions.

First, we could extend the partially linear model considered in this thesis by allowing additional univariate smooth terms to enter the model. Such models arise frequently in practical applications. Developing inferential methodology for these models is therefore important. To carry out inferences on the linear effects in such models we would need to simultaneously choose the amounts of smoothing for estimating all the non-linear effects. These amounts should be appropriate for accurate estimation of the linear effects and should account for correlation between the linear and non-linear variables and correlation between the model errors.

Second, we could extend the partially linear model considered in this thesis to responses that are not continuous. For instance, the responses could follow a Poisson distribution. Incorporating correlation in such models could be a challenge.

Third, we could extend the partially linear model considered in this thesis by allowing the non-linear variable to be a spatial coordinate, in which case m is a spatial effect. Such a model is termed a spatial partially linear model. Clearly, in many contexts, the errors would be correlated. Spatial partially linear models with correlated errors can be used, for instance, to analyze spatial data observed in epidemiological studies of particulate air pollution and mortality. Typically, in these applications, the linear effects $\boldsymbol{\beta}$ are of main interest, while the spatial effect m is treated as a nuisance. Ramsay et al. (2003b) considered spatial partially linear models with uncorrelated errors and estimated $\boldsymbol{\beta}$ and m using the S-Plus function *gam* with loess as a smoother. They used *gam*'s default choice of smoothing to control the degree of smoothness of the estimated m. They showed via simulation that the correlation between the linear and spatial terms in the model can lead to underestimation of the true standard errors associated with the estimated linear effects, both when using S-Plus standard errors and so-called asymptotically unbiased standard errors. They cautioned that using such standard errors can compromise the validity of inferences concerning the linear effects, but did not propose a solution for alleviating this problem. Their findings highlight the fact that carrying out inferences on the linear effects in spatial partially linear models with uncorrelated errors is challenging in the presence of correlation between the linear and spatial terms in the model. Obviously, error correlation will further compound the challenges involved in conducting valid inferences on the linear effects in spatial partially linear models. Of course, this work would be relevant in the non-spatial context as well.

Bibliography

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

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

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

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

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

[6] Broersen, P.M.T. (2000). Finite Sample Criteria for Autoregressive Order Selection. *IEEE Transactions on Signal Processing,* **48**, 3550-3558.

[7] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics,* **17**, 453-555.

[8] Chatfield, C. (1989). *The Analysis of Time Series: An Introduction.* Fourth Edition. New York: Chapman and Hall.

[9] Chu, C.-K., Marron, J.S. (1991). Comparison of two bandwidth selectors with dependent errors. *Annals of Statistics,* **19**, 1906-1918.

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

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

[12] Engle, R.F., Granger, C. W.J., Rice, J. and Weiss, A. (1983). Nonparametric estimates of the relation between weather and electricity demand. Technical report, U.C. San Diego

[13] Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *The Journal of the American Statistical Association,* **81**, 310-320.

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

[15] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* New York: Chapman and Hall.

[16] Fan, J. and Gijbels, I. (1992). Variable Bandwidth and Local Linear Regression Smoothers. *The Annals of Statistics,* **20**, 2008-2036.

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

[18] Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics,* **11**, 171-185.

[19] Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B,* **47**, 299-315.

[20] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* New York: Chapman and Hall.

[21] Härdle,W. and Vieu, P. (1992). Kernel regression smoothing of time series. *Journal of Time Series Analysis,* **13**, 209-232.

[22] Heckman, N.E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Association, Series B,* **48**, 244-248.

[23] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables.* Groningen: Wolters Noordhoff.

[24] Katsouyanni, K., Toulomi, G. and Samoli, E., et al. (1997). Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. *Epidemiology,* **12**, 521-531.

[25] Kelsall, J.E., Samet, J.M. and Zeger, S.L. (1997). Air pollution and mortality in Philadelphia, 1974-1988. *American Journal of Epidemiology,* **146**, 750-762.

[26] Moolgavakar, S. (2000). Air pollution and hospital admissions for diseases of the circulatory system in three U.S. metropolitan areas. *Journal of the Air Waste Management Association,* **50**, 1199-1206.

[27] Moyeed, R.A. and Diggle, P.J. (1994). Rate of convergence in semiparametric modelling of longitudinal data. *Australian Journal of Statistics,* **36**, 75-93.

[28] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and Its Applications* , **9**, 141-142.

[29] Opsomer, J.D. and Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *The Journal of the American Statistical Association,* **93**, 605-620.

[30] Opsomer, J.D. and Ruppert, D. (1999). A root-n consistent estimator for semiparametric additive modelling. *Journal of Computational and Graphical Statistics,* **8**, 715-732.

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

[32] Ramsay, T., Burnett, R., Krewski, D. (2003b). Exploring bias in a generalized additive model for spatial air pollution data. *Environmental Health Perspectives,* **111**, 1283-1288.

[33] Rice, J.A. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters,* **4**, 203-208.

[34] Robinson, P.M. (1988). Root-n-consistent semiparametric regression. *Econometrica,* **56**, 931-954.

[35] Samet, J.M., Dominici, F., Curriero,F., et al. (2000). Fine particulate air pollution and mortality in 20 U.S. cities: 1987-1994 (with discussion). *New England Journal of Medicine,* **343**, 1742-1757.

[36] Schwartz, J. (1994). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *The Canadian Journal of Statistics,* **22**, 471-488.

[37] Schwartz, J. (1999). Air pollution and hospital admissions for heart disease in eight US counties. *Epidemiology,* **10**, 17-22.

[38] Schwartz, J. (2000). Assessing confounding, effect modification, and thresholds in the associations between ambient particles and daily deaths. *Environmental Health Perspectives,* **108**, 563-568.

[39] Shick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic Processes and their Applications,* **61**, 339-361.

[40] Shick, A. (1999). Efficient estimation in a semiparametric additive regression model with ARMA errors. *Stochastic Processes and their Applications,* **61**, 339-361.

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

[42] Sy, H. (1999). Automatic bandwidth choice in a semiparametric regression model. *Statistica Sinica,* **9**, 775-794.

[43] Truong, Y.K. (1991). Nonparametric curve estimation with time series errors. *Journal of Statistical Planning and Inference,* **28**, 167-183.

[44] Wahba, G. (1984). Cross-validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H. A. David, ed.) Iowa State University Press, 205-235.

[45] Watson, G.S. (1964). Smooth regression analysis. *Sankhya A,* **26**, 359-372.

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

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

Appendix A

MSE Comparisons

In this appendix, we provide plots to help assess and compare the MSE properties of the estimators of the linear effect β_1 in model (8.1) that were discussed in Section 8.2.

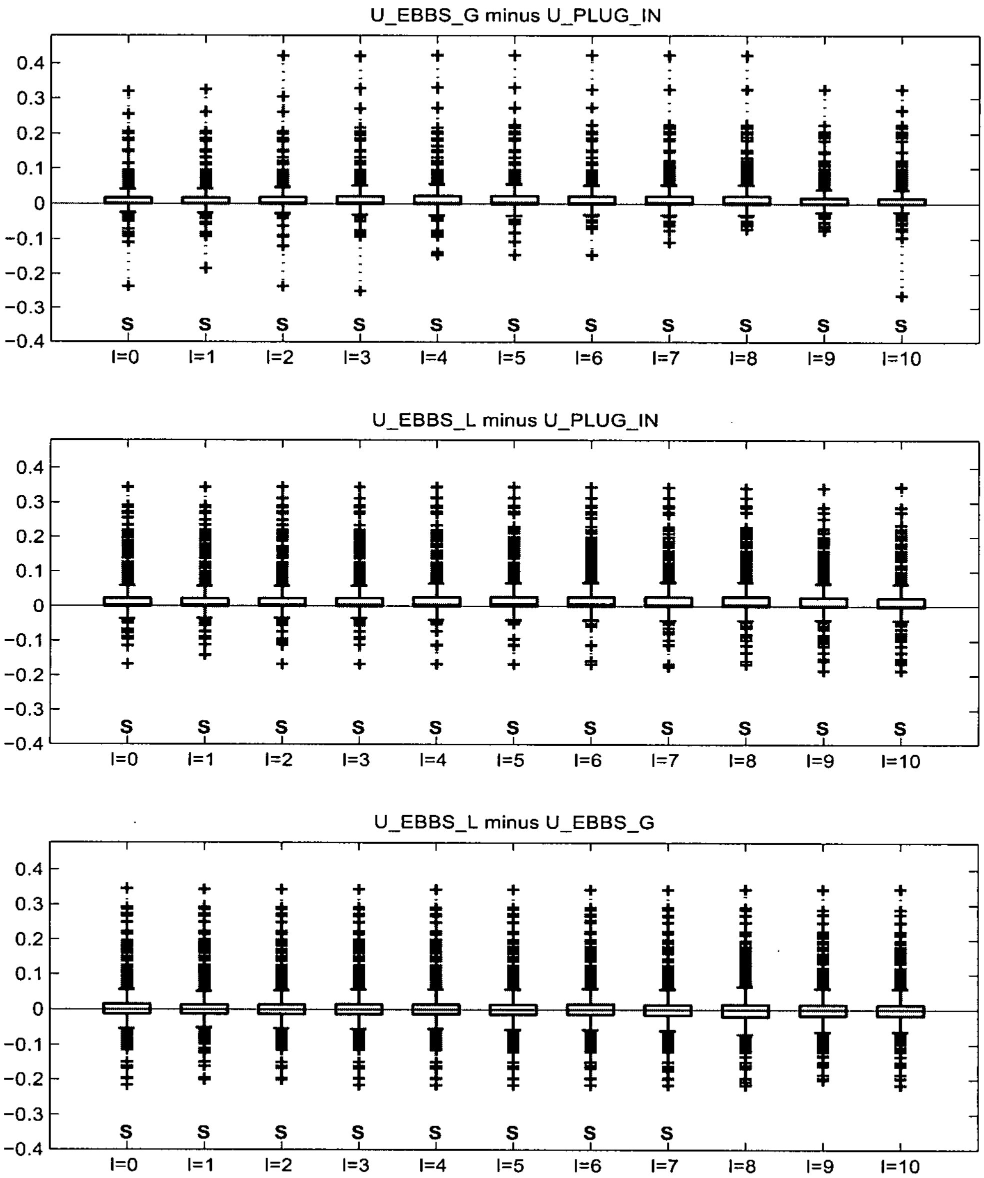

Figure A.1: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_1(z)$.

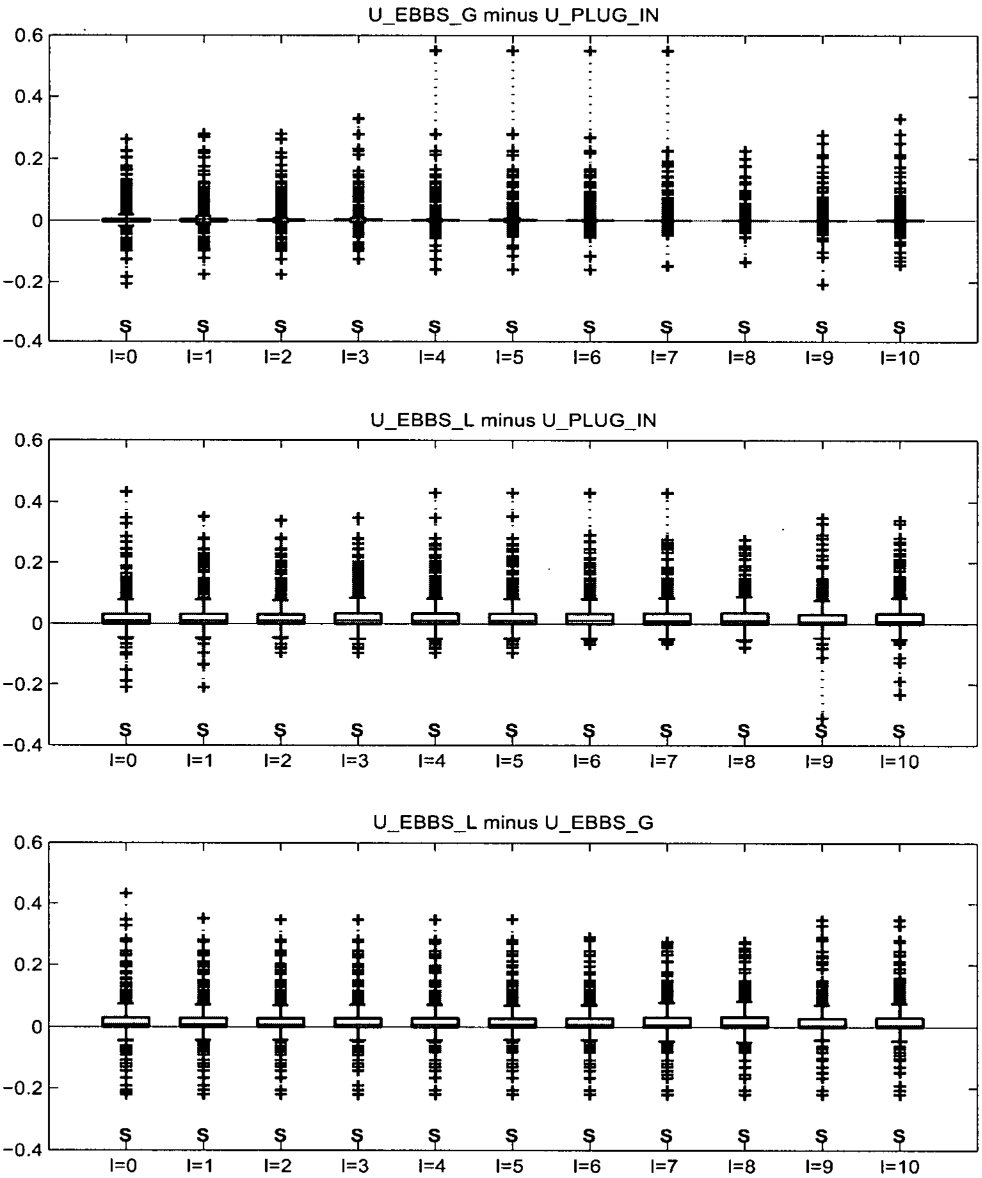

Figure A.2: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an S. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_1(z)$.

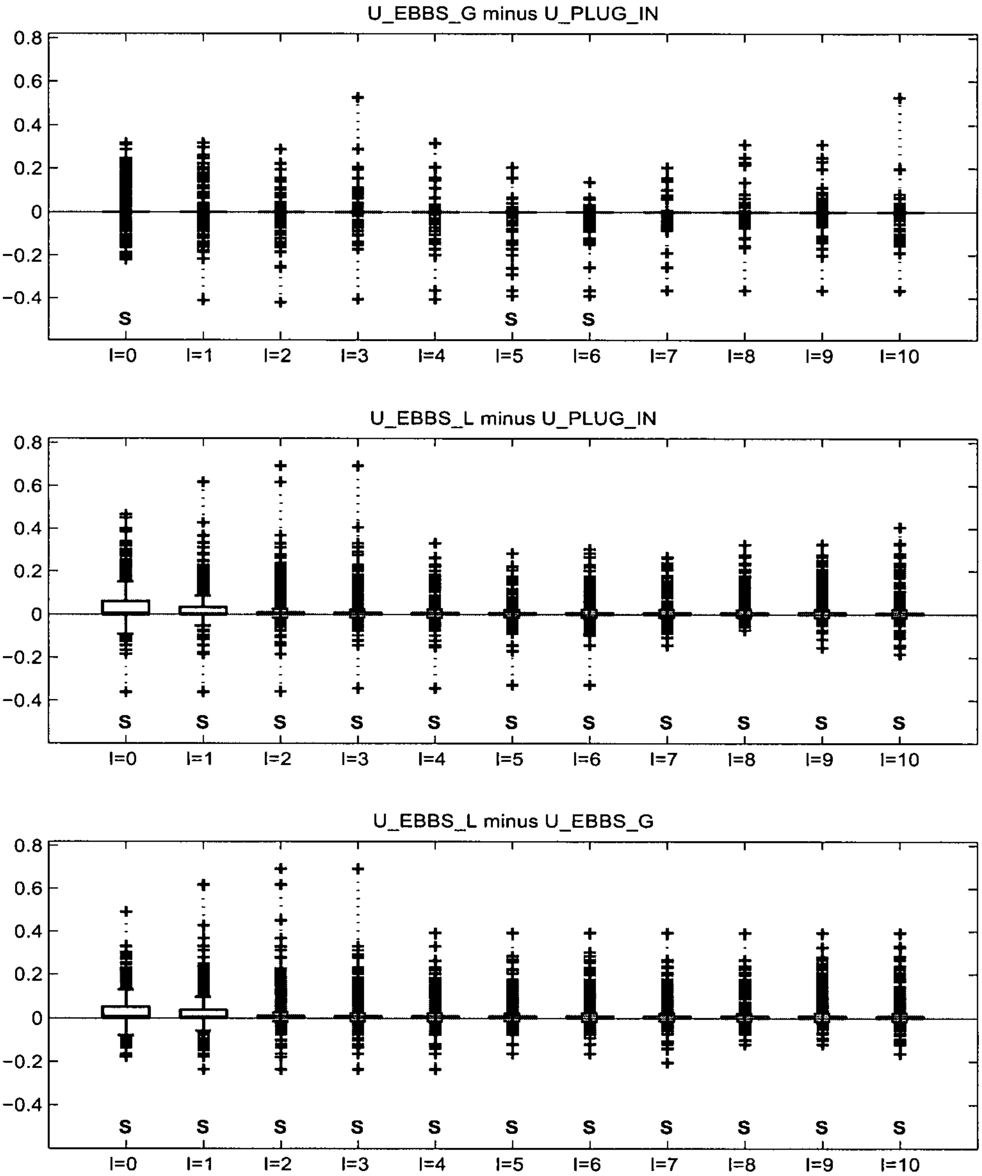

Figure A.3: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1)*, where* $l = 0, 1, \ldots, 10$*. Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S*. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_1(z)$*.*

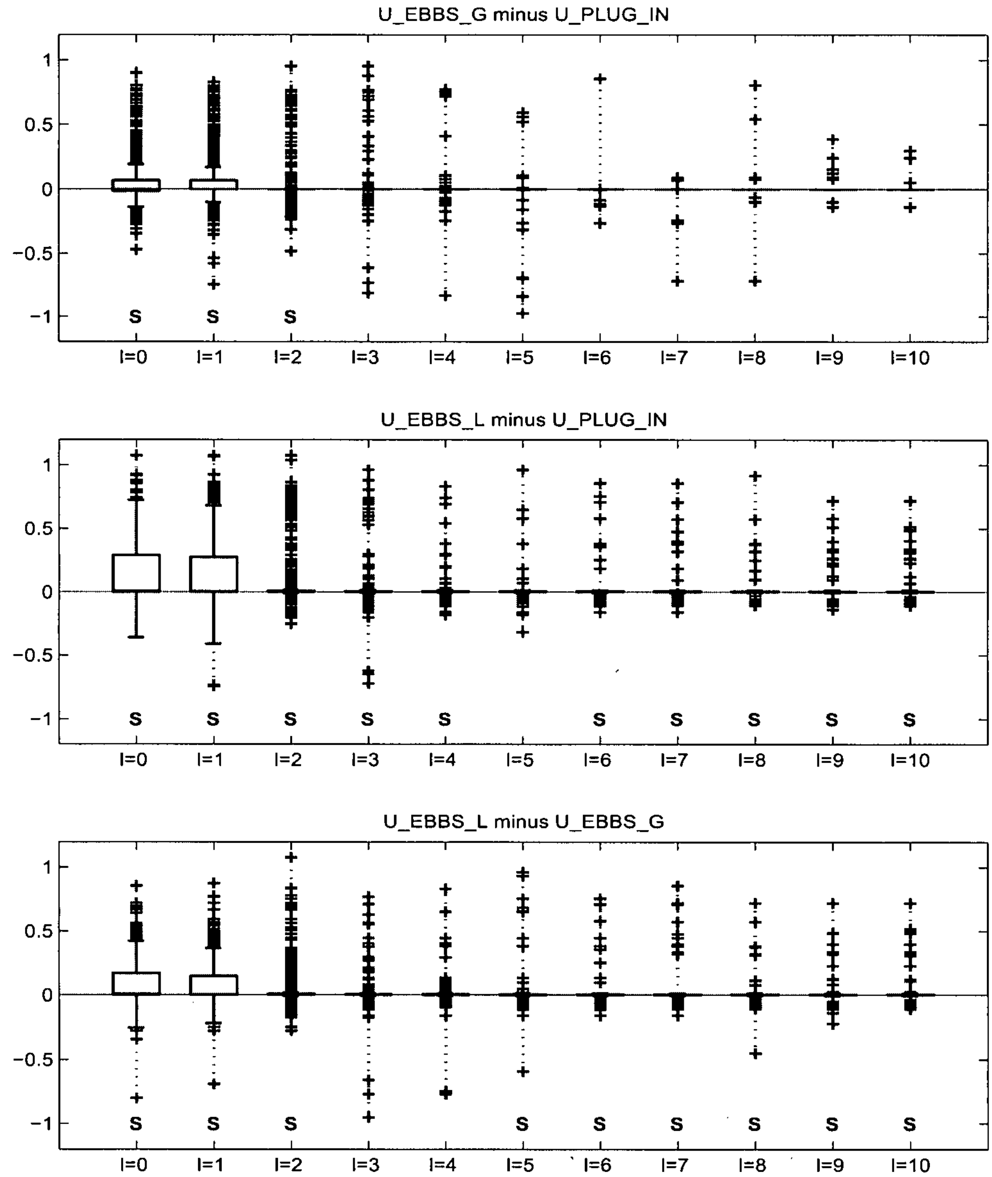

Figure A.4: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.6$ *and* $m(z) = m_1(z)$.

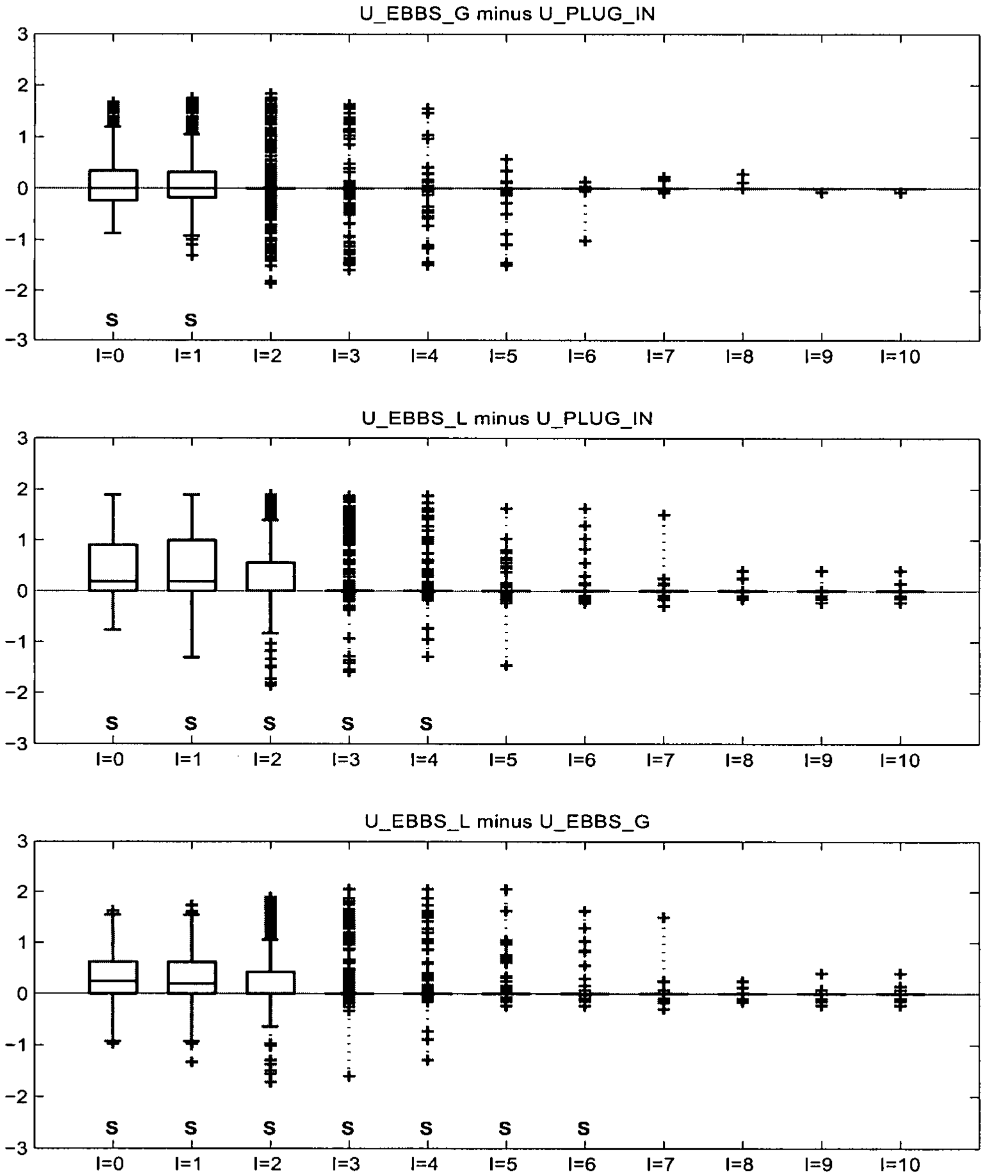

Figure A.5: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$*,* $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1)*, where* $l = 0, 1, \ldots, 10$*. Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an S. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_1(z)$*.*

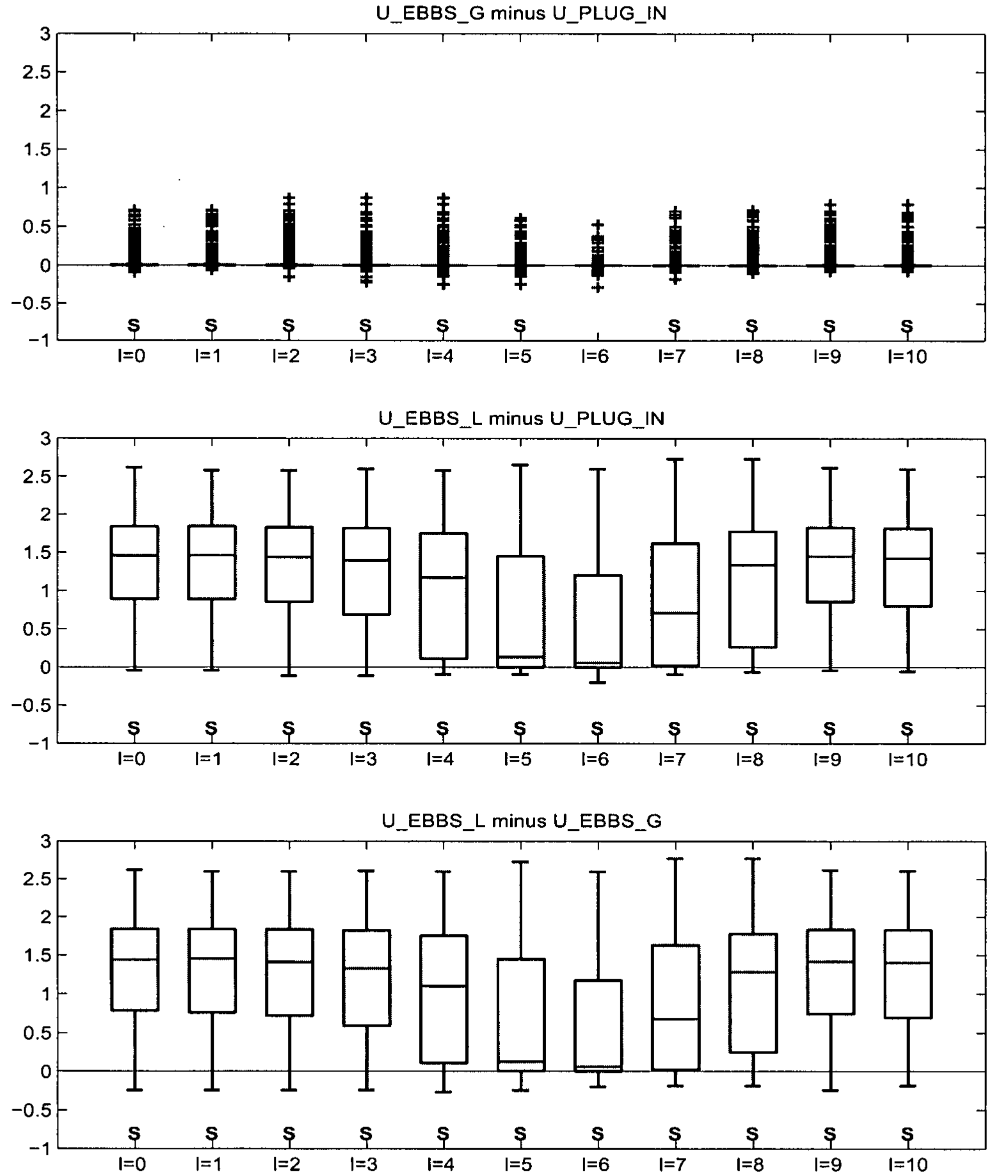

Figure A.6: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1)*, where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_2(z)$.

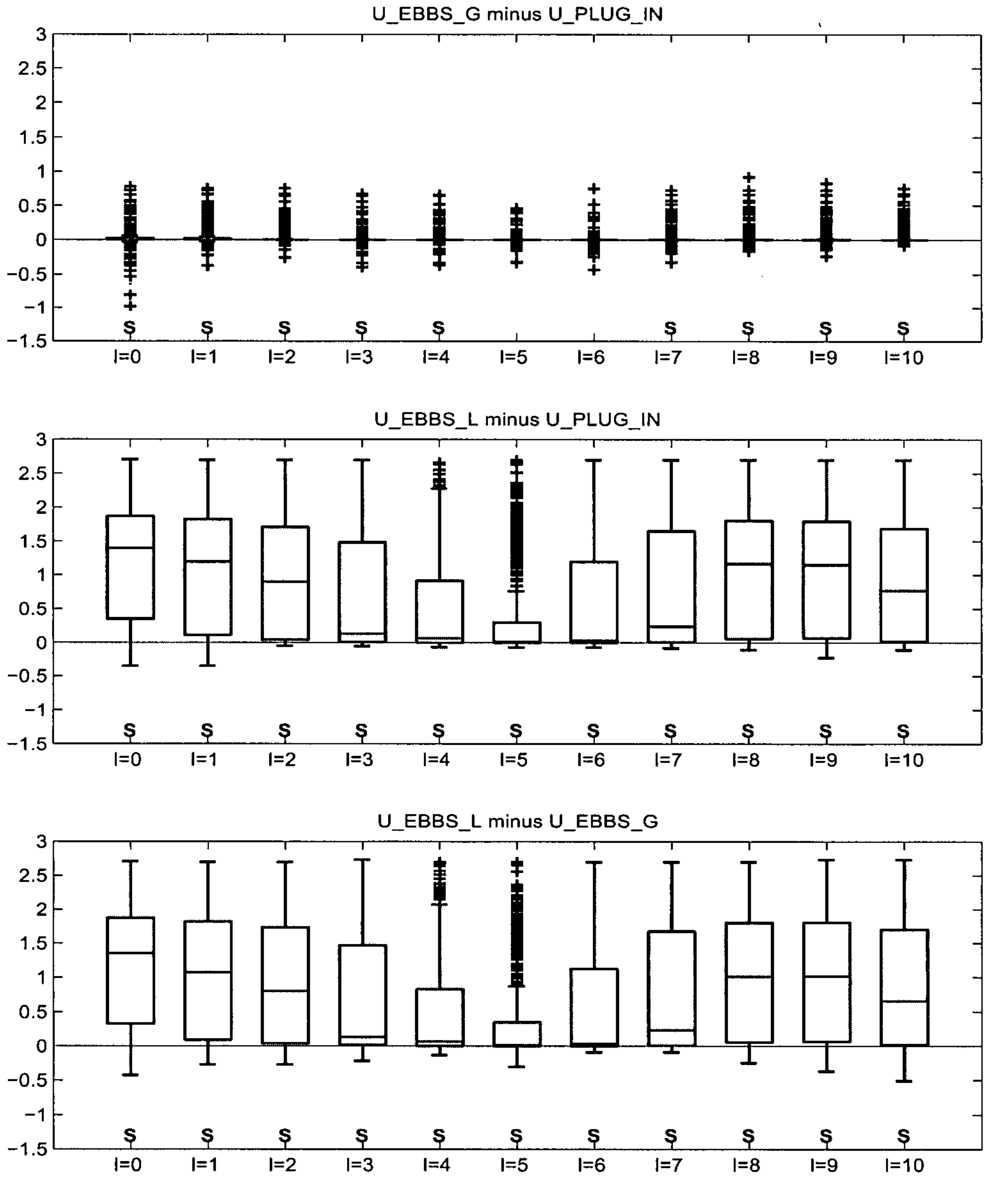

Figure A.7: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_2(z)$.

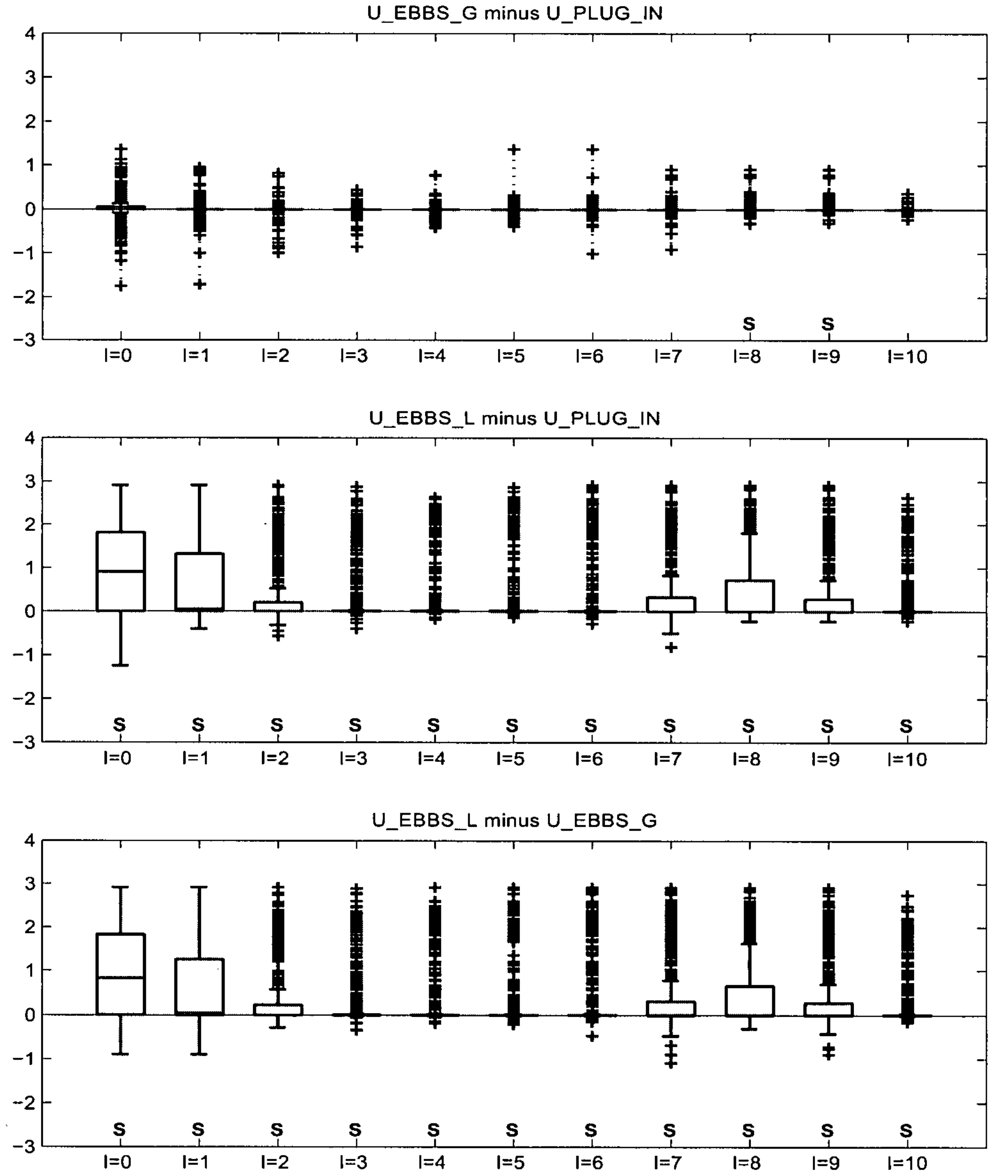

Figure A.8: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_2(z)$.

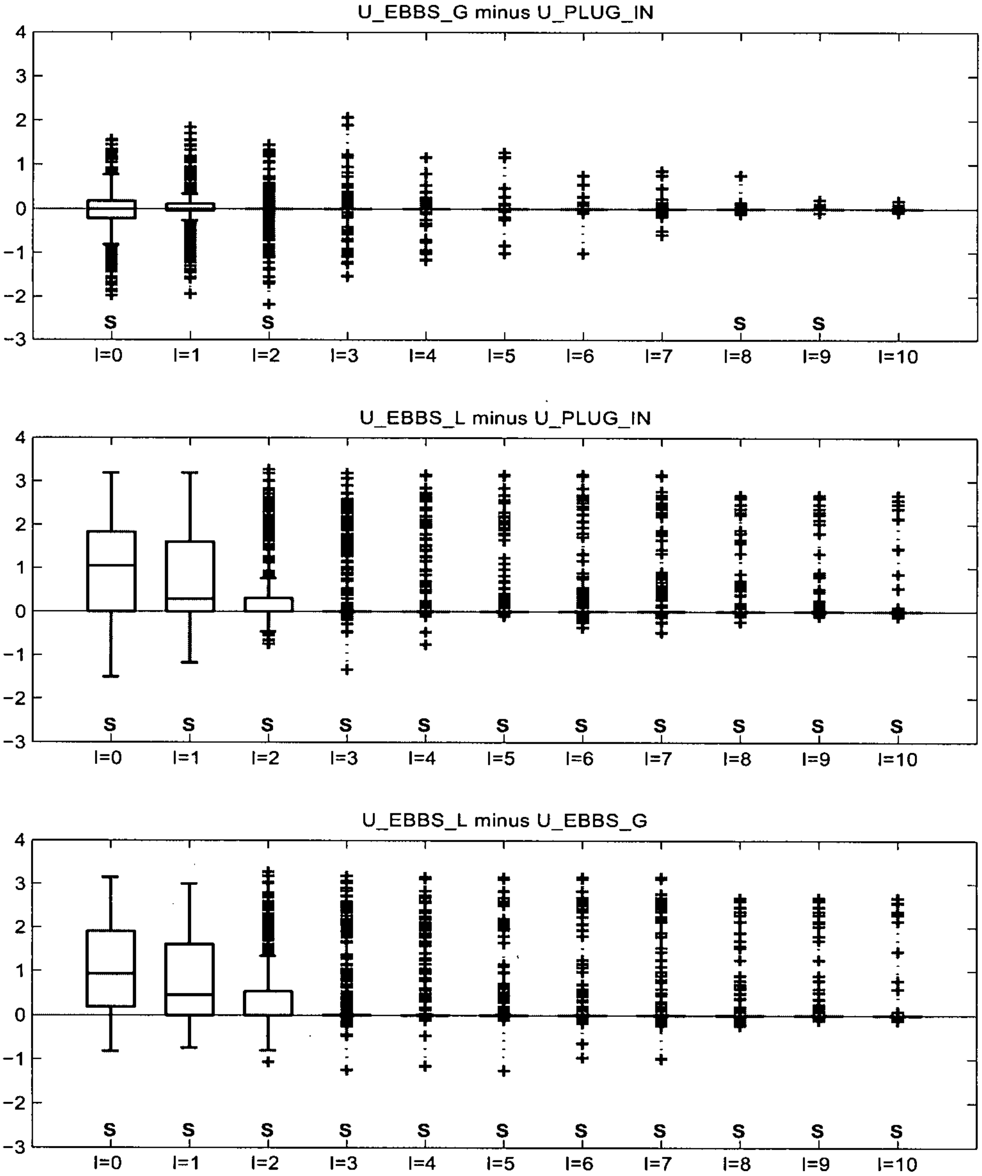

Figure A.9: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.6$ *and* $m(z) = m_2(z)$.

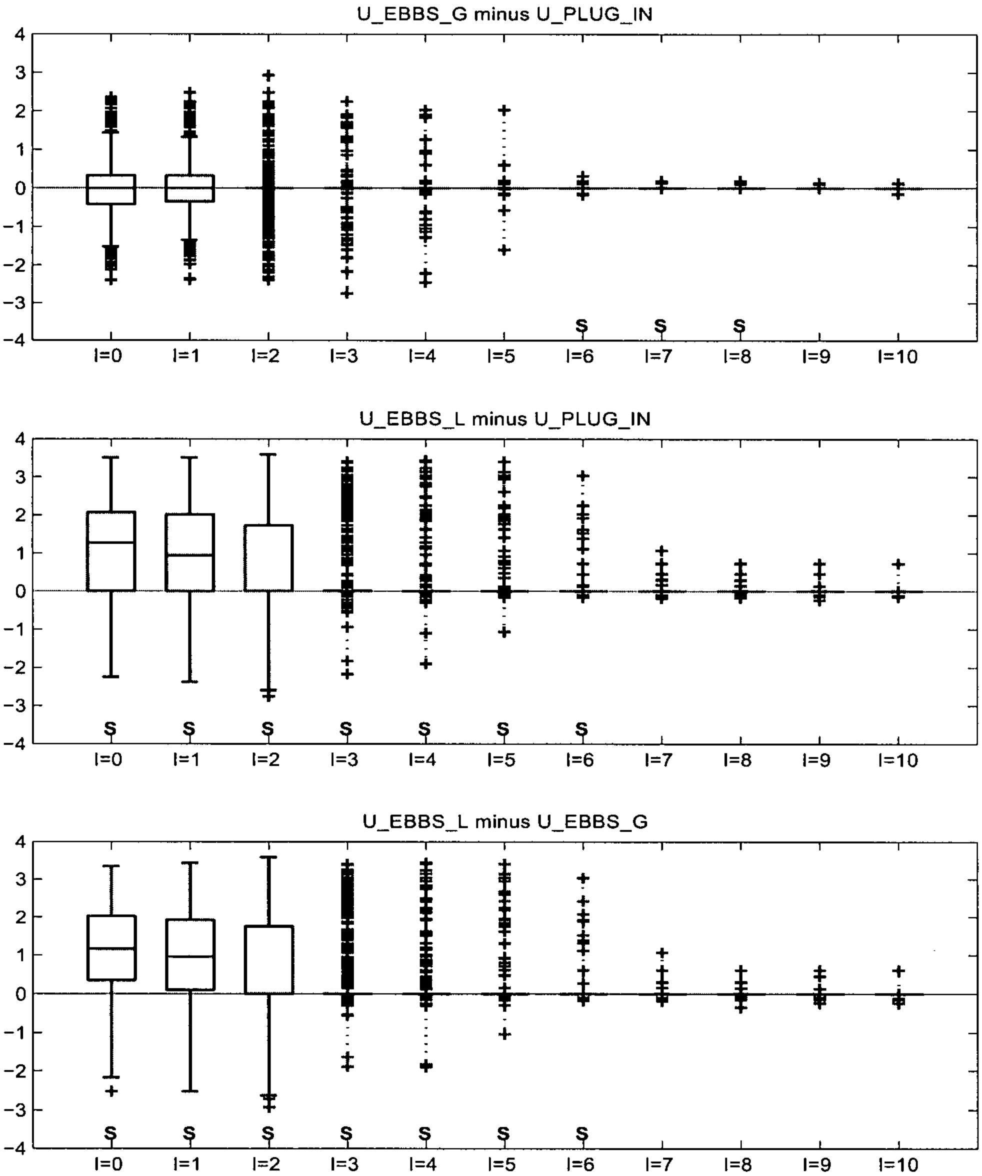

Figure A.10: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{U,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an S. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_2(z)$.

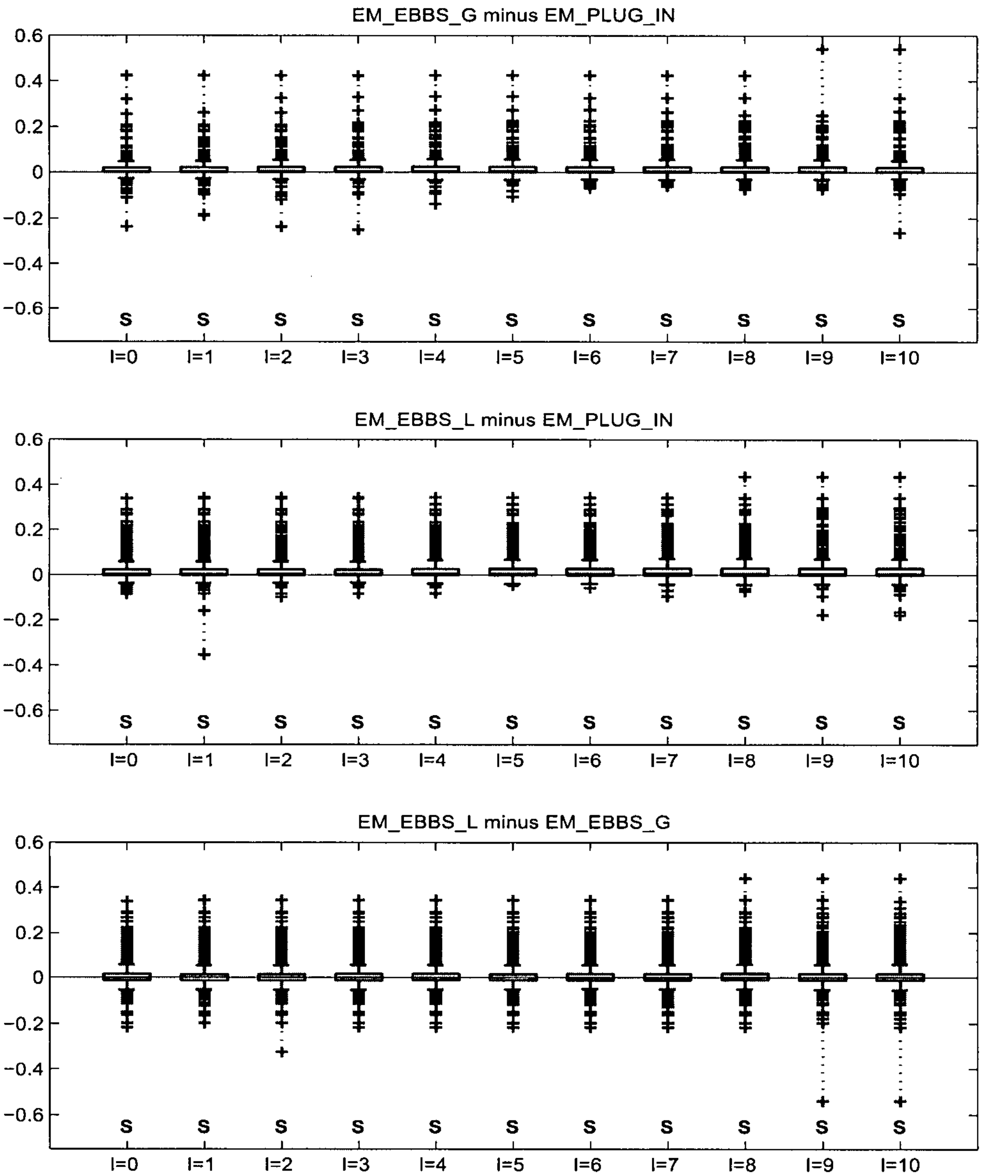

Figure A.11: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an S. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_1(z)$.

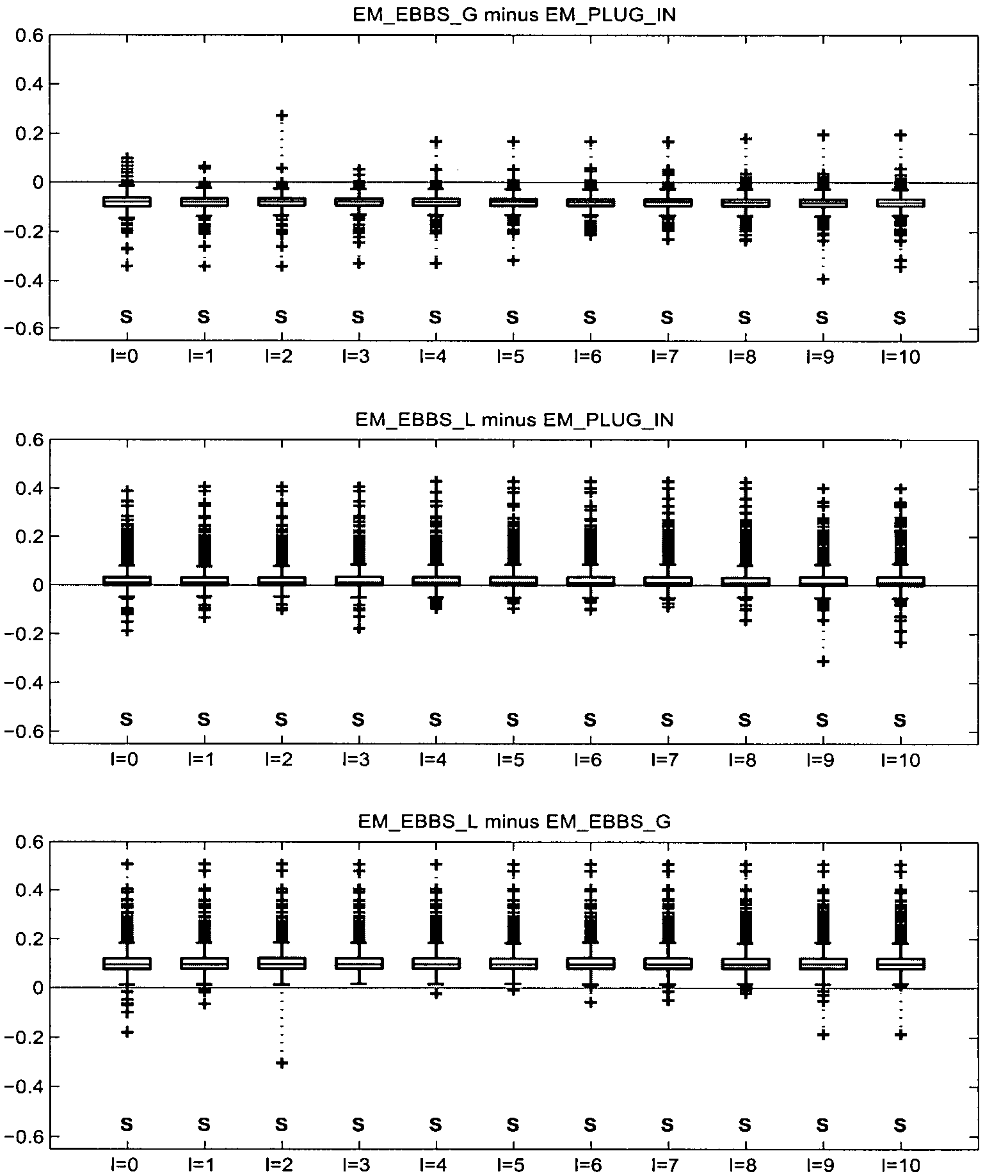

Figure A.12: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an S. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_1(z)$.

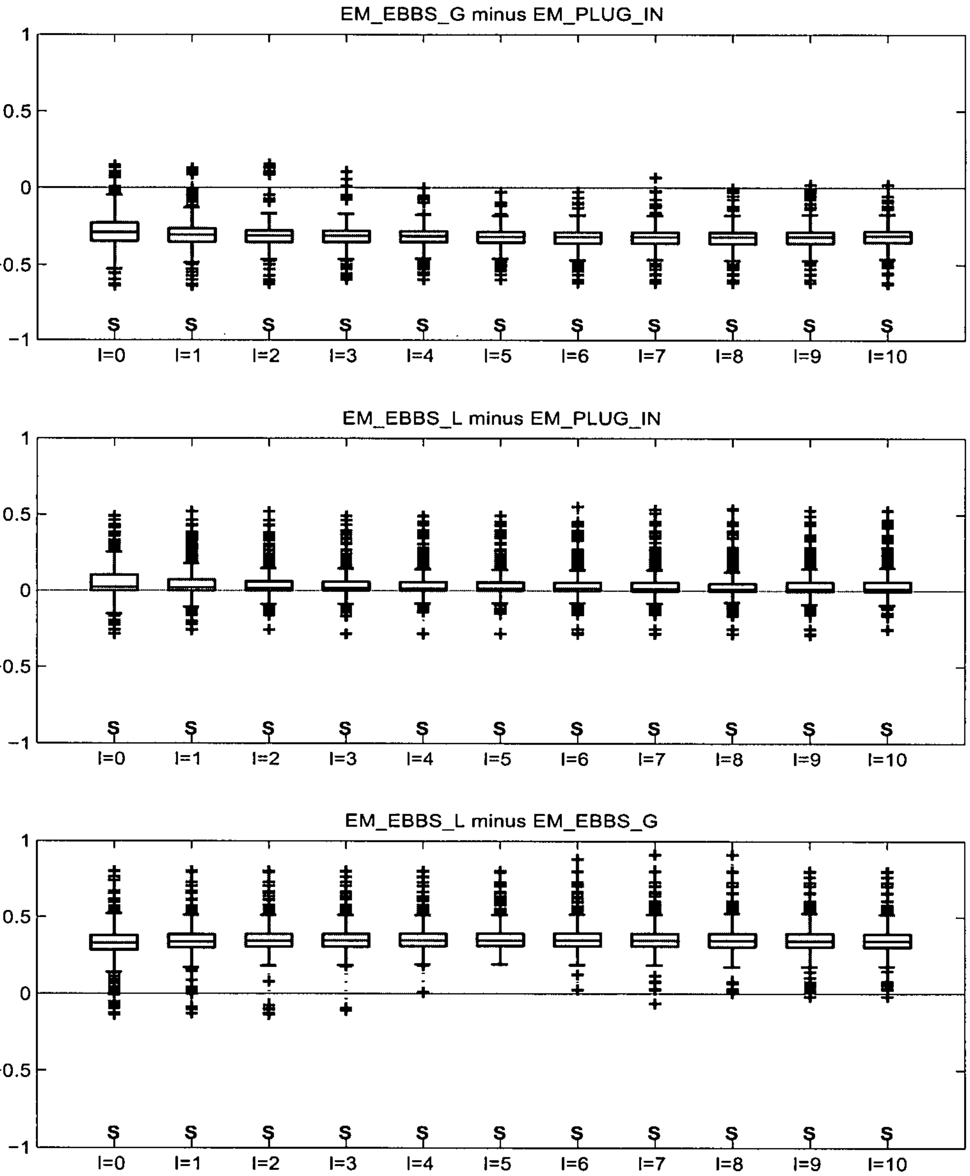

Figure A.13: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_1(z)$.

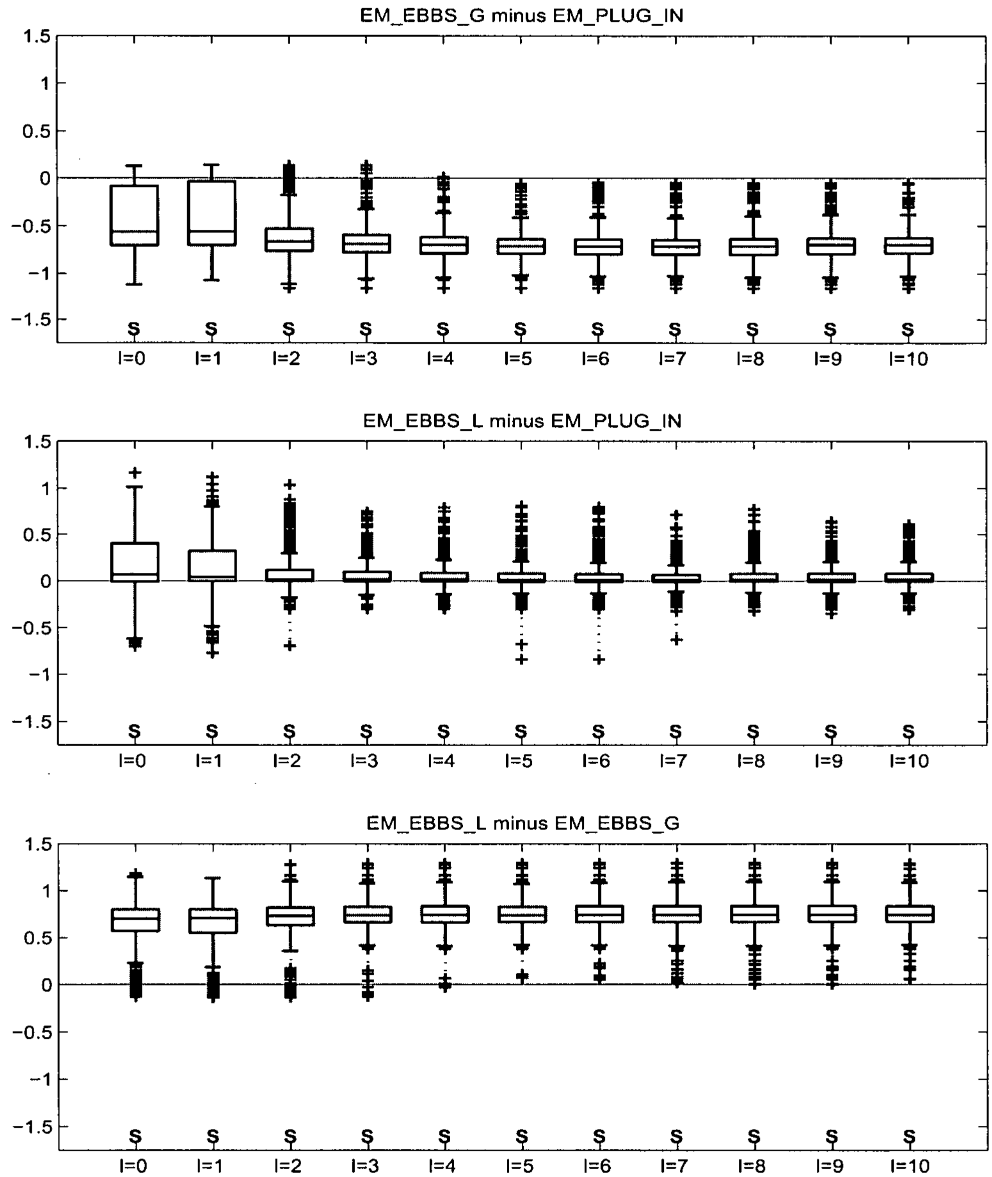

Figure A.14: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE's is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.6$ *and* $m(z) = m_1(z)$.

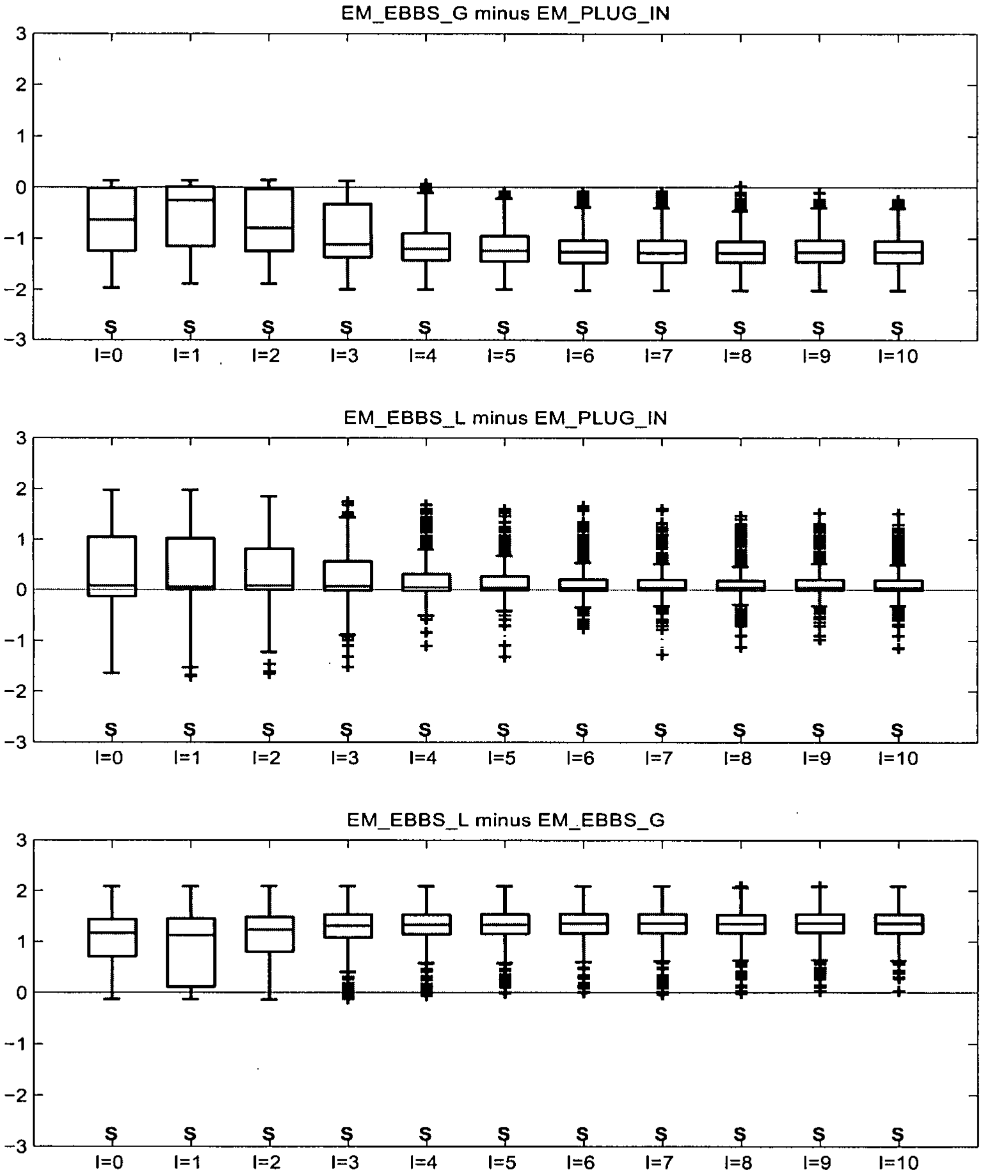

Figure A.15: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1)*, where* $l = 0, 1, \ldots, 10$*. Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S*. Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_1(z)$*.*

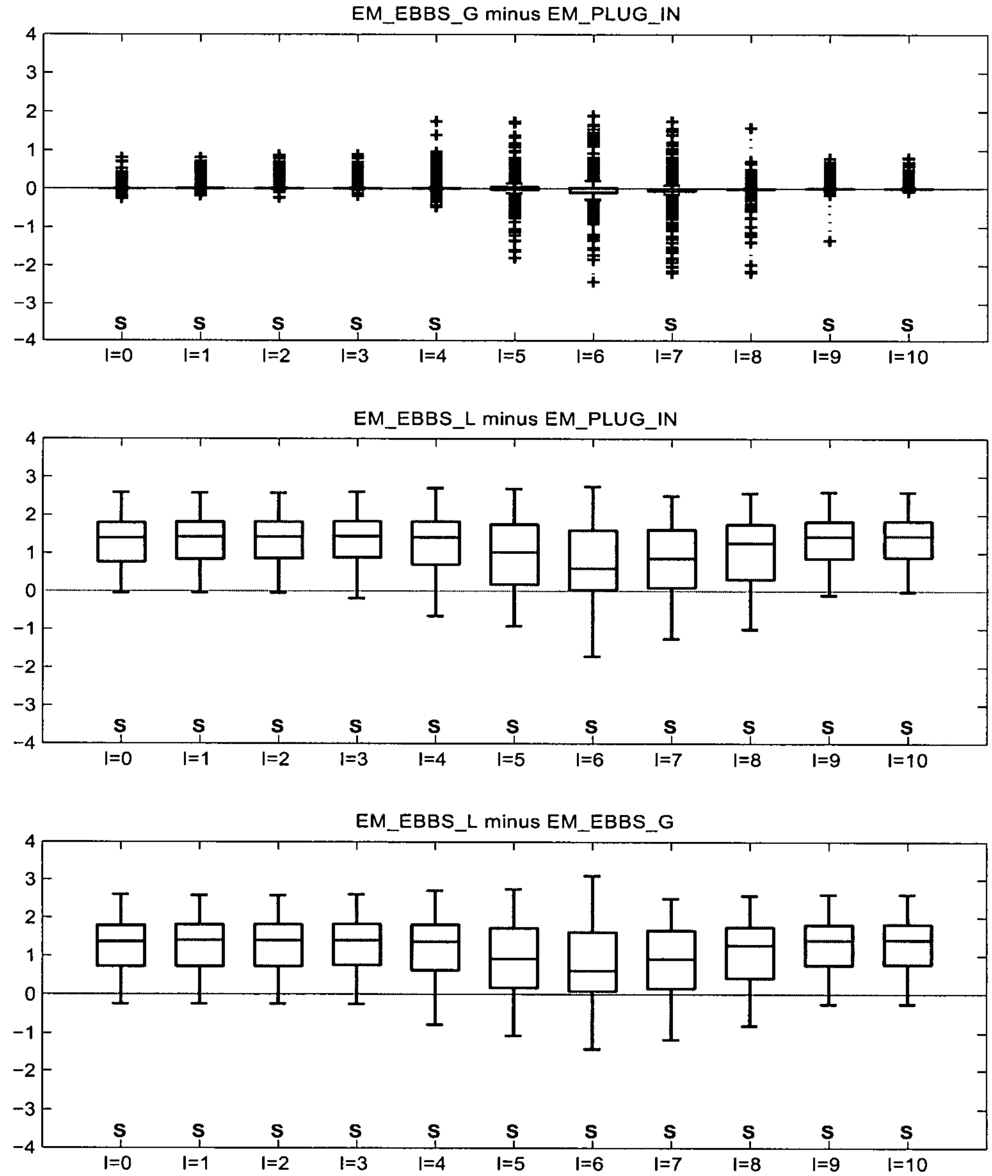

Figure A.16: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *significance level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_2(z)$.

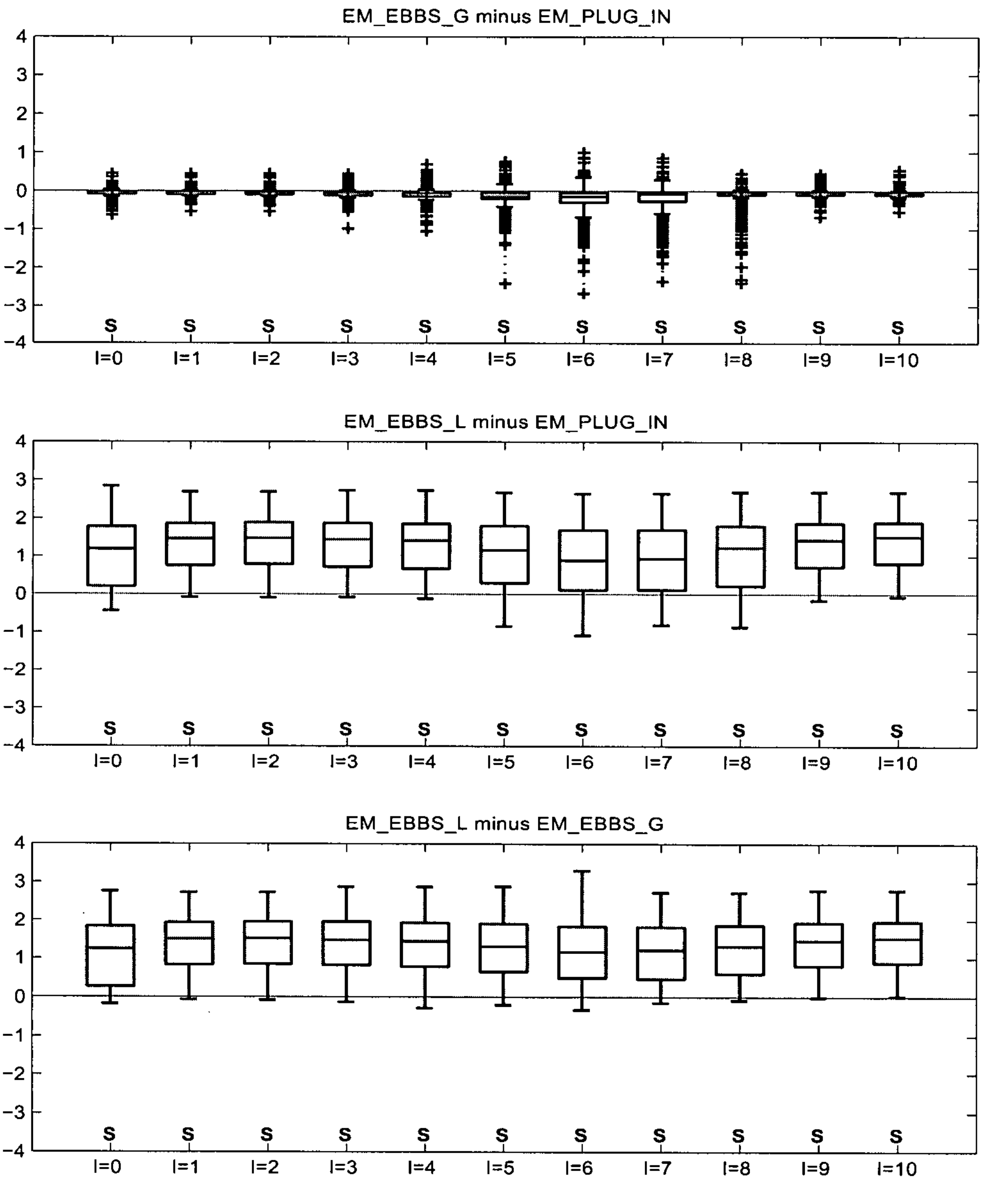

Figure A.17: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_2(z)$.

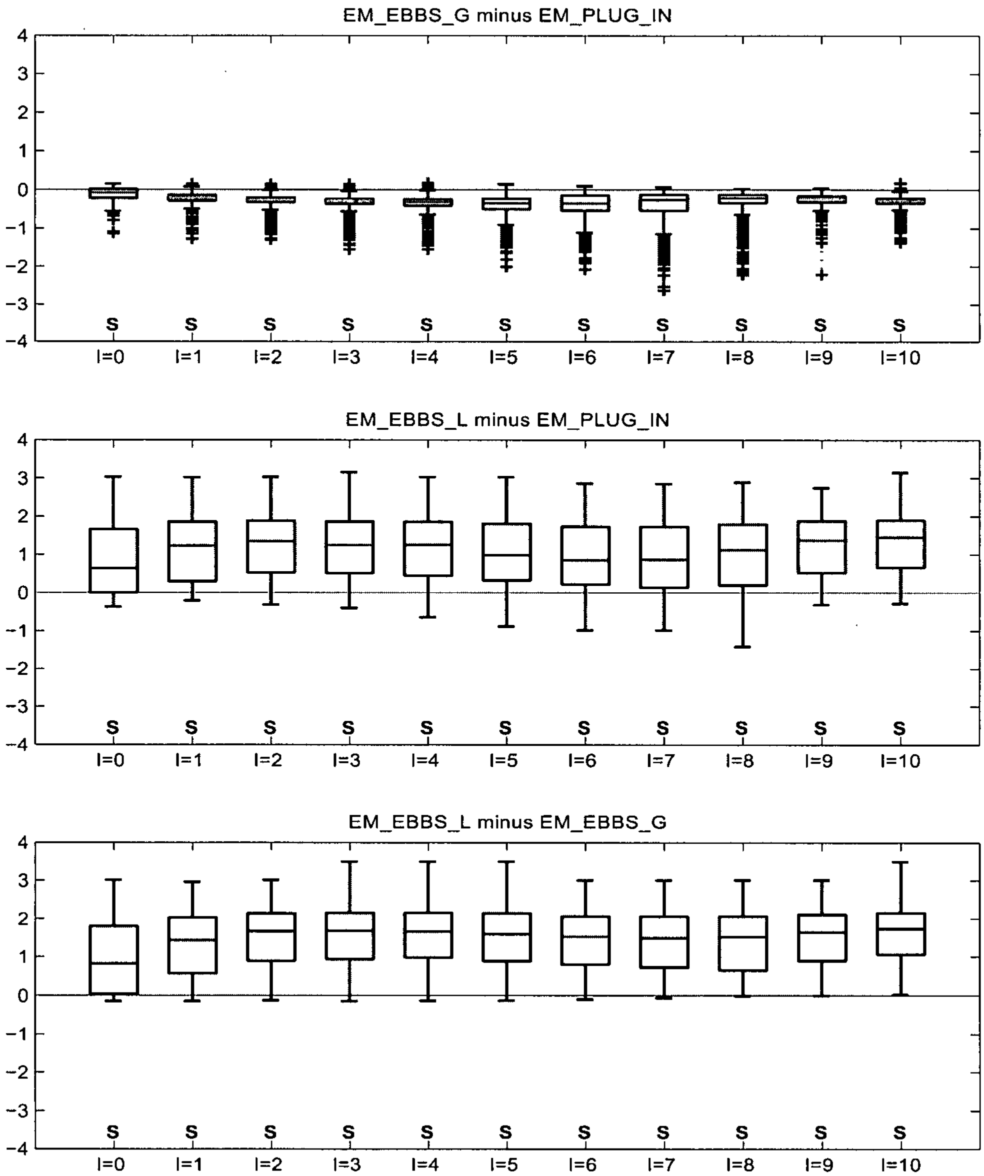

Figure A.18: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_2(z)$.

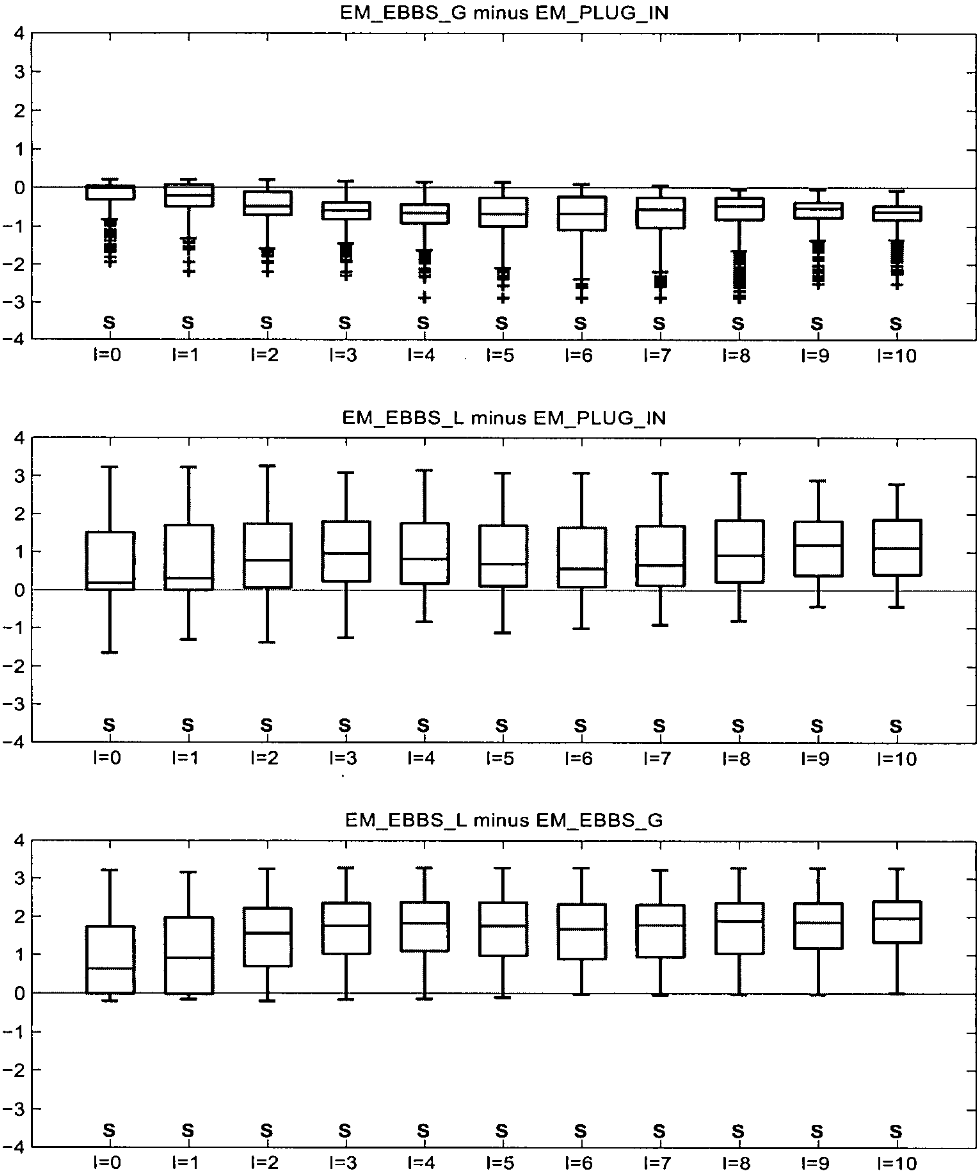

Figure A.19: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.6$ *and* $m(z) = m_2(z)$.

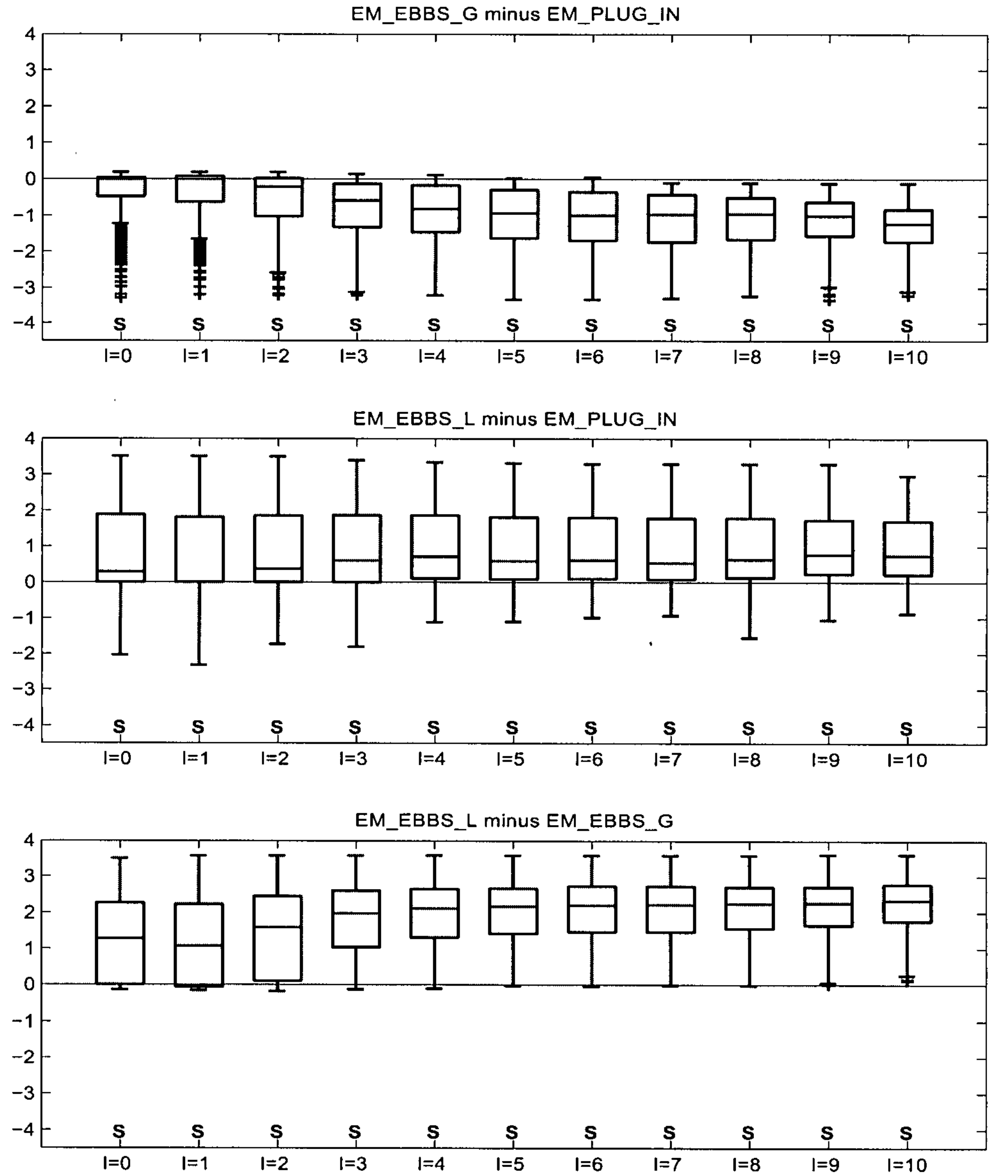

Figure A.20: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{EM,PLUG-IN}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{EM,EBBS-L}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_2(z)$.

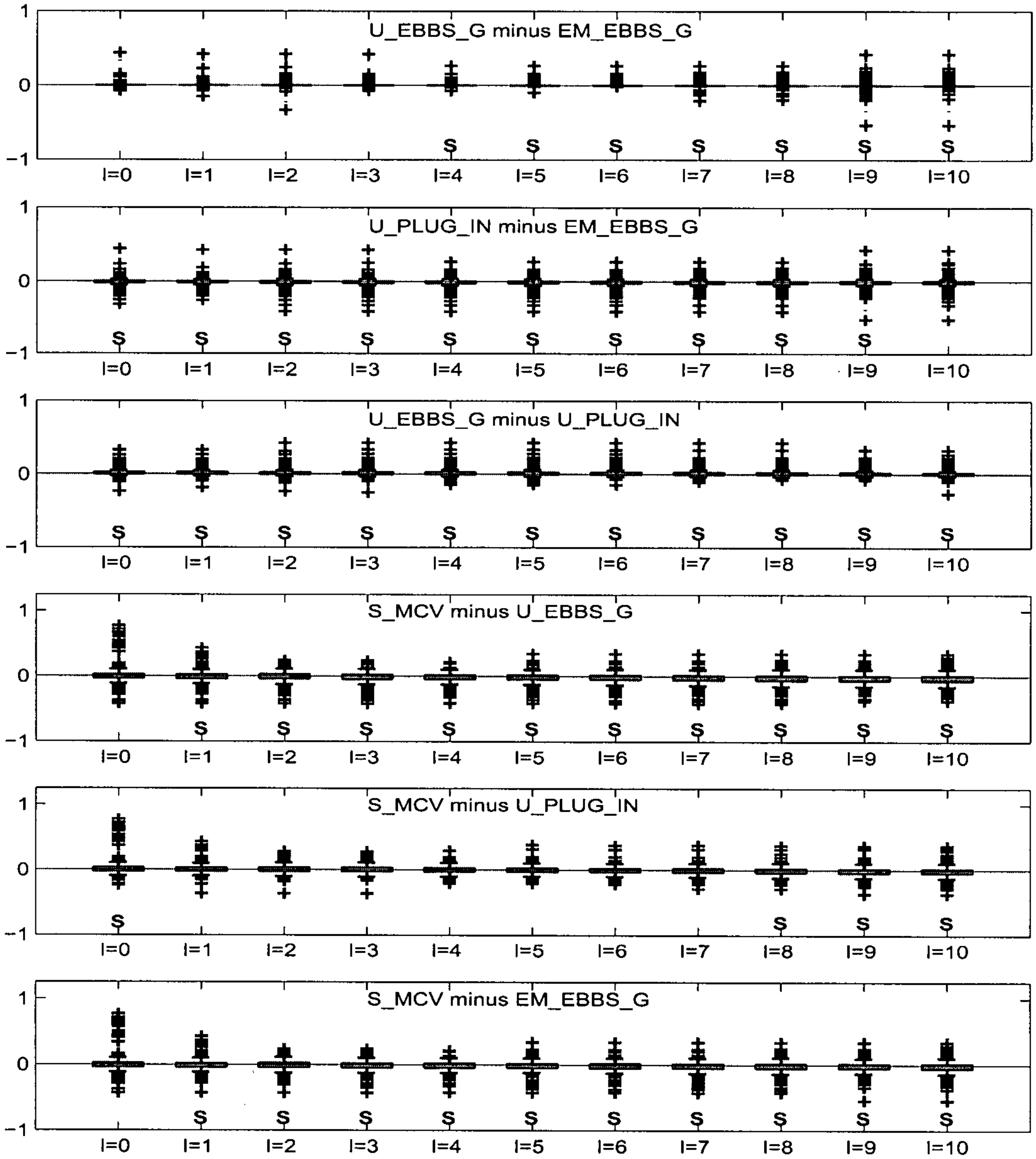

Figure A.21: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_1(z)$.

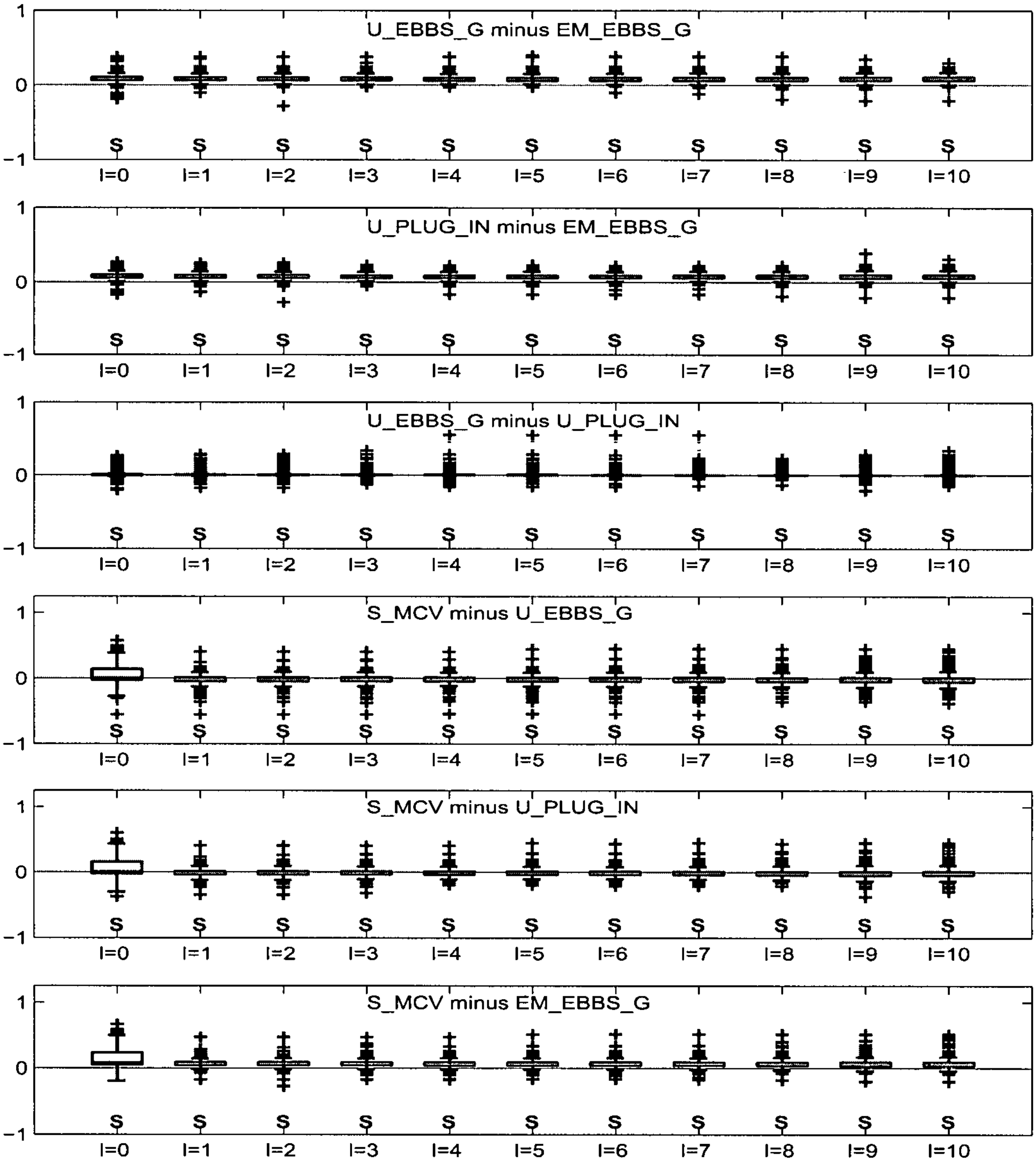

Figure A.22: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_1(z)$.

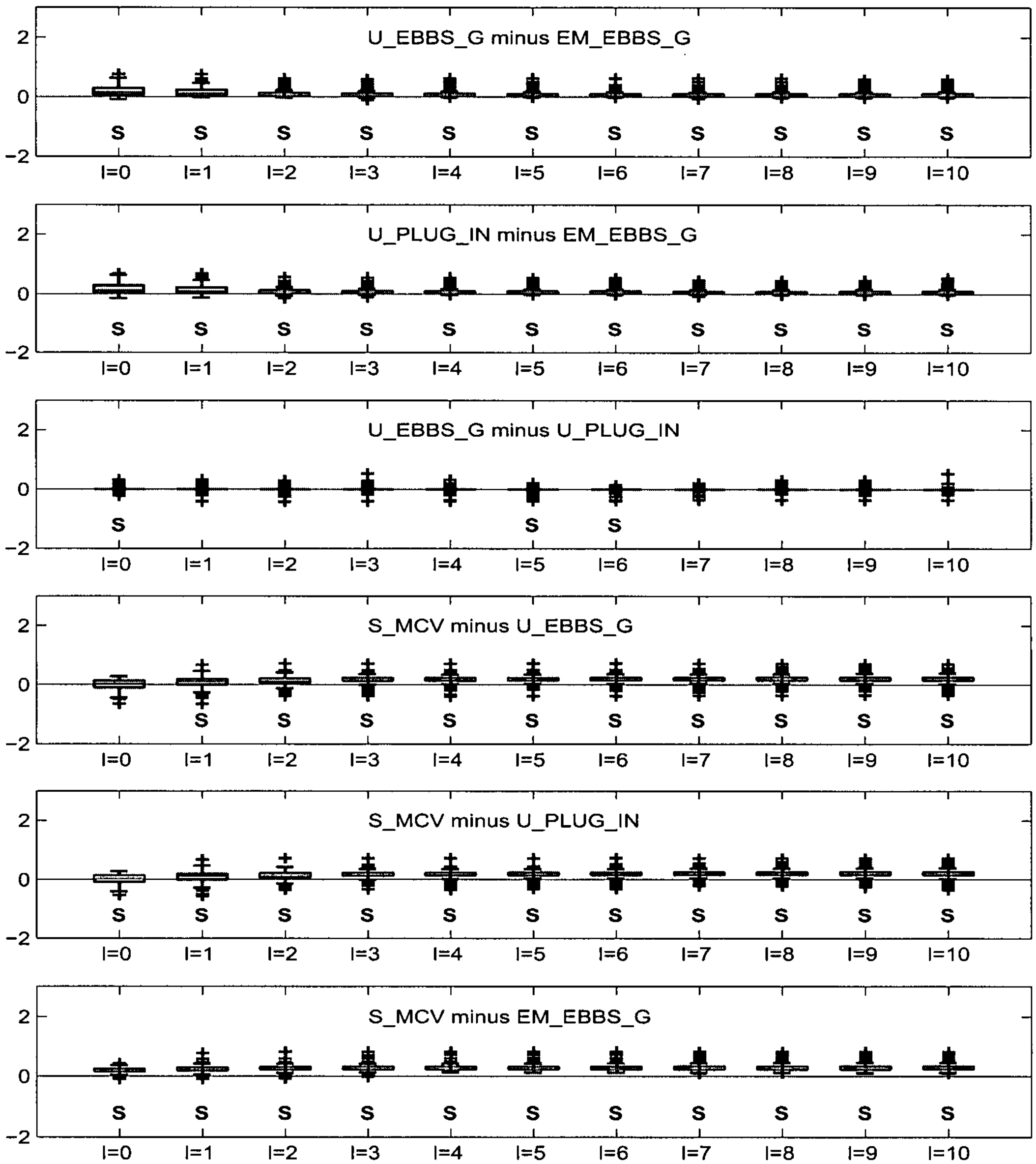

Figure A.23: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_1(z)$.

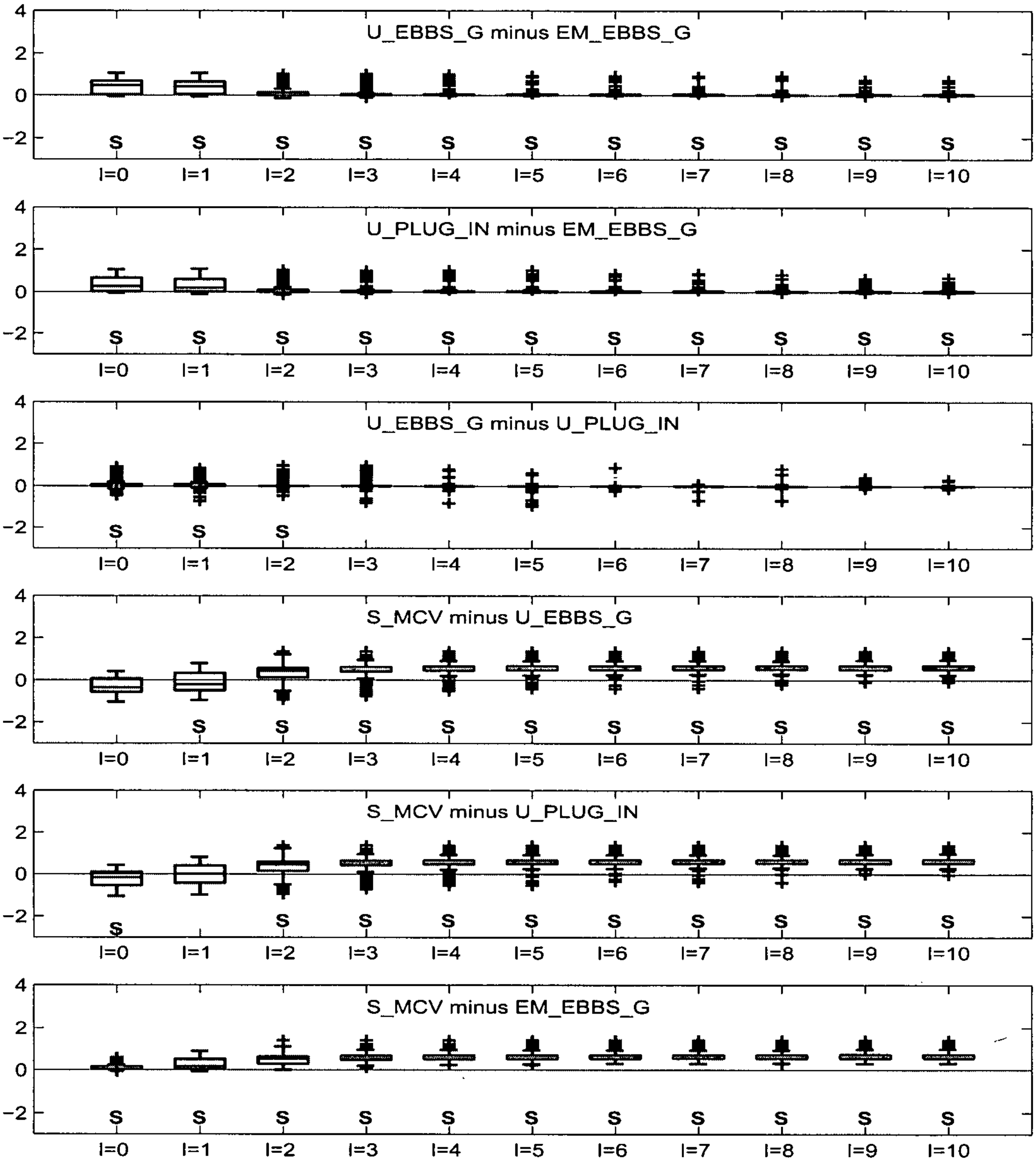

Figure A.24: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.6$ *and* $m(z) = m_1(z)$.

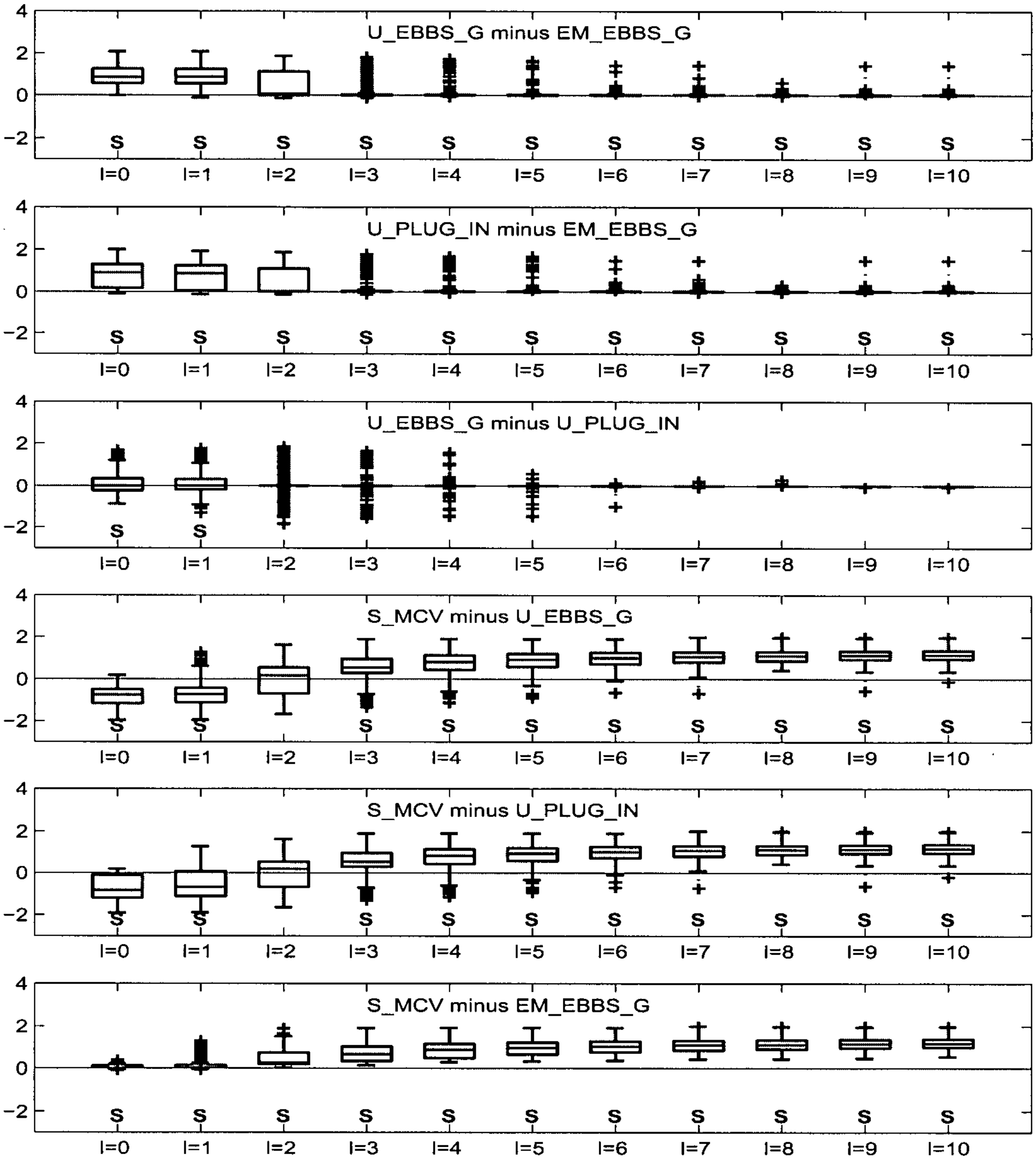

Figure A.25: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_1(z)$.

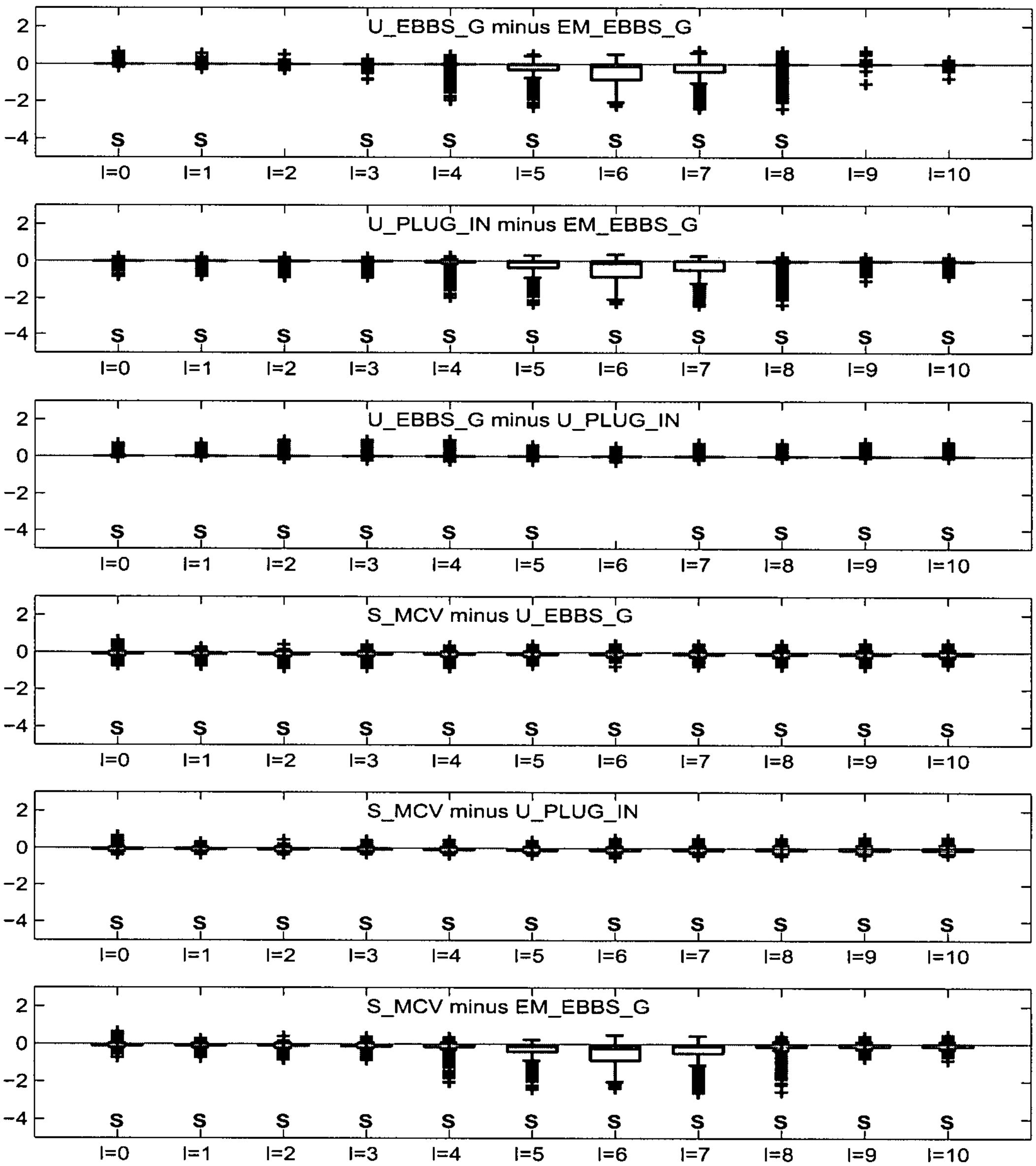

Figure A.26: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0$ *and* $m(z) = m_2(z)$.

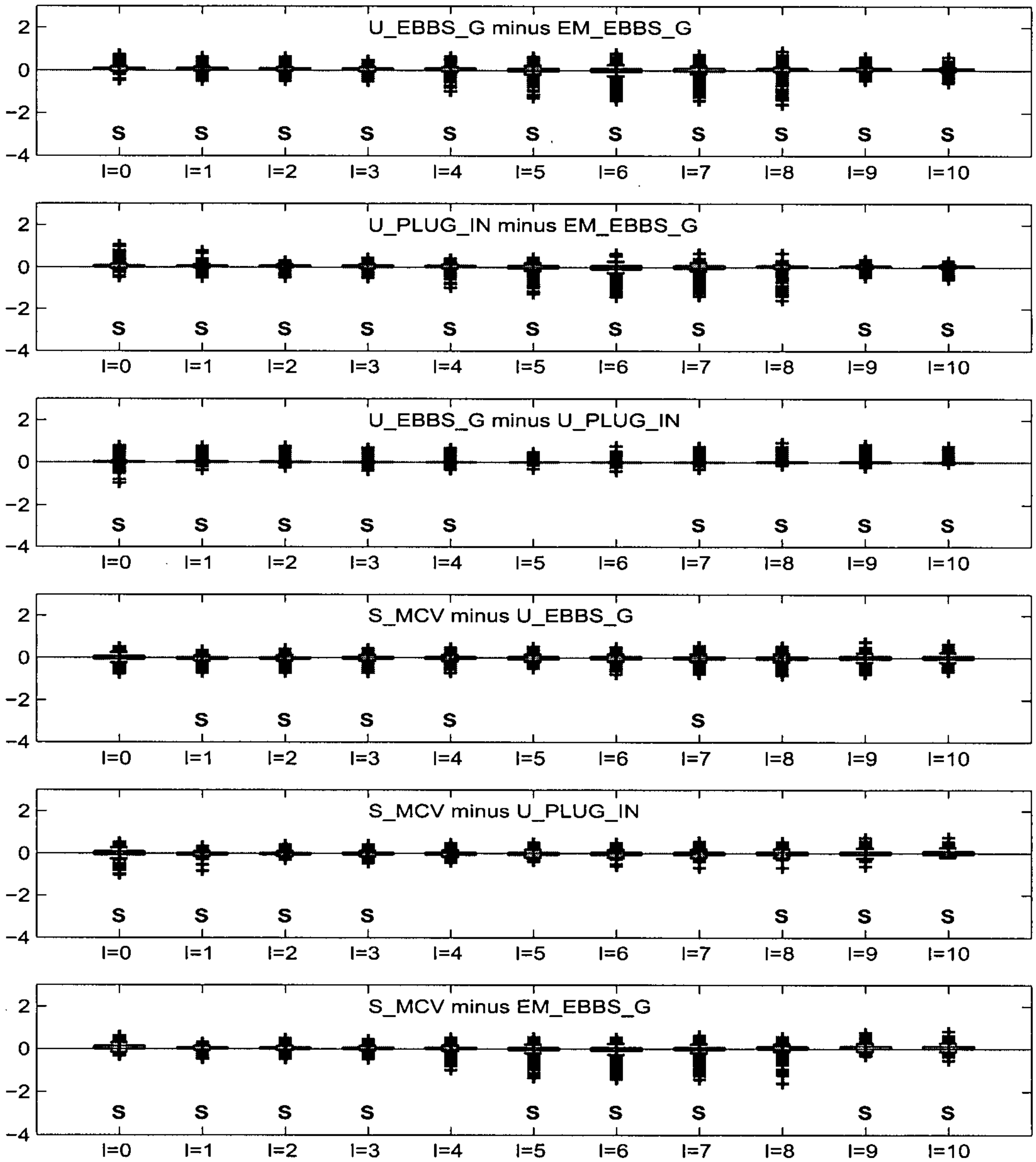

Figure A.27: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.2$ *and* $m(z) = m_2(z)$.

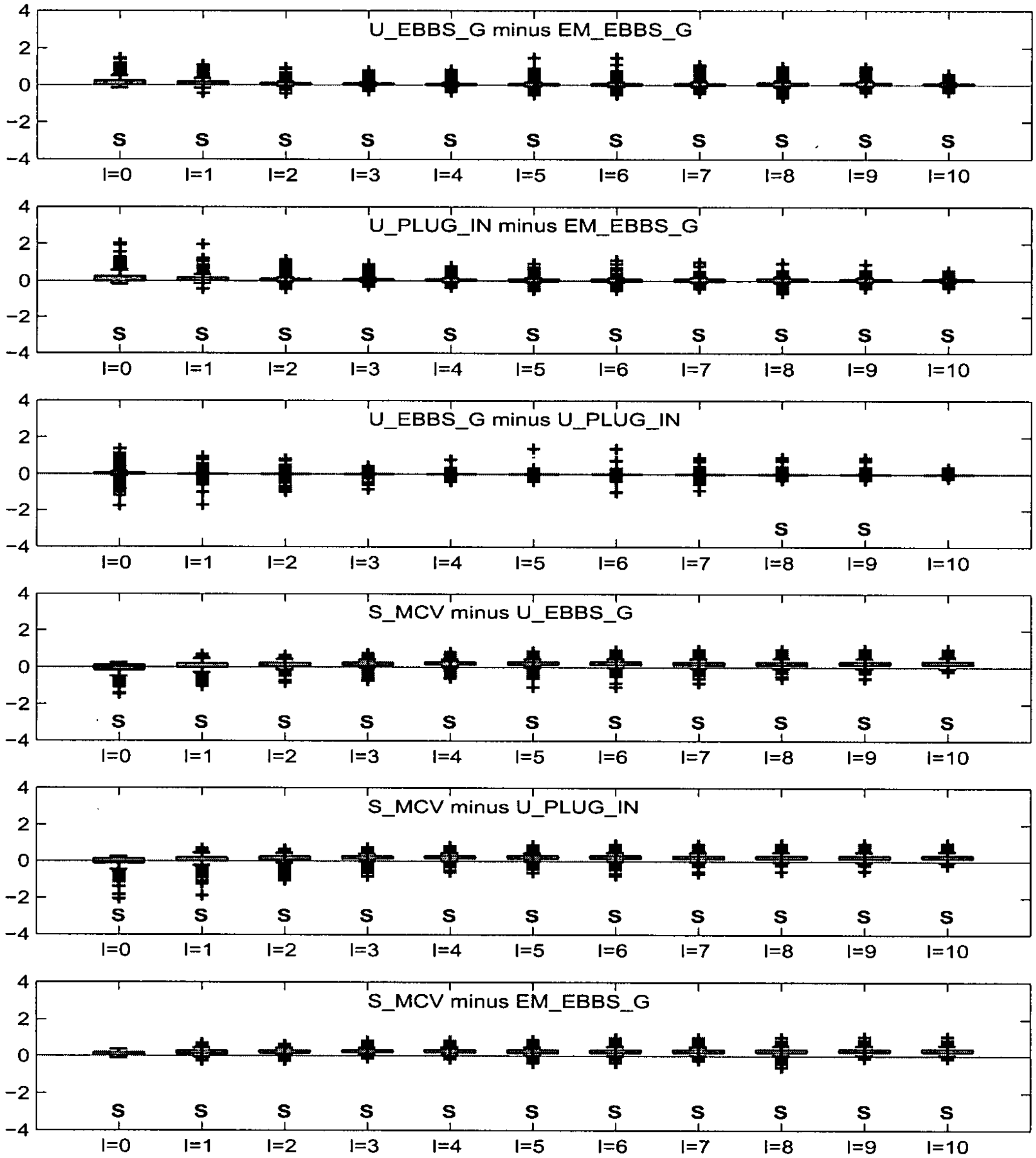

Figure A.28: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.4$ *and* $m(z) = m_2(z)$.

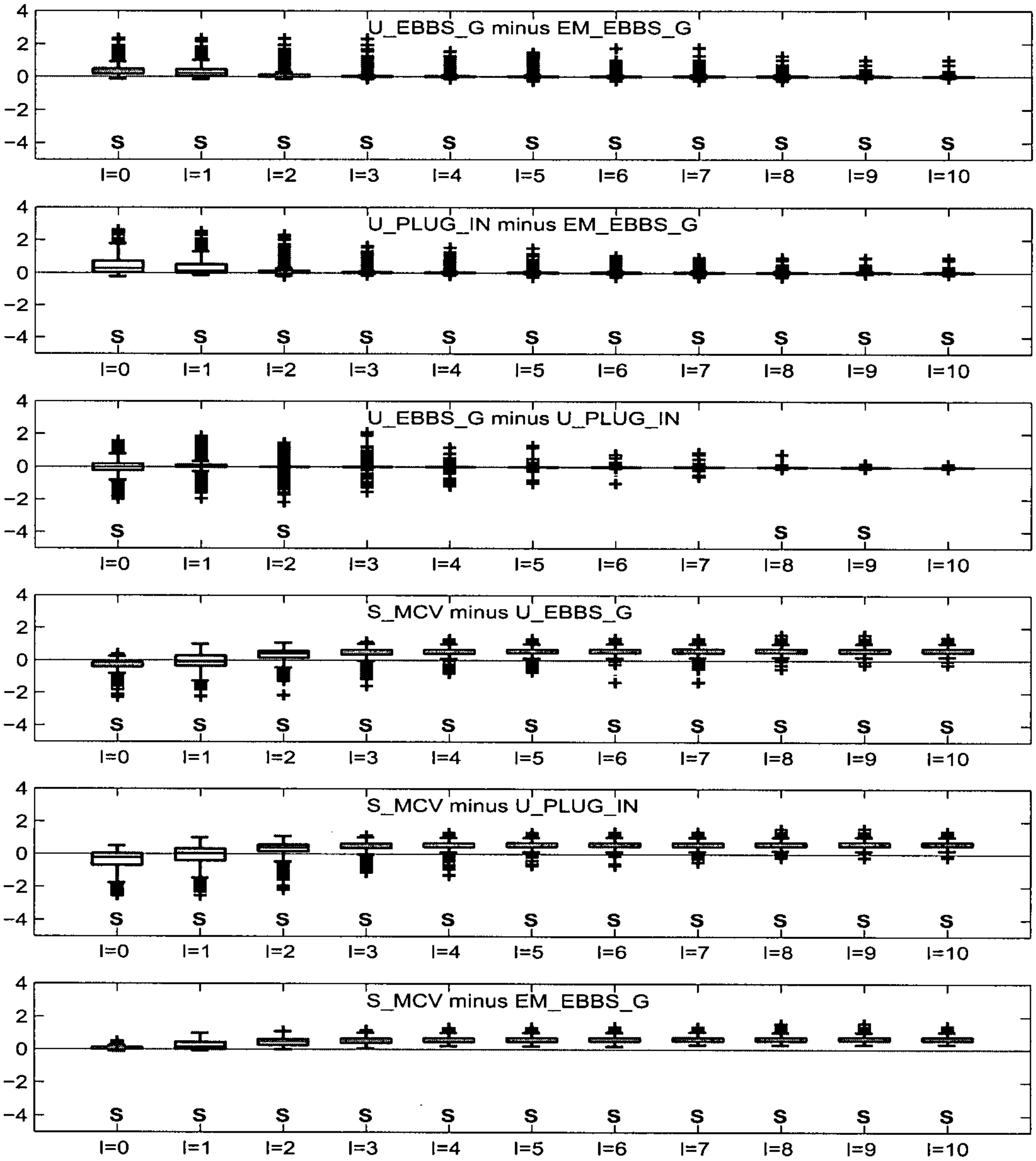

Figure A.29: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *for which* $\rho = 0.6$ *and* $m(z) = m_2(z)$.

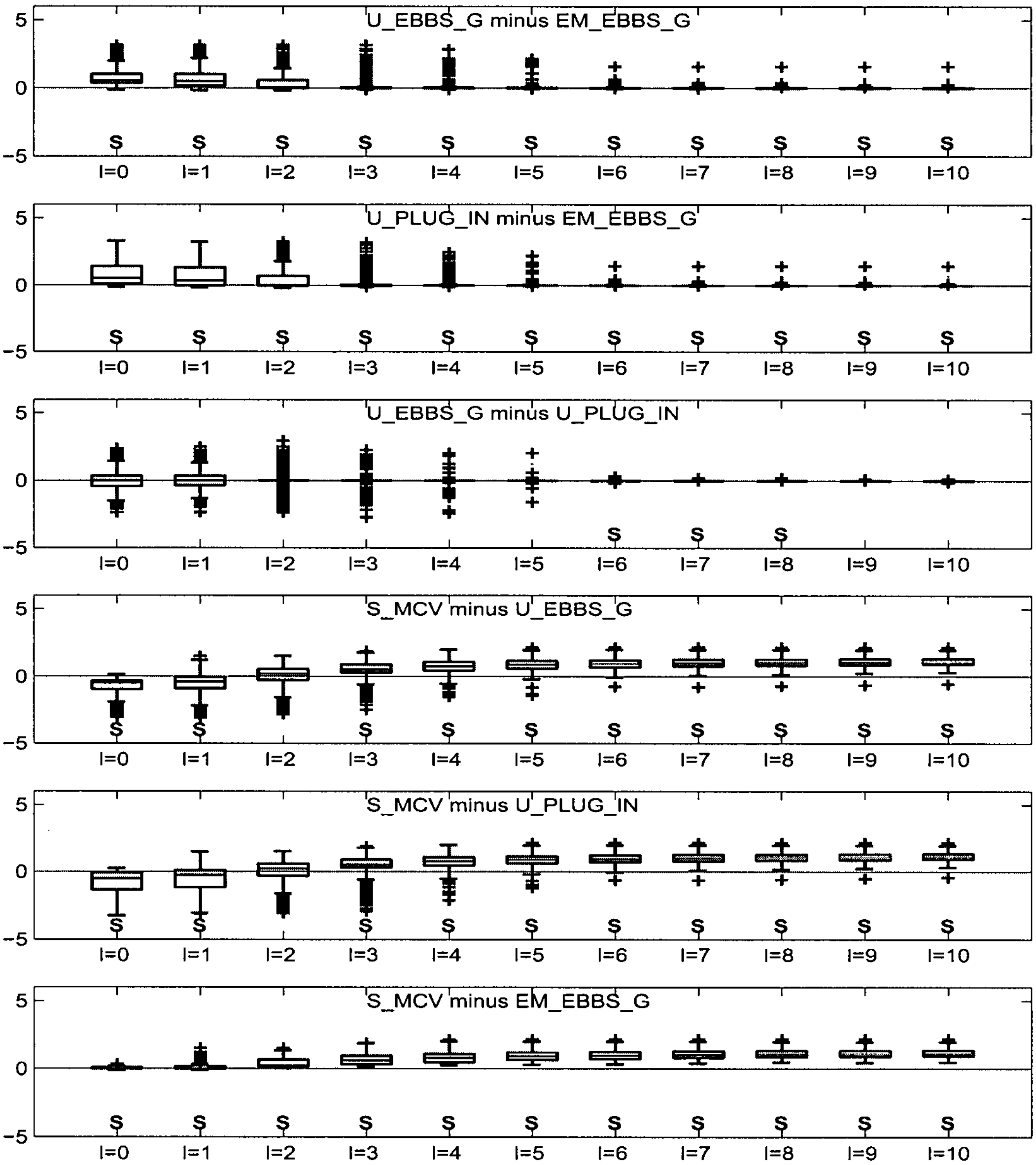

Figure A.30: *Boxplots of pairwise differences in log MSE for the estimators* $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$, $\widehat{\beta}^{(l)}_{EM,EBBS-G}$ *and* $\widehat{\beta}^{(l)}_{S,MCV}$ *of the linear effect* β_1 *in model* (8.1), *where* $l = 0, 1, \ldots, 10$. *Boxplots for which the average difference in log MSE is significantly different than* 0 *at the* 0.05 *level are labeled with an* S. *Differences were obtained by evaluating the log MSE's of the estimators for* 500 *data sets simulated from model* (8.1) *with* $\rho = 0.8$ *and* $m(z) = m_2(z)$.

Appendix B

Validity of Confidence Intervals

In this appendix, we provide plots that help assess and compare the coverage properties of various methods for constructing standard 95% confidence intervals for β_1, the linear effect in model (8.1). For each method, we visualize point estimates and 95% confidence interval estimates for the true coverage achieved by that method.

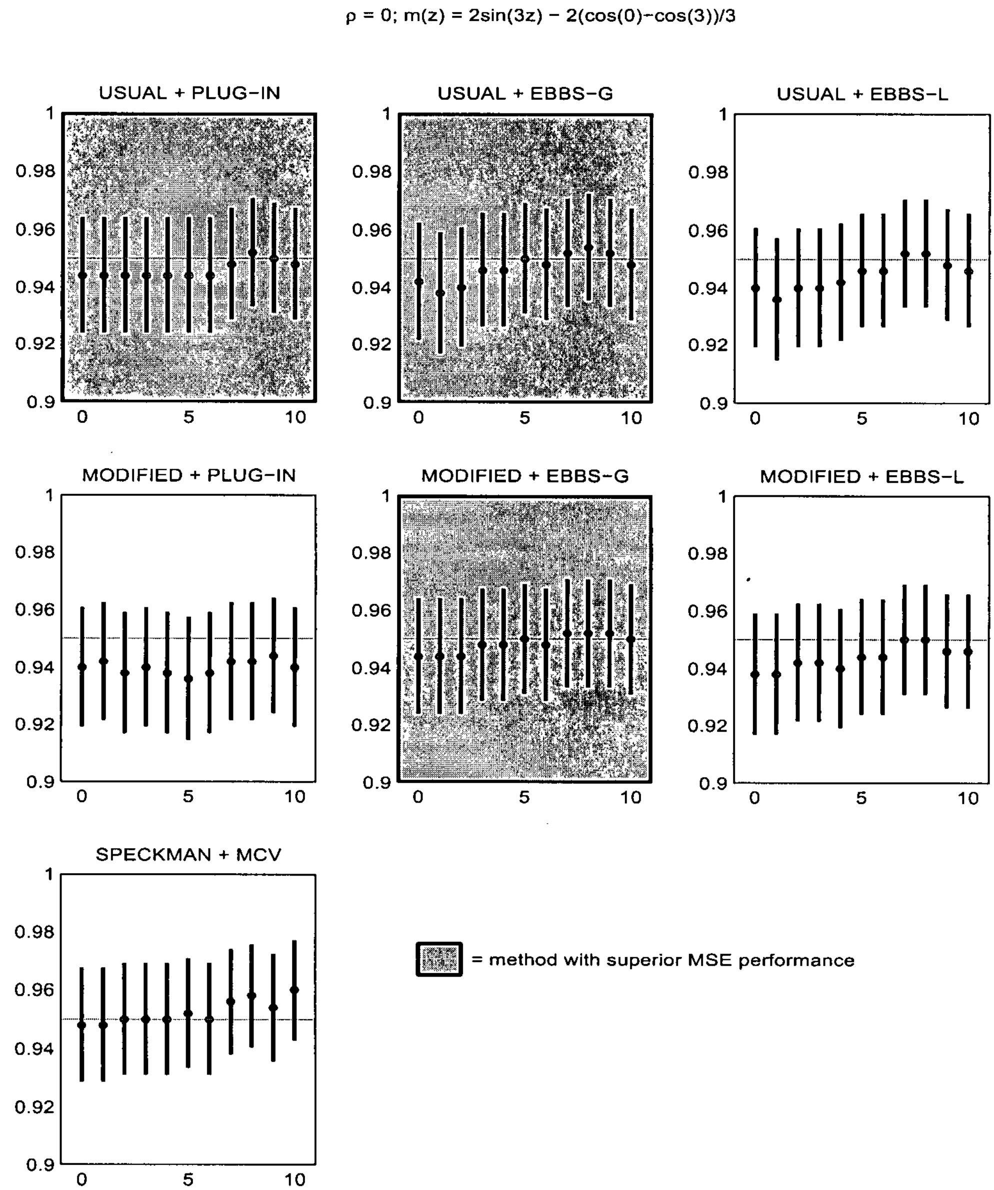

Figure B.1: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0$ *and* $m(z) = m_1(z)$.

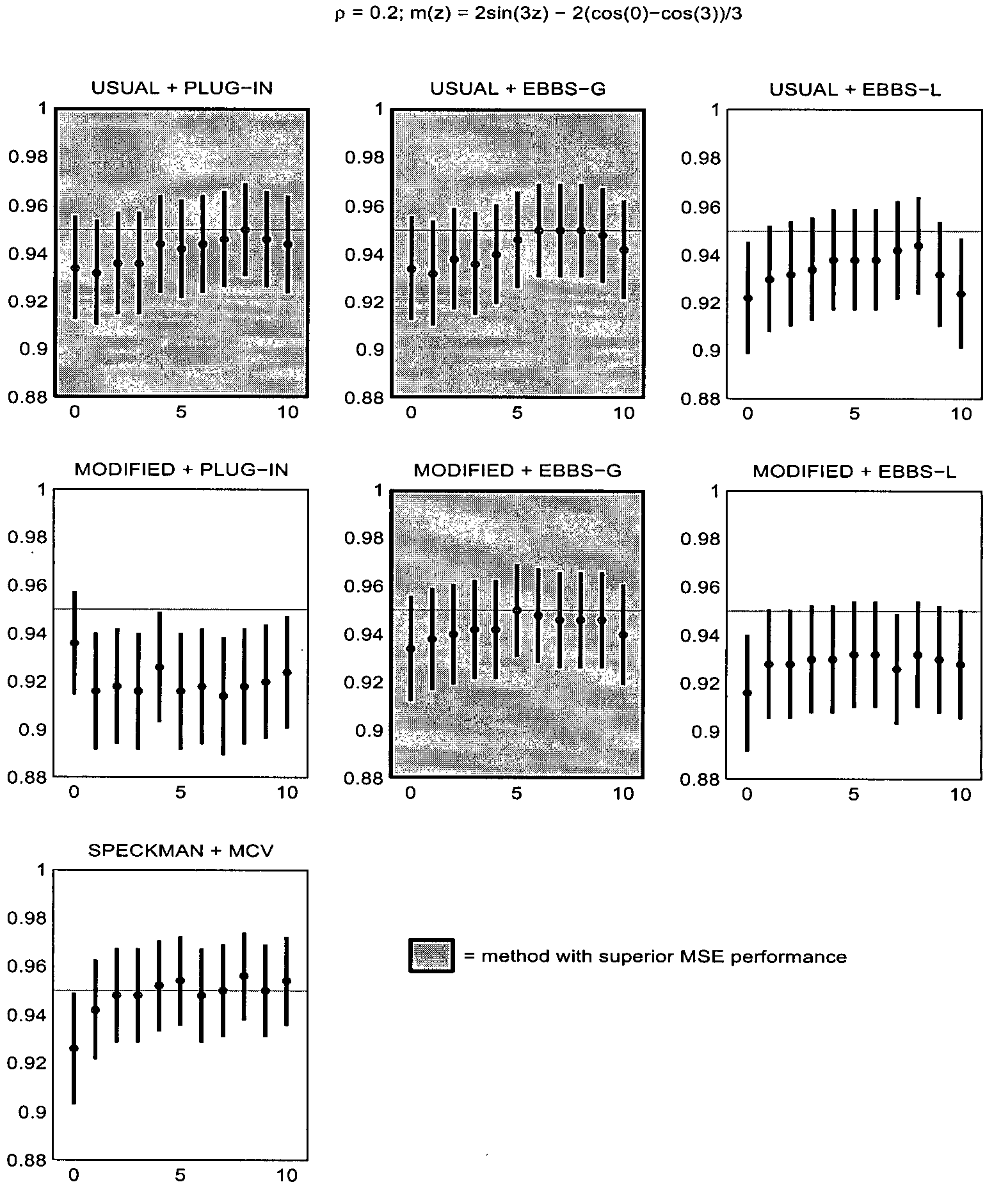

Figure B.2: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.2$ *and* $m(z) = m_2(z)$.

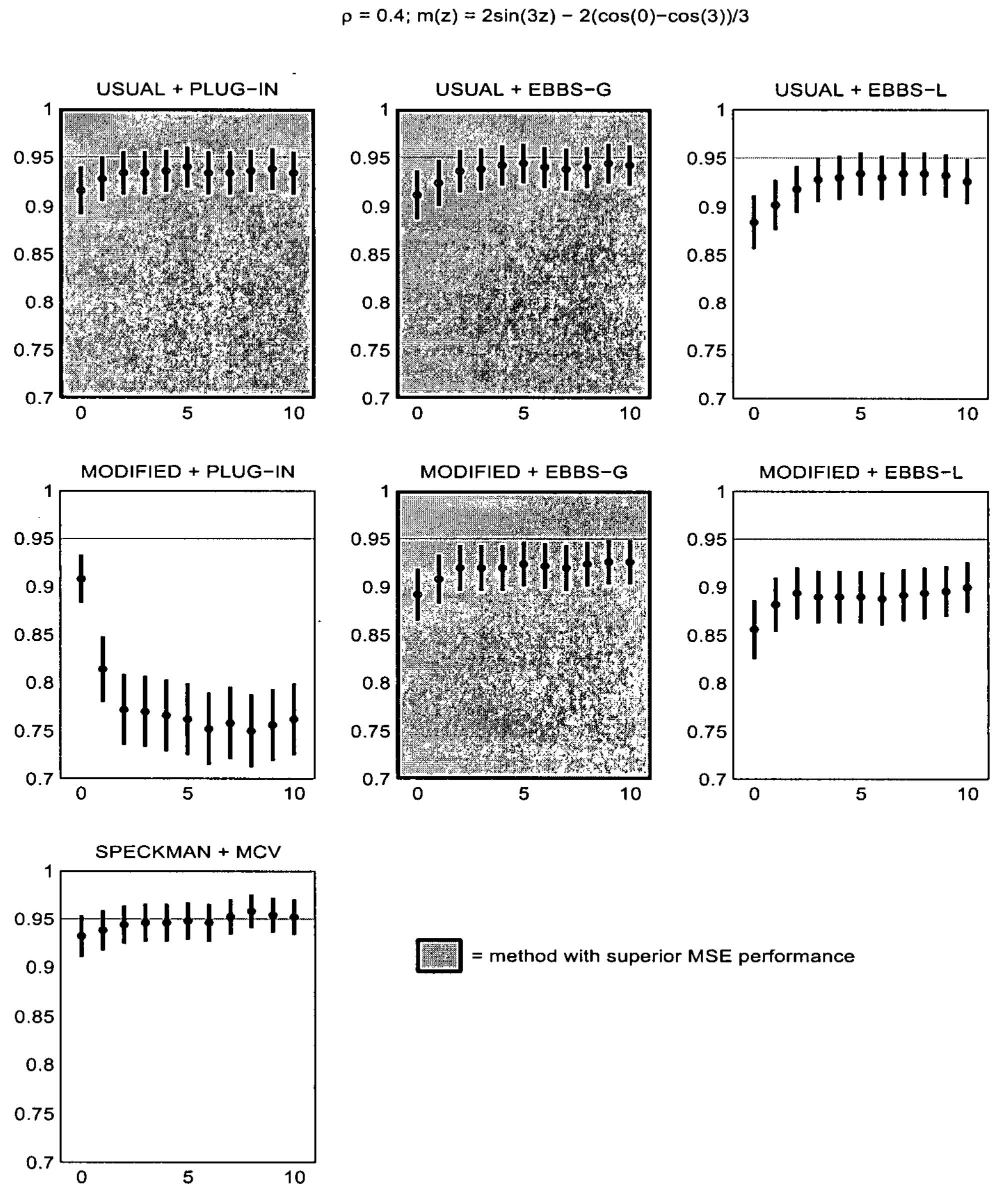

Figure B.3: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.4$ *and* $m(z) = m_1(z)$.

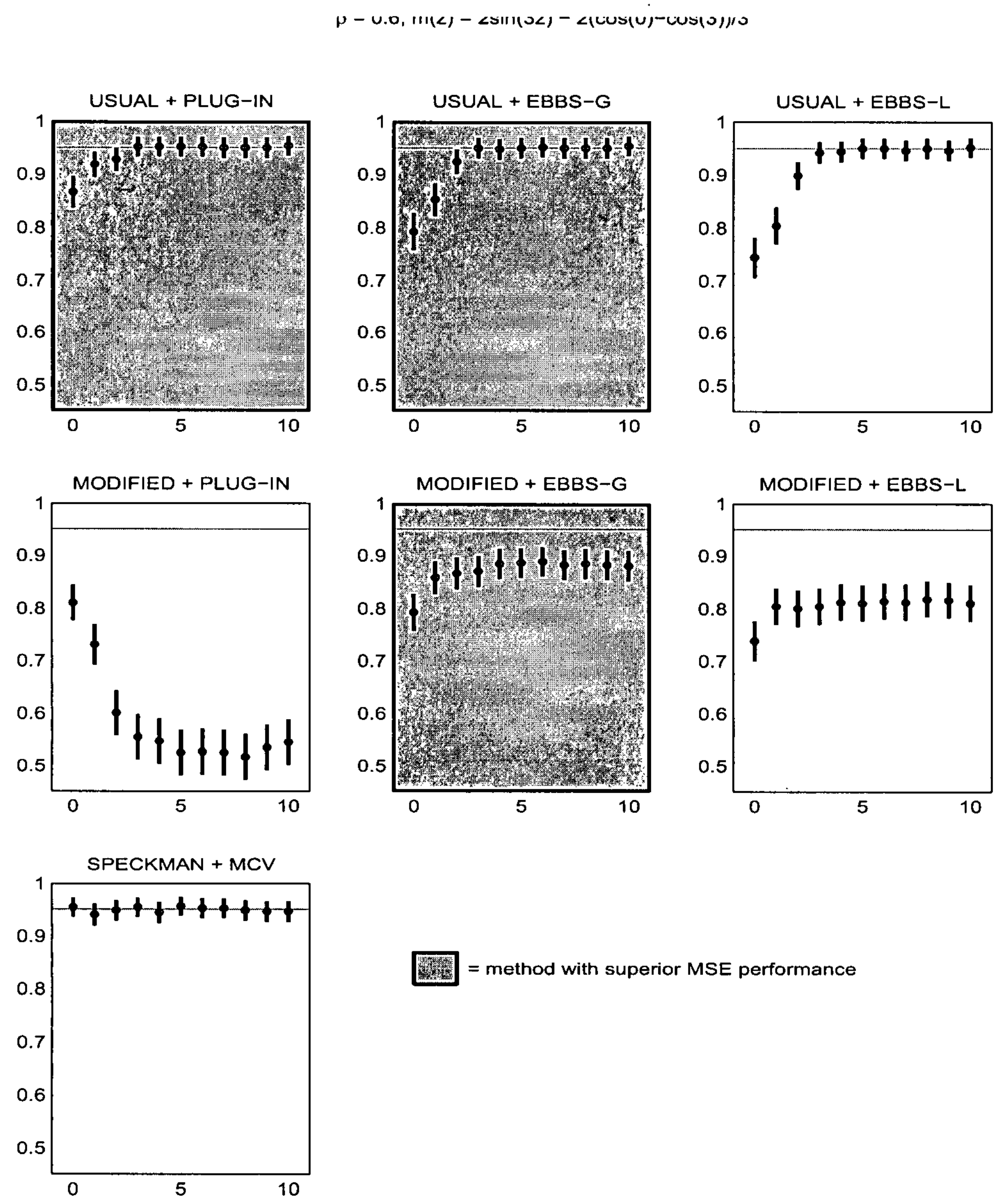

Figure B.4: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.6$ *and* $m(z) = m_1(z)$.

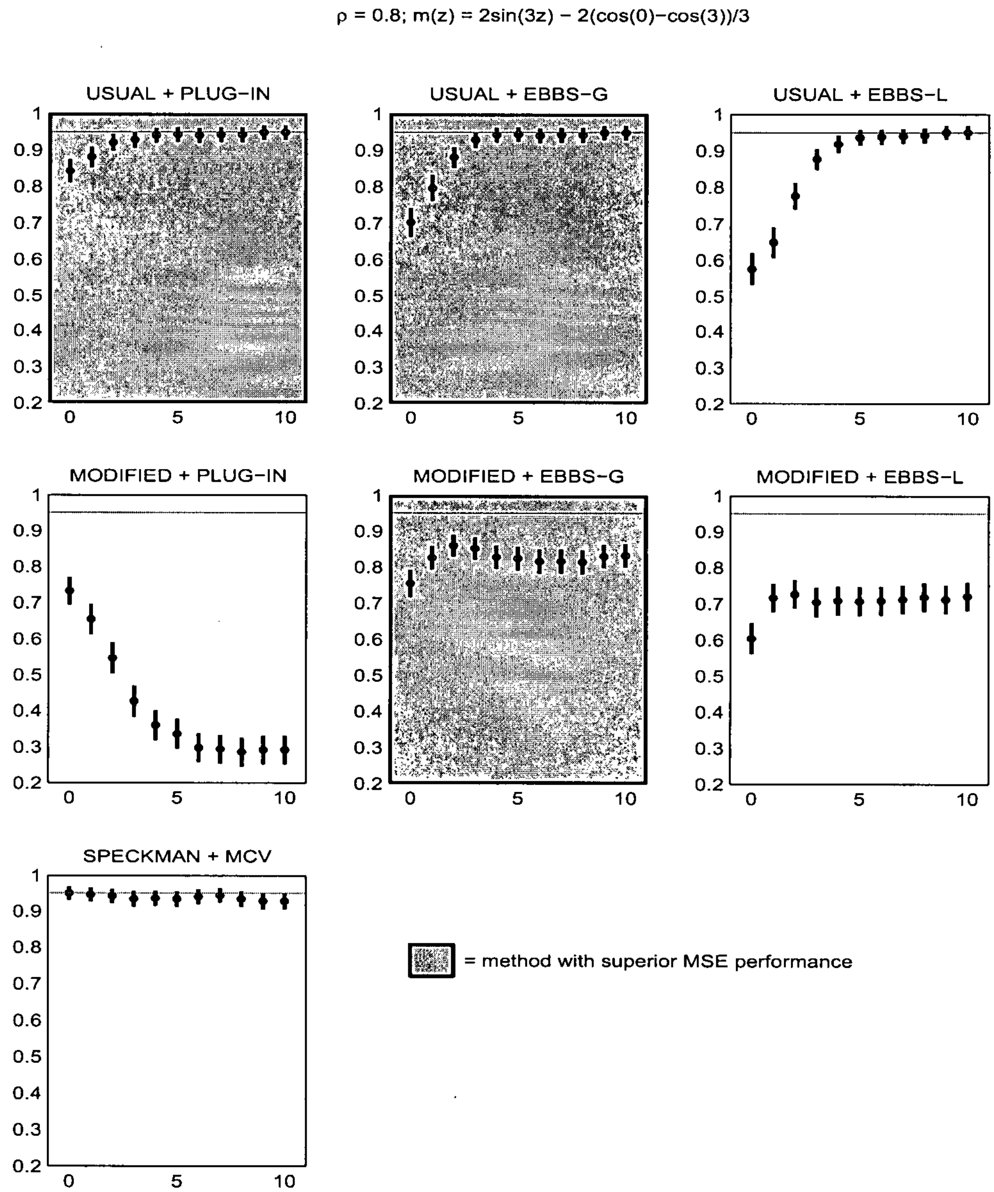

Figure B.5: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.8$ *and* $m(z) = m_1(z)$.

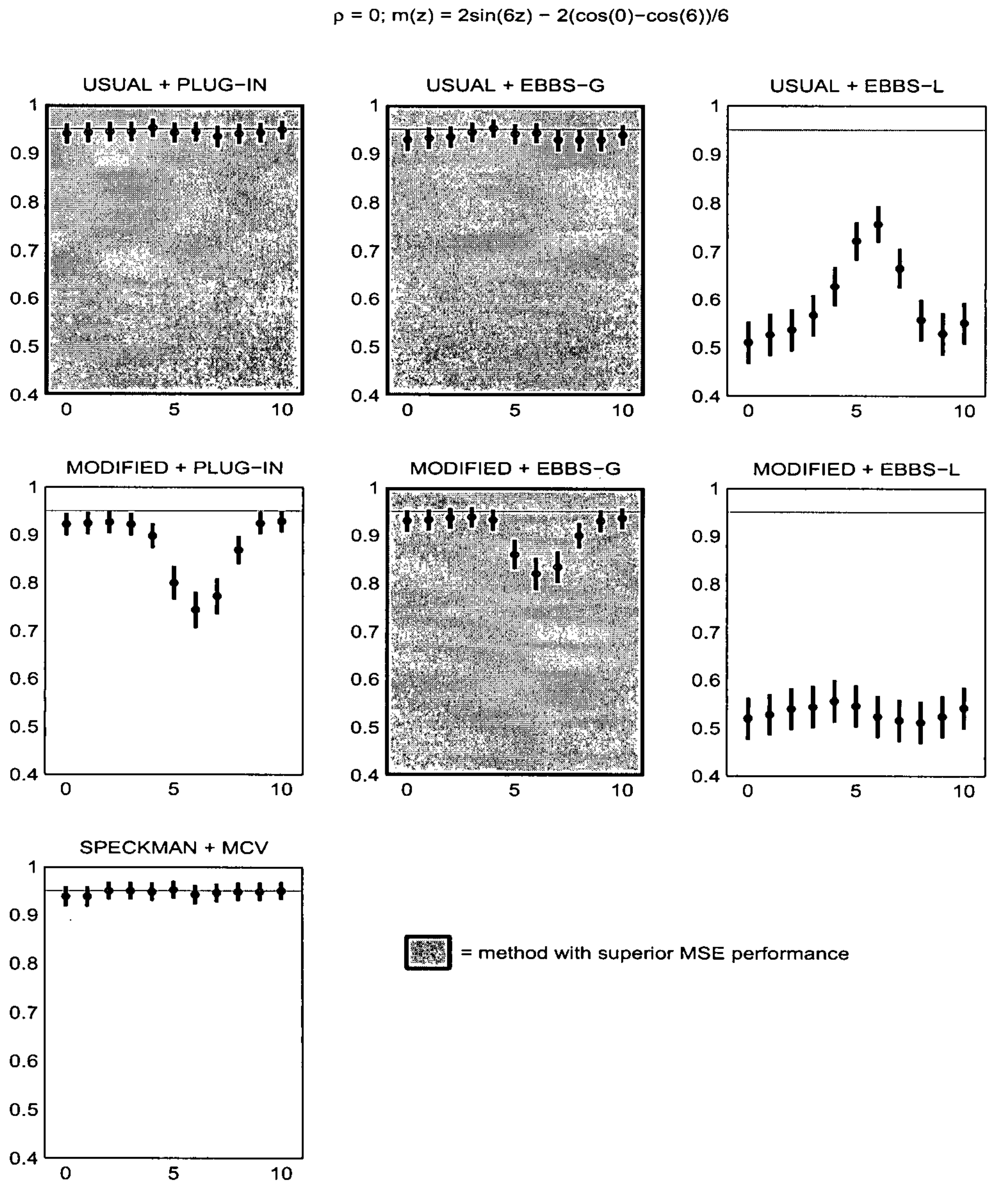

Figure B.6: *Point estimates (circles) and 95% confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing 95% confidence intervals for the linear effect β_1 in model (8.1). Each method depends on a tuning parameter $l = 0, 1, \ldots, 10$. The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with $\rho = 0$ and $m(z) = m_2(z)$.*

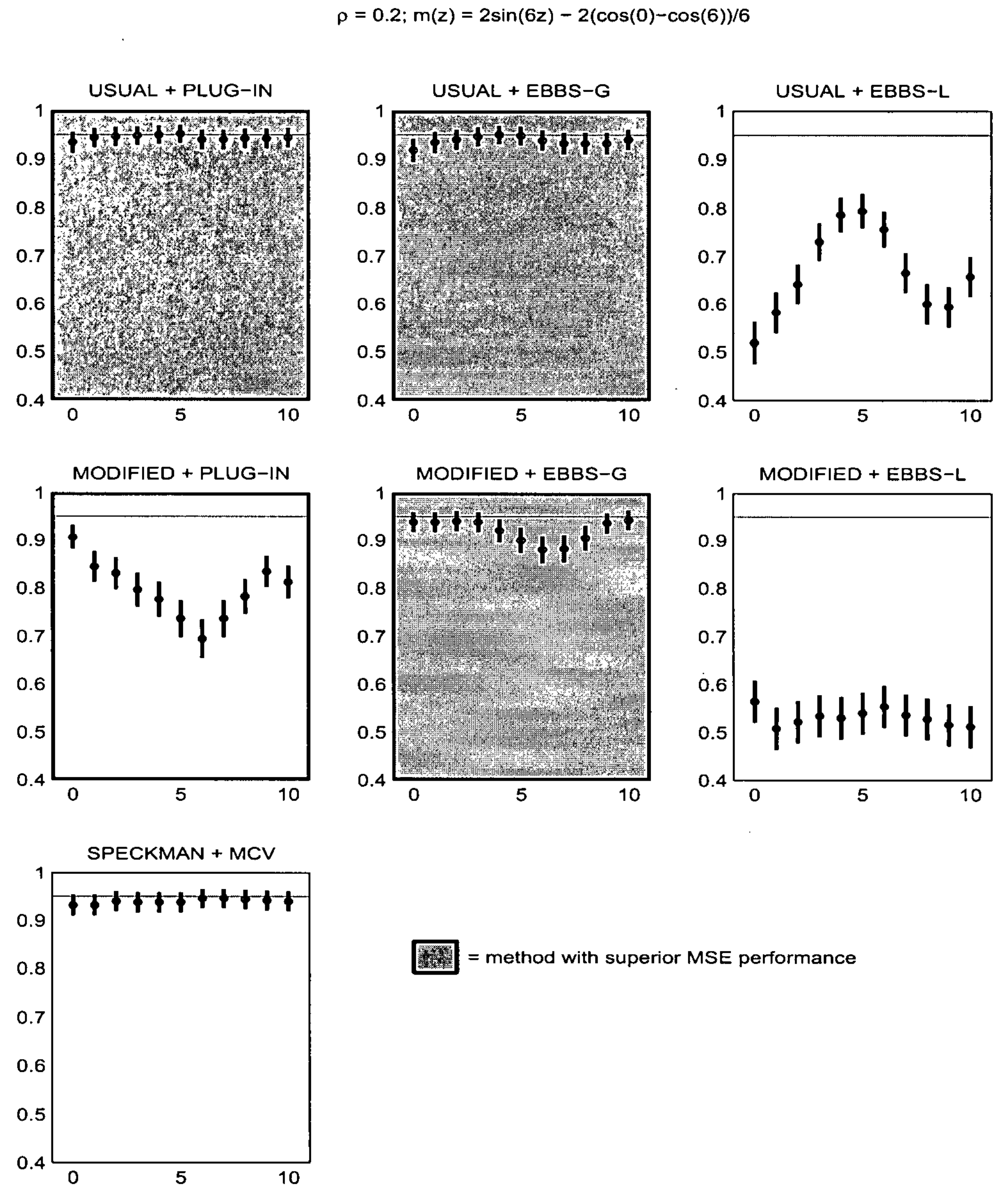

Figure B.7: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.2$ *and* $m(z) = m_2(z)$.

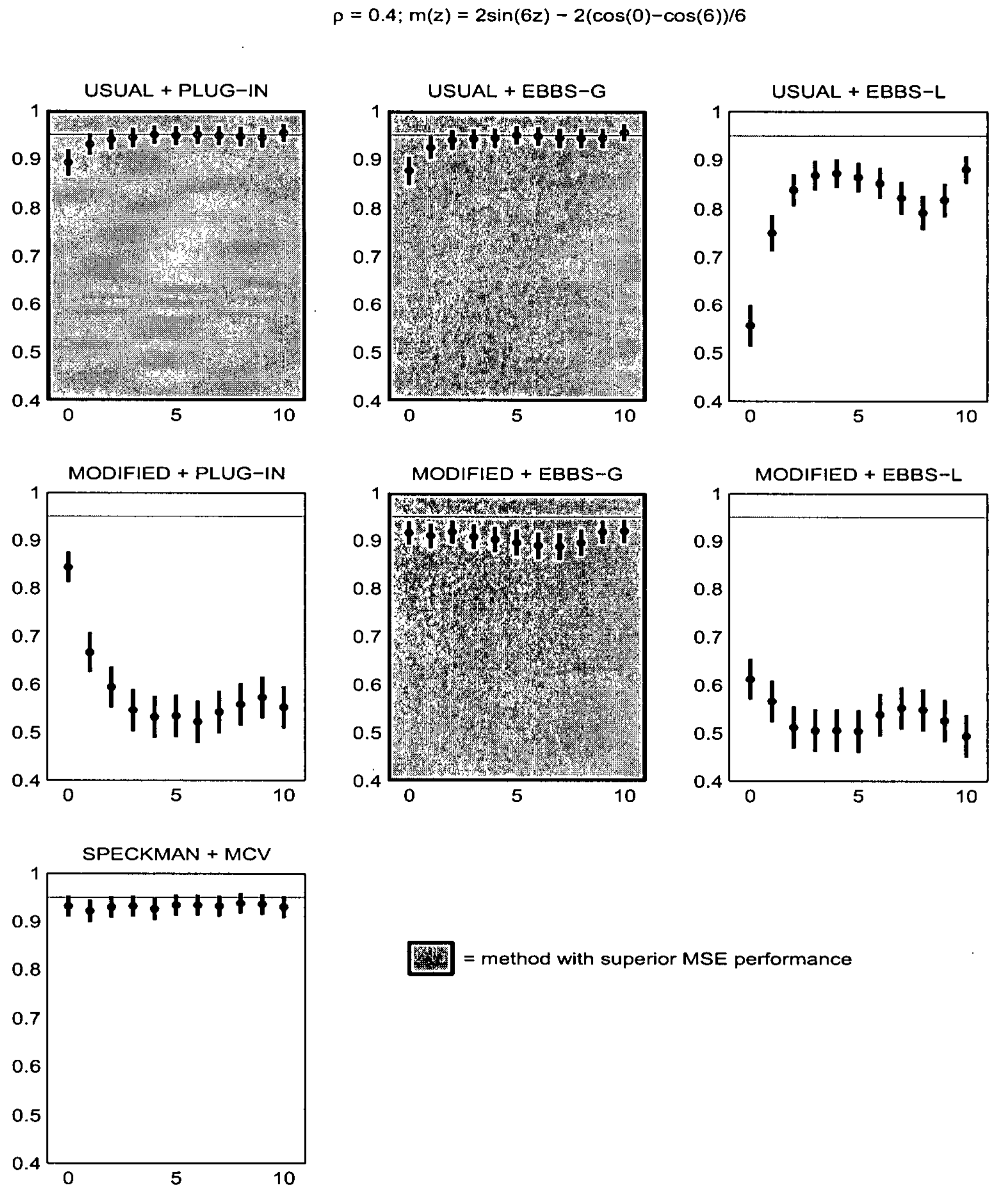

Figure B.8: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.4$ *and* $m(z) = m_2(z)$.

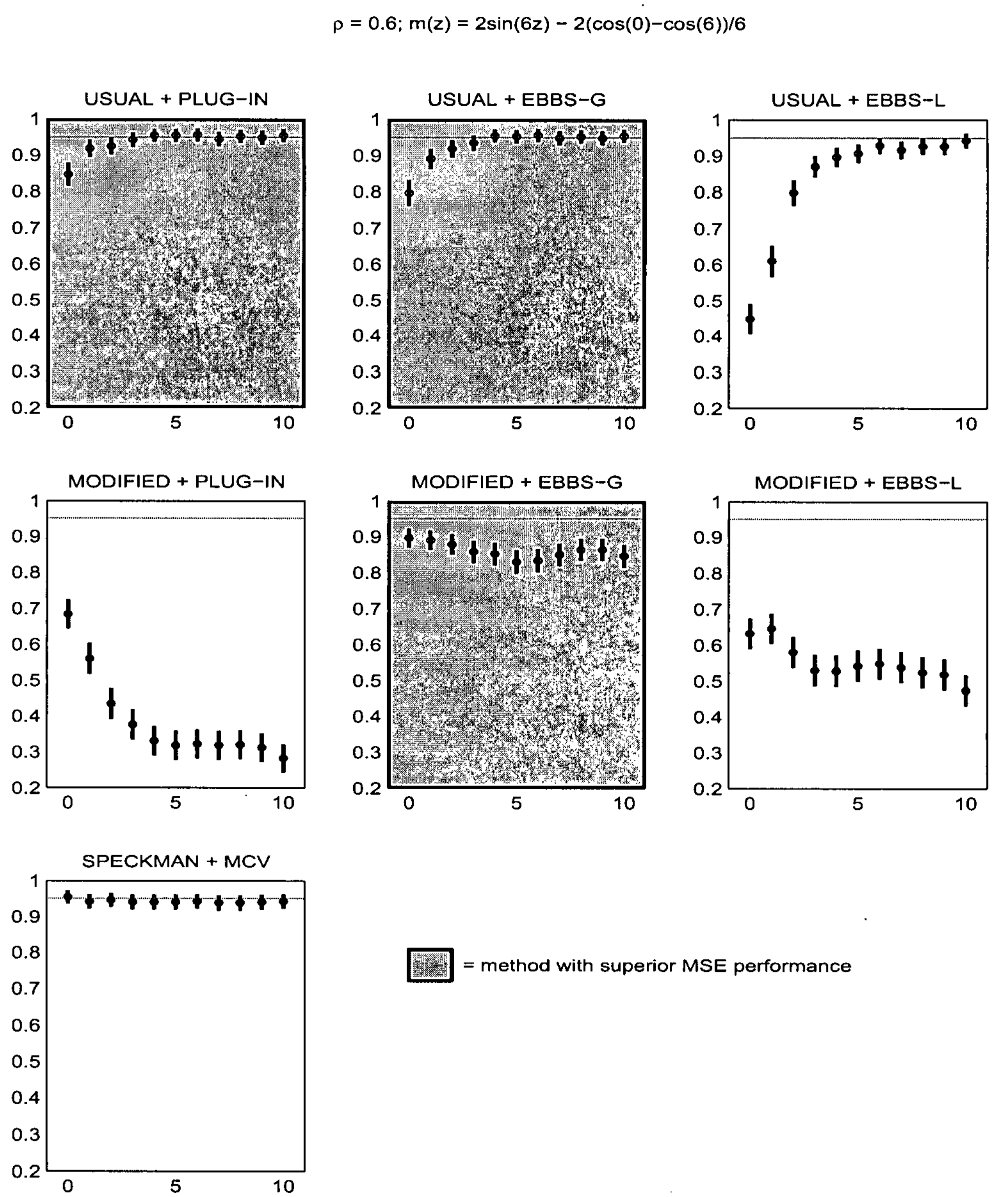

Figure B.9: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1). *Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$. *The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.6$ *and* $m(z) = m_2(z)$.

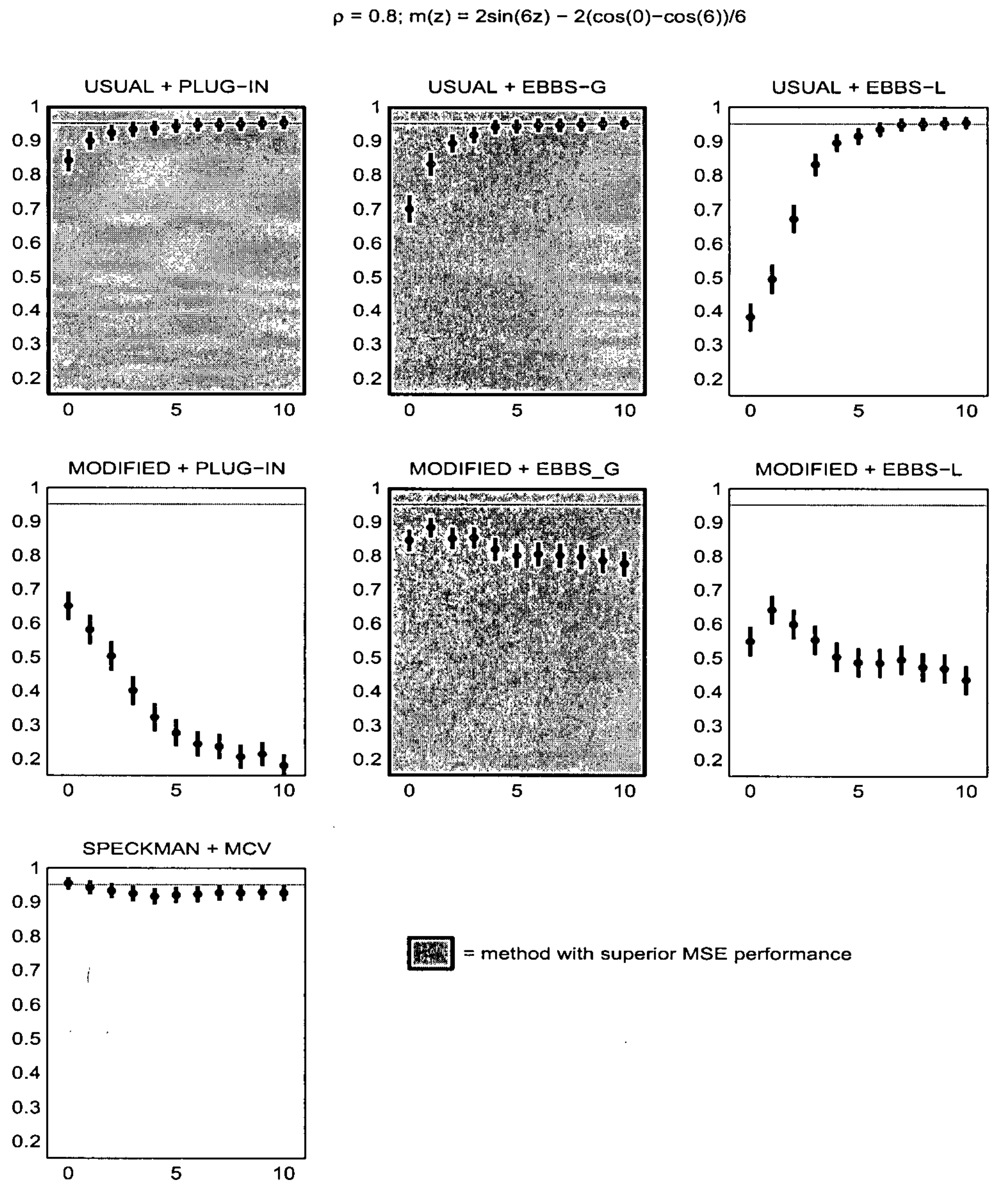

Figure B.10: *Point estimates (circles) and* 95% *confidence interval estimates (segments) for the true coverage achieved by seven different methods for constructing* 95% *confidence intervals for the linear effect* β_1 *in model* (8.1)*. Each method depends on a tuning parameter* $l = 0, 1, \ldots, 10$*. The nominal coverage of each method is indicated via a horizontal line. Estimates were obtained with* $\rho = 0.8$ *and* $m(z) = m_2(z)$.

Appendix C

Confidence Interval Length Comparisons

In this appendix, we provide plots that help assess and compare the length properties of three methods for constructing standard 95% confidence intervals for $\boldsymbol{\beta}_1$, the linear effect in model (8.1). These methods rely on the estimators $\widehat{\beta}^{(l)}_{U,PLUG-IN}$, $\widehat{\beta}^{(l)}_{U,EBBS-G}$ and $\widehat{\beta}^{(l)}_{S,MCV}$, and their associated standard errors. We remind the reader that the finite sample properties of these estimators were investigated via simulation in Chapter 8.

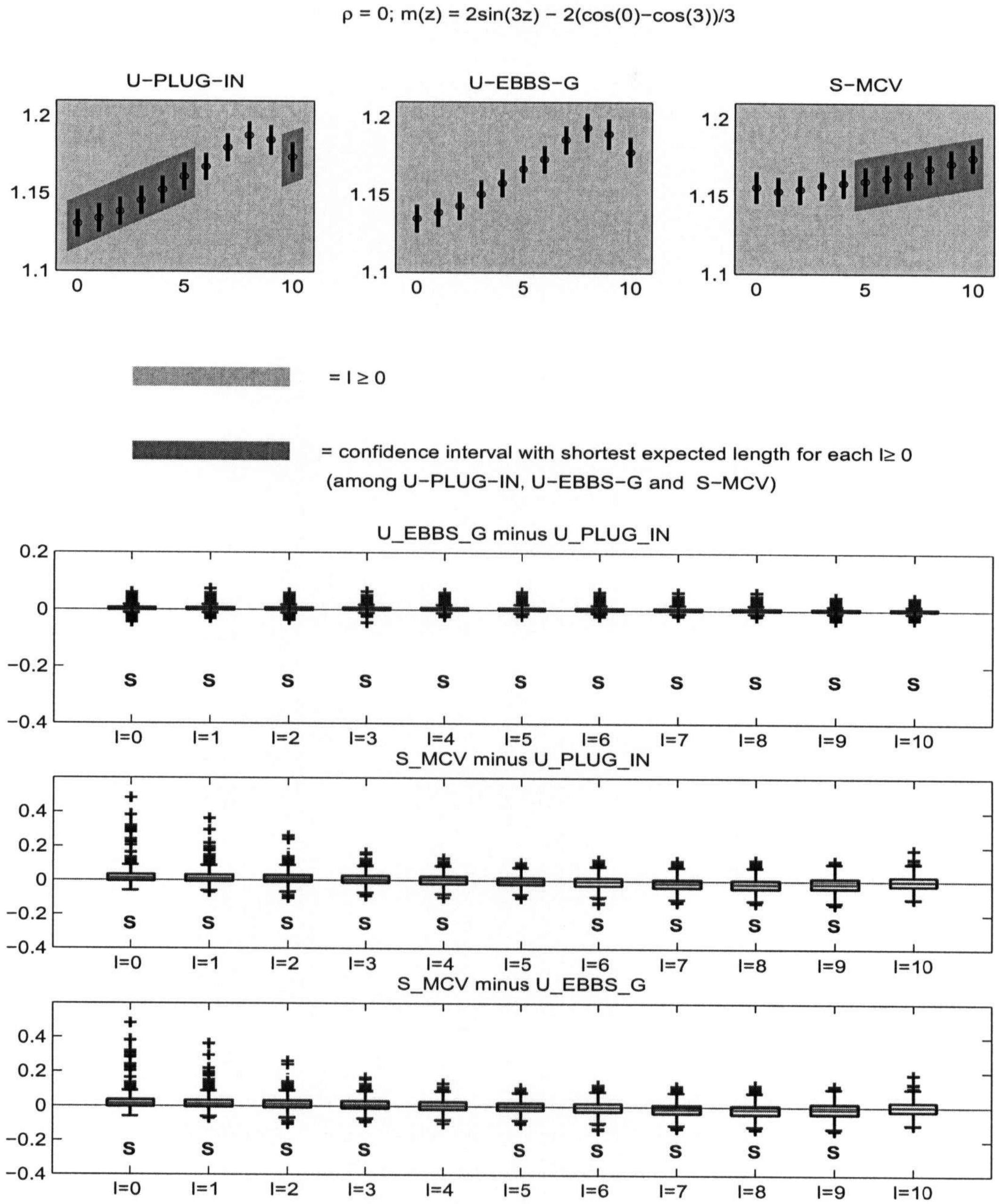

Figure C.1: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0$ and $m(z) = m_1(z)$.*

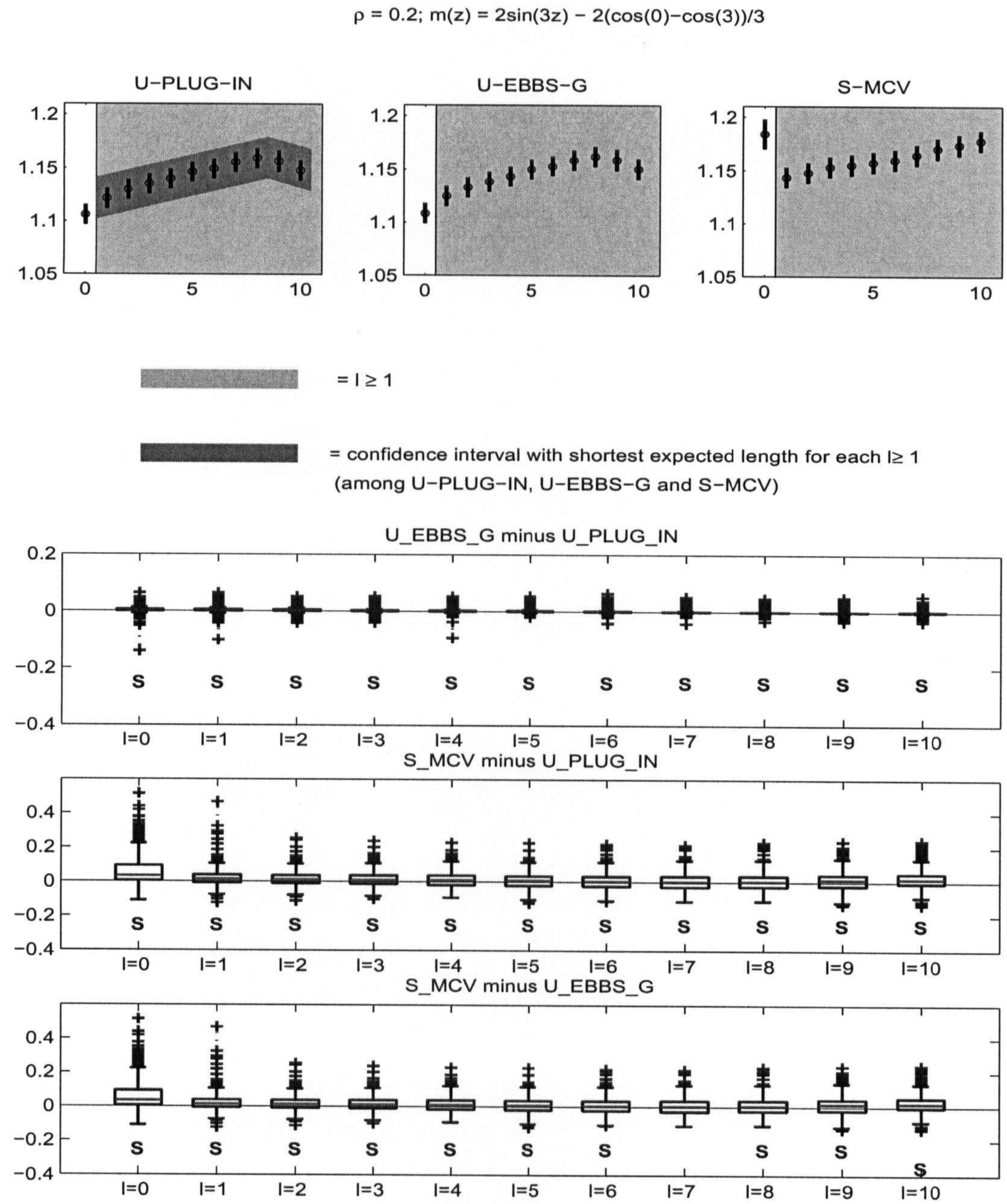

Figure C.2: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.2$ and $m(z) = m_1(z)$.*

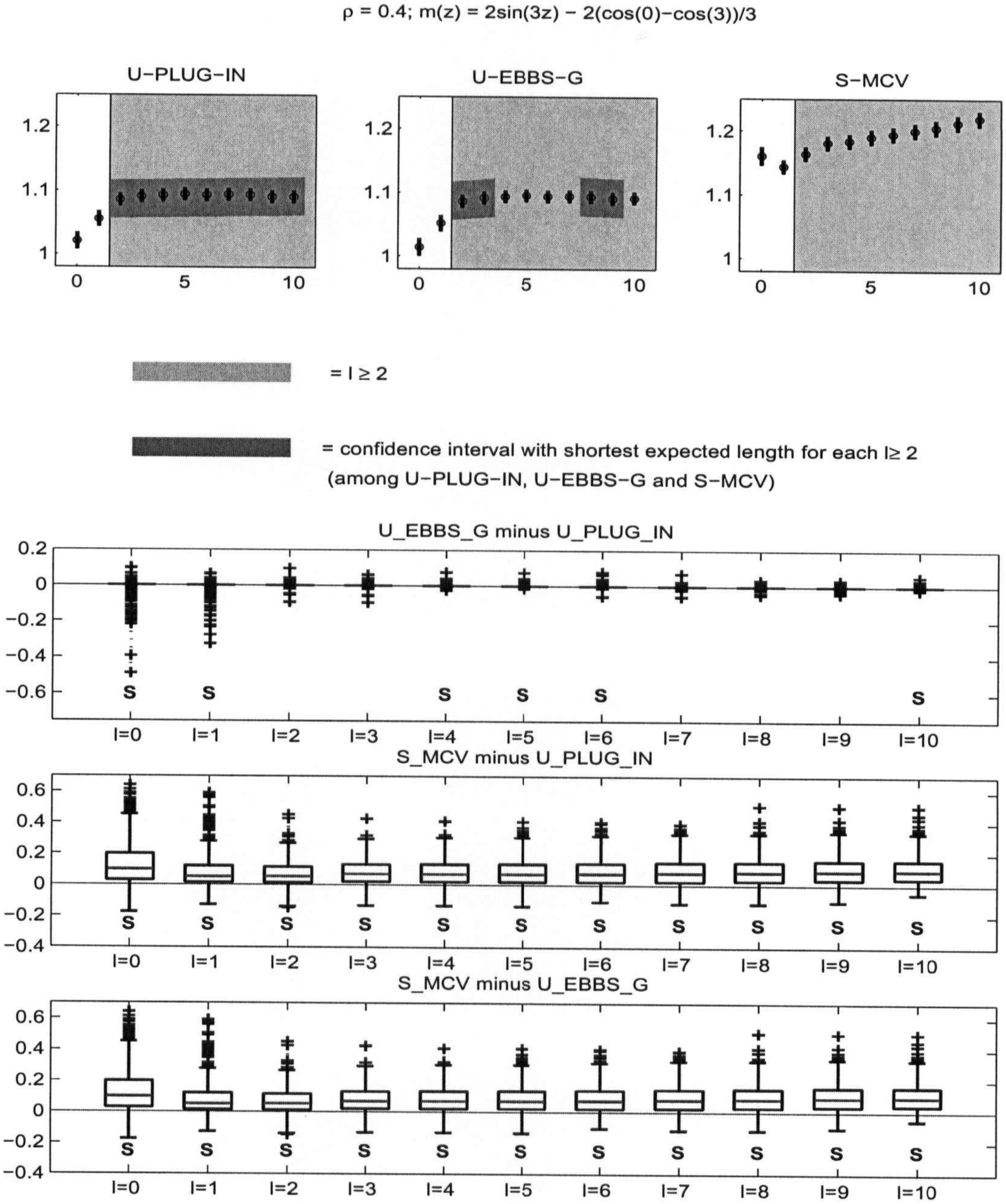

Figure C.3: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.4$ and $m(z) = m_1(z)$.*

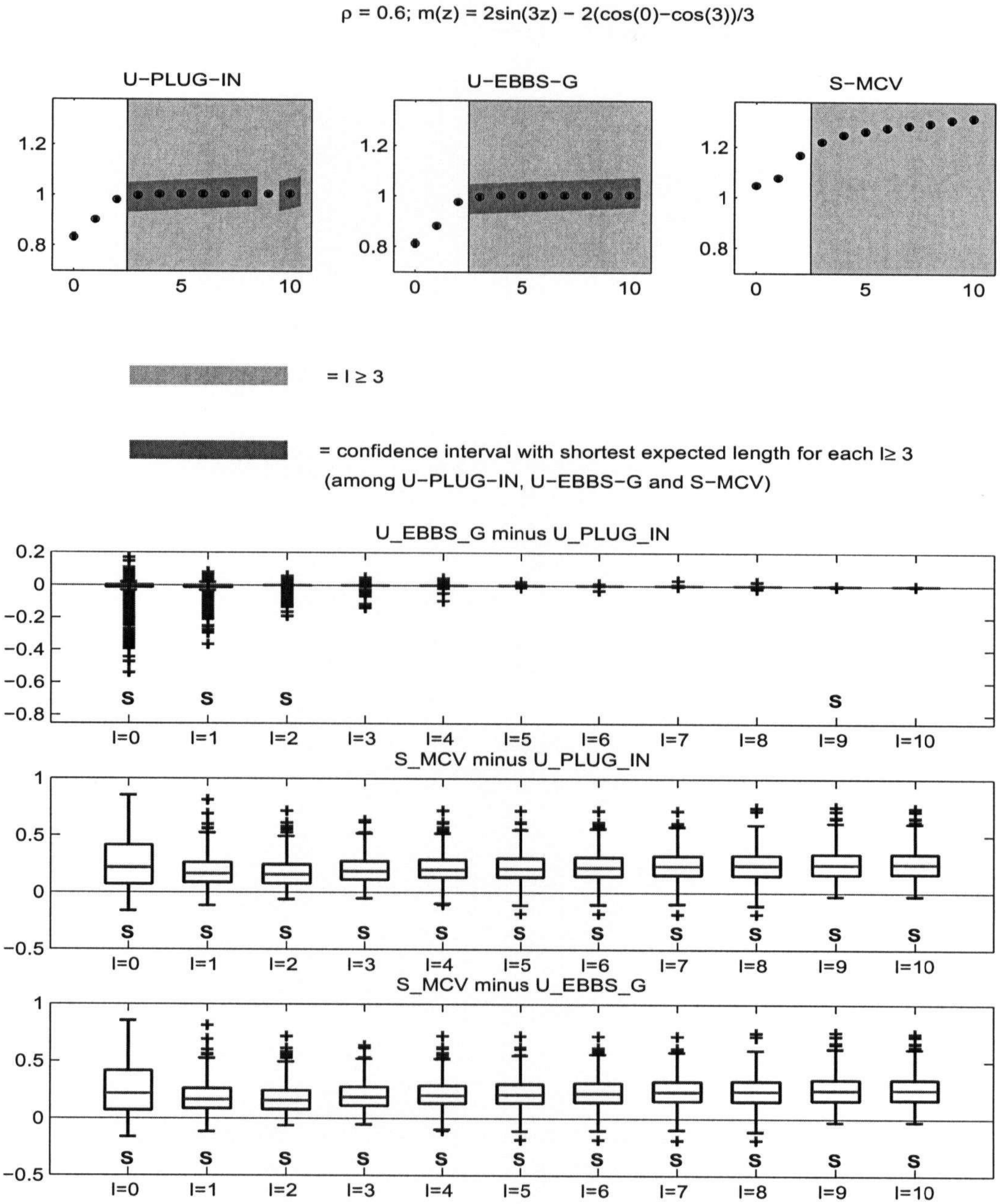

Figure C.4: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.6$ and $m(z) = m_1(z)$.*

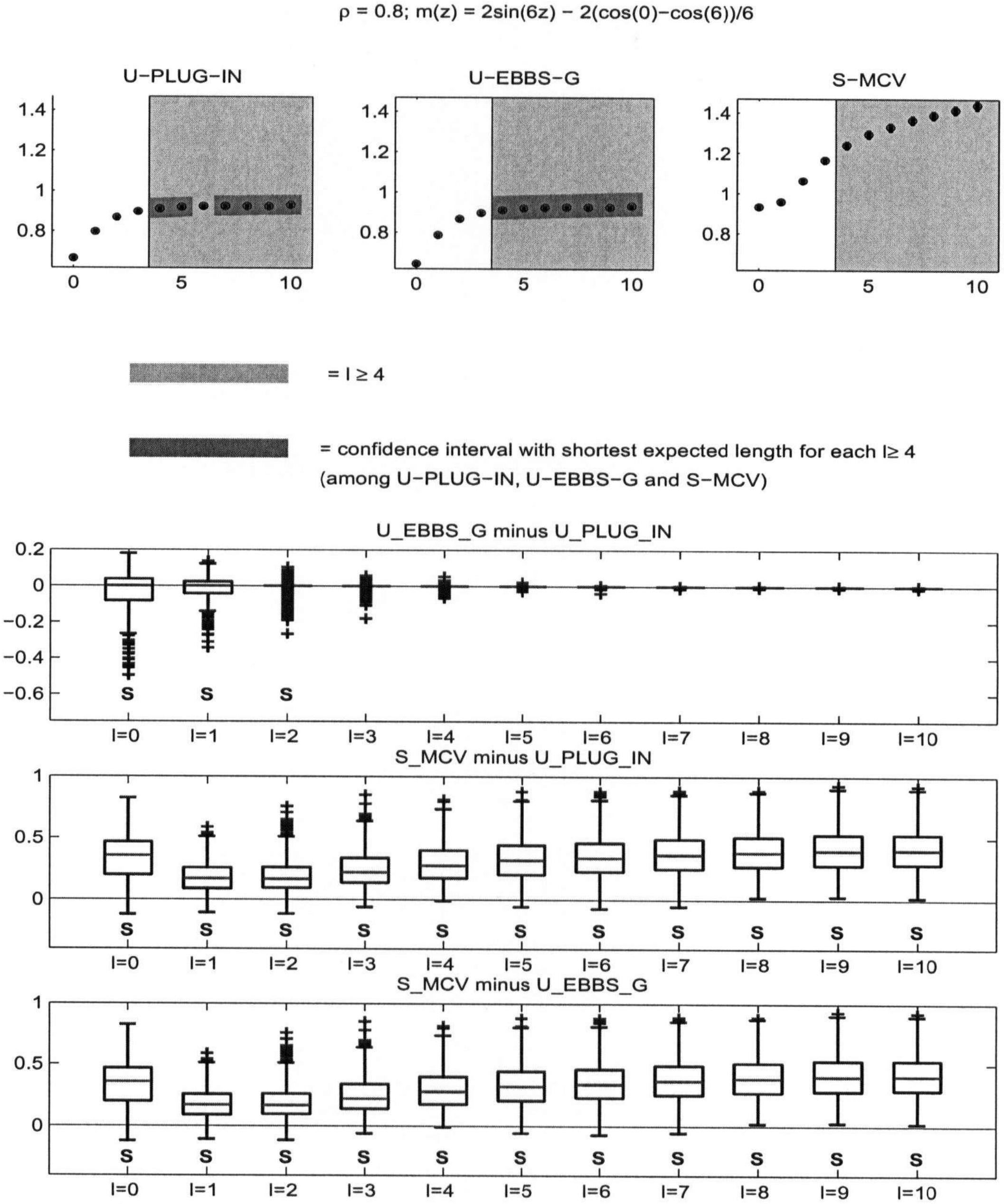

Figure C.5: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.8$ and $m(z) = m_1(z)$.*

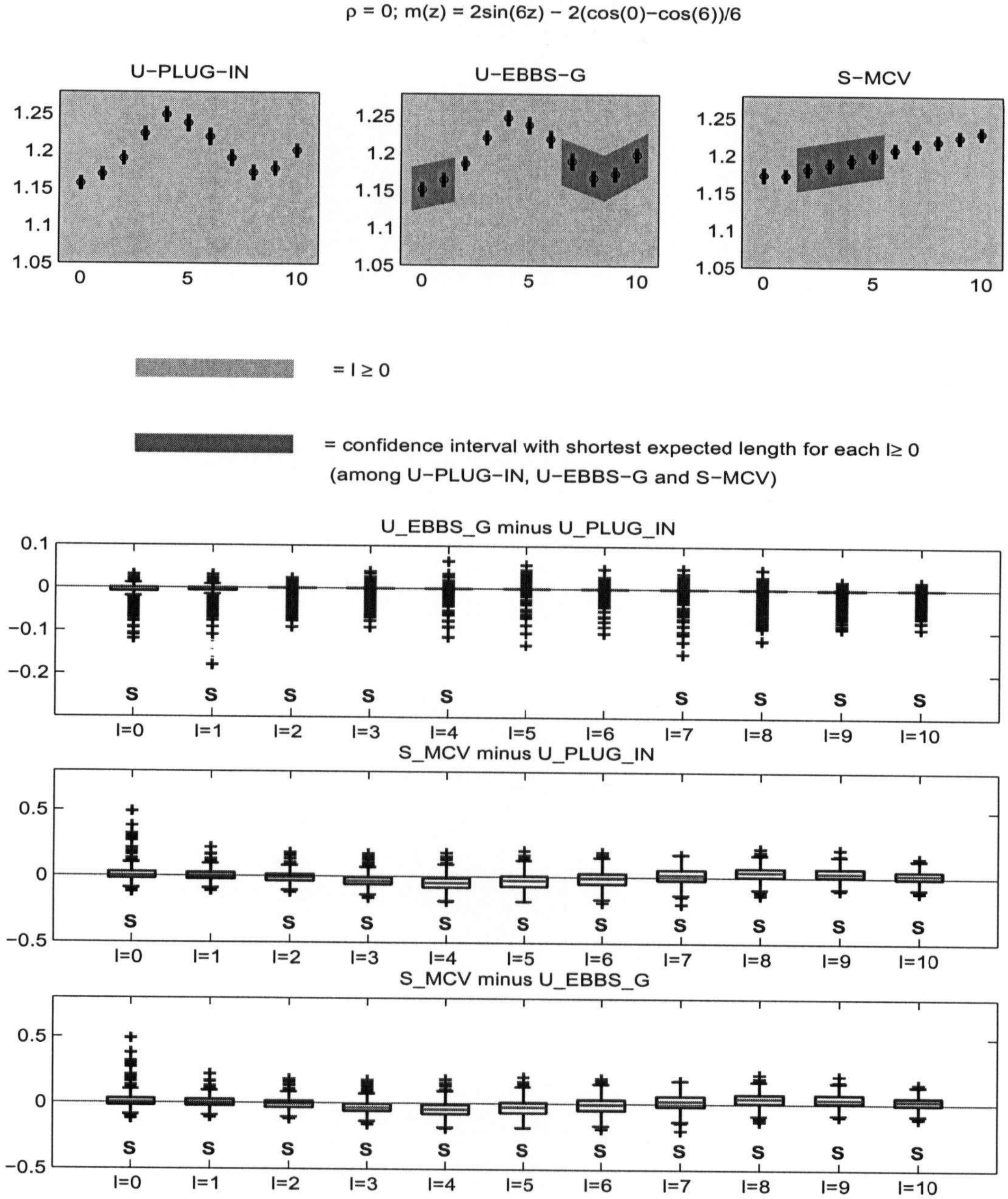

Figure C.6: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0$ and $m(z) = m_2(z)$.*

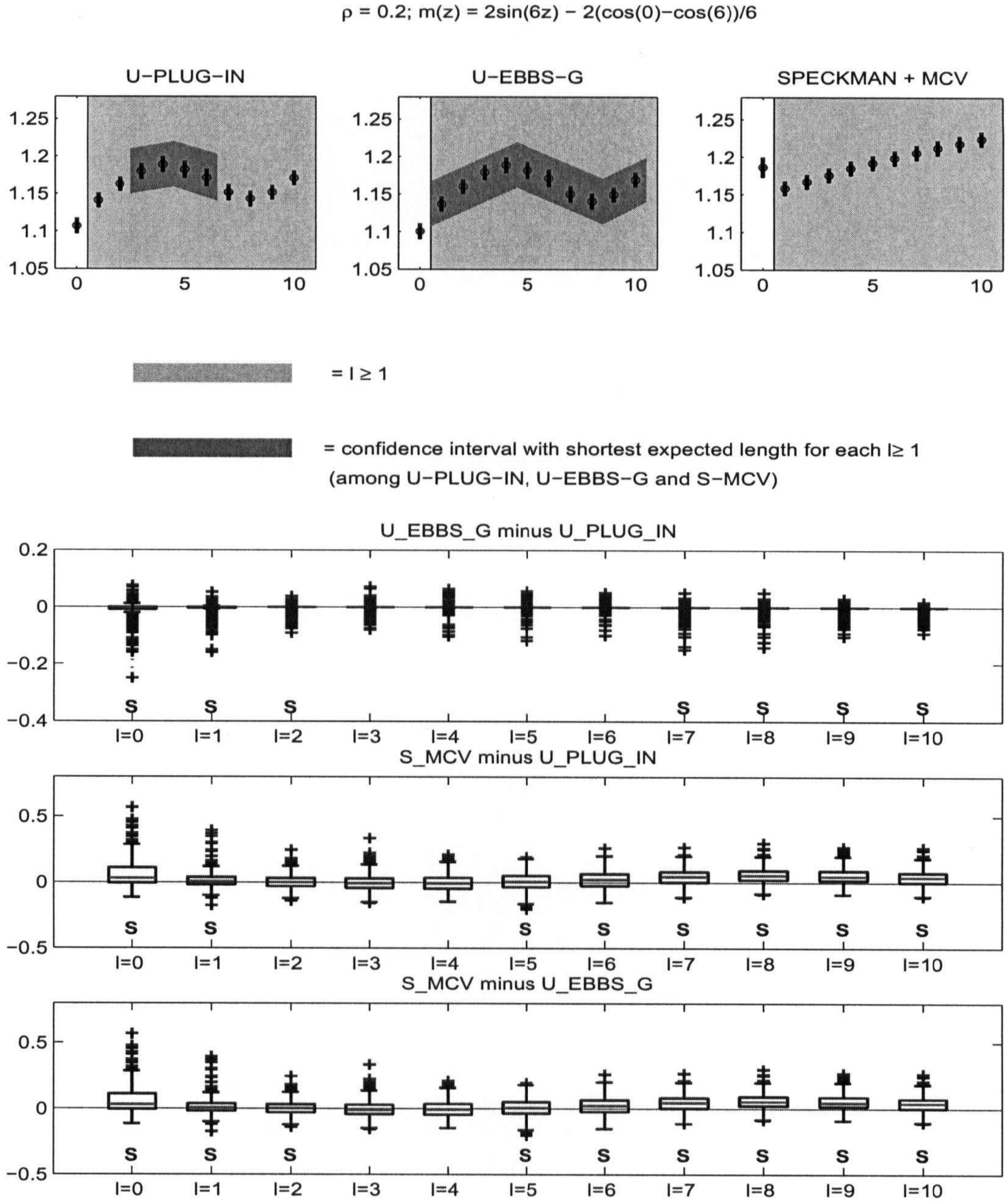

Figure C.7: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.2$ and $m(z) = m_2(z)$.*

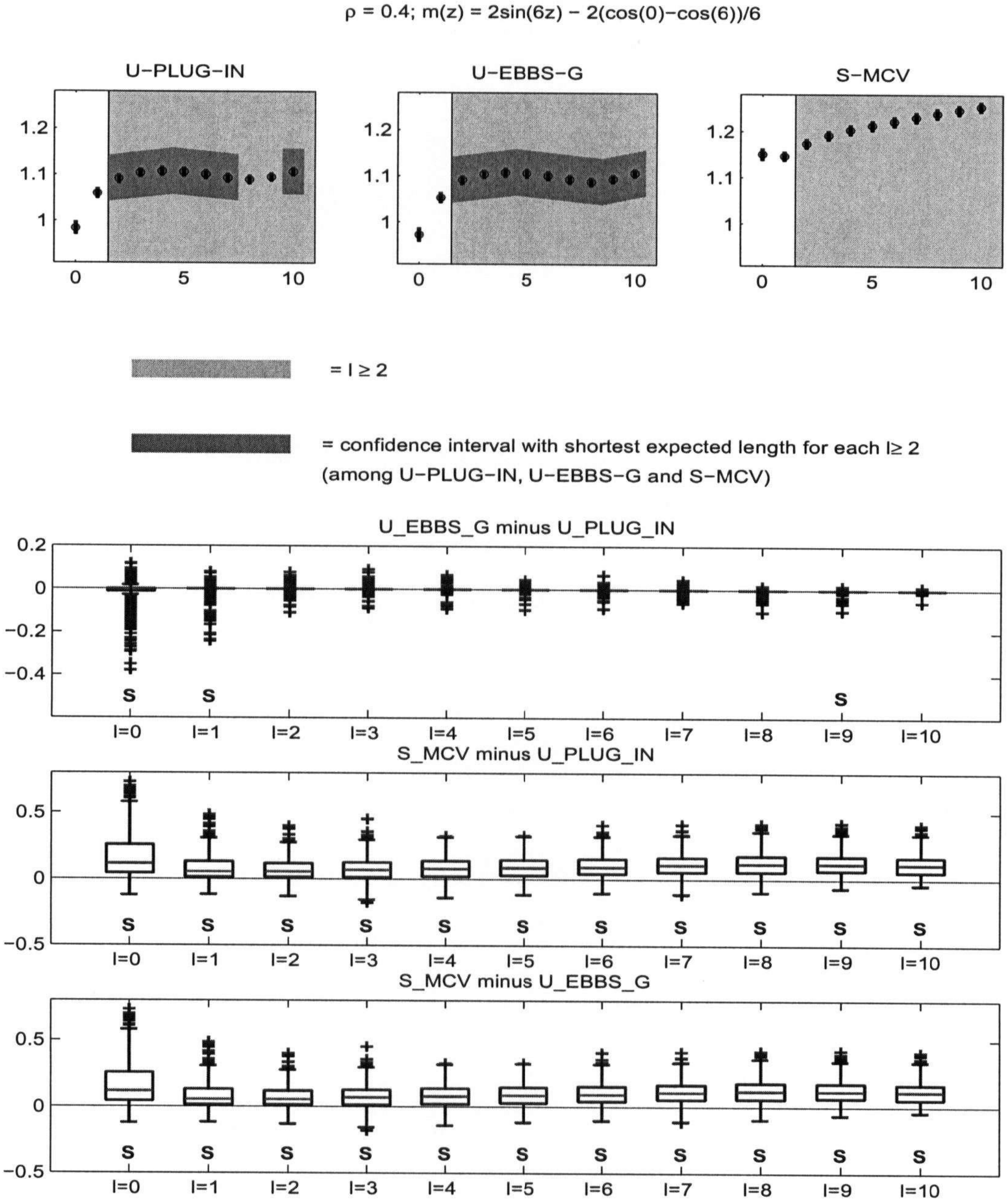

Figure C.8: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.4$ and $m(z) = m_2(z)$.*

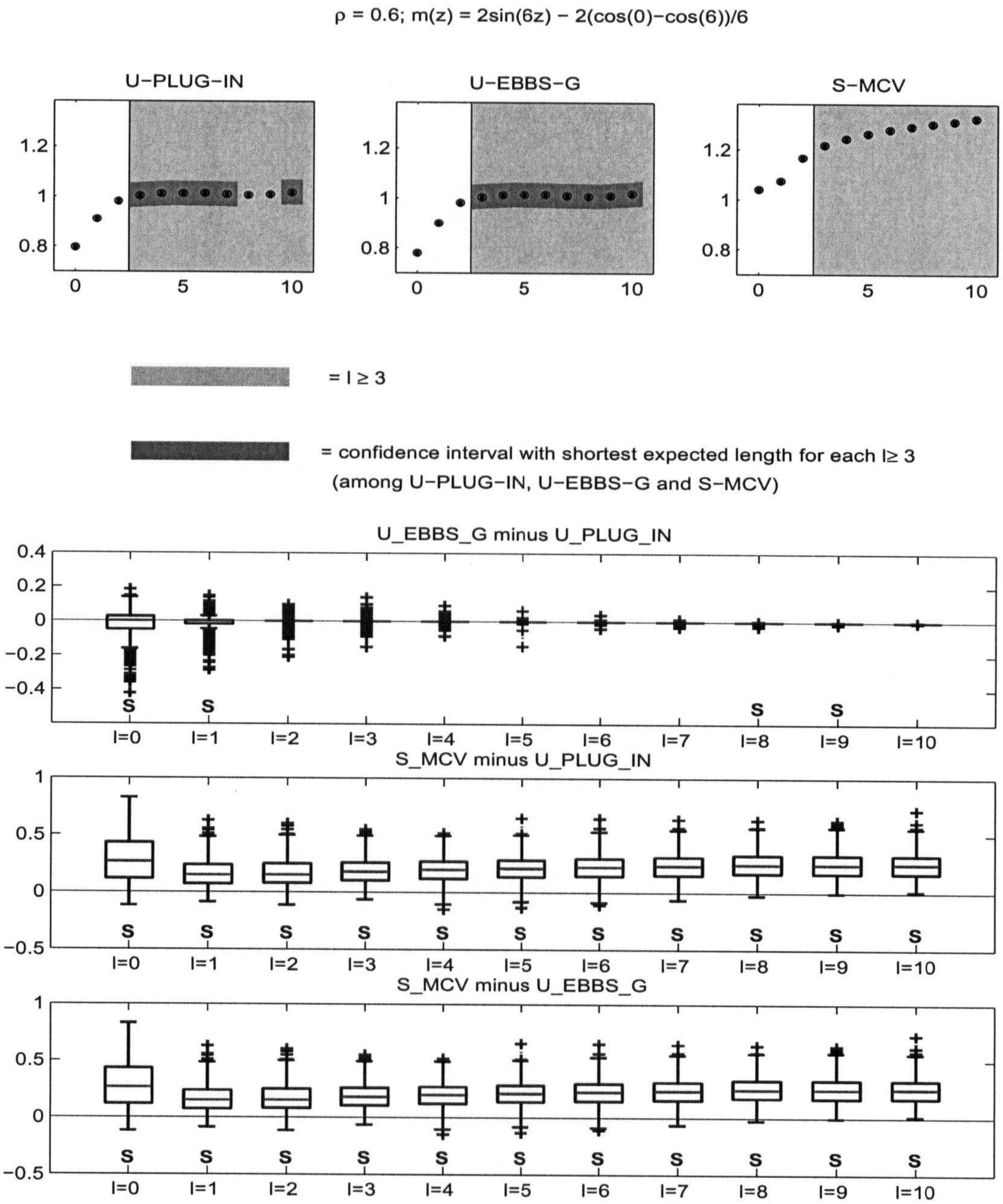

Figure C.9: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.6$ and $m(z) = m_2(z)$.*

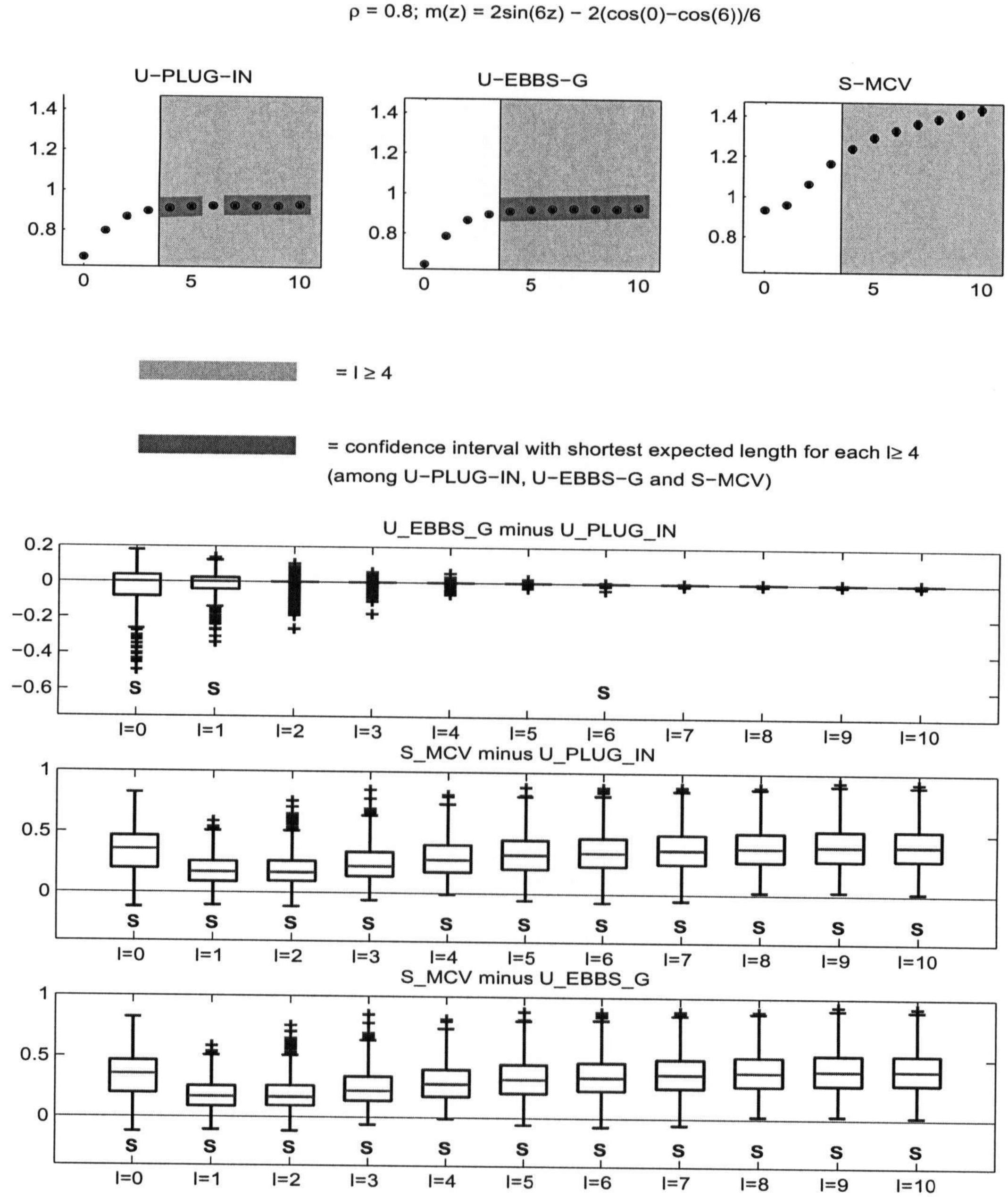

Figure C.10: *Top row: Average length of the standard confidence intervals for the linear effect β_1 in model (8.1) as a function of $l = 0, 1, \ldots, 10$. Standard error bars are attached. Bottom three rows: Boxplots of pairwise differences in the lengths of the standard confidence intervals for β_1. Boxplots for which the average difference in lengths is significantly different than 0 at the 0.05 level are labeled with an S. Lengths were computed with $\rho = 0.8$ and $m(z) = m_2(z)$.*