Open access • Journal Article • DOI:10.1111/J.1467-9868.2005.00535.X

On properties of functional principal components analysis — Source link

Peter Hall, Mohammad Hosseini-Nasab

Institutions: Australian National University

Published on: 01 Feb 2006 - Journal of The Royal Statistical Society Series B-statistical Methodology (Aiden Press)

Topics: Functional principal component analysis, Functional data analysis, Principal component analysis, Eigenvalues and eigenvectors and Estimator

ON PROPERTIES OF FUNCTIONAL PRINCIPAL COMPONENTS ANALYSIS

Mohammad Hosseini-Nasab

March 2006

A thesis submitted for the degree of Doctor of Philosophy of Australian National University

Declaration

The results in this thesis were produced by myself under the supervision of Professor Peter Hall, and with his help. All other sources used have been acknowledged.

Mohammad Hosseini-Nasab

Acknowledgements

First of all, I would like to express my full appreciation to my supervisor, Professor Peter Hall, for what I have learned as a PhD student from him about Functional Data Analysis (FDA). My greatest thanks goes to my supervisor for all his guidance and help throughout the duration of my candidature, including his comments on my drafts. I do appreciate his patience during preparation of this thesis, and owe my knowledge of FDA to his kind help during my PhD. I also gratefully acknowledge his financial support for attending a number of conferences.

I would also like to thank my friends, Dr Sergey Ajiev and Dr Christian Rau for their helpful discussions on some problems; John Maindonald for his help on R software; Professor Alan McIntosh for his guidance on some Functional Analysis problems and the Centre for Mathematics and its Applications at ANU as a whole for the excellent education that I have received.

My sincere thanks go to Dr Gail Craswell for her kind help in improving my academic writing. I also thank Professor J. O. Ramsay who made me familiar with this topic. I hope that in the future, I shall have a chance to attend McGill university and extend further my knowledge in this field.

I also gratefully acknowledge the financial support of the Iranian Ministry of Science for a PhD scholarship during my candidature.

Most important of all has been the support of my parents during this time.

Abstract

Functional data analysis is intrinsically infinite-dimensional; functional principal component analysis, or PCA, reduces dimension to a finite level, and points to the most significant components of the data. While this technique is often discussed, its properties are not as well understood as they might be. In this study we show how the properties of functional PCA can be elucidated through stochastic expansions and related results. Our approach quantifies the errors that arise through statistical approximation, in successive terms of orders $n^{-1/2}, n^{-1}, n^{-3/2}, \ldots$, where n denotes sample size. The expansions show how spacings among eigenvalues impact on statistical performance. The term of size $n^{-1/2}$ illustrates first-order properties, and leads directly to limit theory which describes the dominant impact of spacings. Thus, for example, spacings are seen to have an immediate, first-order effect on properties of eigenfunction estimators, but only a second-order effect on eigenvalue estimators. Our results can be used to explore properties of existing methods, and also to suggest new techniques. In particular, we suggest bootstrap methods for constructing simultaneous confidence regions for an infinite number of eigenvalues, and also for individual eigenvalues and eigenvectors. Also, the impact of eigenvalue spacings on properties of functional linear regression estimators and the validity of simple accounts of its performance are discussed.

Contents

Notation and terminology

Notation

$\mathcal{I}$	The bounded interval on which we define the random process $X(t)$, i.e. the argument $t \in \mathcal{I}$. We took it $\mathcal{I} = [0,1]$
K	Either covariance operator or the kernel function; $Kf(u) = \int K(u,v)\, f(v)\, dv$, where the kernel $K(u,v) = E\{X(u) - \mu(u)\}\,\{X(v) - \mu(v)\}$, and $\mu(u) = EX(u)$.
$\widehat{K}$	Either empirical covariance operator or its kernel function; $\widehat{K}f(u) = \int_{\mathcal{I}} \widehat{K}(u,v)\, f(v)\, dv$, where the kernel $\widehat{K}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(u) - \bar{X}(u)\}\,\{X_i(v) - \bar{X}(v)\}$.
$\lVert\lVert M \rVert\rVert$	$= (\int\int_{\mathcal{I}^2} M^2)^{1/2}$, called the Hilbert-Schmidt norm.
$\lVert\lVert M \rVert\rVert_{\text{sup}}$	$= \sup_{u \in \mathcal{I}} \lVert M(u,.) \rVert$.
$\int M\,\alpha\,\beta$	$= \int\int_{\mathcal{I}^2} M(u,v)\,\alpha(u)\beta(v)\, du\, dv$.
$\int \alpha\,\beta$	$= \int_{\mathcal{I}} \alpha(u)\beta(u)\, du$.
$\widehat{\Delta}$	$= \lVert\lVert \widehat{K} - K \rVert\rVert$.
$\widehat{\Delta}_{\text{sup}}$	$= \lVert\lVert \widehat{K} - K \rVert\rVert_{\text{sup}}$.
θ_j	The jth eigenvalue of the covariance operator K.

ψ_j	The jth eigenfunction of the covariance operator K.
$\hat{\theta}_j$	The jth eigenvalue of the empirical covariance operator $\widehat{K}$.
$\widehat{\psi}_j$	The jth eigenfunction of the empirical covariance operator $\widehat{K}$.
ζ_j	$\zeta_j \in (0,1)$ denotes the infimum of $1-(\theta_k/\theta_j)$ over k such that $\theta_k < \theta_j$
η_j	$\eta_j \in (0,1)$ denotes the infimum of $(\theta_k/\theta_j)-1$ over k such that $\theta_k > \theta_j$.
ρ_j	$\rho_j = \min_{k\neq j} \lvert\theta_k - \theta_j\rvert$.
δ_j	$\delta_j = \min_{1\le k\le j}(\theta_k - \theta_{k+1})$.
F	The space of all Hilbert-Schmidt operators on E.
$\langle T_1, T_2\rangle_F$	The inner product in F, which is defined by $\langle T_1, T_2\rangle_F = \sum_j \langle T_1\, e_j, T_2\, e_j\rangle_E$, where $T_1, T_2 \in F$ and $\{e_j\}$ is any complete orthonormal basis in E.
E	The Hilbert space $L^2_H(\mathcal{I}, \mathcal{G}, \mu)$ of functions $g : \mathcal{I} \to H$ such that $\int_{\mathcal{I}} g(u)^2\, d\mu(u) < \infty$.
$L_2(\mathcal{I})$	The space of square-integrable functions from $\mathcal{I}$ to the real line.
l_2	The space of all sequences $\{a_i\}$ such that $\sum_{i=1}^{\infty} \lvert a_i\rvert^2 < \infty$.
$\lVert\cdot\rVert$	The L_2-norm; i.e. $\lVert f\rVert^2 = \int f^2$, for each $f \in L_2$.
$\lVert\cdot\rVert_{\sup}$	The sup-norm; i.e. $\lVert f\rVert_{\sup} = \sup_{t\in\mathcal{I}} \lvert f(t)\rvert$.
$\mathcal{X}$	$= \{X_1, \cdots, X_n\}$ a random sample from population.

$\mathcal{X}^*$	$= \{X_1^*, \cdots, X_n^*\}$ a resample obtained by sampling from $\mathcal{X}$ with replacement.
$\hat{\theta}_j^*$	The bootstrap version of $\hat{\theta}_j$, computed from $\mathcal{X}^*$, rather than $\mathcal{X}$.
$\widehat{\psi}_j^*$	The bootstrap version of $\widehat{\psi}_j$, computed from $\mathcal{X}^*$, rather than $\mathcal{X}$.
ϕ, Φ	Standard Normal density and distribution functions, respectively.

Chapter 1

Functional Principal Components

1.1 Introduction

In recent years, research on functional data analysis has increased. This is partly due to technological advancements which have made it possible to measure an object over a dense grid by high frequency automatic sensing equipment. For example, Chemometrics, a field of chemistry that studies the application of statistical methods to chemical data analysis, has been one of the first areas of research to move towards analysing these kinds of data. Some of the data in this field, observed mostly from organic and analytical chemistry and food research, have important features which make the data different from other kinds. Each observation is distinguished by many measured variables (sometimes more than the number of the observations) which tend to have high autocorrelation. For example, when applying near infrared reflectance (NIR) spectroscopy information to obtain percentages of fat or of other constituents in biscuit dough, NIR information for one signal consists of hundreds of digitizations (Marx and Eilers, 1999). Another example of the relationship between log-spectra of sequences of spoken syllables and phoneme classification can be found in Marx and Eilers (1996).

Also, further applications of these data may be found in estimating the link between a real random response and a random function, $X(t)$ say, being digitized at many points. For instance, by constructing a linear regression, we can predict the percentage of fat in biscuit, using NIR wavelengths as the argument of the independent variable X, which is in the form of a discrete representation of the observed signal (Marx and Eilers, 1999). Another example arises when using near-infrared spectroscopy to obtain information about the level of a particular protein in different varieties of wheat. Let $X_i(t)$ be the recorded intensity of reflected radiation, when the wavelength equals t and Y_i is the level of the protein for the ith wheat type. By construction of a linear regression model, we may use the model with new values of $X(t)$ to predict the level of the protein in that cereal type (Hall and Horowitz, 2004). In both examples, prediction of the value of the dependent variable, for a new X, from the constructed model, could be very useful due to difficulties in measuring the dependent variable. While measuring the dependent variable could be expensive and slow in a laboratory, the covariate X_i (NIR photometric measurements) can be measured faster with less expenditure (Marx and Eilers, 1999, Hall and Horowitz 2004).

1.1.1 Random and Age Effects

One of the aspects of functional data is to show changes over time (the "age" effect), separated from the effects caused by differences among subjects which are chosen from the population for the study. This is due to the nature of the collected data, consisting of repeated measurements of subjects through time. Unlike cross-sectional studies, in which we measure a single quantity for each object, here we are able to use the capacity of data to explore the "age" effect

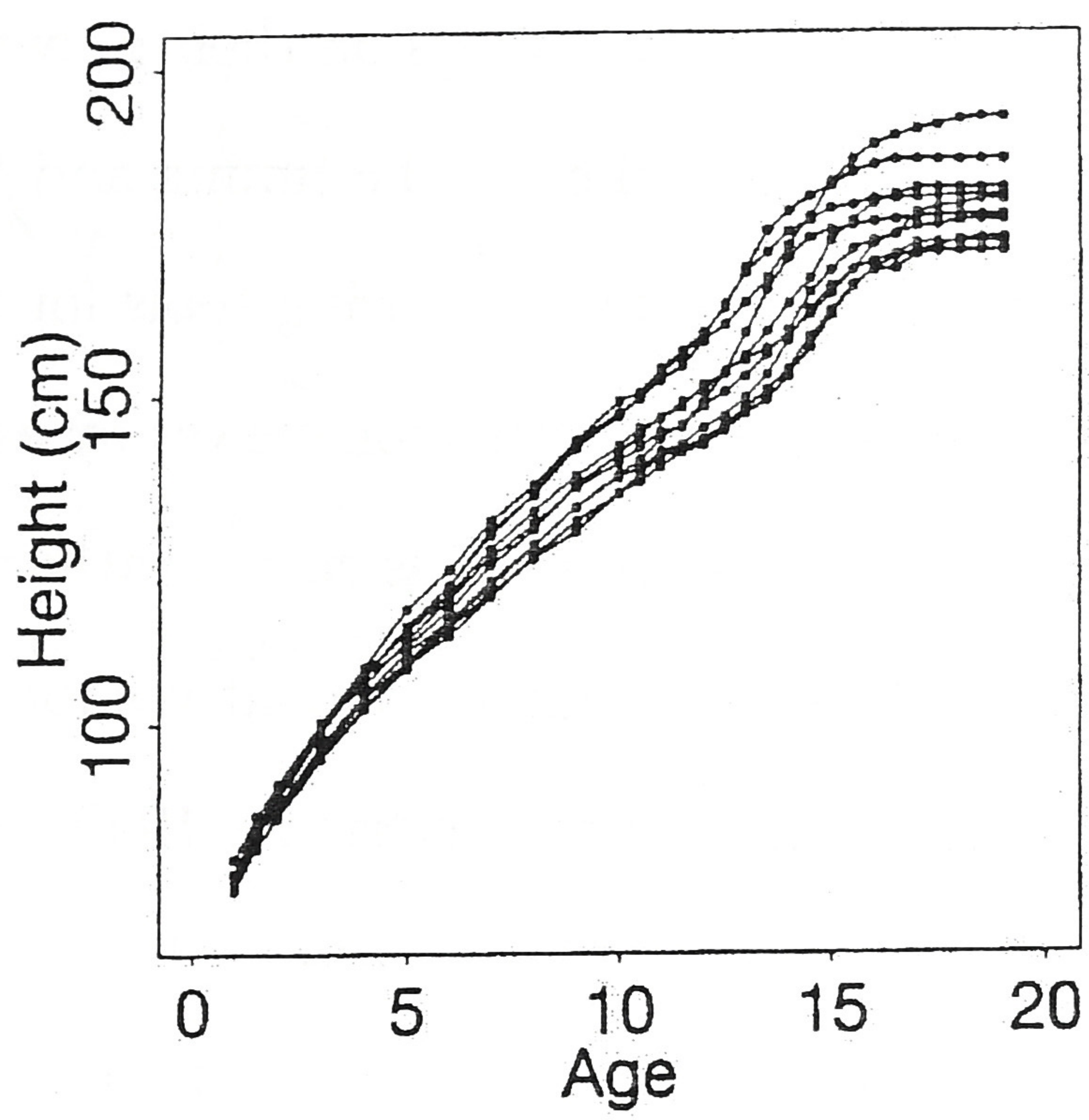

Figure 1.1: The heights of 10 Swiss boys measured at 29 ages. The points indicates the unequally spaced ages of measurement. See Ramsay and Silverman (1997, page 2).

by analyzing the data. Separation of changes over time within objects from those among them can be beneficial for revealing useful characterizations of the population from which the sample was drawn. In a sample like $X_1(t), \cdots, X_n(t)$, the former variation refers to the variable t, time, which belongs to an interval, say $[a, b]$, and the second can be seen through the essential randomness of X, in which for a certain time, we have different values of X when running from the first individual $(X_1(t))$ to the last one in the sample $(X_n(t))$. In cross-sectional data, however, we can see the differences among individuals by measuring a quantity over sampled individuals, showing $X_1, X_2, \cdots, X_n$.

Figure 1.1 shows the heights of ten boys, obtained by measuring each boy at 29 different points of time (Ramsay and Silverman 1997, page 2). For each boy, measurement was begun at age two and continued annually until age ten, after which it was done biannually for all boys. Therefore, we have 29 records for each person, which can be assumed as a continuous function due to the nature of growth. As the graph shows, it can be easily recognized that the sign of almost all boys' height accelerations tend to change at some points (ages), especially at

12, 14, and 16. This might be due to pubertal effects on their growth. However, the effect is not the same for all boys, and differs in the timing and the intensity. To explore the "age" effect, one can be benefited by using tools for investigating behavior of functions, such as obtaining the acceleration curves by estimating the $D^2Height_i$ from the data. Thus, thinking of records as curves rather than vectors of observations in discrete time enables us to employ derivatives for investigating the "age" effect in functional data (Ramsay and Silverman, 1997, page 2).

1.1.2 High Dimensionality

Compared to other kinds of data, the second difference is in the view of dimension of the spaces to which the data belong. Because our data are functions, they lie in function spaces, which are of infinite dimension. One possible way of explaining this aspect is to start with classical Multivariate Data Analysis (MDA). Assume that we draw n subjects at random from a population and for each subject, p quantities are measured. In other words, we assigned a p-vector of real numbers to each individual. Hence, we have n p-vectors, each belonging to the p-dimensional space R^p. Therefore, we can represent each of them as a linear combination of p orthonormal vectors, $\mathrm{e}_1, \mathrm{e}_2, \cdots, \mathrm{e}_p$ say, assumed to be a complete basis in R^p. Then, the p-vector $\mathrm{X_i}$, related to the ith subject, can be presented as

$$\mathrm{X_i} = \xi_{i1}\,\mathrm{e}_1 + \xi_{i2}\,\mathrm{e}_2 + \cdots + \xi_{ip}\,\mathrm{e}_p\,,$$

where the random variables ξ_{ij} denote the coordinates of the X_i with respect to the basis. In this way, we can represent

$$\mathbf{X}_{n;p} = \begin{pmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1p} \\ \vdots & \vdots & & \vdots \\ \xi_{n1} & \xi_{n2} & \dots & \xi_{np} \end{pmatrix} \begin{pmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ \vdots & \vdots & & \vdots \\ e_{p1} & e_{p2} & \dots & e_{pp} \end{pmatrix},$$

where the first matrix on the right-hand side contains the coordinates of the $\mathbf{X}_i$ with respect to the basis in R^p.

When data tend to be a continuum, they can not be shown as a linear combination of any finite basis, but they may need an infinite number of basis. In FDA we take $X(t)$ to be in a separable Hilbert space (usually $L_2(\mathcal{I})$, the space of square-integrable functions defined on the bounded interval $\mathcal{I}$). This implies that the $X_i(t)$ can be represented as

$$X_i(t) = \sum_{j=1}^{\infty} \xi_{ij}\, e_j(t)\,,$$

where the sequence $\{e_1(t), e_2(t), \cdots\}$ is a complete basis in the Hilbert space and the random variables $\xi_{i1}, \xi_{i2}, \dots$ are its coordinates corresponding to the basis. If the associated series of $X(t)$ is truncated at $j = p$, then we can represent a sample of size n as

$$\mathbf{X}_{n;p} = \begin{pmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1p} \\ \vdots & \vdots & & \vdots \\ \xi_{n1} & \xi_{n2} & \dots & \xi_{np} \end{pmatrix} \begin{pmatrix} e_1(t_1) & e_1(t_2) & \dots & e_1(t_q) \\ \vdots & \vdots & & \vdots \\ e_p(t_1) & e_p(t_2) & \dots & e_p(t_q) \end{pmatrix},$$

where the second matrix on the right-hand side is a discretization of the basis at some points, say q points. If we let p go forward, meaning we are increasing the dimension, the number of columns of the first matrix and thus the number of rows of the second matrix will be increased. Therefore, in some sense, FDA can be imagined similarly to MDA as the number of variables (p) increases. In this way, also the concepts and tools in MDA can be easily justified for FDA. For example, in the case of the inner product, the two p-vectors $\mathbf{u}_1, \mathbf{u}_2$ in R^p can be multiplied as $\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \mathbf{u}_1^T \mathbf{u}_2 = \sum_{i=1}^{p} u_{1i} u_{2i}$, where $\mathbf{u}^T$ denotes the transpose of $\mathbf{u}$. However, as p tends to infinity, the vectors tend to have infinite number of coordinates, which can be interpreted as measurements of $u_i(t)$ at time t_j, for many t_j. Thus, the inner product here is changed to the integral $\langle u_1, u_2 \rangle = \int u_1(t) u_2(t) dt$. Then, the concept of orthonormality of two elements is the same in both MDA and FDA, i.e. the corresponding inner product of the two elements is zero. As a result, the norm obtained from their inner product is defined as $||x||^2 = \langle x, x \rangle$. So, it can be seen that majority of concepts in FDA have similar interpretation in the MDA, as the dimensionality goes to infinity. These features were pointed out in Ramsay's first article about FDA (1983) as well as by Ramsay and Silverman (1997).

In regard to high dimensionality of functional data, they may challenge classical methods of data analysis and need theoretical justification. Some theoretical justification for PCA in functional data analysis (FDA) is provided by limit theory. See, for example, the work of Dauxois, Pousse and Romain (1982) and Bosq (2000).

1.2 Principal Component Analysis for Functional Data

1.2.1 Introduction

Classical principal component analysis (PCA) is amongst the oldest of the multivariate statistical methods of data reduction. A Multivariate Analysis problem could start out with a substantial number of correlated variables. In such situations, PCA produces a small number of constructed variables from the original data that are uncorrelated and account for most of the variation in the original data set. The main reason for reducing the number of variables in this way, is that it helps us to understand the underlying structure of the data. For this reason, PCA has found application in fields such as signal processing, face recognition, image compression and so on (Jolliffe, 2002). Similarly, PCA is widely used in the study of functional data, since it allows finite-dimensional analysis of a problem that is intrinsically infinite-dimensional. See, for example, Chapter 6 of Ramsay and Silverman (1997), and several of the examples treated by Ramsay and Silverman (2002). In traditional PCA the effects of truncating to a finite number of dimensions are often explored in terms of a finite-dimensional parametric model, for example when the data are Gaussian. However, this approach is often not feasible in the case of functional data analysis (FDA), and as a result the justification for methodology there tends to be more ad hoc.

Early work on PCA for FDA includes that of Besse and Ramsay (1986), Ramsay and Dalzell (1991), Rice and Silverman (1991), Pezzulli and Silverman (1993) and Silverman (1995, 1996). Accounts in monographs include those of Ramsay and Silverman (1997), especially Chapter 6, and Ramsay and Silverman (2002). Work of Dauxois et al. (1982), Bosq (1989), Besse (1992), Huang et al. (2002) and Mas (2002), for example, addresses empirical basis function approximation and approximations of covariance operators.

Recent work includes contributions to techniques for functional PCA (see e.g. Brumback and Rice, 1998; Cardot, 2000; Cardot et al., 2000, 2003; Girard, 2000; James et al., 2000; Boente and Fraiman, 200; He et al., 2003.

1.2.2 Definition

Let $(\Omega, \mathcal{F})$ and $(\mathcal{I}, \mathcal{G})$ be measurable spaces with the probability measure P and the bounded measure μ, respectively. Suppose that $\{X(.,t)\,;\, t \in \mathcal{I}\}$ denotes a vector of random functions mapping from $(\Omega, \mathcal{F}, P)$ into $(H, \mathcal{B}_H)$, where H is a separable Hilbert space and $\mathcal{B}_H$ is its Borel field. Furthermore, let $L^2_H(P \otimes \mu)$ denote the separable Hilbert space of the equivalent classes of measurable functions defined from the product measurable space $(\Omega \times \mathcal{I}, \mathcal{F} \otimes \mathcal{G})$ to $(H, \mathcal{B}_H)$ such that their squared norm is $P \otimes \mu$-integrable. It is assumed that $X \in L^2_H(P \otimes \mu)$, i.e. $\int_{\mathcal{I}} \int_\Omega X(\omega, t)^2\, dP(\omega)\, d\mu(t) = \int_{\mathcal{I}} E\big[X(\omega,t)^2\big]\, d\mu(t) < \infty$, and without loss of generality $\eta(t) = E[X(.,t)] = \int_\Omega X(\omega,t)\, dP(\omega) = 0$.

Let the Hilbert space $L^2_R(\Omega, \mathcal{F}, P)$ (denoted by $L^2(P)$) be square P-integrable functions $f : \Omega \to R$, and the Hilbert space $L^2_H(\mathcal{I}, \mathcal{G}, \mu)$ (denoted by E) of functions $g : \mathcal{I} \to H$ such that $\int_{\mathcal{I}} g(u)^2\, d\mu(u) < \infty$. Suppose that the bounded linear operator V from $L^2(P)$ to E is defined by

$$(Vf)(t) = \int_\Omega X(\omega, t)\, f(\omega)\, dP(\omega) = E[X(.,t)\, f], \quad \mu \text{ a.e.} \quad \text{for each } f \in L^2(P).$$

The adjoint operator V^* is defined through $\langle Vf, \upsilon\rangle = \langle f, V^*\upsilon\rangle$, for each $f \in L^2(P)$ and $\upsilon \in E$ as follows:

$$\begin{aligned}
\int_{\mathcal{I}} \int_\Omega X(\omega,t)\, f(\omega)\, \upsilon(t)\, dP(\omega)\, d\mu(t) &= \int_\Omega \int_{\mathcal{I}} X(\omega,t)\, f(\omega)\, \upsilon(t)\, d\mu(t)\, dP(\omega) \\
&= \int_\Omega \left\{ \int_{\mathcal{I}} \langle X(\omega,t), \upsilon(t)\rangle_H\, d\mu(t) \right\} f(\omega)\, dP(\omega) \\
&= \int_\Omega (V^*\upsilon)(\omega)\, f(\omega)\, dP(\omega)\,,
\end{aligned}$$

where $\langle .,\, . \rangle_H$ denotes the inner product of H. Hence, for each $g \in E$,

$$(V^*g)(\omega) = \int_{\mathcal{I}} \langle X(\omega, t),\, g(t) \rangle_H \, d\mu(t) = \langle X(\omega, .),\, g \rangle, \quad P \text{ a.e.},$$

where $\langle ., . \rangle$ denotes the inner product of E.

Let ϕ be a continuous linear functional on a Hilbert space H. There exists a unique $y \in H$ such that $\phi(x) = \langle x\,,\, y \rangle$ for all $x \in H$, and $\|\phi\|_{\mathcal{L}} = \|y\|$, where $\|\phi\|_{\mathcal{L}} = \sup \big\{ \, \|\phi x\| \, : \, x \in H \, , \, \|x\| \leq 1 \big\}$. Moreover, the mapping, which associates to ϕ the unique element y, is one-to-one, onto, norm-preserving conjugate-linear map of the dual of H onto H (a result of the Riesz Representation Theorem, see Theorem 3.3.4 of Ash 1972). Thus, the dual of H is identified (via this map) with H itself.

The associated schema of duality for these operators is

$$\begin{array}{ccc} E & \xleftarrow{\;V\;} & L^2(P) \\ {\scriptstyle K}\Big\uparrow\Big\downarrow{\scriptstyle I} & & {\scriptstyle I}\Big\uparrow\Big\downarrow{\scriptstyle W} \\ E & \xrightarrow{\;V^*\;} & L^2(P) \end{array}$$

With this schema, the identity operator I identifies the two Hilbert spaces with their corresponding duals, as discussed before. Furthermore, K and W are $V \circ V^*$ and $V^* \circ V$, respectively. For each $g \in E$ and $u \in \mathcal{I}$, the former can be expressed as

$$(Kg)(u) = \Big(V \circ (V^*g) \Big)(u) = \Big(V \circ (\langle X(\omega, .),\, g \rangle) \Big)(u) = E\big[X(.,\, u) \langle X(\omega, .),\, g \rangle \big].$$

This is actually a kernel operator with symmetric kernel $K(u, v) = E\big[X(., u)\, X(., v) \big]$

such that

$$(Kg)(u) = \int_{\mathcal{I}} K(u,v)\, g(u)\, du\,, \tag{1.1}$$

where

$$\begin{aligned}\int_{\mathcal{I}}\int_{\mathcal{I}} \left|K(u,v)\right|^2 d\mu(u)\, d\mu(v) &= \int_{\mathcal{I}}\int_{\mathcal{I}} \Big|E\big[X(.,u)\, X(.,v)\big]\Big|^2 d\mu(u)\, d\mu(v)\\ &\le \int_{\mathcal{I}}\int_{\mathcal{I}} E\big[X(.,u)^2\big]\, E\big[X(.,v)^2\big]\, d\mu(u)\, d\mu(v)\\ &= \left\{\int_{\mathcal{I}} E\big[X(.,u)^2\big]\, d\mu(u)\right\}\left\{\int_{\mathcal{I}} E\big[X(.,v)^2\big]\, d\mu(v)\right\} < \infty\,.\end{aligned} \tag{1.2}$$

It is also clear that the operator $K : E \to E$ is self-adjoint and non-negative.

Theorem 1.1. Let $C(u,v) : \mathcal{I}\times\mathcal{I} \to R$ be a symmetric kernel function. Then the linear integral operator

$$(Cg)(u) = \int_{\mathcal{I}} C(u,v)\, g(u)\, d\mu(u)\,, \quad \text{for each } g \in L^2(d\mu)\,,$$

is a compact self-adjoint Hilbert-Schmidt operator on $L^2(d\mu)$ if

$$\int_{\mathcal{I}}\int_{\mathcal{I}} \left|C(u,v)\right|^2 d\mu(u)\, d\mu(v) < \infty\,.$$

Furthermore, if the above condition is satisfied then $C(u,v)$ is diagonalizable and can be expressed as follows:

$$C(u,v) = \sum_{j=1}^{\infty} \lambda_j\, \phi_j(u)\, \phi_j(v)\,,$$

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ is an enumeration of the eigenvalues of C, and the corresponding orthonormal eigenfunctions are $\phi_1, \phi_2, \ldots$.

Proof: See Conway (1985), page 47, or Dunford and Schwartz (1963), pages

1009 and 1130. ■

Therefore, using the above theorem with (1.1) and (1.2) implies that K is a Hilbert-Schmidt operator with

$$K(u,v) = \sum_{j=1}^{\infty} \theta_j\, \psi_j(u)\, \psi_j(v)\,, \tag{1.3}$$

where $\theta_1 \geq \theta_2 \geq \cdots \geq 0$ is an enumeration of the eigenvalues of K, and the corresponding orthonormal eigenfunctions are $\psi_1, \psi_2, \ldots$. Moreover, if all θ_j are positive, the sequence $\{\psi_j\}$ forms a complete orthonormal sequence in E (Riesz and SZ. Nagy (1955), page 234). In analogy to K, it can be shown that $W : L^2(P) \to L^2(P)$ is a non-negative, self-adjoint Hilbert-Schmidt operator. We will not investigate further the operator W since it is not related to PCA.

It should be noted that $F = \sigma_2(E)$, the space of all Hilbert-Schmidt operators on E, with the inner product

$$\langle T_1, T_2 \rangle_F = \sum_j \langle T_1\, e_j, T_2\, e_j \rangle_E\,, \tag{1.4}$$

is a separable Hilbert space, where $T_1, T_2 \in F$ and $\{e_j\}$ is any complete orthonormal basis in E. Thus, the Hilbert-Schmidt norm induced by the inner product does not depend on the choice of the basis (Conway 1990, page 273), and for $K \in F$, $\|K\|^2 \equiv \langle K, K \rangle_F = \sum_{j=1}^{\infty} \|K\,\psi_j\|^2 = \sum_{j=1}^{\infty} \theta_j^2 < \infty$.

Suppose that $h_1 : E \to E \times E$ maps $x \in E$ to $(x,x) \in E \times E$ and $h_2 : E \times E \to F$ maps $(x,y) \in E \times E$ to $x \otimes y \in F$, where $x \otimes y$ is defined by $(x \otimes y)f = \langle x, f \rangle\, y$ for each $f \in E$. The two maps h_1, h_2 are continuous. So, we have $h_2 \circ h_1(X) = h_2\big((X,X)\big) = X \otimes X$ is a random variable on $(F, \mathcal{B}_F)$ ($\mathcal{B}_F$ denotes the corresponding Borel field of F), by which each $\omega \in \Omega$ is mapped to $X(\omega,.) \otimes X(\omega,.) \in F$. Therefore, K can be shown in the simple form $K =$

$E(X \otimes X)$ in F.

The above arguments can be easily justified for the classical multivariate analysis of PCA (simply, take $H = R$ and $\mathcal{I} = \{1, 2, \cdots, p\}$). Furthermore, if we take $\Omega = \{\omega_1, \cdots, \omega_n\}$ and the probability measure P such that $P(\{\omega_i\}) = \frac{1}{n}$, for $i = 1, \cdots, n$, it can be shown that $L^2(P)$ is isomorphic to R^n with metric $\frac{1}{n} I_n$, where the n by n matrix I_n is the identity matrix. Define a map to carry any element $f \in L^2(P)$ to $\frac{1}{\sqrt{n}} \left(f(\omega_1), \cdots, f(\omega_n)\right)^T \in R^n$. The map is isomorphic, and we have the following schema of duality:

$$\begin{array}{ccc} E & \xleftarrow{V_n} & (R^n)^* \\ {\scriptstyle K_n}\uparrow\downarrow{\scriptstyle I} & & \uparrow{\scriptstyle I_n} \\ E & \xrightarrow{V_n^*} & R^n \end{array}$$

where $(R^n)^*$ denotes the dual space of R^n, and

$$V_n f = \frac{1}{n} \sum_{i=1}^{n} X(\omega_i, .) f(\omega_i), \qquad \text{for each } f \in \left(R^n\right)^*,$$

$$\left(V_n^* u\right)(\omega_i) = \langle X(\omega_i, .), u\rangle, \quad \text{for each } u \in E,$$

and

$$\begin{aligned} K_n g &= \frac{1}{n} \sum_{i=1}^{n} \langle X(\omega_i, .), g\rangle \, X(\omega_i, .) \\ &= \left(\frac{1}{n} \sum_{i=1}^{n} X(\omega_i, .) \otimes X(\omega_i, .)\right)(g), \text{ for each } g \in E. \end{aligned} \tag{1.5}$$

Suppose that $(\prod_{j=1}^{\infty} \Omega, \prod_{j=1}^{\infty} \mathcal{F}, \prod_{j=1}^{\infty} P)$ is the measure product space, where $\omega = (\omega_1, \omega_2, \cdots) \in \prod_{j=1}^{\infty} \Omega$, and $\prod_{j=1}^{\infty} \mathcal{F}$ is the minimal σ-field over the measurable rectangles. We denote the above measure product space by $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$. Let Π_i be the ith canonical projection from $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$ into $(\Omega, \mathcal{F}, P)$ such that $\Pi_i(\omega) = \omega_i$ for each $i = 1, 2, \cdots$.

Theorem 1.2. Let $(\Omega_j, \mathcal{F}_j, P_j)$, $j = 1, 2, \cdots$ be an arbitrary sequence of probability spaces. If $\Omega' = \prod_{j=1}^\infty \Omega_j$, $\mathcal{F}' = \prod_{j=1}^\infty \mathcal{F}_j$, and $P' = \prod_{j=1}^\infty P_j$, then the canonical projections $\Pi_i(\omega) = \omega_i$, $i = 1, 2, \cdots$ are a sequence of independent random variables with the distribution $P'(\Pi_i \in A_i) = P_i(A_i)$ for each $A_i \in \mathcal{F}_i$.

Proof: See Theorem 5.11.1 of Ash (1972). ■

The above theorem implies that $\{\Pi_1, \Pi_2, \cdots\}$ is a sequence of independent and identically distributed random variables. Moreover, the random variables $X_i = X \circ \Pi_i$; $i = 1, \cdots, n$, for all n, are independent, identically distributed (iid) as X with the common distribution induced by P, P_X say. Therefore, for each $\omega = (\omega_1, \omega_2, \cdots) \in \Omega^\infty$, the random variables $X_i(\omega) = X(\omega_i)$ are iid from the distribution P_X.

It should be mentioned that the above construction of the X_i on Ω^∞ is only needed when we are dealing with limit theory. Dauxois, Pousse and Romain (1982) showed that as $n \to \infty$, the random variables $K_n(\omega) = \frac{1}{n}\sum_{i=1}^n X_i(\omega) \otimes X_i(\omega)$, defined from $(\Omega^\infty, \mathcal{F}^\infty, P^\infty)$ into $(F, \mathcal{B}_F)$, converge almost surely to K in F.

To analyse a sample of size n of trajectories $X(\omega_i)$, however, we need to consider the subspace Ω^n of Ω^∞ whose elements are in the form of finite sequences $(\omega_1, \cdots, \omega_n)$. For the time being, we restrict X on Ω^n. So, we ignore the index n in $K_n(\omega)$, and denote it by

$$\widehat{K}(\omega) = \frac{1}{n}\sum_{i=1}^n X_i(\omega) \otimes X_i(\omega),$$

as the empirical estimator of K.

The above arguments can be easily changed for the situations when $\eta(t) \neq 0$. All we need is to replace $X(\omega, t) - \eta(t)$ instead of $X(\omega, t)$ in the above formulas, by which we obtain

$$K(u,v) = E\big[\{X(u) - \eta(u)\}\,\{X(v) - \eta(v)\}\big]\,, \tag{1.6}$$

$$\widehat{K}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(u) - \bar{X}(u)\}\,\{X_i(v) - \bar{X}(v)\}\,, \tag{1.7}$$

where $\bar{X} = n^{-1}\sum_i X_i$. Also, for simplicity in notation we ignored the component ω of X in the equations (1.6) and (1.7) as the randomness of the function can be reflected by X itself. Furthermore, without loss of generality, we take $\mathcal{I} = [0,1]$. Also, the bounded measure $\mu(dt)$ is replaced by Lebesge measure on the interval $[0,1]$, and random functions are regarded as real-valued. In this case, we denote the Hilbert space E by $L_2(\mathcal{I})$. Analogously to (1.1) and (1.3) we can define the non-negative, self-adjoint Hilbert-Schmidt operator $\widehat{K}$ on $L_2(\mathcal{I})$ by $(\widehat{K}g)(u) = \int_{\mathcal{I}} \widehat{K}(u,v)\, g(v)\, dv$ for each $g \in L_2(\mathcal{I})$, where

$$\widehat{K}(u,v) = \sum_{j=1}^{\infty} \hat{\theta}_j\, \widehat{\psi}_j(u)\, \widehat{\psi}_j(v)\,. \tag{1.8}$$

In (1.8) the random variables $\hat{\theta}_1 \geq \hat{\theta}_2 \geq \cdots \geq 0$ are eigenvalues of the operator $\widehat{K}$, and $\widehat{\psi}_1, \widehat{\psi}_2, \ldots$ is the corresponding sequence of eigenvectors.

1.2.3 Karhunen-Loève expansion

An expansion of the function $X - \eta$ with respect to the orthonormal basis ψ_j (in $L_2(\mathcal{I})$ sense) is its Karhunen-Loève expansion:

$$X(u) - \eta(u) = \sum_{j=1}^{\infty} \xi_j\, \psi_j(u)\,, \tag{1.9}$$

where the principal component scores $\xi_1, \xi_2, \ldots$ are given by $\xi_j = \int_{\mathcal{I}}(X - \eta)\,\psi_j$. As regards the kernel $K(u,v)$, it follows that

$$E(\xi_j\,\xi_k) = \int_{\mathcal{I}}\int_{\mathcal{I}} \psi_j(u)\,K(u,v)\,\psi_k(v)\,du\,dv = \theta_j\,\delta_{jk}\,, \tag{1.10}$$

where δ_{jk} is the Kronecker delta (recall that the θ_j are eigenvalues and ψ_j are the corresponding orthonornal eigenfunctions of the operator K). Equation (1.10) implies that the random variables ξ_j are uncorrelated. Furthermore, they have zero means and variance $\theta_j = E(\xi_j^2)$. Moreover, $\int_{\mathcal{I}} E(X - \eta)^2 = \sum_{j\geq 1}\theta_j < \infty$. We call the expansion (1.9) the Karhunen-Loève expansion of $X - \eta$. It is also known that if the kernel $K(u,v)$ is a continuous function on $\mathcal{I}\times\mathcal{I}$, then the series on the right-hand side of (1.9) converges uniformly to $X(u) - \eta(u)$ (Theorem 1.5 of Bosq 2000). However, we do not need this restriction on $K(u,v)$, as long as we work with L_2 convergence of the series to $X(u) - \eta(u)$, satisfied by the condition $\int_{\mathcal{I}} E[(X - \eta)^2] < \infty$.

Also, in regard to $\widehat{K}(u,v)$ as the standard empirical approximation to $K(u,v)$, we write

$$X_i - \bar{X} = \sum_{j=1}^{\infty} \widehat{\xi}_{ij}\,\widehat{\psi}_j\,,$$

where $\widehat{\xi}_{ij} = \int_{\mathcal{I}}(X_i - \bar{X})\,\widehat{\psi}_j$ is the jth empirical principal component score of X_i. In analogy to (1.10),

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{\xi}_{ij}\,\widehat{\xi}_{ik} = \int_{\mathcal{I}}\int_{\mathcal{I}} \widehat{\psi}_j(u)\,\widehat{K}(u,v)\,\widehat{\psi}_k(v)\,du\,dv = \hat{\theta}_j\,\delta_{jk}\,.$$

1.2.4 Stochastic Expansions for Eigenvalues and Eigenfunctions

In this Section we give expansions for eigenvalues and eigenfunctions which explicitly include terms of sizes $n^{-1/2}$ and n^{-1}, where n denotes sample size, and a remainder of order $n^{-3/2}$. This work shows that eigenvalue spacings have only a second-order effect on properties of eigenvalue estimators, but a first-order effect on properties of eigenfunction estimators. Our expansions are immediately valid for any finite number of principal components, but they are also available uniformly in increasingly many components; the issue of uniformity is addressed in the next Chapter.

Assume that the eigenvalues θ_j are all distinct. The case where there is only a finite number of ties among the θ_j's can be treated without much difficulty, but other settings are more awkward. Distinctness of eigenvalues implies that the operator K is strictly positive definite, i.e. each $\theta_j > 0$. Also, since ψ_j and $-\psi_j$ are both eigenfunctions associated with θ_j, the statistical parameter ψ_j is not well defined. To overcome problems arising from the fact that ψ_j and $\widehat{\psi}_j$ are defined only up to a sign change, and to ensure that $\widehat{\psi}_j$ is viewed as an estimator of ψ_j rather than of $-\psi_j$, we shall tacitly assume, below, that the sign of $\widehat{\psi}_j$ is chosen so that $\int_{\mathcal{I}} \psi_j \, \widehat{\psi}_j \geq 0$.

Lemma 1.1. Let $\widehat{\psi}_j = \sum_{k \geq 1} a_{jk} \, \psi_k$, where the generalised Fourier coefficients a_{jk} are functionals of the data $\mathcal{X}$. Then,

$$a_{jj} = 1 - \tfrac{1}{2} n^{-1} \sum_{\ell : \ell \neq j} (\theta_j - \theta_\ell)^{-2} \left(\int Z \, \psi_j \, \psi_\ell \right)^2 + O_p(n^{-3/2}) \,, \tag{1.11}$$

and for each $k \neq j$,

$$
\begin{aligned}
a_{jk} &= n^{-1/2}\,(\theta_j - \theta_k)^{-1} \int Z\,\psi_j\,\psi_k \\
&\quad + n^{-1}\Bigg\{(\theta_j - \theta_k)^{-1} \sum_{\ell\,:\,\ell \neq j} (\theta_j - \theta_\ell)^{-1} \left(\int Z\,\psi_j\,\psi_\ell\right)\left(\int Z\,\psi_k\,\psi_\ell\right) \\
&\quad - (\theta_j - \theta_k)^{-2}\left(\int Z\,\psi_j\,\psi_j\right)\left(\int Z\,\psi_j\,\psi_k\right)\Bigg\} + O_p(n^{-3/2})\,, \qquad (1.12)
\end{aligned}
$$

where $Z = n^{1/2}\,(\widehat{K} - K)$ and $\int Z\,\psi_r\,\psi_s$ denotes $\int\int_{\mathcal{I}^2} Z(u,v)\,\psi_r(u)\,\psi_s(v)\,du\,dv$.

Proof of Lemma: It should be mentioned that the current proof is not rigorous. We shall give a rigorous proof of it later in Chapter 2.

The eigenfunctions $\widehat{\psi}_1, \widehat{\psi}_2, \cdots$ are defined recursively. For example, supposing we have defined $\widehat{\psi}_1, \cdots, \widehat{\psi}_{j-1}$, we define $\widehat{\psi}_j$ so as to be orthogonal to these previous functions, to have unit length, to be such that $\int \widehat{\psi}_j\,\psi_j \geq 0$, and to maximise $\int \widehat{K}\,\phi\,\phi$ over all functions ϕ that satisfy these constraints. That is, $\widehat{\psi}_j$ is obtained by maximizing

$$
\begin{aligned}
&\int\int_{\mathcal{I}^2} \widehat{K}(u,v)\,\phi_j(u)\,\phi_j(v)\,du\,dv \\
&\qquad\qquad \text{subject to}\,, \qquad (1.13) \\
&\qquad\qquad \|\phi_j\|^2 = 1,\ \ \langle \phi_j, \widehat{\psi}_k \rangle = 0 \ \text{ for } k < j\,,
\end{aligned}
$$

where $\langle .\,,\,. \rangle$ and $\|.\|$ are the inner product and norm in $L_2(\mathcal{I})$, respectively. In fact the above problem is equivalent to maximization of the scale-invariant ratio:

$$
\frac{\int\int_{\mathcal{I}^2} \widehat{K}(u,v)\,\phi_j(u)\,\phi_j(v)\,du\,dv}{\|\phi_j\|^2}\,, \qquad (1.14)
$$

with respect to ϕ_j, and under the orthogonality constraint. Assume that

$$\phi_\ell = \sum_{j=1}^{\infty} a_{\ell j} \psi_j \,, \tag{1.15}$$

where $\psi_1, \psi_2, \ldots$ are the eigenfunctions of the covariance operator K, and make up an orthonormal basis for $L_2(\mathcal{I})$ if its eigenvalues θ_j are all positive. After substituting (1.15) into (1.13), the problem is reduced to seeking an extremum of the Lagrangian function

$$\begin{aligned} \sum_{j=1}^{\infty} a_{\ell j}^2 \,\theta_j + \lambda_0 \left(\sum_{j=1}^{\infty} a_{\ell j}^2 - 1 \right) + 2 \sum_{r=1}^{\ell} \sum_{s=1}^{r-1} \lambda_{rs} \sum_{j=1}^{\infty} a_{rj}\, a_{sj} \\ + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_{\ell j}\, a_{\ell k} \left(n^{-1/2} \int Z\, \psi_j\, \psi_k \right), \end{aligned} \tag{1.16}$$

where λ_0 and λ_{rs} are Lagrange multipliers. It is assumed that $a_{\ell j}, \lambda_0$, and λ_{rs} have the following expansions:

$$\begin{aligned} \lambda_0 &= \lambda_0^{(0)} + n^{-1/2}\, \lambda_0^{(1)} + n^{-1}\, \lambda_0^{(2)} + O_p(n^{-3/2}) \\ a_{rj} &= a_{rj}^{(0)} + n^{-1/2}\, a_{rj}^{(1)} + n^{-1}\, a_{rj}^{(2)} + O_p(n^{-3/2}) \\ \lambda_{rs} &= \lambda_{rs}^{(0)} + n^{-1/2}\, \lambda_{rs}^{(1)} + n^{-1}\, \lambda_{rs}^{(2)} + O_p(n^{-3/2})\,. \end{aligned} \tag{1.17}$$

Differentiating with respect to $a_{\ell j}$, and equating to zero for an extremum, we have

$$a_{\ell j}\,(\theta_j + \lambda_0) + \sum_{s=1}^{\ell-1} \lambda_{\ell s}\, a_{sj} + \sum_{k=1}^{\infty} a_{\ell k} \left(n^{-1/2} \int Z\, \psi_j\, \psi_k \right) = 0\,. \tag{1.18}$$

The constraint $\|\phi_\ell\| = 1$ implies

$$\sum_{j=1}^{\infty}(a_{\ell j}^{(0)})^2 + 2\,n^{-1/2}\sum_{j=1}^{\infty} a_{\ell j}^{(0)}\,a_{\ell j}^{(1)} + n^{-1}\sum_{j=1}^{\infty}\left(2\,a_{\ell j}^{(0)}\,a_{\ell j}^{(2)} + (a_{\ell j}^{(1)})^2\right) + O_p(n^{-3/2}) = 1\,, \tag{1.19}$$

and the orthogonality constraint gives us, for $s < \ell$,

$$\begin{aligned}\sum_{j=1}^{\infty} a_{\ell j}\,a_{sj} = \sum_{j=1}^{\infty} a_{\ell j}^{(0)}\,a_{sj}^{(0)} + n^{-1/2}\sum_{j=1}^{\infty}\left(a_{\ell j}^{(0)}\,a_{sj}^{(1)} + a_{\ell j}^{(1)}\,a_{sj}^{(0)}\right)\\ + n^{-1}\sum_{j=1}^{\infty}\left(a_{\ell j}^{(1)}\,a_{sj}^{(1)} + a_{\ell j}^{(0)}\,a_{sj}^{(2)} + a_{\ell j}^{(2)}\,a_{sj}^{(0)}\right) + O_p(n^{-3/2}) = 0\,.\end{aligned} \tag{1.20}$$

The two equations (1.19) and (1.20) give the following results:

$$\sum_{j=1}^{\infty}(a_{\ell j}^{(0)})^2 = 1, \quad \sum_{j=1}^{\infty} a_{\ell j}^{(0)}\,a_{sj}^{(0)} = 0\,, \tag{1.21}$$

$$a_{s\ell}^{(1)} + a_{\ell s}^{(1)} = 0, \quad \text{for } s < \ell \qquad a_{\ell\ell}^{(1)} = 0\,, \tag{1.22}$$

$$2\,a_{\ell\ell}^{(2)} = -\sum_{j=1}^{\infty}\left(a_{\ell j}^{(1)}\right)^2, \qquad a_{s\ell}^{(2)} + a_{\ell s}^{(2)} = -\sum_{j=1}^{\infty} a_{\ell j}^{(1)}\,a_{sj}^{(1)} \tag{1.23}$$

Result (1.21) points to $a_{\ell j}^{(0)} = \delta_{\ell j}$. Also, from (1.18) we obtain

$$\delta_{\ell j}(\theta_j + \lambda_0^{(0)}) + \sum_{k=1}^{\ell-1} \lambda_{\ell k}^{(0)}\,\delta_{kj} = 0\,, \tag{1.24}$$

which implies $\lambda_0^{(0)} = -\theta_\ell$, and $\lambda_{\ell j}^{(0)} = 0$, for all $j < \ell$. Substituting from (1.17) into (1.18) leads us to

$$n^{-1/2}\left\{a_{\ell j}^{(1)}\,(\theta_j-\theta_\ell)+\lambda_0^{(1)}\,\delta_{\ell j}+\sum_{k=1}^{\infty}\delta_{\ell k}\Big(\int Z\,\psi_j\,\psi_k\Big)+\sum_{s=1}^{\ell-1}\lambda_{\ell s}^{(1)}\,\delta_{sj}\right\}$$
$$+\,n^{-1}\left\{\delta_{\ell j}\,\lambda_0^{(2)}+a_{\ell j}^{(1)}\,\lambda_0^{(1)}+a_{\ell j}^{(2)}\,(\theta_j-\theta_\ell)+\sum_{k=1}^{\infty}a_{\ell k}^{(1)}\Big(\int Z\psi_j\,\psi_k\Big)+\sum_{s=1}^{\ell-1}\lambda_{\ell s}^{(1)}\,a_{sj}^{(1)}\right.$$
$$\left.+\sum_{s=1}^{\ell-1}\delta_{sj}\,\lambda_{\ell s}^{(2)}\right\}+O_p(n^{-3/2})\,. \tag{1.25}$$

The above equation shows that

$$a_{\ell j}^{(1)}\,(\theta_j-\theta_\ell)+\lambda_0^{(1)}\,\delta_{\ell j}+\sum_{k=1}^{\infty}\delta_{\ell k}\Big(\int Z\,\psi_j\,\psi_k\Big)+\sum_{s=1}^{\ell-1}\lambda_{\ell s}^{(1)}\,\delta_{sj}=0\,,$$
$$\delta_{\ell j}\,\lambda_0^{(2)}+a_{\ell j}^{(1)}\,\lambda_0^{(1)}+a_{\ell j}^{(2)}\,(\theta_j-\theta_\ell)+\sum_{k=1}^{\infty}a_{\ell k}^{(1)}\Big(\int Z\psi_j\,\psi_k\Big)+\sum_{s=1}^{\ell-1}\lambda_{\ell s}^{(1)}\,a_{sj}^{(1)}+\sum_{s=1}^{\ell-1}\delta_{sj}\,\lambda_{\ell s}^{(2)}=0\,, \tag{1.26}$$

where the first equation, when $j=\ell$, implies

$$\lambda_0^{(1)}=-\int Z\,\psi_\ell\,\psi_\ell\,, \tag{1.27}$$

and, if $j\neq\ell$, $a_{\ell j}^{(1)}\,(\theta_j-\theta_\ell)+\int Z\,\psi_j\,\psi_\ell+\lambda_{\ell j}^{(1)}I(j\leq\ell-1)=0$, implying, by the second part of (1.22),

$$a_{\ell j}^{(1)}=(\theta_j-\theta_\ell)^{-1}\left[\int Z\,\psi_j\,\psi_\ell+\lambda_{\ell j}^{(1)}I(j\leq\ell-1)\right].$$

Using the first part of (1.22) with this result gives $\lambda_{\ell s}^{(1)}=0$, for each $s<\ell$. Therefore,

$$a_{\ell j}^{(1)}=(\theta_\ell-\theta_j)^{-1}\Big(\int Z\,\psi_j\,\psi_\ell\Big);\quad\text{for } j\neq\ell. \tag{1.28}$$

When $j \neq \ell$, the second equation in (1.26) implies

$$\begin{aligned} a_{\ell j}^{(2)} &= (\theta_j - \theta_\ell)^{-2}\Big(\int Z\,\psi_j\,\psi_\ell\Big)\Big(\int Z\,\psi_\ell\,\psi_\ell\Big) \\ &\quad - (\theta_j - \theta_\ell)^{-1} \sum_{k\,:\,k\neq\ell} (\theta_k - \theta_\ell)^{-1}\Big(\int Z\,\psi_k\,\psi_\ell\Big)\Big(\int Z\,\psi_j\,\psi_k\Big) \\ &\quad + (\theta_\ell - \theta_j)^{-1}\lambda_{\ell j}^{(2)}\, I(j < \ell)\,. \end{aligned} \tag{1.29}$$

Results (1.28), second part of (1.22), and first part of (1.23) lead us to:

$$a_{\ell\ell}^{(2)} = -\tfrac{1}{2} \textstyle\sum_{j\,:\,j\neq\ell} (\theta_j - \theta_\ell)^{-2}\Big(\int Z\,\psi_j\,\psi_\ell\Big)^2 . \tag{1.30}$$

Also, from the second part of (1.23), (1.26), and (1.29) we have

$$\begin{aligned} a_{\ell s}^{(2)} + a_{s\ell}^{(2)} &= \sum_{k\neq s,\,\ell} (\theta_k - \theta_\ell)^{-1}\,(\theta_k - \theta_s)^{-1}\Big(\int Z\,\psi_k\,\psi_s\Big)\Big(\int Z\,\psi_\ell\,\psi_k\Big) \\ &\qquad + (\theta_\ell - \theta_s)^{-1}\,\lambda_{\ell s}^{(2)} \\ &= -\sum_{j=1}^{\infty} a_{\ell j}^{(1)}\,a_{sj}^{(1)} \\ &= \sum_{k\neq s,\,\ell} (\theta_k - \theta_\ell)^{-1}\,(\theta_k - \theta_s)^{-1}\Big(\int Z\,\psi_k\,\psi_s\Big)\Big(\int Z\,\psi_\ell\,\psi_k\Big), \end{aligned}$$

which gives $\lambda_{\ell s}^{(2)} = 0$, for each $s < \ell$. Consequently,

$$\begin{aligned} a_{\ell j}^{(2)} &= (\theta_j - \theta_\ell)^{-2}\Big(\int Z\,\psi_j\,\psi_\ell\Big)\Big(\int Z\,\psi_\ell\,\psi_\ell\Big) \\ &\quad - (\theta_j - \theta_\ell)^{-1} \sum_{k\,:\,k\neq\ell} (\theta_k - \theta_\ell)^{-1}\Big(\int Z\,\psi_k\,\psi_\ell\Big)\Big(\int Z\,\psi_j\,\psi_k\Big); \quad \text{for } j \neq \ell\,. \end{aligned} \tag{1.31}$$

Combining second part of (1.22), (1.28), (1.30) and (1.31) finishes the proof of the Lemma.■

Results (1.11) and (1.12) point to the following expansion:

$$
\begin{aligned}
&\widehat{\psi}_j(t) - \psi_j(t) \\
&= n^{-1/2} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k(t) \int Z\, \psi_j\, \psi_k - \tfrac{1}{2}\, n^{-1}\, \psi_j(t) \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \left(\int Z\, \psi_j\, \psi_k \right)^2 \\
&\quad + n^{-1} \sum_{k\,:\,k\neq j} \psi_k(t) \left\{ (\theta_j - \theta_k)^{-1} \sum_{\ell\,:\,\ell\neq j} (\theta_j - \theta_\ell)^{-1} \left(\int Z\, \psi_j\, \psi_\ell \right) \left(\int Z\, \psi_k\, \psi_\ell \right) \right. \\
&\quad \left. - (\theta_j - \theta_k)^{-2} \left(\int Z\, \psi_j\, \psi_j \right) \left(\int Z\, \psi_j\, \psi_k \right) \right\} + O_p\big(n^{-3/2}\big)\,.
\end{aligned}
\tag{1.32}
$$

Analogously to (1.11) and (1.12), it can be shown that

$$
\hat{\theta}_j - \theta_j = n^{-1/2} \int Z\, \psi_j\, \psi_j + n^{-1} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \left(\int Z\, \psi_j\, \psi_k \right)^2 + O_p\big(n^{-3/2}\big)\,.
\tag{1.33}
$$

Of course, there are analogues of (1.11) and (1.12) with remainders $O_p(n^{-r/2})$ for any positive integer r. We have taken $r = 3$ only for brevity and simplicity.

1.2.5 Properties of Stochastic Expansions

We develop theory based on stochastic expansions of eigenvalue and eigenvector estimators, providing not only a new understanding of the effects of truncating to a finite number of principal components, but also pointing to new methodology. We show how stochastic, bootstrapped versions of these ideas can be used to construct simultaneous confidence regions for literally all eigenvalue estimates, and for increasing numbers of eigenfunction estimates. The developed theory makes it possible to justify bootstrap methods of that type, in terms of asymptotic theory. In particular, we are able to obtain coverage accuracy of bootstrap confidence regions. These results have appeared in Chapter 3. The theory also provides new insight into more conventional FDA methods, including those which are used for linear regression. These results are used to explore the validity of simple accounts of the performance of functional linear regression. We shall discuss it in Chapter 4 and the Appendix.

In the next Chapter we shall discuss conditions under which the stochastic expansions hold, and in particular under which the infinite series in (1.11), (1.12), (1.32) and (1.33) converge. However, it should be mentioned that the relationship between $L_2(\mathcal{I})$, the space of all square-integrable functions from $\mathcal{I}$ to the real line, and ℓ_2, the space of all sequences $\{a_j\}$ such that $\sum_{j=1}^{\infty} |a_j|^2 < \infty$, may be used to illustrate a method for proving convergency of the infinite series in the expansions. For example, using this method, below we show that in (1.12), $S_{1j} \equiv \sum_{\ell:\, \ell \neq j} (\theta_j - \theta_\ell)^{-1}\, b_{j\ell}\, b_{k\ell} < \infty$, where $b_{rs} = \int Z\, \psi_r\, \psi_s$. For this purpose, we first prove that $S_{2j} \equiv \sum_{\ell:\, \ell \neq j} \{(\theta_j - \theta_\ell)^{-1} - \theta_j^{-1}\}\, b_{j\ell}\, b_{k\ell} < \infty$. To do that, we show that the series of absolute values, S_{2j}^* say, converges:

$$\theta_j\, S_{2j}^* = \sum_{\ell:\, \ell \neq j} \frac{\theta_\ell}{|\theta_j - \theta_\ell|}\, |b_{j\ell}\, b_{k\ell}| \leq \sum_{\ell:\, \ell \neq j} \frac{\theta_\ell}{\min_{\ell:\, \ell \neq j} |\theta_j - \theta_\ell|}\, |b_{j\ell}|\, |b_{k\ell}|$$

$$= \frac{1}{\min_{\ell:\, \ell \neq j} |\theta_j - \theta_\ell|} \sum_{\ell:\, \ell \neq j} \theta_\ell\, |b_{j\ell}\, b_{k\ell}|\,.$$

Hence, denoting $|\mathcal{I}|$ as the length of $\mathcal{I}$ and $\|Z\|_\infty$ as $\sup_{u,v} |Z(u,v)|$, and using the fact that $|\iint_{\mathcal{I}^2} Z(u,v)\, \psi_r(u)\, \psi_s(v)\, du\, dv| \leq \|Z\|_\infty\, |\mathcal{I}|$, we deduced

$$\theta_j \big(\min_{\ell:\, \ell \neq j} |\theta_j - \theta_\ell| \big)\, S_{2j}^* \leq \|Z\|_\infty^2\, |\mathcal{I}|^2 \sum_{\ell=1}^{\infty} \theta_\ell < \infty\,.$$

Thus, the series S_2 converges absolutely. Moreover,

$$S_{1j} = \sum_{\ell:\,\ell\neq j} (\theta_j - \theta_\ell)^{-1}\, b_{j\ell}\, b_{k\ell} = \sum_{\ell:\,\ell\neq j} \{(\theta_j - \theta_\ell)^{-1} - \theta_j^{-1}\}\, b_{j\ell}\, b_{k\ell} + \theta_j^{-1} \sum_{\ell:\,\ell\neq j} b_{j\ell}\, b_{k\ell}\,. \tag{1.34}$$

We have already proved that the first series on the right-hand side of (1.34) converges, so it suffices to prove that $S_{3j} \equiv \sum_{\ell:\,\ell\neq j} b_{j\ell}\, b_{k\ell}$ converges.

Below we will prove that S_{3j} is equivalent to the integral

$$\int\!\!\int\!\!\int_{\mathcal{I}^3} Z(u,v_1)\, Z(u,v_2)\, \psi_j(v_1)\, \psi_k(v_2)\, du\, dv_1\, dv_2.$$

Define $(A\, f)(u) = \int Z(u,v)\, f(v)\, dv$. We have $A: L_2(\mathcal{I}) \longrightarrow L_2(\mathcal{I})$. Furthermore,

$$\begin{aligned}(A^2\, f)(w) = \int\!\!\int Z(w,u)\, Z(u,v)\, f(v)\, du\, dv &= \int \left\{ \int Z(w,u)\, Z(u,v)\, du \right\} f(v)\, dv \\ &= \int Z_2(w,v)\, f(v)\, dv\,,\end{aligned}$$

where $Z_2(w,v) = \int Z(w,u)\, Z(u,v)\, du$ is called the kernel of the operator. We know that $B: \{\theta_k\} \in \ell_2 \longrightarrow \{\mu_j\} \in \ell_2$ such that $\sum_{k\geq 1} B_{jk}\, \theta_k = \mu_j$. Moreover, define $J: f \in L_2(\mathcal{I}) \longrightarrow \int f(u)\, \psi_j(u)\, du \in \ell_2$. This map is an isometric isomorphism (a one-to-one-onto, linear, norm-preserving map) between $L_2(\mathcal{I})$ and ℓ_2 (Theorem 3.2.15 of Ash 1972). These arguments lead us to

$$\begin{array}{ccc} L_2(\mathcal{I}) & \xrightarrow{A} & L_2(\mathcal{I}) \\ {\scriptstyle J}\downarrow & & \downarrow{\scriptstyle J} \\ \ell_2 & \xrightarrow{B} & \ell_2 \end{array}$$

We also have

$$\{\theta_k\} \in \ell_2 \xrightarrow{J^{-1}} \Big(\sum_{k\geq 1} \theta_k \, \psi_k(v)\Big) \in L_2(\mathcal{I}) \xrightarrow{A} \Big(\int Z(u,v) \sum_{k\geq 1} \theta_k \, \psi_k(v)\, dv\Big) \in L_2(\mathcal{I}) \,,$$

and

$$\begin{aligned}\sum_{k\geq 1} \theta_k \int Z(u,v)\, \psi_k(v)\, dv \in L_2(\mathcal{I}) \xrightarrow{J} & \sum_{k\geq 1} \theta_k \int\int Z(u,v)\, \psi_k(v)\, \psi_j(u)\, du\, dv \\ &= \sum_{k\geq 1} \theta_k \, B_{jk} = \mu_j \,, \qquad (1.35)\end{aligned}$$

where $B_{jk} = \int\int Z(u,v)\, \psi_k(v)\, \psi_j(u)\, du\, dv$. Thus,

$$B = J \circ A \circ J^{-1} \text{ and } \quad B^2 = J \circ A \circ J^{-1} \circ J \circ A \circ J^{-1} = J \circ A^2 \circ J^{-1}$$

In other words, we have $B^2 : \ell_2 \longrightarrow \ell_2$, where

$$\begin{aligned} B^2(\{\theta_k\}) &= J \circ A^2 \Big(\sum_{k\geq 1} \theta_k \, \psi_k(v)\Big) = J\Big(\sum_{k\geq 1} \theta_k \int Z_2(w,v)\, \psi_k(v)\, dv\Big) \\ &= \sum_{k\geq 1} \theta_k \int\int Z_2(w,v)\, \psi_k(v)\, \psi_j(w)\, dv\, dw \\ &= \sum_{k\geq 1} \theta_k \int\int\int_{\mathcal{I}^3} Z(w,u)\, Z(u,v)\, \psi_k(v)\, \psi_j(w)\, du\, dv\, dw \,. \qquad (1.36)\end{aligned}$$

Comparing B_{jk} in (1.35) with (1.36), we can write B^2_{jk} in analogy with B_{jk} as follows:

$$B^2_{jk} = \sum_{\ell\geq 1} b_{j\ell}\, b_{k\ell} = \int\int\int_{\mathcal{I}^3} Z(w,u)\, Z(u,v)\, \psi_k(v)\, \psi_j(w)\, du\, dv\, dw \,.$$

Because the integral of the absolute value of the integrand is bounded by $\|Z\|^2_\infty \, |\mathcal{I}|^2$, S_3 converges.

It can be shown that the expected value of the term in $n^{-1/2}$ on the right-hand side of (1.12), is zero. It then follows from (1.11) and (1.12) that, for all $k \geq 1$, $a_{jk} = \delta_{jk} + O_p(n^{-1/2})$ and $E(a_{jk}) = \delta_{jk} + O(n^{-1})$, where δ_{jk} is the Kronecker delta (see Lemma 4.1). For a general discussion about the stochastic expansions and their properties see Chapter 2.

1.3 Weak Convergence Results

In this Section, we give asymptotic results for eigenvalues and eigenfunctions. In order to study asymptotic behavior of our estimators, we need first the asymptotic distribution of the bivariate random process $Z(u, v)$.

Lemma 1.2. If the random process $X(t)$ satisfies the following condition:

for all $C > 0$ and some $\epsilon > 0$,

$$\sup_{t \in \mathcal{I}} E\big|X(t)\big|^C < \infty, \quad \sup_{s,t \in \mathcal{I}} E\Big[\big\{|s-t|^{-\epsilon}\,|X(s) - X(t)|\big\}^C\Big] < \infty, \quad (1.37)$$

then $Z(u, v) = n^{1/2}\,(\widehat{K} - K)(u, v) \longrightarrow \zeta(u, v)$ in distribution, where ζ is a Gaussian process.

Proof of Lemma: We have

$$\begin{aligned} Z(u, v) &= n^{1/2}\,(\widehat{K} - K)(u, v) \\ &= n^{1/2}\left[\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(u)\,Y_i(v) - K(u, v)\right) - \{\bar{X}(u) - \eta(u)\}\,\{\bar{X}(v) - \eta(v)\}\right] \\ &= n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n} W_i(u, v) - \{\bar{X}(u) - \eta(u)\}\,\{\bar{X}(v) - \eta(v)\}\right], \qquad (1.38) \end{aligned}$$

where $\eta(u) = E\{X(u)\}$, $Y_i = X_i - \eta$ and $W_i(u, v) = Y_i(u)\,Y_i(v) - K(u, v)$. The first term on the right-hand side of (1.38) is the sample mean of the n independent terms. Furthermore, by using the fact that $\bar{X}(v) - \eta(v) = O_p(n^{-1/2})$ (Theorem

3.2.5 of Sen and Singer (1993) with tightness of $X(v)$ which is discussed later), we can write $Z(u,v) = Z_n(u,v) + O_p(n^{-1/2})$, where $Z_n(u,v) = n^{-1/2} \sum_{i=1}^{n} W_i(u,v)$.

Theorem 1.3. Let P_n and P be probability measures on $(C_{[0,1]}, \mathcal{F})$, where $C_{[0,1]}$ is the space of continuous functions on $[0,1]$ with the uniform metric $\rho(x,y) = \sup_t |x(t) - y(t)|$, for each $x, y \in C_{[0,1]}$, and $\mathcal{F}$ is the σ-field constructed on $C_{[0,1]}$. If the finite-dimensional distributions of P_n converge weakly to those of P, and if $\{P_n\}$ is tight, then P_n converges weakly to P.

Proof: See Theorem 8.1 of Billingsley (1968). ■

Using the above theorem, if X_n are random elements of $C_{[0,1]}$, then $\{X_n\}$ is tight if $\{P_n\}$ is tight, where P_n is the distribution of X_n, as we identify the finite-dimensional distribution of X_n with those of P_n in the above theorem. Therefore, Theorem 1.3 is equivalent to the following argument.

If the finite-dimensional distributions of X_n converge weakly to those of X, and $\{X_n\}$ is tight, then $X_n \to X$ in distribution. Regarding the k points $(u_1, v_1), \cdots, (u_k, v_k)$, for each point if $\int_{\mathcal{I}} E(X^4) < \infty$, then the classical Central Limit Theorem with the Slutsky Theorem (Theorem 3.3.1 and 3.4.2 of Sen and Singer 1993) implies that, for each $j = 1, \cdots, k$,

$$Z_n(u_j, v_j) + o_p(1) \longrightarrow \zeta(u_j, v_j), \text{ in distribution}, \tag{1.39}$$

where $\zeta(u_j, v_j)$ is the weak limit of $Z(u_j, v_j)$, for each $j = 1, \cdots, k$. Now, for the k points $(u_1, v_1), \cdots, (u_k, v_k)$ in $[0,1] \times [0,1]$, let $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}$ be the mapping that carries the point h of $C_{[0,1]\times[0,1]}$ to the point $\big(h(u_1, v_1), \cdots, h(u_k, v_k)\big)$ of R^k. Since $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}$ is continuous, we have $\Pi_{(u_1,v_1),\cdots,(u_k,v_k)}(Z_n) \to \Pi_{(u_1,v_1),\cdots,(u_k,v_k)}(\zeta)$ in distribution (Corollary 1 of Theorem 5.1 of Billingsley, 1968), i.e.

$$\big(Z_n(u_1,v_1),\cdots,Z_n(u_k,v_k)\big)\rightarrow\big(\zeta(u_1,v_1),\cdots,\zeta(u_k,v_k)\big),\ \text{in distribution}. \tag{1.40}$$

(in particular, (1.40) follows via the Cramér-Wold device.)

Theorem 1.4. The sequence $\{X_n\}$ is tight if it satisfies these two conditions:

(i) The sequence $\{X_n(0)\}$ is tight.

(ii) There exist constants $\gamma \geq 0$ and $\alpha > 1$ and a nondecreasing, continuous function F on $[0,1]$ such that

$$P\{|X_n(t_2)-X_n(t_1)|\geq\lambda\}\leq\frac{1}{\lambda^\gamma}|F(t_2)-F(t_1)|^\alpha, \tag{1.41}$$

holds for all t_1, t_2 and n and all positive λ.

Proof: See Theorem 12.3 of Billingsley (1968). ■

We know that the moment condition

$$E\{|X_n(t_2)-X_n(t_1)|^\gamma\}\leq|F(t_2)-F(t_1)|^\alpha, \tag{1.42}$$

implies (1.41). Furthermore, we can immediately obtain tightness of $\{X_n\}$ from condition (1.37), (1.42) and Theorem 1.4. Also, for some $\gamma = 2k$, where $k \geq 1$ is an integer, using Rosenthal's inequality (Hall and Heyde 1980, p. 23) for fixed $u, v, s, t \in [0,1]$ results in

$$\begin{aligned}
E\big[\big|n^{-1/2}\sum_{i=1}^n W_i(u,v)-n^{-1/2}\sum_{i=1}^n W_i(s,t)\big|^\gamma\big] &= n^{-k}E\big|\sum_{i=1}^n\{W_i(u,v)-W_i(s,t)\}\big|^{2k}\\
&\leq C_{1\gamma}\,n^{-k}\Big\{\sum_{i=1}^n E\big|W_i(u,v)-W_i(s,t)\big|^{2k}+\Big(\sum_{i=1}^n E\big|W_i(u,v)-W_i(s,t)\big|^2\Big)^k\Big\}\\
&\leq C_{2\gamma}\,E\big|W(u,v)-W(s,t)\big|^\gamma,
\end{aligned} \tag{1.43}$$

and

$$\begin{aligned}
E\big|W(u,v)-W(s,t)\big|^{\gamma} &= E\big|\{Y(u)\,Y(v)-Y(s)\,Y(t)\}-\{K(u,v)-K(s,t)\}\big|^{\gamma} \\
&\le C_{3\gamma}\Big(E\big|Y(u)\,Y(v)-Y(s)\,Y(t)\big|^{\gamma}+E\big|Y(u)\,Y(v)-Y(s)\,Y(t)\big|^{\gamma}\Big) \\
&\le C_{4\gamma}\,E\big|Y(u)\,Y(v)-Y(s)\,Y(t)\big|^{\gamma}, \qquad (1.44)
\end{aligned}$$

where $C_{1\gamma}$, $C_{2\gamma}$, $C_{3\gamma}$ and $C_{4\gamma}$ are constants depending only on γ, and W and Y denote a generic W_i and Y_i, respectively. Also,

$$\begin{aligned}
E\Big[\big|Y(u)\,Y(v)-Y(s)\,Y(t)\big|^{\gamma}\Big] &= E\Big[\big|Y(u)\big(Y(v)-Y(t)\big)+Y(t)\big(Y(u)-Y(s)\big|^{\gamma}\Big] \\
&\le C_{\gamma}\Big\{E\big[|Y(u)|^{\gamma}\,|Y(v)-Y(t)|^{\gamma}\big]+E\big[|Y(t)|^{\gamma}\,|Y(u)-Y(s)|^{\gamma}\big]\Big\} \\
&\le C_{\gamma}\Big\{\Big(E\big[|Y(u)|^{2\gamma}\big]\Big)^{1/2}\Big(E\big[|Y(v)-Y(t)|^{2\gamma}\big]\Big)^{1/2} \\
&\qquad +\Big(E\big[|Y(t)|^{2\gamma}\big]\Big)^{1/2}\Big(E\big[|Y(u)-Y(s)|^{2\gamma}\big]\Big)^{1/2}\Big\}, \\
&\qquad\qquad (1.45)
\end{aligned}$$

where C_γ is a constant depending only on γ. If condition (1.37) holds, then (1.43)–(1.45) imply that for each two points $(u,v),(s,t)\in[0,1]\times[0,1]$,

$$E\Big[\Big|n^{-1/2}\sum_{i=1}^{n}W_i(u,v)-n^{-1/2}\sum_{i=1}^{n}W_i(s,t)\Big|^{\gamma}\Big]\le C_{\gamma}\left\{|v-t|^{\alpha}+|u-s|^{\alpha}\right\}, \quad (1.46)$$

where γ can be chosen such that $\alpha=\epsilon\gamma>2$. Hence, using Markov's inequality, for each $\lambda>0$, each two points $(u,v),(s,t)\in\mathcal{I}\times\mathcal{I}$ and all n we have

$$P_n\Big(\Big|n^{-1/2}\sum_{i=1}^{n}W_i(u,v)-n^{-1/2}\sum_{i=1}^{n}W_i(s,t)\Big|>\lambda\Big)\le C_{\gamma}\lambda^{-\gamma}\left\{|v-t|^{\alpha}+|u-s|^{\alpha}\right\}. \quad (1.47)$$

The proof of the above theorem, given by Billingsley (1968), with condition

(1.47) instead of (1.41) may be followed to show that $n^{-1/2}\sum_{i=1}^n W_i(u,v)$ is tight. To appreciate why, fix n, δ, j and k, then for a positive integer m consider the random variables

$$\kappa_\ell = n^{-1/2}\sum_{i=1}^n \left(W_i(j\,\delta+\frac{\ell}{m}\,\delta, k\,\delta+\frac{\ell}{m}\,\delta) - W_i(j\,\delta+\frac{(\ell-1)}{m}\,\delta, k\,\delta+\frac{(\ell-1)}{m}\,\delta)\right),$$

for $\ell = 1,\cdots,m$. The random variables κ_ℓ with $S_k = \sum_{\ell=1}^k \kappa_\ell$ satisfy

$$\begin{aligned} S_s - S_r &= \sum_{r<\ell\le s} \kappa_\ell \\ &= n^{-1/2}\sum_{i=1}^n \left(W_i(j\,\delta+\frac{s}{m}\,\delta, k\,\delta+\frac{s}{m}\,\delta) - W_i(j\,\delta+\frac{r}{m}\,\delta, k\,\delta+\frac{r}{m}\,\delta)\right). \end{aligned}$$

So, for $0\le r\le s\le m$ and $\lambda>0$ we have

$$\begin{aligned} P\big(|S_s - S_r|\ge\lambda\big) &= P\big(|\sum_{r<\ell\le s}\kappa_\ell|\ge\lambda\big) \\ &\le C_\gamma\,\lambda^{-\gamma}\left[(\frac{(s-r)\,\delta}{m})^\alpha + (\frac{(s-r)\,\delta}{m})^\alpha\right] \\ &\le C_\gamma\,\lambda^{-\gamma}\,[(\sum_{r<\ell\le s}\delta\,m^{-1})^\alpha + (\sum_{r<\ell\le s}\delta\,m^{-1})^\alpha] \\ &\le C_\gamma\,\lambda^{-\gamma}\,\delta^\alpha, \end{aligned}$$

where we have used (1.47) to obtain the first inequality above. By using Theorem 12.2 of Billingsley (1968), we have

$$P\left(\max_{0\le\ell\le m}\left|n^{-1/2}\sum_{i=1}^n\left(W_i(j\,\delta+\frac{\ell}{m}\,\delta, k\,\delta+\frac{\ell}{m}\,\delta) - W_i(j\,\delta, k\,\delta)\right)\right| > \epsilon\right) \le \frac{B}{\epsilon^\gamma}\,\delta^\alpha,$$

where B depends on γ and α ($B = B_{\gamma,\alpha}$). Since the W_i for each $1\le i\le n$ are continuous functions, if $m \to \infty$ we have

$$P\Big(\sup_{\|(u,v)-(j\,\delta,k\,\delta)\|_E<\delta} \Big| n^{-1/2} \sum_{i=1}^{n} \Big(W_i(u,v) - W_i(j\,\delta,k\,\delta) \Big) \Big| > \epsilon \Big) \le \frac{B}{\epsilon^\gamma}\,\delta^\alpha . \tag{1.48}$$

where $\|.\|_E$ denotes the Euclidian norm in R^2. If δ^{-1} is integer, the above inequality leads to

$$\sum_{j<\delta^{-1}} \sum_{k<\delta^{-1}} P\Big(\sup_{\|(u,v)-(j\,\delta,k\,\delta)\|_E<\delta} \Big| n^{-1/2} \sum_{i=1}^{n} \Big(W_i(u,v) - W_i(j\,\delta,k\,\delta) \Big) \Big| > \epsilon \Big) \le \frac{B}{\epsilon^\gamma}\,\delta^{\alpha-2}. \tag{1.49}$$

Define the modulus of continuity of an element x of $C_{[0,1]\times[0,1]}$ by

$$w_x^{(2)}(\delta) = \sup_{\|(u,v)-(s,t)\|_E<\delta} |x(u,v) - x(s,t)|,$$

where $0 < \delta \le 1$. Let

$$A_{s,t} = \left\{ n^{-1/2} \sum_{i=1}^{n} W_i : \sup_{\|(u,v)-(s,t)\|_E<\delta} \Big| n^{-1/2} \sum_{i=1}^{n} \Big(W_i(u,v) - W_i(s,t) \Big) \Big| \ge \epsilon \right\}.$$

If we want to lie (u,v) and (s,t) each in rectangles of the form $[j\,\delta,\,(j+1)\,\delta] \times [k\,\delta,\,(k+1)\,\delta]$, then if $\|(u,v)-(s,t)\|_E < \delta$, these rectangles either coincide or abut. Therefore,

$$\begin{aligned} P\Big(n^{-1/2} \sum_{i=1}^{n} W_i : w^{(2)}_{n^{-1/2}\sum_{i=1}^{n} W_i}(\delta) \ge 3\,\epsilon \Big) &\le P\big(\cup_{j,k<\delta^{-1}} A_{j\,\delta,k\,\delta} \big) \\ &\le \sum_{k<\delta^{-1}} \sum_{j<\delta^{-1}} P(A_{j\,\delta,k\,\delta}) \le \frac{B}{\epsilon^\gamma}\,\delta^{\alpha-2}, \end{aligned}$$

where we have used (1.49) to obtain the last inequality above. Because we can choose γ such that $\alpha = \epsilon\gamma > 2$, we may take δ as the reciprocal of a large integer, and in this way make $(B/\epsilon^\gamma)\,\delta^{\alpha-2}$ very small. Moreover, for all $\alpha > 0$

and adequate choice of δ we have

$$\begin{aligned} P\big(\big|n^{-1/2}\sum_{i=1}^{n} W_i(0,0)\big| \geq \alpha\big) &\leq \alpha^{-\delta}\, E\big[\big|n^{-1/2}\sum_{i=1}^{n} W_i(0,0)\big|^{\delta}\big] \\ &\leq C_{1\delta}\,\alpha^{-\delta}\Big\{E\big|Y(0)\,Y(0)\big|^{\delta} + \big|K(0,0)\big|\Big\}^{\delta} \\ &\leq C_{2\delta}\,\alpha^{-\delta}\Big\{E\big|Y(0)\times Y(0)\big|^{\delta} + E\big|Y(0)\times Y(0)\big|^{\delta}\Big\} \\ &\leq C_{3\delta}\,\alpha^{-\delta}\Big\{E\big|Y(0)\big|^{2\delta}\Big\}^{1/2} \times \Big\{E\big|Y(0)\big|^{2\delta}\Big\}^{1/2} \\ &\leq C_{4\delta}\,\alpha^{-\delta}, \end{aligned}$$

where $C_{1\delta}, C_{2\delta}, C_{3\delta}, C_{4\delta}$ are constants depending only on δ, and we have used condition (1.37) to obtain the last inequality above. Thus,

$$Z(u,v) \longrightarrow \zeta(u,v), \text{ in distribution}. \blacksquare \tag{1.50}$$

The weak limit of Z, ζ, is a bivariate Gaussian process with mean zero and the covariance function

$$\begin{aligned} C(u,v,s,t) &= \operatorname{cov}\{\zeta(u,v)\,,\zeta(s,t)\} = \operatorname{cov}\{W(u,v)\,, W(s,t)\} \\ &= E\{Y(u)\,Y(v)\,Y(s)\,Y(t)\} - K(u,v)\,K(s,t)\,, \end{aligned}$$

where W and Y denote a generic W_i and Y_i, respectively. Using the Karhunen-Loève expansion $X(u) - \eta(u) = \sum_{j=1}^{\infty} \xi_j\, \psi_j(u)$, where $\xi_j = \int_{\mathcal{I}}(X - \eta)\,\psi_j$, we have:

$$\begin{aligned} C(u,v,s,t) = &\sum_{j_1=1}^{\infty} \cdots \sum_{j_4=1}^{\infty} E\big(\xi_{j_1}\cdots\xi_{j_4}\big)\,\psi_{j_1}(u)\,\psi_{j_2}(v)\,\psi_{j_3}(s)\,\psi_{j_4}(t) \\ &- \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} \theta_{j_1}\,\theta_{j_2}\,\psi_{j_1}(u)\,\psi_{j_1}(v)\,\psi_{j_2}(s)\,\psi_{j_2}(t)\,, \end{aligned} \tag{1.51}$$

where we have used the fact that $E(\xi_{j_k}^2) = \theta_{j_k}$. If the absence of correlation among the ξ_j's is replaced by independence, only for the purpose of calculating the expected values of products of four of the variables ξ_j, then the first series on the right-hand side of (1.51) can be written as

$$
\begin{aligned}
\sum_{j_1=1}^{\infty} \cdots \sum_{j_4=1}^{\infty} E(\xi_{j1} \ldots \xi_{j_4})\, \psi_{j1}(u)\, \psi_{j_2}(v)\, \psi_{j_3}(s)\, \psi_{j_4}(t) = \\
& \sum_{j=1} E(\xi_j^4)\, \psi_j(u)\, \psi_j(v)\, \psi_j(s)\, \psi_j(t) \\
& + \sum\sum_{j_1 \neq j_2} \theta_{j_1}\, \theta_{j_2} \Big\{ \psi_{j_1}(u)\, \psi_{j_1}(v)\, \psi_{j_2}(s)\, \psi_{j_2}(t) \\
& + \psi_{j_1}(u)\, \psi_{j_1}(s)\, \psi_{j_2}(v)\, \psi_{j_2}(t) + \psi_{j_1}(u)\, \psi_{j_1}(t)\, \psi_{j_2}(s)\, \psi_{j_2}(v) \Big\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
C(u, v, s, t) = & \sum_{j=1}^{\infty} \{E(\xi_j^4) - \theta_j^2\}\, \psi_j(u)\, \psi_j(v)\, \psi_j(s)\, \psi_j(t) \\
& + \sum\sum_{j_1 \neq j_2} \theta_{j_1}\, \theta_{j_2} \Big\{ \psi_{j_1}(u)\, \psi_{j_1}(s)\, \psi_{j_2}(v)\, \psi_{j_2}(t) + \psi_{j_1}(u)\, \psi_{j_1}(t)\, \psi_{j_2}(s)\, \psi_{j_2}(v) \Big\}.
\end{aligned}
\tag{1.52}
$$

Under the assumption that random processes X is a Gaussian processes, (i.e. the variables ξ_j are independent and jointly normally distributed, rather than merely uncorrelated, and with zero kurtosis), $E(\xi_j^4) = 3\,\theta_j^2$. In this situation, the asymptotic covariance function is simplified as follows:

$$
\begin{aligned}
C(u, v, s, t) = & \sum\sum_{j_1 \neq j_2} \theta_{j_1}\, \theta_{j_2} \Big\{ \psi_{j_1}(u)\, \psi_{j_2}(v)\, \psi_{j_1}(s)\, \psi_{j_2}(t) + \psi_{j_1}(u)\, \psi_{j_2}(v)\, \psi_{j_2}(s)\, \psi_{j_1}(t) \Big\} \\
& + 2 \sum_{j=1}^{\infty} \theta_j^2\, \psi_j(u)\, \psi_j(v)\, \psi_j(s)\, \psi_j(t).
\end{aligned}
\tag{1.53}
$$

1.3.1 Asymptotic Distribution of Eigenvalues and Eigenfunctions

It should be mentioned that accounts of asymptotic normality of eigenfunctions, eigenvectors and their projections have been given by Dauxois et al. (1982) and Bosq (2000). In connection with the results discussed above, using the expansion (1.32) and (1.33), the following shorter expansions can be derived:

$$n^{1/2}\left(\widehat{\psi}_j(t) - \psi_j(t)\right) = \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\,\psi_k(t) \int Z\,\psi_j\,\psi_k + o_p(1)\,. \tag{1.54}$$

$$n^{1/2}\,(\hat{\theta}_j - \theta_j) = \int Z\,\psi_j\,\psi_j + o_p(1)\,. \tag{1.55}$$

Similarly, it can be seen from equation (1.32) that:

$$\begin{aligned} n\left\|\widehat{\psi}_j - \psi_j\right\|^2 - n\sum_{k=1}^{\infty}(a_{jk} - \delta_{jk})^2 \\ = \sum_{k\,:\,k\neq j}(\theta_j - \theta_k)^{-2}\left(\int Z\,\psi_j\,\psi_k\right)^2 + o_p(1)\,, \end{aligned} \tag{1.56}$$

Results (1.54)–(1.56) lead directly to limit theorems for $\widehat{\psi}_j$ and $\hat{\theta}_j$, as follows.

Let $P_j = \psi_j \otimes \psi_j$ and $Q_j = \sum_{k\,:\,k\neq j}(\theta_j - \theta_k)^{-1}\psi_k \otimes \psi_k$. Define the operator Φ_j such that it maps $Z \in F$ to $Q_j Z P_j \in F$, where F, the space of all Hilbert-Schmidt operators on E with the inner product $\langle .,. \rangle_F$ introduced in (1.4), and the Borel field $\mathcal{B}_F$. The operator Φ_j is linear and continuous. So, (1.50) implies that $\Phi_j(Z) \to \Phi_j(\zeta)$ in distribution (Theorem 5.5 of Billingsley (1968)). In regard to the latter result, if the ψ_j's are continuous (for each $j \geq 1$, $\psi_j \in C_{[0,1]}$), then (1.54) entails that the random function $n^{1/2}\left(\widehat{\psi}_j(t) - \psi_j(t)\right)$ converges weakly to a Gaussian process, $\Psi_j(t)$ say, precisely;

$$n^{1/2}\,\big(\widehat{\psi}_j(t)-\psi_j(t)\big) \longrightarrow \Psi_j(t) = \sum_{k\,:\,k\neq j} (\theta_j-\theta_k)^{-1}\,\psi_k(t) \int \zeta\,\psi_j\,\psi_k\,, \quad \text{in distribution,} \tag{1.57}$$

where the non-stationary Gaussian process $\Psi_j(t)$ has zero mean and covariance function

$$\tau_j(u,v) = \sum_{k1:\,k1\neq j}\ \sum_{k2:\,k2\neq j} (\theta_j-\theta_{k1})^{-1}\,(\theta_j-\theta_{k2})^{-1}\,\psi_{k1}(u)\,\psi_{k2}(v)$$
$$\times \int C(u,v,s,t)\,\psi_j(u)\,\psi_{k1}(v)\,\psi_j(s)\,\psi_{k2}(t)\,du\,dv\,ds\,dt\,,$$

where $C(u,v,s,t)$ was introduced in (1.51). After some algebraic calculations, under the assumption of independence the above formula is simplified to

$$\tau_j(u,v) = \sum_{k:\,k\neq j} (\theta_j-\theta_k)^{-2}\,\theta_j\,\theta_k\,\psi_k(u)\,\psi_k(v)\,.$$

In particular, the asymptotic variance of $n^{1/2}\,\widehat{\psi}_j(t)$ equals

$$\sigma_j(t)^2 = \tau_j(t,t) = \sum_{k:\,k\neq j} (\theta_j-\theta_k)^{-2}\,\theta_j\,\theta_k\,\psi_k(t)^2\,. \tag{1.58}$$

Result (1.57) can be extended to a p-tuple of the $\widehat{\psi}_j$. The p-tuple $n^{1/2}\,\big(\widehat{\psi}_j(t)-\psi_j(t)\big)$ for $1\leq j\leq p$ converge jointly and weakly to the non-stationary Gaussian process $\Psi_1,\cdots,\Psi_p$. In particular, for $j_1,j_2\geq 1$, the two random functions $n^{1/2}\,\big(\widehat{\psi}_{j_1}(t)-\psi_{j_1}(t)\big)$ and $n^{1/2}\,\big(\widehat{\psi}_{j_2}(t)-\psi_{j_2}(t)\big)$ have the asymptotic covariance function

$$\tau_{j_1,j_2}(u,v) = \delta_{j_1,j_2} \sum_{k:\, k \neq j_1} (\theta_{j_1} - \theta_k)^{-2}\, \theta_{j_1}\, \theta_k\, \psi_k(u)\, \psi_k(v)$$
$$+ \left(\delta_{j_1,j_2} - 1\right) (\theta_{j_2} - \theta_{j_1})^{-2}\, \theta_{j_1}\, \theta_{j_2} \psi_{j_2}(u)\, \psi_{j_1}(v)\,,$$

where δ_{j_1,j_2} denotes the Kronecker delta. The covariance function shows that for $j_1 \neq j_2$, the two elements $n^{-1/2}\left(\widehat{\psi}_{j_1}(t) - \psi_{j_1}(t)\right)$ and $n^{-1/2}\left(\widehat{\psi}_{j_2}(t) - \psi_{j_2}(t)\right)$ are not asymptotically independent.

Moreover, in connection to (1.56) we have, by (1.50),

$$n^{1/2}\, \|\widehat{\psi}_j - \psi_j\| \to U_j\,, \quad \text{in distribution,} \tag{1.59}$$

where

$$U_j^2 = \sum_{k:\, k \neq j} (\theta_j - \theta_k)^{-2}\, N_{jk}^2\,, \tag{1.60}$$

and the random variables $N_{jk} = \int \zeta\, \psi_j\, \psi_k$ are jointly normally distributed with zero mean. If the random function X is a Gaussian process then $N_{j1}, N_{j2}, \ldots$ are independent as well as normally distributed. Note that, since $\int_{\mathcal{I}} E(X^4) < \infty$,

$$\begin{aligned}
\sum_{k=1}^{\infty} E(N_{jk}^2) &= E\Big(\sum_{k=1}^{\infty} N_{jk}^2\Big) = E\bigg\{\int_{\mathcal{I}}\bigg[\int_{\mathcal{I}} \zeta(u,v)\, \psi_j(v)\, dv\bigg]^2 du\bigg\} \\
&= \int_{\mathcal{I}} E\bigg\{\int_{\mathcal{I}} \zeta(u,v)\, \psi_j(v)\, dv\bigg\}^2 du \le \iint_{\mathcal{I}^2} E(\zeta^2) \\
&= \iint_{\mathcal{I}^2} E(W^2) = \int_{\mathcal{I}}\int_{\mathcal{I}} \left\{E[X(u)^2\, X(v)^2] - K(u,v)^2\right\} du\, dv \\
&\le \int_{\mathcal{I}}\int_{\mathcal{I}} E[X(u)^2\, X(v)^2]\, du\, dv = E\bigg(\int_{\mathcal{I}} X^2\bigg)^2 \le \int_{\mathcal{I}} E(X^4) < \infty\,,
\end{aligned} \tag{1.61}$$

from which it follows that the series defining U_j^2 is finite provided the eigenvalue

θ_j is not repeated. Observe too that (1.55) implies

$$n^{1/2}\,(\hat{\theta}_j - \theta_j) \longrightarrow N_{jj}\,, \quad \text{in distribution}\,, \tag{1.62}$$

where N_{jj} is normally distributed with mean zero and variance

$$\gamma_j^2 = \text{Var}\biggl(\int W\,\psi_j\,\psi_j\biggr) = E(\xi_j^4) - \theta_j^2\,. \tag{1.63}$$

An extension of the result (1.62) is available for any p-tuple of the $\hat{\theta}_j$. The p-tuple $n^{1/2}\,(\hat{\theta}_j - \theta_j)$ for $1 \le j \le p$ converge jointly to a p-variate Normal distribution. In particular, for $j_1, j_2 \ge 1$, the two random quantities $n^{1/2}\,(\hat{\theta}_{j_1} - \theta_{j1})$ and $n^{1/2}\,(\hat{\theta}_{j_2} - \theta_{j_2})$ are asymptotically distributed as a two-variate Normal distribution with the covariance

$$\gamma_{j_1 j_2} = E(\xi_{j_1}^2\,\xi_{j_2}^2) - \theta_{j1}\,\theta_{j2}\,.$$

Comparing result (1.62) with formula (1.60) for the limiting distribution of $n^{1/2}\,\|\widehat{\psi}_j - \psi_j\|$, we see that spacings among the eigenvalues θ_k impact immediately on properties of $\widehat{\psi}_j$, through first-order terms in its limiting distribution; but impact on $\hat{\theta}_j$ only through second-order terms. Note also (1.33), where it is clear that eigenvalue spacings affect only the term in n^{-1}, not that in $n^{-1/2}$.

1.4 Inequalities

Explicit bounds on $\|\widehat{\psi}_j - \psi_j\|$ in terms of spacings, and a spacings-free bound for $|\hat{\theta}_j - \theta_j|$, can be obtained from Theorem 1.5 below. Define $\widehat{\Delta} = (\int |\widehat{K} - K|^2)^{1/2}$,

$$\begin{aligned} \delta_j &= \min\nolimits_{1 \le k \le j}\,(\theta_k - \theta_{k+1})\,, \qquad J = \inf\bigl\{j \ge 1\,:\,\theta_j - \theta_{j+1} \le 2\,\widehat{\Delta}\bigr\}\,, \\ \hat{\delta}_j &= \min\nolimits_{1 \le k \le j}\,(\hat{\theta}_k - \hat{\theta}_{k+1})\,, \qquad \hat{J} = \inf\bigl\{j \ge 1\,:\,\hat{\theta}_j - \hat{\theta}_{j+1} \le 2\,\widehat{\Delta}\bigr\}\,. \end{aligned} \tag{1.64}$$

Theorem 1.5. (a) With probability 1, $\sup_{j\geq 1} |\hat{\theta}_j - \theta_j| \leq \widehat{\Delta}$ and for all $1 \leq j \leq J-1$

$$\|\widehat{\psi}_j - \psi_j\| \leq 2^{1/2} \left[1 - \left\{1 - 4\,(\widehat{\Delta}/\delta_j)^2\right\}^{1/2}\right]^{1/2} \leq 8^{1/2}\,\widehat{\Delta}/\delta_j\,.$$

(b) This result continues to hold if (J, δ_j) is replaced by $(\widehat{J}, \hat{\delta}_j)$ throughout.

Proof: The theorem follows from Bhatia et al. (1983). Let $\theta_{j1} \geq \theta_{j2} \geq \cdots \geq 0$ denote the eigenvalues of the self-adjoint, semi-positive definite Hilbert-Schmidt operator K_j $(j = 1, 2)$ on $L_2(\mathcal{I})$, and $\psi_{11}, \psi_{12}, \psi_{13}, \cdots$ and $\psi_{21}, \psi_{22}, \psi_{23}, \cdots$ be the sequence of respective orthonormal eigenfunctions K_1 and K_2. Sections 4-6 of Bhatia et al. (1983) show that

$$\text{if} \quad \|K_1 - K_2\| \leq \epsilon, \text{ then } \quad |\theta_{1\ell} - \theta_{2\ell}| \leq \epsilon\,, \quad \text{for each } \ell \geq 1. \tag{1.65}$$

In other words, if $\|K_1 - K_2\| \leq \epsilon$, then the eigenvalue sequences of K_1 and K_2, with multiplicities included, are no further apart than ϵ. Furthermore, Theorem 6.1 of Bhatia et al. (1983) implies that if $\|K_1 - K_2\| \leq \epsilon \leq \frac{1}{2}\,\delta_j$ then, for $1 \leq \ell \leq j$, $\frac{1}{2}\,\delta_j\,\|\psi_{1\ell} - \bar{\psi}_{1\ell}\|^2 \leq \epsilon$, where $\delta_j = \inf_{\ell\leq j}(\theta_{1\ell} - \theta_{1,\ell+1})$ and $\bar{\psi}_{1\ell} = \psi_{2\ell} \int \psi_{1\ell}\,\psi_{2\ell}$ denotes the projection of $\psi_{1\ell}$ onto $\psi_{2\ell}$. Hence, if $\|K_1 - K_2\| \leq \epsilon \leq \frac{1}{2}\,\delta_j$, then, for $1 \leq \ell \leq j$,

$$\begin{aligned}\|\psi_{1\ell} - \psi_{1\ell}\|^2 &= 2\left\{1 - (1 - \|\psi_{1\ell} - \bar{\psi}_{1\ell}\|^2)^{1/2}\right\}\\ &\leq 2\left[1 - \{1 - 4(\epsilon/\delta_j)^2\}^{1/2}\right] \leq 8\,\epsilon^2/\delta_j^2\,.\end{aligned} \tag{1.66}$$

Parts (a) and (b) of Theorem 1.5 follow on taking $\epsilon = \widehat{\Delta}$ and letting $(K_1, K_2) = (K, \widehat{K})$ and $(\widehat{K}, K)$, respectively. The only assumptions needed for the theorem are that $X_1, \ldots, X_n$ are square-integrable random functions. Theorem 1.5 can also be obtained from Bosq (2000, Lemma 4.2, and Lemma 4.3). ■

The inequalities are potentially attractive, because they can be used as the basis for simultaneous confidence regions, and for bootstrap confidence procedures. These methods can be justified, and in particular can be shown to have appropriate degrees of accuracy, by using the stochastic expansions mentioned earlier. See Chapter 3.

Numerical experiments show that regions based on the simultaneous bound for $\|\widehat{\psi}_j - \psi_j\|$ are too conservative to be of much practical benefit. However, the simultaneous bound for $|\hat{\theta}_j - \theta_j|$ can give useful results. See Chapter 5.

Chapter 2

Mathematical Properties of Stochastic Expansions

2.1 Introduction

The stochastic expansions for eigenvalues and eigenfunctions given in Chapter 1 are of intrinsic interest. They can be used as the foundation for theory in a particularly wide range of settings, for example bootstrap methods for confidence intervals for eigenvalues and eigenfunctions. However, in order to develop informative theory about the performance of such methodologies, we need a concise account of the accuracy to which $\hat{\theta}_j$ and $\widehat{\psi}_j$ approximate θ_j and ψ_j, respectively. That account can be easily provided using properties of the expansions. Moreover, the problem of determining estimator accuracy, uniformly over many components, prompts consideration of explicit uniform bounds that are obtainable via the mathematical theory of infinite-dimensional operators.

In Section 2.2, we first state the basic theorem about the properties of the stochastic expansions. We introduce the notations used in this Chapter, and also give the two auxiliary Theorems 2.2 and 2.3 in Section 2.3. Section 2.4 provides bounds in both Sup and Hilbert-Schmidt metrics for $\widehat{K} - K$. In Section 2.5, we exploit the auxiliary results, discussed in Section 2.3, for empirical approximations. After that, we give approximations to $\hat{\theta}_j$ and $\widehat{\psi}_j$ in Sections 2.6 and 2.7, respectively. In the latter Section, we obtain bounds in both Sup and L_2 metrics.

2.2 The Basic Theorem

Take $\mathcal{I}$ to be the unit interval. Put $A = X - E(X)$ and assume that

(a) for all $C > 0$ and some $\epsilon > 0$,

$$\sup_{t\in\mathcal{I}} E\big|X(t)\big|^C < \infty\,, \quad \sup_{s,t\in\mathcal{I}} E\Big[\big\{|s-t|^{-\epsilon}\,|X(s) - X(t)|\big\}^C\Big] < \infty\,; \tag{2.1}$$

(b) for each integer $r \geq 1$, $\theta_j^{-r}\, E(\int_\mathcal{I} A\,\psi_j)^{2r}$ is bounded uniformly in j.

For example, (2.1) holds for Gaussian processes with Hölder-continuous sample paths. If X is a Gaussian process, then the variables ξ_j in $X - E(X) = \sum_{j=1}^\infty \xi_j\,\psi_j$ are independent and jointly normally distributed. Therefore, ξ_j^2/θ_j has a Chi-squared distribution with one degree of freedom, and $E(\int_\mathcal{I} A\,\psi_j)^{2r} = E(\xi_j^{2r}) \leq C_r\,\theta_j^r$, where the constant C_r only depends on r. Generally, if $\xi_j/(E\xi_j^2)^{1/2}$ has the distribution of ξ, for each j, then part (b) of (2.1) holds whenever $E(\xi_j^{2r}) < \infty$.

Recall that the eigenvalues of the covariance operator K are ordered so that $\theta_1 \geq \theta_2 \geq \ldots \geq 0$. Define $\|\widehat{K} - K\|^2 = \int(\widehat{K} - K)^2$, and put $\rho_j = \min_{k\neq j} |\theta_j - \theta_k|$, $s_j = \sup_u |\psi_j(u)|$ and $\zeta_j = \inf_{k\geq j}\{1 - (\theta_k/\theta_j)\}$.

Results (1.11), (1.12), (1.32) and (1.33) are straightforward corollaries of the following theorem.

Theorem 2.1. If (2.1) holds, then for each j for which

$$\left|\left|\left|\widehat{K} - K\right|\right|\right| \leq \tfrac{1}{2} \min(\theta_j - \theta_{j+1}, \theta_{j-1} - \theta_j)\,, \tag{2.2}$$

the absolute values of the "$O_p(n^{-3/2})$" remainders on the right-hand sides of (1.32) and (1.33) are each bounded above by $n^{-3/2}\, U_{nj}\, (1-\zeta_j)^{-1/2}\, \rho_j^{-3}\, \theta_j^{-1/2}\, s_j$, where the random variables U_{nj} satisfy $\sup_{n,j\geq 1} E(U_{nj}^C) < \infty$ for each $C > 0$. In the case of (1.32), this bound is also valid uniformly in t. Moreover, the "$O_p(n^{-3/2})$" remainders on the right-hand sides of (1.33) are bounded above in the L_2 metric by $n^{-3/2}\, U_{nj}\, \rho_j^{-3}$, where the U_{nj} have the same properties as before.

We may paraphrase (2.2) by saying that that condition holds for all j for which the distance of θ_j to the nearest other eigenvalue does not fall below $2\,|||\widehat{K} - K|||$, which in turn equals $O_p(n^{-1/2})$. Moreover, using Markov's inequality for each $\epsilon, C > 0$ we have

$$P(U_{nj} > j^{\epsilon}) \leq j^{-C\epsilon}\, E|U_{nj}|^C\,.$$

Then, using the Borel-Cantelli Lemma for appropriate values of C and ϵ implies that $j^{-\epsilon}\, U_{nj} = O(1)$, with probability one. Therefore, the bounds given in Theorem 2.1 imply that the "$O_p(n^{-3/2})$" remainders in (1.32) and (1.33) equal $O_p\{n^{-(3/2)}\, j^{\epsilon}\, (1-\xi_j)^{-1/2}\, \rho_j^{-3}\, \theta_j^{-1/2}\, s_j\}$ uniformly in j for which (2.2) holds and $1 \leq j \leq n^C$ for each $C, \epsilon > 0$. In the case of (1.32) the bound is also uniform in $t \in \mathcal{I}$.

2.3 Notations and Auxiliary Results

In this Section we state and prove two theorems used to derive Theorem 2.1. Derivations of the parts of Theorem 2.1, pertaining to expansions (1.32) and (1.33), will be given in Sections 2.8 and 2.7, respectively.

Define $|\mathcal{I}| = \int_{\mathcal{I}} dx$, denoting the content of the compact region $\mathcal{I}$. Given univariate functions α and β, and a symmetric bivariate function M, put $\|\alpha\| = (\int_{\mathcal{I}} \alpha^2)^{1/2}$ denoting the L_2 norm of α, $\|\!|M|\!\| = (\int\int_{\mathcal{I}^2} M^2)^{1/2}$ and $\|\!|M|\!\|_{\sup} = \sup_{u\in\mathcal{I}} \|M(u,.)\|$, and write $\int \alpha\,\beta$ and $\int M\,\alpha\,\beta$ for

$$\int_{\mathcal{I}} \alpha(u)\beta(u)\,du, \qquad \int\int_{\mathcal{I}^2} M(u,v)\,\alpha(u)\beta(v)\,du\,dv,$$

respectively. Furthermore, let $\int M\,\alpha$ denote the function of which the value at u is $\int_{\mathcal{I}} M(u,v)\,\alpha(v)\,dv$.

Let K and L be two self-adjoint, positive semi-definite, Hilbert-Schmidt operators on $\mathcal{I}$, with respective kernels which we also write as K and L. By the singular value decomposition theorem, we can write the respective spectral decompositions as

$$K(u,v) = \sum_{j=1}^{\infty} \theta_j\,\psi_j(u)\psi_j(v), \qquad L(u,v) = \sum_{j=1}^{\infty} \lambda_j\,\phi_j(u)\phi_j(v), \tag{2.3}$$

where the terms are ordered in such a way that

$$\theta_1 \geq \theta_2 \geq \cdots \geq 0, \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq 0. \tag{2.4}$$

We wish to describe the closeness of θ_j to λ_j, and of ψ_j to ϕ_j, in terms of the nearness of K to L. The results discussed below apply only to Hilbert-Schmidt operators, of which the covariance operator K and its estimator, the sample covariance operator $\widehat{K}$, are examples.

Theorem 2.2. If K and L are Hilbert-Schmidt operators with corresponding kernels that have the expansions (2.3); and if we write ϕ_j as " ψ_j plus a first-order term χ_j plus a remainder Δ_j", i.e. as

$$\phi_j - \psi_j = \chi + \Delta_j \,; \tag{2.5}$$

then,

$$\left|(\theta_j - \lambda_j)\Big(1 + \int \chi_j \psi_j\Big) - \int (K-L)(\psi_j + \chi_j)\psi_j\right|$$
$$\leq \|\Delta_j\| \left(|\theta_j - \lambda_j| + \mathcal{I}^{1/2} \, |\!|\!| K - L |\!|\!|_{\text{sup}}\right). \tag{2.6}$$

Proof: Formula (2.3) directly implies a series expansion of $K - L$. Multiplying both sides of this by $\phi_j(u)\,\psi_j(v)$, and integrating over u and v, we deduce that

$$(\theta_j - \lambda_j) \int \phi_j \, \psi_j - \int (K - L)\, \phi_j \, \psi_j = 0. \tag{2.7}$$

In view of (2.5),

$$\left| \int \phi_j \, \psi_j - 1 - \int \chi_j \, \psi_j \right| = \left| \int \Delta_j \, \psi_j \right| \leq \|\Delta_j\|,$$

$$\left| \int (K - L)(\phi_j - \psi_j - \chi_j)\, \psi_j \right|^2 = \left| \int (K - L)\, \Delta_j \, \psi_j \right|^2$$
$$\leq \Big(\int \Delta_j^2 \Big) \int_{\mathcal{I}} \left[\int_{\mathcal{I}} \{K(u,v) - L(u,v)\}\, \psi_j(u)\, du \right]^2 dv$$
$$\leq |\mathcal{I}| \, \|\Delta_j\|^2 \, |\!|\!| K - L |\!|\!|^2_{\text{sup}}. \tag{2.8}$$

Results (2.7) and (2.8) imply (2.6) as follows:

$$
\begin{aligned}
&\left|(\theta_j-\lambda_j)\left(1+\int\chi_j\psi_j\right)-\int(K-L)(\psi_j+\chi_j)\psi_j\right| \\
&\quad=\left|(\theta_j-\lambda_j)\left(1+\int\chi_j\psi_j-\int\phi_j\,\psi_j\right)-\int(K-L)(\psi_j+\chi_j-\phi_j)\psi_j\right| \\
&\quad\le|\theta_j-\lambda_j|\left|\int\Delta_j\,\psi_j\right|+\left|\int(K-L)\,\Delta_j\,\psi_j\right| \\
&\quad\le|\theta_j-\lambda_j|\,|\Delta_j\|+|\mathcal{I}|^{1/2}\,\|\Delta_j\|\,\|K-L\|_{\sup} \\
&\quad=\|\Delta_j\|\left(|\theta_j-\lambda_j|+\mathcal{I}^{1/2}\,\|K-L\|_{\sup}\right).\ \blacksquare
\end{aligned}
$$

Theorem 2.3. Assume K and L are Hilbert-Schmidt operators for which the corresponding kernels have the expansions (2.3), and that the eigenvalues θ_j are all distinct. Then, provided (for a given value of j) $\inf_{k:\,k\neq j}|\lambda_j-\theta_k|>0$,

$$
\phi_j-\psi_j=\sum_{k\,:\,k\neq j}(\lambda_j-\theta_k)^{-1}\,\psi_k\int(L-K)\phi_j\,\psi_k+\psi_j\int(\phi_j-\psi_j)\,\psi_j \tag{2.9}
$$

$$
\begin{aligned}
&=\sum_{s=0}^{\infty}(\theta_j-\lambda_j)^s\sum_{k\,:\,k\neq j}(\theta_j-\theta_k)^{-(s+1)}\,\psi_k\int(L-K)\phi_j\,\psi_k \\
&\quad+\psi_j\int(\phi_j-\psi_j)\,\psi_j
\end{aligned} \tag{2.10}
$$

$$
\begin{aligned}
&=\theta_j^{-1}\sum_{s=0}^{\infty}\left(\frac{\theta_j-\lambda_j}{\theta_j}\right)^s\sum_{k\,:\,k\neq j}\psi_k\int(L-K)\phi_j\,\psi_k \\
&\quad+\sum_{s=0}^{\infty}(\theta_j-\lambda_j)^s\sum_{k\,:\,k\neq j}\left\{(\theta_j-\theta_k)^{-(s+1)}-\theta_j^{-(s+1)}\right\}\psi_k\int(L-K)\phi_j\,\psi_k \\
&\quad+\psi_j\int(\phi_j-\psi_j)\,\psi_j.
\end{aligned} \tag{2.11}
$$

Let $\zeta_j\in(0,1)$ denote the infimum of $1-(\theta_k/\theta_j)$ over k such that $\theta_k<\theta_j$, and let $\eta_j\in(0,1)$ denote the infimum of $(\theta_k/\theta_j)-1$ over k such that $\theta_k>\theta_j$. If $|\theta_j-\lambda_j|\le\theta_j\min(\zeta_j,\eta_j)$, and if K and L are uniformly bounded, then the series in s at (2.10) and (2.11) are absolutely convergent. Moreover,

$$\Big|\sum_{k\,:\,k\neq j} \psi_k \int (L-K)\,\phi_j\,\psi_k\Big| \leq |||L-K|||_{\sup} \{1+\sup_{u\in\mathcal{I}} |\psi_j(u)|\}, \tag{2.12}$$

$$\begin{aligned}|\theta_j-\lambda_j|^s \sum_{k\,:\,k\neq j} &\Big|\{(\theta_j-\theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\}\,\psi_k \int (L-K)\,\phi_j\,\psi_k\Big| \\ &\leq \max\{\zeta_j^{-(s+1)}(1-\zeta_j)^{-1/2}, \eta_j^{-(s+1)}+1\}\,\theta_j^{-3/2} \\ &\quad\times \left|\frac{\theta_j-\lambda_j}{\theta_j}\right| \{\sup_{u\in\mathcal{I}} K(u,u)\}^{1/2}\,|||L-K|||. \end{aligned}\tag{2.13}$$

Proof: It can be shown directly from (2.3) that

$$\lambda_j\,(\phi_j-\psi_j) = \int K(\phi_j-\psi_j) + \int (L-K)\,\phi_j - (\lambda_j-\theta_j)\,\psi_j.$$

Multiplying both sides by ψ_k, and integrating, we deduce that

$$\begin{aligned}\lambda_j \int (\phi_j-\psi_j)(v)\,\psi_k(v)\,dv &= \iint_{\mathcal{I}^2} K(u,v)(\phi_j-\psi_j)(u)\,\psi_k(v)\,du\,dv \\ &\quad + \iint_{\mathcal{I}^2} (L-K)(u,v)\,\phi_j(u)\,\psi_k(v)\,du\,dv \\ &\quad - (\lambda_j-\theta_j) \int \psi_j(v)\,\psi_k(v)\,dv \\ &= \theta_k \int (\phi_j-\psi_j)(u)\,\psi_k(u)\,du \\ &\quad + \int (L-K)\,\phi_j\,\psi_k - (\lambda_j-\theta_j)\,\delta_{jk}\,, \end{aligned}\tag{2.14}$$

where δ_{jk} denotes the Kronecker delta. Equivalently, provided $\lambda_j \neq \theta_k$,

$$\int (\phi_j-\psi_j)\,\psi_k = (\lambda_j-\theta_j)^{-1} \int (L-K)\,\phi_j\,\psi_k - \delta_{jk}.$$

Results (2.9) and (2.10) follow from this formula and the fact that

$$\phi_j - \psi_j = \sum_{k=1}^{\infty} \psi_k \int (\phi_j - \psi_j)\, \psi_k,$$

$$(\lambda_j - \theta_k)^{-1} = (\theta_j - \theta_k)^{-1} \left(1 - \frac{\theta_j - \lambda_j}{\theta_j - \theta_k}\right)^{-1} = \sum_{s=0}^{\infty} \frac{(\theta_j - \lambda_j)^s}{(\theta_j - \theta_k)^{s+1}}.$$

To derive (2.12), note that

$$\left| \sum_{k\,:\,k \neq j} \psi_k \int (L - K)\, \phi_j\, \psi_k \right| = \left| \int (L - K)\phi_j - \psi_j \int (L - K)\, \phi_j\, \psi_j \right|$$
$$\leq \|L - K\|_{\sup} \left\{1 + \sup_u |\psi_j(u)|\right\}.$$

To obtain (2.13), we need the following lemma.

Lemma 2.1. For each integer $s \geq 1$, we have:

$$|\{1 - (\theta_k/\theta_j)\}^{-(s+1)} - 1|\, (\theta_j/\theta_k)^{1/2} \leq \max\{\zeta_j^{-(s+1)}\, (1 - \zeta_j)^{-1/2}, \eta_j^{-(s+1)} + 1\} \equiv A_j\,,$$

uniformly in values of k such that $k \neq j$.

Proof of Lemma: Define $\nu_{jk} = \{1 - (\theta_k/\theta_j)\}$, $\zeta_j = \inf_{k:\,k>j}\{1 - (\theta_k/\theta_j)\}$ and $\eta_j = \inf_{k:\,k<j}\{(\theta_k/\theta_j) - 1\}$. For $k > j$, $\zeta_j \leq \nu_{jk} < 1$,

$$\begin{aligned}
|\{1 - (\theta_k/\theta_j)\}^{-(s+1)} - 1|\, (\theta_j/\theta_k)^{1/2} &= |\nu_{jk}^{-(s+1)} - 1|\, (1 - \nu_{jk})^{-1/2} \\
&= (\nu_{jk}^{-(s+1)} - 1)\, (1 - \nu_{jk})^{-1/2} \\
&= \nu_{jk}^{-(s+1)}\, (1 - \nu_{jk}^{s+1})(1 - \nu_{jk})^{-1/2} \\
&= \nu_{jk}^{-(s+1)}\, (1 - \nu_{jk})\, (1 + \nu_{jk} + \cdots + \nu_{jk}^{s})\, (1 - \nu_{jk})^{-1/2} \\
&= \nu_{jk}^{-(s+1)}\, (1 - \nu_{jk})^{1/2}\, (1 + \nu_{jk} + \cdots + \nu_{jk}^{s}) \\
&= (1 - \nu_{jk})^{1/2}\, (\nu_{jk}^{-1} + \nu_{jk}^{-2} + \cdots + \nu_{jk}^{-(s+1)}).
\end{aligned}$$

The left hand-side is a decreasing function of ν_{jk}, as the terms on the right hand-side are decreasing functions of ν_{jk}. So, its maximum is obtained at $\nu_{jk} = \zeta_j$ for all $k > j$, implying that

$$|\{1-(\theta_k/\theta_j)\}^{-(s+1)}-1|\,(\theta_j/\theta_k)^{1/2} \le (\zeta_j^{-(s+1)}-1)\,(1-\zeta_j)^{-1/2} \le \zeta_j^{-(s+1)}\,(1-\zeta_j)^{-1/2}. \tag{2.15}$$

For $k < j$, we also have

$$\begin{aligned} |\{1-(\theta_k/\theta_j)\}^{-(s+1)}-1|\,(\theta_j/\theta_k)^{1/2} &\le |\{1-(\theta_k/\theta_j)\}^{-(s+1)}-1| \\ &= |\{1-(\theta_k/\theta_j)\}|^{-(s+1)}+1 \\ &= \{(\theta_k/\theta_j)-1\}^{-(s+1)}+1 \\ &\le \eta_j^{-(s+1)}+1\,. \end{aligned}$$

Combining this result and (2.15) finishes the proof of the lemma. ■

Therefore, by using lemma 2.1, we deduce that

$$\begin{aligned} \theta_j^{s+(3/2)} \sum_{k\,:\,k\neq j} &\left|\{(\theta_j-\theta_k)^{-(s+1)}-\theta_j^{-(s+1)}\}\,\psi_k \int (L-K)\phi_j\,\psi_k\right| \\ &\le A_j \sum_{k\,:\,k\neq j}\left|\theta_k^{1/2}\,\psi_k \int (L-K)\,\phi_j\,\psi_k\right| \\ &\le A_j\Big(\sum_{k\,:\,k\neq j}\theta_k\,\psi_k^2\Big)^{1/2}\left[\sum_{k\,:\,k\neq j}\left\{\int (L-K)\,\phi_j\,\psi_k\right\}^2\right]^{1/2} \\ &\le A_j\{\sup_u K(u,u)\}^{1/2}\,|||L-K|||\,. \end{aligned} \tag{2.16}$$

This finishes the proof of (2.13). ■

Theorem 2.4. If there are no ties for the eigenvalue θ_j, then

$$\sup_{j\ge 1}\max\left\{\left|\hat\theta_j-\theta_j\right|,\,8^{-1/2}\rho_j\left\|\widehat\psi_j-\psi_j\right\|\right\} \le \left|\left|\left|\widehat K-K\right|\right|\right|, \tag{2.17}$$

and

$$\left|\hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j\right| \leq 2\, \|\widehat{\psi}_j - \psi_j\|\, \|\widehat{K} - K\|_{\sup}$$
$$\leq 8\, \rho_j^{-1}\, \|\widehat{K} - K\|\, \|\widehat{K} - K\|_{\sup} \,. \tag{2.18}$$

Proof: The bound (2.17) follows from Lemmas 4.2 and 4.3 of Bosq (2000). To obtain the first inequality in (2.18), using (2.14) with $L = \widehat{K}$, $\lambda_j = \widehat{\theta}_j$ and $\phi_j = \widehat{\psi}_j$, we can write

$$\hat{\theta}_j - \theta_j = \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_j - (\widehat{\theta}_j - \theta_j) \int (\widehat{\psi}_j - \psi_j)\, \psi_j \,. \tag{2.19}$$

Thus,

$$\begin{aligned}
&\left|\hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j\right| \\
&\quad = \left|\int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_j - (\hat{\theta}_j - \theta_j) \int (\widehat{\psi}_j - \psi_j)\, \psi_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j\right| \\
&\quad \leq \left|\int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_j\right| + \left|(\hat{\theta}_j - \theta_j)\right| \left|\int (\widehat{\psi}_j - \psi_j)\, \psi_j\right|,
\end{aligned} \tag{2.20}$$

where the equality above resulted from substituting (2.19) instead of $\hat{\theta}_j - \theta_j$. We have

$$\begin{aligned}
\left|\int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_j\right| &\leq \left[\int \left\{\int (\widehat{K} - K)(u, v)\, (\widehat{\psi}_j - \psi_j)(v)\, dv\right\}^2 du\right]^{1/2} \|\psi_j\| \\
&\leq \left[\int \left\{\int (\widehat{K} - K)^2(u, v)\, dv\right\}\left\{\int (\widehat{\psi}_j - \psi_j)^2(v)\, dv\right\} du\right]^{1/2} \\
&\leq \|\widehat{\psi}_j - \psi_j\|\, |\mathcal{I}|^{1/2} \left[\sup_u \left\{\int (\widehat{K} - K)^2(u, v)\, dv\right\}\right]^{1/2} \\
&= \|\widehat{\psi}_j - \psi_j\|\, |\mathcal{I}|^{1/2}\, \|\widehat{K} - K\|_{\sup} \,.
\end{aligned} \tag{2.21}$$

Hence, (2.20) leads us to

$$\left|\hat{\theta}_j - \theta_j - \int (\widehat{K}-K)\,\psi_j\,\psi_j\right| \leq \|\widehat{\psi}_j - \psi_j\|\,(|\hat{\theta}_j - \theta_j| + \|\widehat{K}-K\|_{\sup}) \tag{2.22}$$

$$\leq 2\|\widehat{\psi}_j - \psi_j\|\,\|\widehat{K}-K\|_{\sup}\,, \tag{2.23}$$

where we have obtained the last inequality from the fact that $|\hat{\theta}_j - \theta_j| \leq \||\widehat{K} - K\|| \leq \|\widehat{K}-K\|_{\sup}$. The second inequality in (2.18) follows on applying (2.17) to $\|\widehat{\psi}_j - \psi_j\|$. This finishes the proof of Theorem 2.4. ■

2.4 Bounds in Sup and H-S Metrics for $\widehat{K}-K$

Lemma 2.2. Let $X(t)$ satisfy part (a) of (2.1). Then, for each $C > 0$,

$$E\{\sup_{t\in\mathcal{I}} |X(t)|^C\} < \infty\,.$$

Proof of Lemma: If $X(t)$ satisfies part (a) of (2.1) for a given $C > 0$, then we show that

$$E\{\sup_{0\leq t\leq 1} |X(t)|^{C_1}\} < \infty\,, \quad \text{for all } C_1 < C\,.$$

Assume $X(0) = 0$. Let $Q_m = \{\frac{i}{2^m} | i = 1, 2, 3, \cdots, 2^m\}$, and $0 < \lambda < 1$ such that $\lambda^{-C}\,2^{1-\epsilon} < 1$. Define $b_0, b_1, \cdots$ as $b_m = \sum_{i=0}^m \lambda^i$. Then, for any fixed $M > 0$ define $A_m = P(\max_{t\in Q_m} |X(t)| > M\,b_m)$. Now by Chebyshev's inequality

$$A_0 = P(\max_{t\in Q_0} |X(t)| > M\,b_0) = P(|X(1)| > M\,b_0) \leq \frac{E|X(1)|^C}{(M\,b_0)^C} \leq \frac{C^*}{M^C}\,. \tag{2.24}$$

For $m \geq 1$ we have:

$$
\begin{aligned}
A_m &= P(\max_{t\in Q_m}|X(t)| > M\, b_m) \\
&\leq A_{m-1} + P(\max_{t\in Q_m}|X(t)| > M\, b_m,\ \max_{t\in Q_{m-1}}|X(t)| \leq M\, b_{m-1}) \\
&\leq A_{m-1} + P(\max_{t\in Q_{m-1}}|X(t-2^{-m}) - X(t)| > M\,(b_m - b_{m-1})) \\
&\leq A_{m-1} + \sum_{t\in Q_{m-1}} P(|X(t-2^{-m}) - X(t)| > M\,(b_m - b_{m-1})) \\
&\leq A_{m-1} + 2^{m-1}\frac{C^* 2^{-m\,\epsilon}}{M^C\,\lambda^{m\,C}} \\
&= A_{m-1} + \frac{C^*}{2} M^{-C}\,(2^{1-\epsilon}\,\lambda^{-C})^m\,.
\end{aligned} \tag{2.25}
$$

Let $k = C^* + \sum_{m=1}^{\infty} \frac{C^*}{2}\,(2^{1-\epsilon}\,\lambda^{-C})^m$ which is finite by construction of λ. Now as $m \to \infty$, $b_m \to \frac{1}{1-\lambda}$. Therefore,

$$
\left\{\max_{t\in\cup_{m=0}^{\infty}Q_m}|X(t)| > \frac{M}{1-\lambda}\right\} \Rightarrow \text{there exists } m^* \text{ and there exists } t^* \in Q_{m^*} \text{ such that}
$$

$$
|X(t)| > \frac{M}{1-\lambda} \Rightarrow \max_{t\in Q_{m^*}}|X(t)| > \frac{M}{1-\lambda} \Rightarrow \bigcup_{m=0}^{\infty}\left\{\max_{t\in Q_m}|X(t)| > \frac{M}{1-\lambda}\right\}. \tag{2.26}
$$

Furthermore,

$$
\begin{aligned}
&\left\{\max_{t\in Q_i}|X(t)| > \frac{M}{1-\lambda}\right\} \subseteq \bigcap_{j\geq i}\left\{\max_{t\in Q_j}|X(t)| > \frac{M}{1-\lambda}\right\} \\
&\subseteq \bigcap_{j\geq i}\left\{\max_{t\in Q_j}|X(t)| > M\, b_n\right\} \text{for all } n\geq 1 \subseteq \bigcap_{j\geq i}\left\{\max_{t\in Q_j}|X(t)| > M\, b_j\right\}.
\end{aligned} \tag{2.27}
$$

Substituting (2.27) into (2.26) results in

$$
\begin{aligned}
\max_{t\in\cup_{m=0}^{\infty}Q_m}|X(t)| > \frac{M}{1-\lambda} &\subseteq \bigcup_{m=0}^{\infty}\bigcap_{j\geq m}\left\{\max_{t\in Q_m}|X(t)| > M\, b_m\right\} \\
&= \liminf_m \left\{\max_{t\in Q_m}|X(t)| > M\, b_m\right\}.
\end{aligned}
$$

Using (2.24) and (2.25) implies that

$$P\Big(\max_{t\in\bigcup_{m=0}^{\infty}Q_m} |X(t)| > \frac{M}{1-\lambda}\Big) \le \limsup_{m\to\infty} A_m \le M^{-C}\, C^* + \sum_{m=1}^{\infty} M^{-C}\, \frac{C^*}{2}\, (2^{1-\epsilon}\, \lambda^{-C})^m$$
$$= M^{-C}\, k\,.$$

Since $\bigcup_{m=0}^{\infty} Q_m$ is dense in $\mathcal{I} = [0,1]$, we have $P(\sup_{0\le t\le 1} |X(t)| > \frac{M}{1-\lambda}) \le M^{-C}\, k$. Now, define $V = \sup_{0\le t\le 1} |X(t)|$. We have

$$E|V|^{C_1} \le \sum_{n=0}^{\infty} P(|V|^{C_1} > n) = \sum_{n=0}^{\infty} P(|V| > n^{1/C_1}) \le k\,(1-\lambda)^C \sum_{n=0}^{\infty} n^{-(C/C_1)}\,.$$

It follows that $E\{\, \sup_{0\le t\le 1} |X(t)|^{C_1}\} < \infty$ when $C > C_1$.■

Theorem 2.5. If the random process $X(t)$ satisfies part (a) of (2.1), then for each integer $C > 0$, with

$$\widehat{\Delta}^2_{\text{sup}} \equiv \|\widehat{K} - K\|^2_{\text{sup}} = \sup_u \int \{\widehat{K}(u,v) - K(u,v)\}^2\, dv,$$

we have

$$E(\widehat{\Delta}^C_{\text{sup}}) < \text{const.}\, n^{-C/2}\,. \tag{2.28}$$

Moreover, from the first part of (2.1) we see that

$$E(\widehat{\Delta}^C) < \text{const.}\, n^{-C/2}\,. \tag{2.29}$$

(Recall that $\widehat{\Delta}^2 \equiv \|\widehat{K} - K\|^2 = \int \{\widehat{K}(u,v) - K(u,v)\}^2$.)

Proof: We have: $\widehat{K}(u,v) = \widetilde{K}(u,v) - \{\bar{X}(u) - \eta(u)\}\{\bar{X}(v) - \eta(v)\}$, where $\widetilde{K}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\{\bar{X}_i(u) - \eta(u)\}\{\bar{X}_i(v) - \eta(v)\}$. Also, $\widehat{\Delta}_{\text{sup}} \le \widehat{\Delta}^{(1)}_{\text{sup}} + \widehat{\Delta}^{(2)}_{\text{sup}}$, where

$\widehat{\Delta}^{(1)}_{\text{sup}} = \|\tilde{K} - K\|_{\text{sup}}$ and $\widehat{\Delta}^{(2)}_{\text{sup}} = \|\{\bar{X}(u) - \eta(u)\}\{\bar{X}(v) - \eta(v)\}\|_{\text{sup}}$. So,

$$E\big[\widehat{\Delta}_{\text{sup}}\big]^C \leq \text{const.}\, E\big[\widehat{\Delta}^{(1)}_{\text{sup}}\big]^C + E\big[\widehat{\Delta}^{(2)}_{\text{sup}}\big]^C, \tag{2.30}$$

where the constant depends on C. Regarding $\widehat{K}_1 \equiv \big(\widetilde{K} - K\big) = \frac{1}{n}\sum_{i=1}^{n} W_i(u,v)$, where $W_i(u,v) = \{X_i(u) - \eta(u)\}\{X_i(v) - \eta(v)\} - K(u,v)$, we have

$$E\big[\widehat{\Delta}^{(1)}_{\text{sup}}\big]^C = E\big[\|\widehat{K}_1\|_{\text{sup}}\big]^C = n^{-C/2}\, E\big[\sup_u \widehat{L}(u)\big]^C,$$

where $\widehat{L}(u) = \left(n \int \widehat{K}_1(u,v)^2\, dv\right)^{1/2}$. Thus, it is enough to prove that

$$E\big[\sup_u \widehat{L}(u)\big]^C < \infty. \tag{2.31}$$

To do that, we need the following Lemma.

Lemma 2.3. For each $s, t \in [0,1]$ and each $C \geq 1$,

$$E\big|\widehat{L}(s)\big|^C \leq C_1, \quad E\big|\widehat{L}(s) - \widehat{L}(t)\big|^C \leq C_1\, |s-t|^{C\epsilon}, \tag{2.32}$$

where $\epsilon > 0$ is as at (2.1) and $C_1 = C_1(C)$ does not depend on s or t.

Proof of Lemma: For each fixed $u \in [0,1]$ and for $C = 2k$, by using Holder's inequality, we have

$$\begin{aligned}
E\big|\widehat{L}(s)\big|^C &= n^k\, E\Big[\int \widehat{K}_1(s,v)^2\, dv\Big]^k \\
&\leq n^k \int \cdots \int E\big|\widehat{K}_1(s,v_1)^2 \times \cdots \times \widehat{K}_1(s,v_k)^2\big|^k\, dv_1 \cdots dv_k \\
&\leq n^k \int \cdots \int \big\{E\big|\widehat{K}_1(s,v_1)\big|^{2k} \times \cdots \times E\big|\widehat{K}_1(s,v_k)\big|^{2k}\big\}^{1/k}\, dv_1 \cdots dv_k\,,
\end{aligned}$$

and then by the relationship between arithmetic and geometric means, we conclude that

$$
\begin{aligned}
E\big|\widehat{L}(s)\big|^{C} &\le n^k \frac{1}{k}\int\cdots\int \big\{E\big|\widehat{K}_1(s,v_1)\big|^{2k}+\cdots+E\big|\widehat{K}_1(s,v_k)\big|^{2k}\big\}\,dv_1\cdots dv_k \\
&= n^k\left\{\int E|\widehat{K}_1(s,v)|^{2k}\,dv\right\}. \qquad (2.33)
\end{aligned}
$$

Using Rosenthal's inequality for each fixed $s, v \in [0,1]$ gives

$$
\begin{aligned}
E\Big|\frac{1}{n}\sum_{i=1}^{n} W_i(s,v)\Big|^{2k} &\le n^{-2k}\,C_{1k}\left\{\sum_{i=1}^{n} E\big|W_i(s,v)\big|^{2k}+\left(\sum_{i=1}^{n} E\big|W_i(s,v)\big|^{2}\right)^{k}\right\} \\
&\le n^{-k}\,C_{2k}\left\{E\big|W(s,v)\big|^{2k}+\left(E\big|W(u,v)\big|^{2}\right)^{k}\right\} \\
&\le n^{-k}\,C_{3k}\,E\big|W(s,v)\big|^{2k}\le n^{-k}\,C_{4k}\,,
\end{aligned}
$$

where W is a generic version of the W_i, and we have used (2.1) to obtain the last inequality above. Combining this result and (2.33) implies the first part of (2.32). For the second part of (2.32), we define $f(s,t,v)=\sum_{i=1}^{n}\{W_i(s,v)-W_i(t,v)\}$, then

$$
\begin{aligned}
E\big|\widehat{L}(s)-\widehat{L}(t)\big|^{2k} &\le n^k\,E\Big|\big[\int \widehat{K}_1(s,v)^2\,dv\big]^{1/2}-\big[\int \widehat{K}_1(t,v)^2\,dv\big]^{1/2}\Big|^{2k} \\
&\le n^{-k}\,E\Big|\int\big[\sum_{i=1}^{n}\{W_i(s,v)-W_i(t,v)\}\big]^2\,dv\Big|^{k} \\
&\le n^{-k}\,E\Big[\int\cdots\int f(s,t,v_1)^2\times\cdots\times f(s,t,v_k)^2\,dv_1\cdots dv_k\Big] \\
&\le n^{-k}\int\cdots\int E\big|f(s,t,v_1)^2\times\cdots\times f(s,t,v_k)^2\big|\,dv_1\cdots dv_k\,.
\end{aligned}
$$

Using Holder's inequality and the relationship between arithmetic and geometric means for the right-hand side of the last inequality above leads to

$$
E\big|\widehat{L}(s)-\widehat{L}(t)\big|^{2k}\le n^{-k}\int E\big|f(s,t,v)\big|^{2k}\,dv\,. \qquad (2.34)
$$

Furthermore, using Rosenthal's inequality for the integrand on the right-hand side of (2.34) for fixed $s, t, v \in [0,1]$ results in

$$\begin{aligned}
E\big|f(s,t,v)\big|^{2k} &\le C_{1k}\Big\{\sum_{i=1}^{n} E\big|W_i(s,v) - W_i(t,v)\big|^{2k} + \big(\sum_{i=1}^{n} E\big|W_i(s,v) - W_i(t,v)\big|^2\big)^k\Big\} \\
&\le C_{1k}\, n^k \Big\{E\big|W(s,v) - W(t,v)\big|^{2k} + \big(E\big|W(s,v) - W(t,v)\big|^2\big)^k\Big\} \\
&\le C_{2k}\, n^k\, E\big|W(s,v) - W(t,v)\big|^{2k}. \qquad (2.35)
\end{aligned}$$

Moreover, we have $W(s,v) - W(t,v) = Y(v)\,[Y(s) - Y(t)] + E\big[Y(v)\{Y(t) - Y(s)\}\big]$. Thus,

$$\begin{aligned}
E\big|W(s,v) - W(t,v)\big|^{2k} &\le C_{1k}\Big\{\big(E|Y(v)|^{4k}\big)^{1/2}\big(E|Y(s) - Y(t)|^{4k}\big)^{1/2} \\
&\qquad + \big(E\big[Y(v)\{Y(t) - Y(s)\}\big]\big)^{2k}\Big\} \\
&\le C_{2k}\Big\{\big(E|Y(v)|^{4k}\big)^{1/2}\big(E|Y(s) - Y(t)|^{4k}\big)^{1/2}\Big\} \\
&\le C_{3k}\,\big\{E|Y(s) - Y(t)|^{4k}\big\}^{1/2} \\
&\le C_{4k}\,|s - t|^{2k\epsilon}, \qquad (2.36)
\end{aligned}$$

where we have used (2.1) and the Cauchy-Schwarz inequality to obtain the above results. Finally, combining (2.34)-(2.36) gives the second part of (2.32). ∎

Markov's inequality and the second part of (2.32) imply that, for each $u, C > 0$,

$$P\{\big|\widehat{L}(s) - \widehat{L}(t)\big| > u\,|s - t|^{\epsilon/2}\} \le C_1\, u^{-C}|s - t|^{C\epsilon/2}. \qquad (2.37)$$

Put $t_{ki} = i\,2^{-k}$. In view of (2.37), we have for each $0 \le i \le 2^k - 1$,

$$P\{\big|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\big| > u\,|t_{ki} - t_{k,i+1}|^{\epsilon/2}\} \le C_1\, u^{-C}|t_{ki} - t_{k,i+1}|^{C\epsilon/2} = C_1\, u^{-C}\, 2^{-C\epsilon k/2}.$$

If we define $A_{ik} = \left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right|$, then

$$\begin{aligned} P\Big(\max_{0\le i<2^k-1} A_{ik} \ge u\,2^{\epsilon k/2}\Big) &\le P\Big(\bigcup_{i=0}^{2^k-1}\{A_{ik} \ge u\,2^{\epsilon k/2}\}\Big) \\ &\le \sum_{i=0}^{2^k-1} P\big(A_{ik} \ge u\,2^{\epsilon k/2}\big) \\ &\le C_1\,u^{-C}\sum_{i=0}^{2^k-1} 2^{-C\,\epsilon\,k/2}\,. \end{aligned}$$

If we choose $C \ge 4/\epsilon$, then $\sum_{i=0}^{2^k-1} 2^{-C\,\epsilon\,k/2} \le 2^{-C\,\epsilon\,k/4}$. Therefore,

$$P\Big(\max_{0\le i<2^k-1}\left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right| \ge u\,2^{\epsilon k/2}\Big) \le C_1\,u^{-C}\,2^{-C\,\epsilon\,k/4}\,.$$

Also, define $B_k \equiv \max_{0\le i<2^k-1}\left\{2^{\epsilon\,k/2}\left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right|\right\}$. Then, using the above result gives

$$\begin{aligned} P\Big(\sup_{0\le i<2^k-1,\,k\ge 0}\left\{2^{\epsilon\,k/2}\left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right|\right\} \ge u\Big) &\le P\Big(\bigcup_{k=0}^{\infty}\{B_k \ge u\}\Big) \le \sum_{k\ge 0} P(B_k \ge u) \\ &= \sum_{k\ge 0} P\Big(\max_{0\le i<2^k-1}\left\{2^{\epsilon\,k/2}\left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right|\right\} \ge u\Big) \\ &\le C_1\,u^{-C}\sum_{k=0}^{\infty} 2^{-C\,\epsilon\,k/4}. \end{aligned}$$

Thus, if we define $C_2 = C_1\sum_{k=0}^{\infty} 2^{-C\,\epsilon\,k/4}$, then

$$P\Big(\sup_{0\le i<2^k-1,\,k\ge 0}\left\{2^{\epsilon\,k/2}\left|\widehat{L}(t_{ki}) - \widehat{L}(t_{k,i+1})\right|\right\} \ge u\Big) \le C_2\,u^{-C}\,. \tag{2.38}$$

If $v \in [0,1)$ then we may express v in a dyadic expansion, say $v = \sum_{j=1}^{\infty} r_j\,2^{-j}$, where each $r_j = 0$ or 1. Write $\sum_{1\le j\le k} r_j\,2^{-j} = i_k(v)\,2^{-k}$, where the integers $i_1(v), i_2(v), \cdots$ satisfy $0 \le i_k(v) \le 2^k - 1$ and $i_{k+1}(v) = 2\,i_k(v) + r_{k+1}(v)$. Define $i_0(v) = 0$. Then, since X is left-continuous with probability 1, and

$v = \lim_{k\to\infty} t_{k,i_k(v)} = \lim_{k\to\infty} i_k(v)\, 2^{-k} = \lim_{k\to\infty} \sum_{j=1}^{k} r_j\, 2^{-j}$,

$$\begin{aligned}
\big|\widehat{L}(v) - \widehat{L}(0)\big| &\le \sum_{k=0}^{\infty} \big|\widehat{L}(t_{k,i_k(v)}) - \widehat{L}(t_{k+1,i_{k+1}(v)})\big| \\
&\le \sum_{k=0}^{\infty} \big|\widehat{L}(t_{k,2i_k(v)}) - \widehat{L}(t_{k+1,2i_k(v)+1})\big| \\
&\le \sum_{k=0}^{\infty} \max_{0\le i\le 2^{k+1}-1} \big|\widehat{L}(t_{k+1,i}) - \widehat{L}(t_{k+1,i+1})\big| .
\end{aligned}$$

Hence, by (2.38) we have

$$\begin{aligned}
&P\big(\big|\widehat{L}(v) - \widehat{L}(0)\big| > \sum_{k=0}^{\infty} u\, 2^{-\epsilon k/2}\big) \\
&\quad\le P\big(\sum_{k=0}^{\infty} \max_{0\le i\le 2^{k+1}-1} \big|\widehat{L}(t_{k+1,i}) - \widehat{L}(t_{k+1,i+1})\big| > \sum_{k=0}^{\infty} u\, 2^{-\epsilon k/2}\big) \\
&\quad\le P\big(\bigcup_{k\ge 0} \{\max_{0\le i\le 2^{k+1}-1} \big|\widehat{L}(t_{k+1,i}) - \widehat{L}(t_{k+1,i+1})\big| > u\, 2^{-\epsilon k/2}\}\big) \\
&\quad\le \sum_{k=0}^{\infty} P\big(\{\max_{0\le i\le 2^{k+1}-1} \big|\widehat{L}(t_{k+1,i}) - \widehat{L}(t_{k+1,i+1})\big| > u\, 2^{-\epsilon k/2}\}\big) \\
&\quad\le \sum_{k=0}^{\infty} C_1\, u^{-C}\, 2^{-C\epsilon k/2} \le C_2\, u^{-C} ,
\end{aligned}$$

where $C_2 = C_1 \sum_{k=0}^{\infty} 2^{-C\epsilon k/4}$. So, with $C_3 = \sum_{k=0}^{\infty} 2^{-\epsilon k/2}$ we obtained that with probability at least $1 - C_2\, u^{-C}$, $\big|\widehat{L}(v) - \widehat{L}(0)\big| \le C_3\, u$. That is, for each $C > 0$ there exists $C_3 > 0$ such that

$$P\big(\sup_{0\le v<1} \big|\widehat{L}(v) - \widehat{L}(0)\big| > C_3\, u\big) \le C_2\, u^{-C} .$$

Define $T = \sup_{0\le v<1} \big|\widehat{L}(v) - \widehat{L}(0)\big|$. Then,

$$E\big[T^C\big] \le \sum_{n=1} P(|T| \ge n^{1/C}) \le \sum_{n=1}^{\infty} C_4\, n^{-C_1/C} .$$

If we choose $C_1 > C$, then $\sum_{n=1}^{\infty} n^{-C_1/C} < \infty$. Therefore, by Theorem 3.2.1 of Chung 1974, it follows that $E\{\sup_{0\le v<1} |\widehat{L}(v) - \widehat{L}(0)|\}^C < \infty$ for each $C > 0$, and hence, (2.31) holds by the first part of (2.32). On the other hand,

$$\begin{aligned} E[\widehat{\Delta}^{(2)}_{\text{sup}}]^C &= E\big[\sup_u |\bar{X}(u) - \eta(u)|^2 \big\{ \int (\bar{X}(u) - \eta(u))^2\, dv\big\}\big]^{C/2} \\ &= \Big\{ E\big[\sup_u |\bar{X}(u) - \eta(u)|\big]^{2C} E\big[\big\{ \int (\bar{X}(u) - \eta(u))^2\, dv\big\}\big]^C \Big\}^{1/2}. \end{aligned} \tag{2.39}$$

If we define $Y_i(u) = X_i(u) - \eta(u)$, then the second term on the right-hand side of (2.39) can be bounded as follows:

$$\begin{aligned} E\big[\int \bar{Y}(v)^2\, dv\big]^{2k} &= E\big[\int \cdots \int \bar{Y}(v_1)^2 \cdots \bar{Y}(v_{2k})^2\, dv_1 \cdots dv_{2k}\big] \\ &\le \int \cdots \int \Big\{ E|\bar{Y}(v_1)|^{4k} \times \cdots \times E|\bar{Y}(v_{2k})|^{4k} \Big\}^{1/2k} dv_1 \cdots dv_{2k} \\ &\le \frac{1}{2k} \int \cdots \int \Big\{ E|\bar{Y}(v_1)|^{4k} + \cdots + E|\bar{Y}(v_{2k})|^{4k} \Big\}\, dv_1 \cdots dv_{2k} \\ &= \int E|\bar{Y}(u)|^{4k}\, dv\,. \end{aligned} \tag{2.40}$$

Moreover, for each fixed $v \in [0, 1]$ we can write

$$\begin{aligned} E|\bar{Y}(u)|^{4k} &= n^{-4k} E\Big|\sum_{i=1}^n Y_i\Big|^{4k} \le C_{1k}\, n^{-4k} \Big\{ \sum_{i=1}^n E|Y_i(v)|^{4k} + \big(\sum_{i=1}^n E|Y_i(v)|^2\big)^{2k} \Big\} \\ &\le C_{2k}\, n^{-2k} \{E|Y(v)|^{4k} + (E|Y(v)|^2)^{2k}\} \\ &\le C_{3k}\, n^{-2k} E|Y(v)|^{4k} \le C_{4k}\, n^{-2k}\,, \end{aligned} \tag{2.41}$$

where we have used Rosenthal's inequality and (2.1) to obtain the above results. For the first term on the right-hand side of (2.39) we have:

$$
\begin{aligned}
E\big[\sup_u |\bar{Y}(u)|\big]^{2C} &= E\big[\sup_u \big|\frac{1}{n}\sum_{i=1}^{n} Y_i(u)\big|^{2C}\big] \\
&\le E\big[\sup_u \big(\frac{1}{n}\sum_{i=1}^{n} |Y_i(u)|^{2C}\big)\big] \\
&\le E\big[\frac{1}{n}\sum_{i=1}^{n} \sup_u |Y_i(u)|^{2C}\big] \\
&= E\big[\sup_u |Y(u)|^{2C}\big] < \infty\,. \qquad (2.42)
\end{aligned}
$$

Combining (2.39)– (2.42) implies that $E\big[\widehat{\Delta}^{(2)}_{\sup}\big]^C = O(n^{-C/2})$. This finishes the proof of Theorem 2.5 for $C = 2k$, where $k > 0$ is an integer. Finally, if $C = 2k - 1$, then

$$
E[\widehat{\Delta}^{2k-1}] \le \Big(E[\widehat{\Delta}^{2k}]\Big)^{(2k-1)/2k} \le \big(\text{const. } n^{-k}\big)^{(2k-1)/2k} \le \text{const. } n^{-(2k-1)/2}\,.
$$

A similar but simpler argument can be used to prove (2.29). This finishes the proof of Theorem 2.5. ∎

2.5 Application of the Auxiliary Results to Empirical Approximations

In Section 2.3 one can interpret L, λ_j and ϕ_j as $\widehat{K}$, $\hat{\theta}_j$ and $\widehat{\psi}_j$, respectively. Recall that $\sup_j |\hat{\theta}_j - \theta_j| \le \widehat{\Delta}$, and note that, since $\mathcal{I} = [0,1]$, we have $|\mathcal{I}| = 1$. Then, for example, (2.6) and (2.11)–(2.13) imply that, for each candidate for χ_j,

$$
\begin{aligned}
&\Big|(\hat{\theta}_j - \theta_j)\Big(1 + \int \chi_j\,\psi_j\Big) - \int (\widehat{K} - K)(\psi_j + \chi_j)\,\psi_j\Big| \\
&\qquad\qquad \le \|\widehat{\psi}_j - \psi_j - \chi_j\|\,(\widehat{\Delta} + \widehat{\Delta}_{\sup})\,, \qquad (2.43)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\psi}_j - \psi_j = \theta_j^{-1} \sum_{s=0}^{\infty} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j} \right)^s \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \\
+ \sum_{s=0}^{\infty} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\} \\
\times\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j \,, \qquad (2.44)
\end{aligned}
$$

and, provided $\widehat{\Delta} \leq \min(\theta_j - \theta_{j+1}, \theta_{j-1} - \theta_j)$,

$$
\begin{aligned}
\left| \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \right| &= \left| \int (\widehat{K} - K)\, \widehat{\psi}_j - \psi_j \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_j \right| \\
&\leq \left| \int (\widehat{K} - K)\, \widehat{\psi}_j \right| + \left| \psi_j \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_j \right| \\
&\leq \left\{ \int (\widehat{K} - K)^2(u,v)\, dv \right\}^{1/2} \left\{ \int \widehat{\psi}_j^2(v)\, dv \right\}^{1/2} \\
&\quad + \sup_{u\in\mathcal{I}} |\psi_j(u)| \left\{ \int \psi_j^2(u) \right\}^{1/2} \\
&\qquad \times \left[\int \left\{ \int (\widehat{K} - K)(u,v)\, \widehat{\psi}_j(v)\, dv \right\}^2 du \right]^{1/2} .
\end{aligned}
$$

Thus,

$$
\left| \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \right| \leq \left\{ 1 + \sup_{u\in\mathcal{I}} |\psi_j(u)| \right\} \widehat{\Delta}_{\sup} \,. \qquad (2.45)
$$

Also, using (2.16), we deduce that

$$
\begin{aligned}
|\theta_j - \hat{\theta}_j|^s \sum_{k\,:\,k\neq j} \left| \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \right| \leq \\
\max\{\zeta_j^{-(s+1)}(1 - \zeta_j)^{-1/2}, \eta_j^{-(s+1)} + 1\}\, \theta_j^{-3/2} \left\{ \sup_{u\in\mathcal{I}} K(u,u) \right\}^{1/2} (\widehat{\Delta}/\theta_j)^s\, \widehat{\Delta} \,.
\end{aligned}
$$

(2.46)

For $\ell \geq 0$, put

$$\chi_{j\ell} = \theta_j \sum_{s=0}^{\ell} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j} \right)^s \sum_{k\,:\,k \neq j} \psi_j \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ \sum_{s=0}^{\ell} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k \neq j} \left\{ (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)} \right\}$$
$$\times \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j\,. \tag{2.47}$$

For a function α, define $\|\alpha\|_{\sup} = \sup_u |\alpha(u)|$. We know from (2.44)–(2.46) and the bound $|\hat{\theta}_j - \theta_j| \leq \widehat{\Delta}$ that, provided

$$(\widehat{\Delta}/\theta_j)\, \max(\zeta_j^{-1}, \eta_j^{-1}) \leq \tfrac{1}{2}, \tag{2.48}$$

we have

$$\|\widehat{\psi}_j - \psi_j - \chi_{j\ell}\|_{\sup} = \Bigg\| \theta_j^{-1} \sum_{s=\ell+1}^{\infty} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j} \right)^s \sum_{k\,:\,k \neq j} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ \sum_{s=\ell+1}^{\infty} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k \neq j} \left\{ (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)} \right\} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \Bigg\|_{\sup}.$$

Therefore,

$$\|\widehat{\psi}_j - \psi_j - \chi_{j\ell}\|_{\sup} \leq \theta_j^{-1} \sum_{s=\ell+1}^{\infty} \left| \frac{\theta_j - \hat{\theta}_j}{\theta_j} \right|^s \{1 + \sup_{u \in \mathcal{I}} |\psi_j|\}\, \widehat{\Delta}_{\sup}$$
$$+ \sum_{s=\ell+1}^{\infty} \max\{\zeta_j^{-(s+1)}\, (1 - \zeta_j)^{-1/2}, \eta_j^{-(s+1)} + 1\}\, \theta_j^{-3/2}$$
$$\times \Big\{ \sup_{u \in \mathcal{I}} K(u, u) \Big\}^{1/2} (\widehat{\Delta}/\theta_j)^s\, \widehat{\Delta}$$
$$\leq v_{j\ell}\, V_\ell, \tag{2.49}$$

where $V_\ell = \widehat{\Delta}^{\ell+1}\,\widehat{\Delta}_{\rm sup}$, and, for a constant C not depending on j,

$$v_{j\ell} = C\,\{\max(\zeta_j^{-1}, \eta_j^{-1})\}^{\ell+2}\,(1-\zeta_j)^{-1/2}\,(1+\theta_j^{-3/2})\,\theta_j^{-(\ell+1)}$$
$$\times\,\{1 + \sup_{u\in\mathcal{I}} |\psi_j(u)| + \sup_{u\in\mathcal{I}} K(u,u)^{1/2}\}\,.$$

Note that (2.48) is equivalent to the condition (2.2).

2.6 Approximation to $\hat{\theta}_j$

Substituting $\chi_j = \chi_{j\ell}$ at (2.47) into (2.43) and using (2.49) to bound the right-hand side of (2.43), we obtain an approximation to $\hat{\theta}_j - \theta_j$. Provided (2.48) holds,

$$\left|(\hat{\theta}_j - \theta_j)\Big(1 + \int \chi_{j\ell}\,\psi_j\Big) - \int (\widehat{K} - K)(\psi_j + \chi_{j\ell})\,\psi_j\right| \leq v_{j\ell}\,V_\ell\,(\widehat{\Delta} + \widehat{\Delta}_{\rm sup})$$
$$\leq 2\,v_{j\ell}\,V_\ell\,\widehat{\Delta}_{\rm sup}\,. \qquad (2.50)$$

Note too that, by (2.17), we have

$$\|\widehat{\psi}_j - \psi_j\| \leq 8^{-1/2}\,\frac{\widehat{\Delta}}{\theta_j\,\min(\frac{\theta_{j-1}}{\theta_j} - 1, 1 - \frac{\theta_{j+1}}{\theta_j})} \leq 8^{-1/2}\,(\widehat{\Delta}/\theta_j)\,\max(\zeta_j^{-1}, \eta_j^{-1})\,, \qquad (2.51)$$

whence, writing "const." for a generic constant not depending on j or n but depending on ℓ,

$$
\begin{aligned}
&\left| \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\,\psi_k \right| \\
&\qquad = \left| \sum_{k=1}^{\infty} \psi_k \int (\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\,\psi_k - \psi_j \int (\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\,\psi_j \right| \\
&\qquad \le \|(\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\|_{\sup} + \|\psi_j\|_{\sup} \left| \int (\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\,\psi_j \right| \\
&\qquad \le \|\widehat{K}-K\|_{\sup}\,\|\widehat{\psi}_j - \psi_j\| + \|\psi_j\|_{\sup} \|\widehat{K}-K\|_{\sup}\,\|\widehat{\psi}_j - \psi_j\| \\
&\qquad \le \text{const.}\,(\widehat{\Delta}/\theta_j^2)\,\max(\zeta_j^{-1}, \eta_j^{-1})\,\widehat{\Delta}_{\sup}\,. \qquad (2.52)
\end{aligned}
$$

(Here we have used (2.17), and the fact that $\|\psi_j\|_{\sup} \le \|K\|_{\sup}/\theta_j$.) Also, with $c_k = (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}$ for $k \neq j$ and $c_j = 0$, we have:

$$
\begin{aligned}
&\left| \sum_{k\,:\,k\neq j} c_k\,\psi_k \int (\widehat{K}-K)(\widehat{\psi}_j - \psi_j)\,\psi_k \right|^2 \\
&\qquad = \left| c_k \int (\widehat{K}-K)(u,v)(\widehat{\psi}_j - \psi_j)(v)\,dv \right|^2 \\
&\qquad \le c_k^2 \left| \int (\widehat{K}-K)^2(u,v)\,dv \right| \|\widehat{\psi}_j - \psi_j\|^2 \\
&\qquad \le c_k^2\,\widehat{\Delta}_{\sup}^2\,\|\widehat{\psi}_j - \psi_j\|^2 \\
&\qquad \le \widehat{\Delta}_{\sup}^2\,\|\widehat{\psi}_j - \psi_j\|^2 \sum_{k\,:\,k\neq j} c_k^2\,. \qquad (2.53)
\end{aligned}
$$

Below we show that $\sum_{k=1}^{\infty} c_k^2 \le C_s \left\{ \max(\zeta_j^{-1}, \eta_j^{-1})^{s+1}\,\theta_j^{-(s+2)} \right\}^2$.

Lemma 2.4. For an integer $s \ge 1$ if $c_k = (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}$ for $k \neq j$ and $c_j = 0$, then

$$
\sum_{k=1}^{\infty} c_k^2 \le C_s \left\{ \max(\zeta_j^{-1}, \eta_j^{-1})^{s+1}\,\theta_j^{-(s+2)} \right\}^2, \qquad (2.54)
$$

where C_s denotes a constant depending only on s.

Proof of Lemma:

$$
\begin{aligned}
\sum_{k\,:\,k\neq j} c_k^2 &= \sum_{k\,:\,k\neq j} \left\{(\theta_j-\theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\right\}^2 \\
&= \sum_{k<j} \left\{(\theta_j-\theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\right\}^2 + \sum_{k>j} \left\{(\theta_j-\theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\right\}^2 .
\end{aligned} \tag{2.55}
$$

For the first series on the right-hand side of (2.55), related to $k<j \equiv \theta_k \geq \theta_j$, provided $\sum_{k=1}^{\infty} \theta_k^2 \leq C$, by using Chebychev's inequality, we have

$$
\#\{k : k<j\} = \#\{k : \theta_k \geq \theta_j\} \leq \frac{\sum_{k=1}^{\infty} \theta_k^2}{\theta_j} \leq \frac{C}{\theta_j^2} .
$$

Hence,

$$
\begin{aligned}
\sum_{k<j} (\theta_j-\theta_k)^{-2(s+1)} &\leq \frac{C}{\theta_j^2} \frac{C}{\theta_j^{2(s+1)}} \left(\inf_{k<j} \{(\theta_k/\theta_j) - 1\} \right)^{-2(s+1)} \\
&= \frac{C}{\theta_j^{2(s+2)}} \left(\inf_{k<j} \{(\theta_k/\theta_j) - 1\} \right)^{-2(s+1)} \\
&= C\,\theta_j^{-2(s+2)}\,\eta_j^{-2(s+1)} ,
\end{aligned} \tag{2.56}
$$

and

$$
\sum_{k<j} \theta_j^{-2(s+1)} \leq \frac{C}{\theta_j^2} \frac{1}{\theta_j^{2(s+1)}} = C\,\theta_j^{-2\,(s+2)} . \tag{2.57}
$$

For $k>j \equiv \theta_k \leq \theta_j$, define $\nu_{jk} = (1-\frac{\theta_k}{\theta_j})^{-1}$. Then,

$$
\begin{aligned}
\sum_{k>j} \big\{ (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)} \big\}^2 &= \theta_j^{-2(s+1)} \sum_{k:\, k>j} \big\{ (1 - \frac{\theta_k}{\theta_j})^{-(s+1)} - 1 \big\}^2 \\
&= \theta_j^{-2(s+1)} \sum_{k:\, k>j} \big\{ \nu_{jk}^{s+1} - 1 \big\}^2 \\
&= \theta_j^{-2(s+1)} \sum_{k:\, k>j} \big\{ (\nu_{jk} - 1)(1 + \nu_{jk} + \cdots + \nu_{jk}^s) \big\}^2 ,
\end{aligned}
$$

which is simplified to:

$$
\begin{aligned}
\sum_{k>j} \big\{ (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)} \big\}^2 &= \theta_j^{-2(s+1)} \sum_{k:\, k>j} \big\{ (\frac{\frac{\theta_k}{\theta_j}}{1 - \frac{\theta_k}{\theta_j}})^2 (1 + \nu_{jk} + \cdots + \nu_{jk}^s)^2 \big\} \\
&= \theta_j^{-2(s+1)-2} \sum_{k:\, k>j} \big\{ \theta_k^2 \, \nu_{jk}^2 \, (1 + \nu_{jk} + \cdots + \nu_{jk}^s)^2 \big\} \\
&= \theta_j^{-2(s+2)} \sum_{k:\, k>j} \big\{ \theta_k^2 \, \nu_{jk}^2 \, \nu_{jk}^2 \, (\nu_{jk}^{-1} + 1 + \nu_{jk} + \cdots + \nu_{jk}^s)^2 \big\} .
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\sum_{k>j} \big\{ (\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)} \big\}^2 &\le \theta_j^{-2(s+2)} \sum_{k:\, k>j} \big\{ \theta_k^2 \, \nu_{jk}^4 \, (1 + 1 + \nu_{jk} + \cdots + \nu_{jk}^s)^2 \big\} \\
&\le \theta_j^{-2(s+2)} \sum_{k:\, k>j} \big\{ \theta_k^2 \, \nu_{jk}^4 \, (s+1)^2 \, \nu_{jk}^{2(s-1)} \big\} \\
&= (s+1)^2 \, \theta_j^{-2(s+2)} \sum_{k:\, k>j} \big\{ \theta_k^2 \, \nu_{jk}^{2(s+1)} \big\} \\
&\le (s+1)^2 \, \theta_j^{-2(s+2)} \, \zeta_j^{-2(s+1)} \sum_{k:\, k>j} \theta_k^2 \\
&\le (s+1)^2 \, \theta_j^{-2(s+2)} \, \zeta_j^{-2(s+1)} \Big\{ \sum_{k:\, k>j} \theta_k^2 \Big\} , \\
&\le C_s \, \theta_j^{-2(s+2)} \, \zeta_j^{-2(s+1)} , \qquad (2.58)
\end{aligned}
$$

where we have used the fact that $\nu_{jk} = \frac{1}{1-\frac{\theta_k}{\theta_j}} \le \frac{1}{\inf_{k:\, k>j}(1-\frac{\theta_k}{\theta_j})} = \zeta_j^{-1}$. Finally, combining (2.55)–(2.58) finishes the proof. ∎

Therefore, combining (2.53) and (2.54) results in

$$\left| \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\} \, \psi_k \int (\widehat{K} - K)(\widehat{\psi}_j - \psi_j)\, \psi_k \right|$$
$$\leq C_s \, \widehat{\Delta}\, \widehat{\Delta}_{\sup} \, \max(\zeta_j^{-1}, \eta_j^{-1})^{s+2} \, \theta_j^{-(s+3)}. \tag{2.59}$$

Let

$$\begin{aligned}\chi_{j\ell}^{[1]} &= \theta_j^{-1} \sum_{s=0}^{\ell} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j} \right)^s \sum_{k\,:\,k\neq j} \psi_j \int (\widehat{K} - K)\, \psi_j \, \psi_k \\ &\quad + \sum_{s=0}^{\ell} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\} \\ &\qquad \times \psi_k \int (\widehat{K} - K)\, \psi_j \, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j. \end{aligned} \tag{2.60}$$

Then, by (2.47), (2.52), (2.59), (2.60) and (2.17) if (2.48) holds,

$$\begin{aligned}\|\chi_{j\ell} - \chi_{j\ell}^{[1]}\|_{\sup} &= \Bigg\| \theta_j^{-1} \sum_{s=0}^{\ell} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j} \right)^s \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)(\widehat{\psi}_j - \psi_j)\, \psi_k \\ &\quad + \sum_{s=0}^{\ell} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\} \psi_k \int (\widehat{K} - K)(\widehat{\psi}_j - \psi_j)\, \psi_k \Bigg\|_{\sup} \\ &\leq \theta_j^{-1} \sum_{s=0}^{\ell} \left| \frac{\theta_j - \hat{\theta}_j}{\theta_j} \right|^s \left(\frac{\widehat{\Delta}}{\theta_j^2} \right) \max(\zeta_j^{-1}, \eta_j^{-1})\, \widehat{\Delta}_{\sup} \\ &\quad + \sum_{s=0}^{\ell} |\theta_j - \hat{\theta}_j|^s \, \widehat{\Delta}_{\sup} \, \widehat{\Delta} \, \max(\zeta_j^{-1}, \eta_j^{-1})^{s+2} \, \theta_j^{-(s+3)} \\ &\leq \text{const.} \sum_{s=0}^{\ell} \widehat{\Delta}^{s+1} \, \widehat{\Delta}_{\sup} \, \max(\zeta_j^{-1}, \eta_j^{-1})^{s+2} \, \theta_j^{-(s+3)}, \end{aligned} \tag{2.61}$$

Moreover, the following property holds:

if we were to add to $\chi_{j\ell}^{[1]}$ the quantity

$$\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\,\psi_k \int (\widehat{K} - K)(\widehat{\psi}_j - \psi_j)\,\psi_k\,, \tag{2.62}$$

then, using (2.17) and (2.52) implies that

$$\|\chi_{j\ell} - \chi_{j\ell}^{[1]}\|_{\sup} \leq \text{const.} \sum_{s=1}^{\ell} \widehat{\Delta}^{s+1}\,\widehat{\Delta}_{\sup}\,\max(\zeta_j^{-1}, \eta_j^{-1})^{s+2}\,\theta_j^{-(s+3)}\,,$$

which means that (2.61) would continue to hold if the term corresponding to $s = 0$ were dropped from the series on the right-hand side.

Replacing $\chi_{j\ell}$ by $\chi_{j\ell}^{[1]}$ in (2.50), taking $\ell = 1$, and noting (2.17) and (2.61), we deduce that if (2.48) holds then

$$\begin{aligned}
&\left|(\hat{\theta}_j - \theta_j)\left(1 + \int \chi_{j1}^{[1]}\,\psi_j\right) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[1]})\,\psi_j\right| \\
&\quad = \left|(\hat{\theta}_j - \theta_j)\left(1 + \int (\chi_{j1}^{[1]} - \chi_{j1} + \chi_{j1})\,\psi_j\right) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[1]} - \chi_{j1} + \chi_{j1})\,\psi_j\right| \\
&\quad \leq \left|(\hat{\theta}_j - \theta_j)\left(\int (\chi_{j1}^{[1]} - \chi_{j1})\,\psi_j\right) - \int (\widehat{K} - K)(\chi_{j1}^{[1]} - \chi_{j1})\,\psi_j\right| \\
&\qquad + \left|(\hat{\theta}_j - \theta_j)\left(1 + \int \chi_{j1}\,\psi_j\right) - \int (\widehat{K} - K)(\psi_j + \chi_{j1})\,\psi_j\right| \\
&\quad \leq \text{const.}\,\widehat{\Delta}\,\|\chi_{j1}^{[1]} - \chi_{j1}\|_{\sup}\,\sup_u |\psi_j(u)| + \int \left\{\int (\widehat{K} - K)^2(u,v)\,dv\right\}^{1/2} \\
&\qquad\qquad \times \left\{\int (\chi_{j1}^{[1]} - \chi_{j1})^2(v)\,dv\right\}^{1/2} \psi_j(u)\,du + 2\,v_{j1}\,V_1\,\widehat{\Delta}_{\sup} \\
&\quad \leq \text{const.}\,\widehat{\Delta}\,\widehat{\Delta}_{\sup} \sum_{s=0}^{1} \widehat{\Delta}^{s+1}\,\max(\zeta_j^{-1}, \eta_j^{-1})^{s+2}\,\theta_j^{-(s+3)}\,\sup_u |\psi_j(u)| \\
&\qquad + \widehat{\Delta}_{\sup}\,\|\chi_{j1}^{[1]} - \chi_{j1}\|_{\sup}\,\sup_{u\in\mathcal{I}} |\psi_j(u)| + 2\,v_{j1}\,V_1\,\widehat{\Delta}_{\sup}\,.
\end{aligned}$$

Thus,

$$\left|(\hat{\theta}_j - \theta_j)\left(1 + \int \chi_{j1}^{[1]} \psi_j\right) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[1]})\, \psi_j\right|$$

$$\leq \text{const.}\, \widehat{\Delta}^2\, \widehat{\Delta}_{\text{sup}} \max(\zeta_j^{-1}, \eta_j^{-1})^2\, \theta_j^{-3} \sup_u |\psi_j(u)| \sum_{s=0}^{1} \left(\frac{\widehat{\Delta}}{\theta_j} \max(\zeta_j^{-1}, \eta_j^{-1})\right)^s$$

$$+ \widehat{\Delta}_{\text{sup}}^2 \max(\zeta_j^{-1}, \eta_j^{-1})^2\, \widehat{\Delta}\, \theta_j^{-3} \sum_{s=0}^{1} \left(\frac{\widehat{\Delta}}{\theta_j} \max(\zeta_j^{-1}, \eta_j^{-1})\right)^s \sup_{u \in \mathcal{I}} |\psi_j(u)|$$

$$+ 2\, v_{j1}\, V_1\, \widehat{\Delta}_{\text{sup}} \leq D_{nj}\,, \tag{2.63}$$

where

$$D_{nj} = n^{-3/2}\, U_{nj}\, (1 - \zeta_j)^{-1/2} \max(\zeta_j^{-1}, \eta_j^{-1})^3\, \theta_j^{-7/2} \sup_{u \in \mathcal{I}} |\psi_j(u)|\,,$$

and U_{nj} denotes a generic random variable satisfying:

$$\sup_{n,\, j \geq 1} E(U_{nj}^m) < \infty\,, \quad \text{for each integer } m \geq 1. \tag{2.64}$$

In (2.63) it suffices to take $U_{nj} = \text{const.}\, (n^{1/2}\, \widehat{\Delta}_{\text{sup}})^3$, not depending on j; this result uses (2.28) and (2.29). (More generally, U_{nj} will alter from one appearance to the next in versions of D_{nj} below.)

It should be mentioned that since $\theta_j\, |\psi_j|^2 \leq K_{\text{diag}}$, where $K_{\text{diag}} \equiv K(u, u) = \text{Var}(X(u))$, then $\|\psi_j\|_{\text{sup}} \leq \theta_j^{-1/2}\, \|K_{\text{diag}}\|_{\text{sup}}^{1/2}$. Therefore, $\|\psi_j\|_{\text{sup}}$ in D_{nj} can be replaced by $\theta_j^{-1/2}$, because $\|K_{\text{diag}}\|_{\text{sup}} < \infty$ by (2.1).

Note that $\int (\widehat{\psi}_j - \psi_j)\, \psi_j = \int \widehat{\psi}_j\, \psi_j - 1 \leq \|\widehat{\psi}_j\|\, \|\psi_j\| - 1 \leq 1 - 1 = 0$. Bearing this sign in mind, we have:

$$
\begin{aligned}
\tfrac{1}{2}\,\|\widehat{\psi}_j - \psi_j\|^2 = \tfrac{1}{2}\,\langle \widehat{\psi}_j - \psi_j, \widehat{\psi}_j - \psi_j\rangle &= \tfrac{1}{2}\,\{-\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle + \langle \widehat{\psi}_j - \psi_j, \widehat{\psi}_j\rangle\} \\
&= -\tfrac{1}{2}\,\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle + \tfrac{1}{2}\,\{\langle \widehat{\psi}_j, \widehat{\psi}_j\rangle - \langle \widehat{\psi}_j, \psi_j\rangle\} \\
&= -\tfrac{1}{2}\,\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle + \tfrac{1}{2}\,\{\langle \psi_j, \psi_j\rangle - \langle \widehat{\psi}_j, \psi_j\rangle\} \\
&= -\tfrac{1}{2}\,\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle - \tfrac{1}{2}\,\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle \\
&= -\langle \widehat{\psi}_j - \psi_j, \psi_j\rangle = -\int (\widehat{\psi}_j - \psi_j)\,\psi_j \geq 0\,.
\end{aligned}
$$

From this result and (2.51) we see that

$$
\left|\int (\widehat{\psi}_j - \psi_j)\,\psi_j\right| = \tfrac{1}{2}\,\|\widehat{\psi}_j - \psi_j\|^2 \leq \text{const.}\,(\widehat{\Delta}/\theta_j)^2\, \max(\zeta_j^{-1}, \eta_j^{-1})^2\,. \tag{2.65}
$$

Let

$$
\begin{aligned}
\chi_{j\ell}^{[2]} = \theta_j^{-1} \sum_{s=0}^{\ell} \left(\frac{\theta_j - \hat{\theta}_j}{\theta_j}\right)^s \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\,\psi_j\,\psi_k \\
+ \sum_{s=0}^{\ell} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-(s+1)} - \theta_j^{-(s+1)}\}\,\psi_k \int (\widehat{K} - K)\,\psi_j\,\psi_k\,,
\end{aligned}
$$

which is $\chi_{j\ell}^{[1]}$, except for the last term which was dropped. We have

$$
\begin{aligned}
&\left|(\theta_j - \hat{\theta}_j)(1 + \int \chi_{j1}^{[2]}\,\psi_j) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[2]})\,\psi_j\right| \\
&\quad \leq \left|(\theta_j - \hat{\theta}_j)(1 + \int \chi_{j1}^{[1]}\,\psi_j) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[1]})\,\psi_j\right| \\
&\qquad + \left|(\theta_j - \hat{\theta}_j) \int (\widehat{\psi}_j - \psi_j)\,\psi_j + \left\{\int (\widehat{K} - K)\,\psi_j\,\psi_j\right\}\left\{\int (\widehat{\psi}_j - \psi_j)\,\psi_j\right\}\right| \\
&\quad \leq D_{nj} + \left|\int (\widehat{\psi}_j - \psi_j)\,\psi_j\right| \left|(\hat{\theta}_j - \theta_j) - \int (\widehat{K} - K)\,\psi_j\,\psi_j\right| \\
&\quad \leq D_{nj} + \tfrac{1}{2}\,\|\widehat{\psi}_j - \psi_j\|^2 \times 2\,\|\widehat{\psi}_j - \psi_j\|\,\|\widehat{K} - K\|_{\sup} \\
&\quad \leq D_{nj} + \text{const.}\,(\widehat{\Delta}/\theta_j)^3\, \max(\zeta_j^{-1}, \eta_j^{-1})^3\, \widehat{\Delta}_{\sup}\,,
\end{aligned}
$$

where we used (2.63), (2.65) and (2.18) to bound the terms on the right-hand side. We see that the terms on the right-hand side of the above inequality can be bounded by D_{nj} again. (The only change is that the value of D_{nj} is inflated by a multiplicative constant). That is:

$$\left|(\theta_j - \hat{\theta}_j)\left(1 + \int \chi_{j1}^{[2]} \psi_j\right) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[2]})\, \psi_j\right| \leq D_{nj}\,. \tag{2.66}$$

Lemma 2.5. Let $X(t)$ satisfy (2.1). Then, for each integer $k, j\,, m \geq 1$,

$$E\Big[\int (\widehat{K} - K)\, \psi_j\, \psi_k\Big]^{2m} = O(\{n^{-1}\, \theta_j\, \theta_k\}^m)\,,$$

where the θ_j and ψ_j are eigenvalues and eigenfunctions of the operator K.

Proof of Lemma: Define $B_{jk} = \int (\widehat{K} - K)\, \psi_j\, \psi_k$. We have

$$\begin{aligned}(\widehat{K} - K)(u, v) = \frac{1}{n} \sum_{i=1} n\{X_i(u) - \eta(u)\}\, \{X_i(v) - \eta(v)\} - K(u, v) \\ - \big(X_i(u) - \eta(u)\big)\, \big(X_i(v) - \eta(v)\big)\,.\end{aligned}$$

Thus,

$$\begin{aligned}B_{jk} &= \int\int (\widehat{K} - K)(u, v)\, \psi_j(v)\, \psi_k(u)\, du\, dv \\ &= \frac{1}{n}\sum_{i=1}^{n} \Big\{\int \big(X_i(u) - \eta(u)\big)\, \psi_k(u)\, du\Big\}\, \Big\{\int \big(X_i(v) - \eta(v)\big)\, \psi_j(v)\, dv\Big\} \\ &\quad - \theta_j\, \delta_{jk} - \Big\{\int \big(\bar{X}_i(u) - \eta(u)\big)\, \psi_k(u)\, du\Big\}\, \Big\{\int \big(\bar{X}_i(v) - \eta(v)\big)\, \psi_j(v)\, dv\Big\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \xi_{ik}\, \xi_{ij} - \bar{\xi}_k\, \bar{\xi}_j - \theta_j\, \delta_{jk}\,,\end{aligned}$$

where $\xi_{ij} = \int \big(X_i - \eta\big)\, \psi_j$ and $\bar{\xi}_j = \frac{1}{n}\sum_{i=1}^{n} \xi_{ij}$. If $k \neq j$ then,

$$
\begin{aligned}
E\big[B_{jk}^{2m}\big] = E\Big[\int (\widehat{K} - K)\,\psi_j\,\psi_k\Big]^{2m} &= E\Big[\frac{1}{n}\sum_{i=1}^{n} \xi_{ik}\,\xi_{ij} - \bar{\xi}_k\,\bar{\xi}_j\Big]^{2m} \\
&\leq C_m \Big\{ \Big(E\Big[\frac{1}{n}\sum_{i=1}^{n} \xi_{ik}\,\xi_{ij}\Big]^{2m} \Big) + \Big(E\big[\bar{\xi}_k\,\bar{\xi}_j\big]^{2m} \Big) \Big\}.
\end{aligned} \tag{2.67}
$$

By using Rosenthal's inequality and part (b) of (2.1) we have

$$
\begin{aligned}
E\Big|\frac{1}{n}\sum_{i=1}^{n} \xi_{ik}\,\xi_{ij}\Big|^{2m} &\leq C_{1m}\, n^{-2m} \Big\{ \sum_{i=1}^{n} E\big|\xi_{ij}^{2m}\,\xi_{ik}^{2m}\big| + \Big(\sum_{i=1}^{n} E[\xi_{ij}\,\xi_{ik}]^2 \Big)^m \Big\} \\
&\leq C_{1m}\, n^{-2m} \Big\{ \sum_{i=1}^{n} \big\{E\big|\xi_{ij}^{4m}\big|\, E\big|\xi_{ik}^{4m}\big|\big\}^{1/2} + \Big(\sum_{i=1}^{n} \big\{E[\xi_{ij}^4]\, E[\xi_{ik}^4]\big\}^{1/2} \Big)^m \Big\} \\
&\leq C_{2m}\, n^{-2m} \Big\{ \sum_{i=1}^{n} \theta_j^m\,\theta_k^m + \Big(\sum_{i=1}^{n} \theta_j\,\theta_k \Big)^m \Big\} \leq C_{3m}\, n^{-m}\,\theta_j^m\,\theta_k^m\,,
\end{aligned} \tag{2.68}
$$

and also,

$$
\begin{aligned}
E\big[\bar{\xi}_k\,\bar{\xi}_j\big]^{2m} &\leq \Big\{ E\big|\bar{\xi}_j\big|^{4m}\, E\big|\bar{\xi}_k\big|^{4m} \Big\}^{1/2} \\
&= n^{-4m} \Big\{ E\Big|\sum_{i=1}^{n} \xi_{ij}\Big|^{4m}\, E\Big|\sum_{i=1}^{n} \xi_{ij}\Big|^{4m} \Big\}^{1/2} \qquad (2.69) \\
&\leq C_{3m}\, n^{-4m} \Big\{ \sum_{i=1}^{n} E\big|\xi_{ij}\big|^{4m} + \big(\sum_{i=1}^{n} E[\xi_{ij}^2]\big)^{2m} \Big\}^{1/2} \times \\
&\qquad \Big\{ \sum_{i=1}^{n} E\big|\xi_{ik}\big|^{4m} + \big(\sum_{i=1}^{n} E[\xi_{ik}^2]\big)^{2m} \Big\}^{1/2} \\
&\leq C_{4m}\, n^{-2m} \Big\{ \sum_{i=1}^{n} \theta_j^{2m} + \big(\sum_{i=1}^{n} \theta_j\big)^{2m} \Big\}^{1/2} \Big\{ \sum_{i=1}^{n} \theta_k^{2m} + \big(\sum_{i=1}^{n} \theta_k\big)^{2m} \Big\}^{1/2} \\
&\leq C_{4m}\, n^{-2m}\,\theta_j^m\,\theta_k^m\,. \qquad (2.70)
\end{aligned}
$$

Combining (2.67)–(2.70) finishes the proof for $k \neq j$. When $k = j$, we have

$$E\Big\{\int (\widehat{K} - K)\, \psi_j\, \psi_j\Big\}^{2m} = E\Big\{\frac{1}{n}\sum_{i=1}^{n} (\xi_{ij}^2 - E(\xi_j^2)) - (\bar{\xi}_j)^2\Big\}^{2m},$$

where ξ_j denotes a generic ξ_{ij}. Hence, similar to the case $k \neq j$, by moment methods and using part (b) of (2.1), we see that

$$E\Big\{\int (\widehat{K} - K)\, \psi_j\, \psi_j\Big\}^{2m} \leq C_s\, (n^{-1}\, \theta_j^2)^m.$$

This finishes the proof. ■

Define $\|.\|_m \equiv \{E(.)^m\}^{1/m}$. By Minkowski's inequality $\|.\|_m$ is a semi-norm. Then, using Lemma 2.5, we conclude that:

$$\begin{aligned}
E\Big[\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\Big\{\int (\widehat{K} - K)\, \psi_j\, \psi_k\Big\}^2\Big]^m &= \Big(\|\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} B_{jk}^2\|_m\Big)^m \\
&\leq \Big(\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \|B_{jk}^2\|_m\Big)^m \\
&= O(\{n^{-1}\, \rho_j^{-2}\, \theta_j \sum_{k\,:\,k\neq j} \theta_k\}^m) = O(\{n^{-1}\, \rho_j^{-2}\, \theta_j\}^m). \qquad (2.71)
\end{aligned}$$

If (2.48) holds, working out the integrals on the left-hand side of (2.66) and moving a portion of $\int (\widehat{K} - K)\, \chi_{j\ell}^{[2]}\, \psi_j$ onto the right-hand side, we obtain the following result:

$$\begin{aligned}
&\Big|\hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\Big\{\int (\widehat{K} - K)\, \psi_j\, \psi_k\Big\}^2\Big| \\
&\quad \leq \Big|(\hat{\theta}_j - \theta_j)\,(1 + \int \chi_{j1}^{[2]}\, \psi_j) - \int (\widehat{K} - K)(\psi_j + \chi_{j1}^{[2]})\, \psi_j\Big| \\
&\qquad + \Big|\int (\widehat{K} - K)\, \chi_{j1}^{[2]}\, \psi_j - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\Big\{\int (\widehat{K} - K)\, \psi_j\, \psi_k\Big\}^2\Big|,
\end{aligned}$$

and then,

$$\begin{aligned}
&\left| \hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2 \right| \\
&\qquad \leq D_{nj} + |\theta_j - \hat{\theta}_j| \left| \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2 \right| \\
&\qquad \leq D_{nj} + n^{-3/2}\, \rho_j^{-2}\, \theta_j\, U_{nj} \leq \text{const.}\, D_{nj}\,, \qquad (2.72)
\end{aligned}$$

where U_{nj} is a generic random variable satisfying $\sup_{n,\,j\geq 1} E(U_{nj}^m) < \infty$ for all $m \geq 1$. Furthermore, by (2.71), we considered the random variable U_{nj} as

$$n^{3/2}\, \rho_j^2\, \theta_j^{-1}\, \widehat{\Delta} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2,$$

which has the properties ascribed to U_{nj} in (2.64). To obtain the above results, we also used (2.71) and (2.17) as well as (2.28) and (2.29).

2.7 Approximation to $\widehat{\psi}_j$

2.7.1 Bounds in Sup-norm

Let

$$\begin{aligned}
\chi_{j2}^{[3]} &= \theta_j^{-1} \{1 - \theta_j^{-1} \int (\widehat{K} - K)\, \psi_j\, \psi_j\} \{ \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \} \\
&\quad + \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-1} - \theta_j^{-1}\}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \\
&\quad - \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_j \right\} \left[\sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-2} - \theta_j^{-2}\}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \right], \\
\chi_{j2}^{[4]} &= \chi_{j2}^{[3]} + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j\,.
\end{aligned}$$

We have

$$\chi_{j2} = \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ (\theta_j - \hat\theta_j) \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ (\theta_j - \hat\theta_j)^2 \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-3} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j$$

and

$$\chi_{j2}^{[4]} = \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k$$
$$- \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_j \right\} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k$$
$$+ \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j\,.$$

Therefore, provided (2.48) holds,

$$\|\chi_{j2} - \chi_{j2}^{[4]}\|_{\sup} = \|(\theta_j - \hat\theta_j) \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ (\theta_j - \hat\theta_j)^2 \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-3} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$$
$$+ \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_j \right\} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \,\|_{\sup}\,. \tag{2.73}$$

Substitute the Taylor-expansion-with remainder of $\hat\theta_j$, given in (2.72), for $\hat\theta_j$ into the above terms, but omit the "order n^{-1}" term, the absolute value of which is dominated by:

$$T_j = \sum_{k\,:\,k\neq j} |\theta_j - \theta_k|^{-1} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2 . \tag{2.74}$$

Moment methods, together with part (b) of (2.1) and Lemma 2.5, show that

$$E(T_j) = O\big(n^{-1} \sum_{k\,:\,k\neq j} |\theta_j - \theta_k|^{-1}\, \theta_j\, \theta_k\big) = O\big(n^{-1}\, \rho_j^{-1}\, \theta_j \sum_{k\,:\,k\neq j} \theta_k\big) = O\big(n^{-1}\, \rho_j^{-1}\, \theta_j\big)\,,$$

and passing to higher-order moments, similar to (2.71), we obtain for each integer $m \geq 1$,

$$E(T_j^m) = O\big(\{n^{-1}\, \rho_j^{-1}\, \theta_j\}^m\big)\,.$$

In particular, if we take $R_{nj} = n\, \rho_j\, \theta_j^{-1}\, T_j$, then R_{nj} has the properties ascribed to U_{nj} in (2.64). Thus, we deduce that

$$\|\chi_{j2} - \chi_{j2}^{[4]}\|_{\sup} \leq \text{const.}\ \rho_j^{-3} \{\sup_u |\psi_j(u)|\}\, \widehat{\Delta}_{\sup}\, \widehat{\Delta}^2\, (1 - \zeta_j)^{-1/2}\, \theta_j^{-1/2} + D_{nj}\,.$$

Finally, by (2.28) and (2.29) we obtain

$$\begin{aligned}\|\chi_{j2} - \chi_{j2}^{[4]}\|_{\sup} &\leq \text{const.}\ n^{-3/2}\, U_{nj}\, \theta_j^{-1/2}\, \rho_j^{-3}\, (1 - \zeta_j)^{-1/2} \{\sup_u |\psi_j(u)|\} + D_{nj} \\ &\leq \text{const.}\ D_{nj}\,, \end{aligned} \tag{2.75}$$

where the random variable U_{nj} satisfies $\sup_{n,j\geq 1} E(U_{n,j}^m) \leq \infty$, for each integer $m \geq 1$, and the constant not depending on j.

Combining (2.49) and (2.75), we see that, provided (2.48) holds,

$$\begin{aligned} \|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[4]}\|_{\sup} &= \|\widehat{\psi}_j - \psi_j - \chi_{j2} + \chi_{j2} - \chi_{j2}^{[4]}\|_{\sup} \\ &\le \|\widehat{\psi}_j - \psi_j - \chi_{j2}\|_{\sup} + \|\chi_{j2} - \chi_{j2}^{[4]}\|_{\sup} \\ &\le v_{j2}\, V_2 + D_{nj} = v_{j2}\, \widehat{\Delta}^3\, \widehat{\Delta}_{\sup} + D_{nj} \\ &\le \text{const. } D_{nj}\,. \end{aligned} \tag{2.76}$$

Put

$$\alpha_j = \widehat{\psi}_j - \psi_j - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k\,. \tag{2.77}$$

Lemma 2.6. For a constant $C > 0$ and all $k \neq j$, we have $\theta_k^{1/2}/|\theta_k - \theta_j| \le C\, \theta_j^{1/2}/\rho_j$.

Proof of Lemma: For $k > j$ define $1 - \theta_k/\theta_j \equiv x_{jk}$. We have $\zeta_j = \inf_{k:\,k>j}(1 - \theta_k/\theta_j) \le x_{jk} \le 1$, for all $k > j$. So,

$$\frac{\theta_k^{1/2}}{\theta_j - \theta_k} = \frac{\theta_j^{-1/2}(1 - x_{jk})^{1/2}}{x_{jk}} \le \theta_j^{1/2}\, \zeta_j^{-1}\,. \tag{2.78}$$

On the other hand, for $k < j$, define $\theta_k/\theta_j - 1 = \nu_{jk}$. Here, also $\inf_{k:\,k<j}(\theta_k/\theta_j - 1) = \eta_j \le \nu_{jk}$, for all $k < j$. Therefore,

$$\frac{\theta_k^{1/2}}{\theta_k - \theta_j} = \theta_j^{-1/2}\left(\frac{\nu_{jk} + 1}{\nu_{jk}}\right) \le \theta_j^{-1/2}\left(\frac{\eta_j + 1}{\eta_j}\right) = \theta_j^{-1/2}(1 + \eta_j^{-1})\,. \tag{2.79}$$

Combining the two results (2.78) and (2.79), we deduce that

$$\theta_k^{1/2}/|\theta_k - \theta_j| \le C\, \theta_j^{-1/2} \max(\zeta_j^{-1}, \eta_j^{-1}) \le C\, \theta_j^{1/2}\, \rho_j^{-1},$$

which finishes the proof. ■

Writing $K_{\text{diag}}(u) \equiv K(u,u) = \text{Var}(X(u))$ and defining

$$S_1 = \sum_{k\,:\,k\neq j} \frac{\theta_k}{\theta_j\,(\theta_j - \theta_k)}\,\psi_k \int (\widehat{K} - K)\,\alpha_j\,\psi_k\,,$$

we have, by the Cauchy-Schwarz inequality and the above Lemma,

$$\begin{aligned} |S_1| &\le C\,\theta_j^{-1/2}\,\rho_j^{-1}\Big(\sum_{k=1}^{\infty}\theta_k\,\psi_k^2\Big)^{1/2}\Big[\Big\{\int(\widehat{K} - K)\,\alpha_j\,\psi_k\Big\}^2\Big]^{1/2} \\ &\le C\,\theta_j^{-1/2}\,\rho_j^{-1}\,K_{\text{diag}}^{1/2}\,|\!|\!|\widehat{K} - K|\!|\!|\,\|\alpha_j\|\,. \end{aligned} \tag{2.80}$$

More simply,

$$\begin{aligned} S_2 &\equiv \theta_j^{-1}\sum_{k\,:\,k\neq j}\psi_k\int(\widehat{K} - K)\,\alpha_j\,\psi_k \\ &= \theta_j^{-1}\Big\{\int(\widehat{K} - K)\,\alpha_j - \psi_j\int(\widehat{K} - K)\,\alpha_j\,\psi_j\Big\}. \end{aligned}$$

Note too that, since $\theta_j\,|\psi_j|^2 \le K_{\text{diag}}$, then $s_j \le \theta_j^{-1/2}\,\|K_{\text{diag}}\|_{\text{sup}}^{1/2}$. Therefore,

$$\begin{aligned} \|S_2\|_{\text{sup}} &\le \theta_j^{-1}\,\{|\!|\!|\widehat{K} - K)|\!|\!|_{\text{sup}}\,\|\alpha_j\| + s_j\,|\!|\!|\widehat{K} - K|\!|\!|\,\|\alpha_j\|\} \\ &\le \text{const.}\,\theta_j^{-3/2}\,|\!|\!|\widehat{K} - K|\!|\!|_{\text{sup}}\,\|\alpha_j\|\,. \end{aligned} \tag{2.81}$$

Using (2.9), we have

$$\widehat{\psi}_j - \psi_j - \psi_j\int(\widehat{\psi}_j - \psi_j)\,\psi_j = \sum_{k\,:\,k\neq j}(\hat{\theta}_j - \theta_k)^{-1}\,\psi_k\int(\widehat{K} - K)\,\widehat{\psi}_j\,\psi_k.$$

Therefore,

$$
\begin{aligned}
\beta_j \equiv \widehat{\psi}_j - \psi_j - \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \\
= \sum_{k\,:\,k\neq j} (\hat{\theta}_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \\
= \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \\
- \sum_{k\,:\,k\neq j} \{(\hat{\theta}_j - \theta_k)^{-1} - (\theta_j - \theta_k)^{-1}\}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \\
= \beta_{j1} + \beta_{j2}\,, \qquad (2.82)
\end{aligned}
$$

where

$$
\begin{aligned}
\beta_{j1} &= \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k\,, \\
\beta_{j2} &= - \sum_{k\,:\,k\neq j} \frac{(\hat{\theta}_j - \theta_j)}{(\theta_j - \theta_k)\, (\hat{\theta}_j - \theta_k)}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k\,.
\end{aligned}
$$

By (2.17),

$$
\begin{aligned}
\|\beta_{j1}\| &= \left[\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \left\{ \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \right\}^2 \right]^{1/2} \\
&\le \rho_j^{-1} \left[\sum_{k=1}^{\infty} \left\{ \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \right\}^2 \right]^{1/2} \\
&= \rho_j^{-1} \left\| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j) \right\| \\
&\le \rho_j^{-1}\, |\!|\!|\widehat{K} - K|\!|\!|\, \|\widehat{\psi}_j - \psi_j\| \le 8^{1/2}\, \rho_j^{-2}\, |\!|\!|\widehat{K} - K|\!|\!|^2\,; \qquad (2.83)
\end{aligned}
$$

if (2.2) holds, then by (2.17),

$$\begin{aligned}|\theta_j - \theta_k| &\le |\hat{\theta}_j - \theta_j| + |\hat{\theta}_j - \theta_k| \le \widehat{\Delta} + |\hat{\theta}_j - \theta_k| \\ &\le \tfrac{1}{2}\,\theta_j \,\min(\zeta_j, \eta_j) + |\hat{\theta}_j - \theta_k|\,.\end{aligned}$$

Thus,

$$|\hat{\theta}_j - \theta_k| \ge |\theta_j - \theta_k| - \tfrac{1}{2}\,\rho_j \ge \min_{k:\,k\neq j} |\theta_j - \theta_k| - \tfrac{1}{2}\,\rho_j = \tfrac{1}{2}\,\rho_j\,.$$

Hence, if (2.2) holds then by (2.17), $|\hat{\theta}_j - \theta_k|^{-1} \le 2\,\rho_j^{-1}$, and

$$\begin{aligned}\|\beta_{j2}\| &\le 2\,|\hat{\theta}_j - \theta_j|\,\rho_j^{-2}\Bigg[\sum_{k\,:\,k\neq j}\bigg\{\int (\widehat{K} - K)\,\widehat{\psi}_j\,\psi_k\bigg\}^2\Bigg]^{1/2}, \\ &\le 2\,\rho_j^{-2}\,\|\widehat{K} - K\|^2\,; \end{aligned} \tag{2.84}$$

and $\|\psi_j \int (\widehat{\psi}_j - \psi_j)\,\psi_j\|$ equals the left-hand side of (2.65), and so admits the bound there. Combining this result with (2.82)–(2.84), and noting that $\alpha_j = \beta_j + \psi_j \int (\widehat{\psi}_j - \psi_j)\,\psi_j$, we deduce that, if (2.2) holds then

$$\|\alpha_j\| \le \text{const.}\,\rho_j^{-2}\,\|\widehat{K} - K\|^2\,, \tag{2.85}$$

which, in view of (2.80) and (2.81), implies that

$$\|S_1\|_{\sup} + \|S_2\|_{\sup} \le \text{const.}\,\theta_j^{-1/2}\,\rho_j^{-3}\,\|\widehat{K} - K\|^2\,\|\widehat{K} - K\|_{\sup}\,. \tag{2.86}$$

Let

$$\chi_{j2}^{[5]} = \chi_{j2}^{[3]} + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\,\psi_k \int (\widehat{K} - K)\Big\{ \sum_{r\,:\,r\neq j} (\theta_j - \theta_r)^{-1}\psi_r \int (\widehat{K} - K)\,\psi_j\,\psi_r \Big\}\,\psi_k$$
$$- \tfrac{1}{2}\,\|\widehat{\psi}_j - \psi_j\|^2\,\psi_j\,. \tag{2.87}$$

Then it can be seen that $\chi_{j2}^{[4]} - \chi_{j2}^{[5]} = S_1 + S_2$. Consequently, by (2.86) and (2.17),

$$\begin{aligned}\|\chi_{j2}^{[4]} - \chi_{j2}^{[5]}\|_{\sup} &\leq \theta_j^{-1/2}\,\rho_j^{-3}\,|||\widehat{K} - K|||^2\,|||\widehat{K} - K|||_{\sup} \\ &\leq n^{-3/2}\,\theta_j^{-1/2}\,\rho_j^{-3}\,U_{nj}\,, \end{aligned}\tag{2.88}$$

where U_{nj} is a generic random variable satisfying

$$\sup_{n,\,j\geq 1} E(U_{nj}^m) < \infty\,, \quad \text{for each integer } m \geq 1\,,$$

and by (2.76) and (2.88) we have

$$\begin{aligned}\|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[5]}\|_{\sup} &\leq \|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[4]}\|_{\sup} + \|\chi_{j2}^{[4]} - \chi_{j2}^{[5]}\|_{\sup} \\ &\leq \text{const. } D_{nj}\,. \end{aligned}\tag{2.89}$$

Moreover,

$$\begin{aligned}\|\chi_{j2}^{[5]}\|^2 = \bigg\| & \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\,\psi_k \int (\widehat{K} - K)\,\psi_j\,\psi_k \\ & - \Big\{ \int (\widehat{K} - K)\psi_j\,\psi_j \Big\} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\,\psi_k \int (\widehat{K} - K)\,\psi_j\,\psi_k \\ & + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\psi_k \int (\widehat{K} - K)\Big\{ \sum_{r\,:\,r\neq j} (\theta_j - \theta_r)^{-1}\psi_r \int (\widehat{K} - K)\,\psi_j\,\psi_r \Big\}\,\psi_k \\ & - \tfrac{1}{2}\|\widehat{\psi}_j - \psi_j\|^2\,\psi_j \bigg\|^2\,. \end{aligned}\tag{2.90}$$

Now, for each $x, y \in L_2(\mathcal{I})$ it is shown that

$$\begin{aligned} \big|\|x\|^2 - \|y\|^2\big| = \big|\|x\| - \|y\|\big|\,(\|x\| + \|y\|) &\le \big|\|x\| - \|y\|\big|\,(2\|x\| + \|x-y\|) \\ &\le 2\|x\|\big(\|x-y\|\big) + \|x-y\|^2 . \end{aligned} \tag{2.91}$$

Therefore,

$$\begin{aligned} &\Big|\|\chi_{j2}^{[5]}\|^2 - \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big\{ \int (\widehat{K} - K)\,\psi_j\,\psi_k \Big\}^2 \Big| = \\ &\qquad \Big|\|\chi_{j2}^{[5]}\|^2 - \Big\| \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \Big\{ \int (\widehat{K} - K)\,\psi_j\,\psi_k \Big\} \Big\|^2 \Big| \\ &\qquad \le 2 \Big(\sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big\{ \int (\widehat{K} - K)\,\psi_j\,\psi_k \Big\}^2 \Big)^{1/2} \\ &\qquad\qquad \times \Big(\Big| \int (\widehat{K} - K)\psi_j\,\psi_j \Big| \Big\| \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-2}\,\psi_k \int (\widehat{K} - K)\,\psi_j\,\psi_k \Big\| \\ &\qquad\qquad + \Big\| \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k \int (\widehat{K} - K) \Big\{ \sum_{r:\,r\neq j} (\theta_j - \theta_r)^{-1} \psi_r \int (\widehat{K} - K)\,\psi_j\,\psi_r \Big\}\,\psi_k \\ &\qquad\qquad\qquad - \tfrac{1}{2} \|\widehat{\psi}_j - \psi_j\|^2\,\psi_j \Big\| \Big) \\ &\qquad + \Big\{ \int (\widehat{K} - K)\psi_j\,\psi_j \Big\}^2 \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-4} \Big\{ \int (\widehat{K} - K)\,\psi_j\,\psi_k \Big\}^2 \\ &\qquad + \sum_{k:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big[\sum_{r:\,r\neq j} (\theta_j - \theta_r)^{-1} \Big\{ \int (\widehat{K} - K)\,\psi_r\,\psi_k \Big\} \Big\{ \int (\widehat{K} - K)\,\psi_j\,\psi_r \Big\} \Big]^2 \\ &\qquad\qquad + \tfrac{1}{4} \|\widehat{\psi}_j - \psi_j\|^4 . \end{aligned}$$

So, by (2.17) and Lemma 2.5

$$\Big| \|\chi_{j2}^{[5]}\|^2 - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big\}^2 \Big| \leq$$

$$\text{const. } U_{nj} \Big[n^{-1/2}\, \rho_j^{-1}\, \theta_j^{1/2} \times \big(n^{-1/2} \,.\, n^{-1/2}\, \rho_j^{-2} + n^{-1}\, \rho_j^{-2}\big) + n^{-1/2} \,.\, n^{-1}\, \rho_j^{-3}\, \theta_j + n^{-3/2}\, \rho_j^{-3}\, \theta_j + n^{-3/2}\, \rho_j^{-3} \Big]$$

$$\leq \text{const. } n^{-3/2}\, \rho_j^{-3}\, U_{nj}\,, \tag{2.92}$$

where $\sup_{n,j} E(U_{nj}^m) \leq \infty$, for each integer $m \geq 1$.

Let

$$\begin{aligned}
\chi_j^{[6]} &= \theta_j^{-1} \Big\{ 1 - \theta_j^{-1} \int (\widehat{K} - K)\, \psi_j\, \psi_j \Big\} \Big\{ \sum_{k\,:\,k\neq j} \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big\} \\
&\quad + \sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-1} - \theta_j^{-1}\}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \\
&\quad - \Big\{ \int (\widehat{K} - K)\, \psi_j\, \psi_j \Big\} \Big[\sum_{k\,:\,k\neq j} \{(\theta_j - \theta_k)^{-2} - \theta_j^{-2}\}\, \psi_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big] \\
&\quad + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K) \Big\{ \sum_{r\,:\,r\neq j} (\theta_j - \theta_r)^{-1}\, \psi_r \int (\widehat{K} - K)\, \psi_j\, \psi_r \Big\}\, \psi_k \\
&\quad - \tfrac{1}{2}\, \psi_j \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big\}^2 .
\end{aligned} \tag{2.93}$$

We have

$$\|\chi_{j2}^{[5]} - \chi_{j2}^{[6]}\|_{\sup} = \Big\| \tfrac{1}{2}\, \psi_j\, \|\widehat{\psi}_j - \psi_j\|^2 - \tfrac{1}{2}\, \psi_j \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \Big\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big\}^2 \Big\|_{\sup} .$$

Substituting the right-hand side of (2.89) for $\widehat{\psi}_j - \psi_j$ in the previous result, and ignoring the remainder term implies that

$$\begin{aligned}
&\|\chi_{j2}^{[5]} - \chi_{j2}^{[6]}\|_{\sup} \\
&\quad \leq \text{const.} \left| \|\chi_j^{[5]}\|^2 - \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2 \right| \sup_u |\psi_j(u)| \\
&\quad \leq \text{const.}\ n^{-3/2}\, \rho_j^{-3}\, U_{nj} \sup_u |\psi_j(u)| \\
&\quad \leq \text{const.}\ D_{nj}\,, \qquad (2.94)
\end{aligned}$$

where we have used (2.92) to obtain the second inequality above. Therefore, we may replace $\chi_{j2}^{[5]}$, in (2.89), by $\chi_{j2}^{[6]}$ as follows:

$$\begin{aligned}
\|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[6]}\|_{\sup} &\leq \|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[5]}\|_{\sup} + \|\chi_{j2}^{[5]} - \chi_{j2}^{[6]}\|_{\sup} \\
&\leq \text{const.}\ D_{nj}\,, \qquad (2.95)
\end{aligned}$$

where we have used (2.89) and (2.94) to get the last inequality. The right-hand side of (2.93) is identical to the right-hand side of (1.32), except for the remainder term there. Hence, (2.95) implies the part of Theorem 2.1 that pertains to $\widehat{\psi}_j - \psi_j$.

2.7.2 Bounds in L_2-norm

When using the L_2 metric, we can find a better bound for the "$O_p(n^{-3/2})$" remainders on the right-hand sides of (1.33). In this subsection, we shall derive the corresponding bounds in terms of the L_2 norm:

$$\begin{aligned}
\|\chi_{j1}^{[1]} - \chi_{j1}\| &= \|(\theta_j - \hat{\theta}_j) \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k\| \\
&\leq \widehat{\Delta}\, \rho_j^{-2} \left(\| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\| + \left| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_j \right| \right). \qquad (2.96)
\end{aligned}$$

We know that

$$\begin{aligned}
\| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\|^2 &= \int \Big\{ \int (\widehat{K} - K)(u,v)\, (\widehat{\psi}_j - \psi_j)(v)\, dv \Big\}^2 du \\
&\le \int \Big\{ \int (\widehat{K} - K)(u,v)^2\, dv \Big\} \Big\{ \int (\widehat{\psi}_j - \psi_j)(v)^2\, dv \Big\}\, du \\
&= \widehat{\Delta}^2\, \|\widehat{\psi}_j - \psi_j\|^2 ,
\end{aligned}$$

$$| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_j | \le \| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\|\, \|\psi_j\| \le \widehat{\Delta}\, \|\widehat{\psi}_j - \psi_j\| .$$

Thus, combining these results with (2.96), and using (2.17) we have

$$\|\chi_{j1}^{[1]} - \chi_{j1}\| \le 2\, \widehat{\Delta}^2\, \rho_j^{-2}\, \|\widehat{\psi}_j - \psi_j\| \le \text{const.}\, \widehat{\Delta}^3\, \rho_j^{-3} . \tag{2.97}$$

Moreover, if (2.48) holds, by (2.44) and (2.47) we can obtain

$$\|\widehat{\psi}_j - \psi_j - \chi_{j1}\| = \| \sum_{s=2}^{\infty} (\theta_j - \hat{\theta}_j)^s \sum_{k\,:\,k\ne j} (\theta_j - \theta_k)^{-(s+1)}\, \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \|$$

$$\le \text{const.} \sum_{s=2}^{\infty} \widehat{\Delta}^{s+1}\, \rho_j^{-(s+1)} \le \text{const.}\, \widehat{\Delta}^3\, \rho_j^{-3} , \tag{2.98}$$

$$\|\widehat{\psi}_j - \psi_j - \chi_{j2}\| \le \text{const.} \sum_{s=3}^{\infty} \widehat{\Delta}^{s+1}\, \rho_j^{-(s+1)} \le \text{const.}\, \widehat{\Delta}^3\, \rho_j^{-3} . \tag{2.99}$$

Also, using (2.43) and (2.98) we conclude that

$$\begin{aligned}
&\big|(\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1}\, \psi_j) - \int (\widehat{K} - K)\, (\psi_j + \chi_{j1})\, \psi_j\big| \\
&\qquad \le \|\widehat{\psi}_j - \psi_j - \chi_{j1}\|\, (\widehat{\Delta} + \widehat{\Delta}_{\text{sup}}) \le \text{const.}\ n^{-3/2}\, \rho_j^{-3}\, U_{nj} ,
\end{aligned} \tag{2.100}$$

where U_{nj} is a generic random variable satisfying $\sup_{n,j\geq 1} E[U_{nj}^m] < \infty$ for each integer $m \geq 1$. So, by (2.100) we have

$$\begin{aligned}
&\Big|(\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1}^{[1]} \psi_j) - \int (\widehat{K} - K)\,(\psi_j + \chi_{j1}^{[1]})\,\psi_j\Big| \\
&\qquad \leq \left|(\hat{\theta}_j - \theta_j)\big(\int (\chi_{j1}^{[1]} - \chi_{j1})\,\psi_j\big) - \int (\widehat{K} - K)\,(\chi_{j1}^{[1]} - \chi_{j1})\,\psi_j\right| \\
&\qquad\quad + \left|(\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1})\,\psi_j - \int (\widehat{K} - K)\,(\psi_j + \chi_{j1})\,\psi_j\right| \\
&\qquad \leq \widehat{\Delta}\,\|\chi_{j1}^{[1]} - \chi_{j1}\| + \widehat{\Delta}\,\|\chi_{j1}^{[1]} - \chi_{j1}\| + \text{const.}\,\widehat{\Delta}^3\,\rho_j^{-3}\,\widehat{\Delta}_{\text{sup}}\,.
\end{aligned}$$

Therefore, by (2.97), this result leads to

$$\left|(\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1}^{[1]}\,\psi_j) - \int (\widehat{K} - K)\,(\psi_j - \chi_{j1}^{[1]})\,\psi_j\right| \leq \text{const.}\; n^{-3/2}\,\rho_j^{-3}\,U_{nj}\,. \quad (2.101)$$

Furthermore, by using (2.101), (2.65) and (2.18) we lead to

$$\begin{aligned}
&\Big|(\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1}^{[2]}\,\psi_j) - \int (\widehat{K} - K)\,(\psi_j + \chi_{j1}^{[2]})\,\psi_j\Big| \\
&\qquad \leq \Big|\hat{\theta}_j - \theta_j)(1 + \int \chi_{j1}^{[1]}\,\psi_j - \int (\widehat{K} - K)\,(\psi_j + \chi_{j1}^{[1]})\,\psi_j\Big| \\
&\qquad\quad + \left|\int (\widehat{\psi}_j - \psi_j)\,\psi_j\right|\,\left|(\theta_j - \hat{\theta}_j) - \int (\widehat{K} - K)\,\psi_j\,\psi_j\right| \\
&\qquad \leq \text{const.}\; n^{-3/2}\,\rho_j^{-3}\,U_{nj} + \text{const.}\;\widehat{\Delta}^3\,\rho_j^{-3}\,\widehat{\Delta}_{\text{sup}} \\
&\qquad \leq \text{const.}\; n^{-3/2}\,\rho_j^{-3}\,U_{nj}\,. \qquad (2.102)
\end{aligned}$$

Now, result (2.72) holds with the bound given in (2.102). Thus, a similar argument given to obtain (2.75), now results in

$$\begin{aligned}
\|\chi_{j2} - \chi_{j2}^{[4]}\| &\leq \text{const.}\;\rho_j^{-3}\,\widehat{\Delta}^2\left(\Big\|\int (\widehat{K} - K)\,\widehat{\psi}_j - \psi_j \int (\widehat{K} - K)\,\widehat{\psi}_j\,\psi_j\Big\|\right) \\
&\leq \text{const.}\;\widehat{\Delta}^3\,\rho_j^{-3} \leq \text{const.}\; n^{-3/2}\,\rho_j^{-3}\,U_{nj}\,, \qquad (2.103)
\end{aligned}$$

where we have used (2.29). In regard to (2.99) and (2.103) we have

$$\|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[4]}\| \le \|\widehat{\psi}_j - \psi_j - \chi_{j2}\| + \|\chi_{j2} - \chi_{j2}^{[4]}\| \le \text{const.}\, n^{-3/2}\, \rho_j^{-3}\, U_{nj}\,. \tag{2.104}$$

Analogously to (2.88), we write

$$\begin{aligned}\|\chi_{j2}^{[4]} - \chi_{j2}^{[5]}\| = \bigg\| & \sum_{k\,:\,k\ne j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \\ & - \sum_{k\,:\,k\ne j} (\theta_j - \theta_k)^{-1}\, \psi_k \bigg[\sum_{r:\,r\ne j} (\theta_j - \theta_r)^{-1} \bigg\{ \int (\widehat{K} - K)\, \psi_j\, \psi_r \bigg\} \bigg\{ \int (\widehat{K} - K)\, \psi_r\, \psi_k \bigg\} \bigg] \bigg\|\,.\end{aligned} \tag{2.105}$$

In regard to (2.77) we have

$$\widehat{\psi}_j - \psi_j = \alpha_j + \sum_{r:\,r\ne j} (\theta_j - \theta_r)^{-1}\, \psi_r \int (\widehat{K} - K)\, \psi_j\, \psi_r\,.$$

Thus, substituting the right-hand side of the previous equation for $\widehat{\psi}_j - \psi_j$ in (2.105) and using (2.85), we conclude that

$$\begin{aligned}\|\chi_{j2}^{[4]} - \chi_{j2}^{[5]}\| &= \bigg\| \sum_{k\,:\,k\ne j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, \alpha_j\, \psi_k \bigg\| \\ &\le \rho_j^{-1} \bigg\| \int (\widehat{K} - K)\, \alpha_j - \psi_j \int (\widehat{K} - K)\, \alpha_j\, \psi_j \bigg\| \\ &\le \rho_j^{-1} \bigg(\bigg\| \int (\widehat{K} - K)\, \alpha_j \bigg\| + \bigg| \int (\widehat{K} - K)\, \alpha_j\, \psi_j \bigg| \bigg) \\ &\le 2\, \rho_j^{-1} \bigg(\int \bigg\{ \int (\widehat{K} - K)(u,v)\, \alpha_j(v)\, dv \bigg\}^2 du \bigg)^{1/2} \\ &\le 2\, \rho_j^{-1} \bigg(\int \bigg\{ \int (\widehat{K} - K)(u,v)^2\, dv \bigg\} \|\alpha_j\|^2\, du \bigg)^{1/2} \\ &\le 2\, \rho_j^{-1}\, \|\alpha_j\|\, \widehat{\Delta} \le \text{const.}\, \rho_j^{-3}\, \widehat{\Delta}^3 = \text{const.}\, n^{-3/2}\, \rho_j^{-3}\, U_{nj}\,.\end{aligned} \tag{2.106}$$

Hence, results (2.104) and (2.106) lead us to

$$\|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[5]}\| \leq \|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[4]}\| + \|\chi_{j2}^{[4]} - \chi_{j2}^{[5]}\| \leq \text{const.}\, n^{-3/2}\, \rho_j^{-3}\, U_{nj}\,. \tag{2.107}$$

Therefore, using (2.92), we may replace $\|\widehat{\psi}_j - \psi_j\|^2$ in $\chi_{j2}^{[5]}$ by the series on the left-hand side of (2.92), without affecting the veracity of the version of (2.107) that holds with $\chi_{j2}^{[6]}$ in place of $\chi_{j2}^{[5]}$. That is, $\|\chi_{j2}^{[5]} - \chi_{j2}^{[6]}\| \leq \text{const.}\ n^{-3/2}\, \rho_j^{-3}\, U_{nj}$ and then by (2.107),

$$\|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[6]}\| \leq \|\widehat{\psi}_j - \psi_j - \chi_{j2}^{[5]}\| + \|\chi_{j2}^{[5]} - \chi_{j2}^{[6]}\| \leq \text{const.}\, n^{-3/2}\, \rho_j^{-3}\, U_{nj}\,. \tag{2.108}$$

Hence, the part of Theorem 2.1 that appertains to $\widehat{\psi}_j - \psi_j$ in the L_2 metric follows from (2.108).

Chapter 3

Bootstrap Confidence Statements for Eigenvalues and Eigenfunctions

3.1 Introduction

In this Chapter we suggest bootstrap methods, justifiable using the expansion approach developed in Chapters 2 and 3, for quantifying the accuracy of $\hat{\theta}_j$ and $\widehat{\psi}_j$ as approximations to θ_j and ψ_j, respectively. We first briefly express the general idea of bootstrap in Section 3.2. Then we explain how to construct bootstrap confidence intervals for eigenvalues and eigenfunctions. Those results appear in Sections 3.3 and 3.4, respectively. In Section 3.5, we propose simultaneous confidence regions for θ_j and ψ_j, by using the inequalities, given in (2.17) and Theorem 1.5, as a basis for bootstrap confidence procedures. These methods can be justified, and in particular can be shown to have appropriate degrees of accuracy, by using the stochastic expansions and their related properties discussed in Chapters 1 and 2.

3.2 General Idea of Bootstrap

The key idea behind the bootstrap in nonparametric problems is very general and can be applied to construct estimates of means, mean squared errors, quantiles, etc. Let $\mathcal{X} = \{X_1, \cdots, X_n\}$ denote a sample drawn at random from a population with distribution function (d.f.) $F = F_0$. The idea is that of replacing the true d.f. $F = F_0$ by the empirical d.f. $\widehat{F} = F_1$ (the probability measure that assigns to a set a measure equal to the proportion of sample values that lie in that set) in a formula that expresses a parameter as a functional of F, say $\theta(F)$. This entails replacing the pair (F_0, F_1) by (F_1, F_2), where F_2 denotes the d.f. of a sample drawn from $F_1 = \widehat{F}$ conditional on F_1. The latter sample (resample) is called a bootstrap dataset, and denoted by $\mathcal{X}^* = \{X_1^*, \cdots, X_n^*\}$. Each resample is collected by sampling randomly, with replacement from the dataset $\mathcal{X}$.

For example, if we are going to estimate a functional of a population d.f. F, such as the population mean, $\theta(F) = \mu = \int x\, dF(x)$, we can replace F by $\widehat{F}$ and obtain $\bar{X} = \int x\, d\widehat{F}(x)$, which is the sample mean. As another example, assume we wish to estimate mean squared error

$$\tau^2 = E(\hat{\theta} - \theta)^2 = E[\{\theta(F_1) - \theta(F_0)\}^2 |\, F_0] \,.$$

Then, the bootstrap estimator of τ^2 is

$$\hat{\tau}^2 = E\{(\hat{\theta}^* - \hat{\theta})^2 | \mathcal{X}\} = E[\{(\theta(F_2) - \theta(F_1)\}^2 |\, F_1]$$

where $\hat{\theta}^* = \theta[\mathcal{X}^*]$ is the version of $\hat{\theta}$ computed from $\mathcal{X}^*$ rather than $\mathcal{X}$.

Hall (1992) argued that construction of a confidence interval is equivalent to solving the population equation

$$E\{f_t(F_0, F_1)|F_0\} = 0\,, \tag{3.1}$$

for $t = T(F_0)$, where F_0 and F_1 are the population distribution function and empirical distribution function, respectively, T is a functional of F_0, and $f_t = I\{\theta(F_1) - t \leq \theta(F_0) \leq \theta(F_1) + t\} - \alpha$ is a functional which determines the relationship between $\theta_1 = \theta(F_1)$ (bootstrap estimation of the parameter $\theta_0 = \theta(F_0)$) and the parameter θ_0.

Since F_0 is unknown in the equation (3.1), we can not solve the population equation. Using the main principle of bootstrap methods, we believe that the relationship between F_0 and F_1 is similar to that between F_2 and F_1, where F_2 is the distribution function of the resample drawn from our sample $\mathcal{X}$ conditional on $\mathcal{X}$. Therefore, we solve the sample equation

$$E\{f_t(F_1, F_2)|F_1\} = P\{\theta(F_2) - t \leq \theta(F_1) \leq \theta(F_2) + t|F_1\} - \alpha = 0\,, \tag{3.2}$$

for t. The solution $\hat{t} = T(F_1)$ of (3.2) is an approximate solution to (3.1), i.e.

$$E\{f_{T(F_1)}(F_0, F_1)|F_0\} \approx 0\,. \tag{3.3}$$

To improve on this approximation, resulting in improvement of the coverage accuracy of the confidence interval, we appeal to double-bootstrap methods for constructing confidence intervals. In this way, we introduce an additive correction term t by defining $U(.,t) \equiv T(.) + t$ such that $U(.,0) = T(.)$. We choose t so as to improve on the coverage accuracy or to find a better approximation for

$$E\{f_{\hat{t}=T(F_1)}(F_0, F_1)|F_0\} = P\{\theta(F_1) - \hat{t} \leq \theta(F_0) \leq \theta(F_1) + \hat{t}|F_0\} - \alpha \approx 0\,.$$

Now, our aim is to solve the equation

$$E\{f_{U(F_1,t)}(F_0,F_1)|F_0\} = 0. \tag{3.4}$$

For this purpose, we define $f_{U(G,t)}(F,G) \equiv g_t(F,G)$. We observe that (3.4) is equivalent to

$$E\{g_t(F_0,F_1)|F_0\} = 0\,. \tag{3.5}$$

Again, using the main principle of bootstrap methods, we would like to solve its associated sample equation,

$$E\{g_t(F_1,F_2)|F_1\} = E\{f_{U(F_2,t)}(F_1,F_2)|F_1\} = E\{f_{T(F_2)+t}(F_1,F_2)|F_1\} = 0$$

for t, leading to the solution $\hat{t} = T_1(F_1)$, for some functional T_1 of F_1. Consequently, we have an approximate solution to (3.5),

$$E\{g_{T_1(F_1)}(F_0,F_1)|F_0\} = E\{f_{U(F_1,T_1(F_1))}(F_0,F_1)|F_0\} \approx 0\,. \tag{3.6}$$

Using $U(F_1,T_1(F_1)) = T(F_1) + T_1(F_1)$ permits the left-hand side of (3.6) to be closer to zero than the left-hand side of (3.3). This allows the coverage accuracy of the confidence interval to be improved. See Sections 1.4 and 1.5 of Hall (1992).

3.3 Bootstrap Confidence Intervals for Individual Eigenvalues

A confidence band for $\widehat{\psi}_j$ provides information about the likelihood that a bump on $\widehat{\psi}_j$ reflects a similar feature in the true component function ψ_j. Also, confidence intervals for $\hat{\theta}_k$ help to quantify the amount of variability that is lost by confining attention to k dimensions.

Draw a resample, $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$, by sampling randomly, with replacement, from the sample $\mathcal{X}$ of random functions. For this resample, compute the analogues $\hat{\theta}_j^*, \widehat{\psi}_j^*$ of $\hat{\theta}_j, \widehat{\psi}_j$. Approximate the unconditional distribution of $\hat{\theta}_j - \theta_j$ by the distribution of $\hat{\theta}_j^* - \hat{\theta}_j$ conditional on $\mathcal{X}$, and approximate the unconditional distribution of the random function $\widehat{\psi}_j - \psi_j$ by the conditional distribution of $\widehat{\psi}_j^* - \widehat{\psi}_j$. In this way, develop confidence statements about the sizes of (a) $\hat{\theta}_j - \theta_j$, (b) $\sup_t |\widehat{\psi}_j(t) - \psi_j(t)|$ or (c) $\|\widehat{\psi}_j - \psi_j\|$. Using (a) we construct percentile bootstrap confidence intervals for θ_j, or for a collection of θ_j's if we address several eigenvalues simultaneously; using (b) we obtain simultaneous bootstrap confidence bands for ψ_j; and using (c) we get confidence intervals for the L_2 distance of $\widehat{\psi}_j$ from ψ_j.

3.3.1 Coverage Accuracy for the Bootstrap Confidence Interval of θ_j

Result (1.62) shows that $n^{1/2}(\hat{\theta}_j - \theta_j)$ is asymptotically Normally distributed. In particular, we obtained in Theorem 2.1 that

$$\hat{\theta}_j - \theta_j = n^{-1/2} \int Z\, \psi_j\, \psi_j + n^{-1} \sum_{k\,:\,k \neq j} (\theta_j - \theta_k)^{-1} \left(\int Z\, \psi_j\, \psi_k \right)^2 + R_{nj;1}\,, \quad (3.7)$$

where $R_{nj;1}$ can be bounded above by $n^{-3/2}\, Z_{nj}$ in which $Z_{nj} = O_p(1)$,

$$Z_{nj} = j^{\epsilon}\, \rho_j^{-3}\, \theta_j^{-1/2}\, (1 - \zeta_j)^{-1/2}\, s_j\,,$$

$\zeta_j = \inf_{k \geq j}\{1 - (\theta_k/\theta_j)\}$, $\rho_j = \min_{k \neq j} |\theta_j - \theta_k|$ and $s_j = \sup_t |\psi_j(t)|$. Therefore, Z_{nj} grows up as j increases.

Edgeworth expansions of the distributions $\hat{\theta}_j - \theta_j$ and the first two terms on the right-hand side of (3.7) generally disagree only in terms of order $n^{-3/2}$ (or smaller).

In order to evaluate the coverage probabilities, we need an Edgeworth expansion of the distribution function of

$$n^{1/2}\,(\hat{\theta}_j - \theta_j) = D_j = I_j + n^{-1/2} \sum_{k\,:\,k\neq j} w_k\, I_{jk}^2 + R_{nj}\,, \tag{3.8}$$

where $I_j = \int Z\,\psi_j\,\psi_j$, $I_{jk} = \int Z\,\psi_j\,\psi_k$, $w_k = (\theta_j - \theta_k)^{-1}$ and R_{nj} can be bounded above by $n^{-1}\,Z_{nj}$. We may write both I_j and I_{jk} in terms of the principal components as follows:

$$I_j = n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\{\xi_{ij}^2 - E(\xi_{ij}^2)\} - (\bar{\xi}_j)^2\right], \qquad I_{jk} = n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\,\xi_{ik} - \bar{\xi}_j\,\bar{\xi}_k\right], \tag{3.9}$$

where $\xi_{ij} = \int_{\mathcal{I}} X_i\,\psi_j$ and $\bar{\xi}_j = \frac{1}{n}\sum_{i=1}^{n}\xi_{ij}$.

A bootstrap confidence interval for θ_j can be constructed by using either a percentile bootstrap or a percentile-t bootstrap method. We will discuss construction of that by a percentile bootstrap method.

Let $k_j(U)$ denote the jth cumulant of a random variable U. Then,

$$\begin{aligned} k_1(D_j) = E(D_j) &= E(I_j + n^{-1/2} \sum_{k\,:\,k\neq j} w_k\, I_{jk}^2 + R_{nj}) \\ &= k_1(I_j) + 2\,n^{-1/2}\,a_j + ER_{nj}\,, \end{aligned} \tag{3.10}$$

where $a_j = \sum_{k\,:\,k\neq j} w_k^{-1}\,E(I_{jk}^2)$. We may write, for fixed j,

$$I_j = n^{-1/2}\sum_{i=1}^{n} U_i + O_p(n^{-1/2})\,, \qquad I_{jk} = n^{-1/2}\sum_{i=1}^{n} V_{i,k} + O_p(n^{-1/2})\,, \tag{3.11}$$

where $U_i = \xi_{ij}^2 - E(\xi_{ij}^2)$, $V_{i,k} = \xi_{ij}\,\xi_{ik}$ and the pairs $(U_i, V_{i,k})$, $1 \leq i \leq n$, are independent and identically distributed with zero mean. Therefore,

$$E(I_j^2) = E(U_1^2) + O(n^{-1}), \quad E(I_{jk}^2) = E(V_{1,k}^2) + O(n^{-1}), \tag{3.12}$$

$$E(I_j I_{jk}^2) = n^{-1/2}\, E(U_1\, V_{1,k}^2) + O(n^{-3/2}), \quad E(I_j^3) = n^{-1/2}\, E(U_1^3) + O(n^{-1}), \tag{3.13}$$

$$E(I_{jk_1}^2 I_{jk_2}^2) = 3\, E(V_{1,k_1})\, E(V_{1,k_2}) + n^{-1}\,\{\delta_{k_1 k_2}\, E(V_{1,k_1}^4) - 3\, E(V_{1,k_1})\, E(V_{1,k_2})\} + O(n^{-2}), \tag{3.14}$$

$$\begin{aligned} E(I_j^3 I_{jk}^2) &= n^{-1/2}\{3\, E(U_1^2)\, E(U_1 V_{1,k}^2) + 6\, E(U_1 V_{1,k})\, E(U_1^2 V_{1,k})\} \\ &\quad + n^{-3/2}\, E(U_1 V_{1,k})\, E(U_1^2 V_{1,k}) + O(n^{-2}). \end{aligned} \tag{3.15}$$

Using the above results we have:

$$\begin{aligned} k_2(D_j) &= E(D_j^2) - (ED_j)^2 = E(I_j^2) + 2\, n^{-1/2} \sum_{k\,:\,k \neq j} w_k\, E(I_j\, I_{jk}^2) \\ &\quad + n^{-1} \sum_{k_1 :\, k1 \neq j}\ \sum_{k_2 :\, k2 \neq j} w_{k_1}\, w_{k2}\, E(I_{jk_1}^2\, I_{jk_2}^2) + 2\, E(I_j\, R_{nj}) + O(n^{-2}) \\ &= k_2(I_j) + n^{-1}\,(b_{1j} + b_{2j}) + O(n^{-2}), \end{aligned} \tag{3.16}$$

where

$$b_{1j} = 2 \sum_{k\,:\,k \neq j} w_k\, E(U_1 V_{1,k}^2) + 3 \sum_{k_1 :\, k_1 \neq j}\ \sum_{k_2 :\, k_2 \neq j} w_{k_1}\, w_{k_2}\, E(V_{1,k_1}^2)\, E(V_{1,k_2}^2),$$

and $n^{-1}\, b_{2j} = 2\, E(I_j\, R_{1j})$. Furthermore,

$$
\begin{aligned}
k_3(D_j) &= E(D_j^3) - 3\, E(D_j)\, E(D_j^2) + 2\,(ED_j)^3 \\
&= k_3(I_j) + 3\, n^{-1/2} \sum_{k\,:\, k\neq j} w_k \big[E(I_j^2 I_{jk}^2) - E(I_j^2)E(I_{jk}^2)\big] + O(n^{-1}) \\
&= k_3(I_j) + n^{-1/2}\, c_j\, + O(n^{-1})\,, \qquad (3.17)
\end{aligned}
$$

where $c_j = 6 \sum_{k\,:\,k\neq j} w_k \left[E(U_1 V_{1,k})\right]^2$. Also,

$$
\begin{aligned}
&k_4(D_j) = E(D_j^4) - 4\, E(D_j)\, E(D_j^3) - 3\,(ED_j^2)^2 + 12\, E(D_j^2)\,(ED_j)^2 - 6\,(ED_j)^4 \\
&= k_4(I_j) + 4\, n^{-1/2} \sum_{k\,:\, k\neq j} w_k \big[E(I_j^3 I_{jk}^2) - E(I_j^3)E(I_{jk}^2) - 3\, E(I_j I_{jk}^2)\, E(I_j^2)\big] + O(n^{-3/2}) \\
&= k_4(I_j) + n^{-1}\, d_j\, + O(n^{-3/2})\,, \qquad (3.18)
\end{aligned}
$$

where $d_j = \sum_{k\,:\,k\neq j} w_k \left\{24\, E(U_1 V_{1,k})\, E(U_1^2 V_{1,k}) - 16\, E(U_1^3)\, E(V_{1,k}^2)\right\}$. Thus, (3.10) and (3.16)–(3.18) imply that

$$
\begin{aligned}
k_{1,2}(D_j) &= k_{1,2}(I_j) + a_j\,, \qquad (3.19)\\
k_{2,2}(D_j) &= k_{2,2}(I_j) + b_j\,, \qquad (3.20)\\
k_{3,1}(D_j) &= k_{3,1}(I_j) + c_j\,, \qquad (3.21)\\
k_{4,1}(D_j) &= k_{4,1}(I_j) + d_j\,, \qquad (3.22)
\end{aligned}
$$

where $b_j = b_{1j} + b_{2j}$.

Standardizing I_j by dividing to σ_j, where σ_j^2 is the asymptotic variance of $n^{1/2}\,\hat\theta_j$, and writing the usual Edgeworth expansion for that, we have:

$$
P(I_j \le x) = \Phi(\sigma_j^{-1}\, x) + n^{-1/2}\, p_1(\sigma_j^{-1}\, x)\, \phi(\sigma_j^{-1}\, x) + n^{-1}\, p_2(\sigma_j^{-1}\, x)\, \phi(\sigma_j^{-1}\, x) + O(n^{-3/2})\,, \qquad (3.23)
$$

where Φ and ϕ denote the distribution function and density function of the Standard Normal, and the functions p_j are polynomials of degree at most $3\,j-1$ and odd/even indexed polynomials p_j are even/odd functions. In particular, p_1 and p_2 are

$$
\begin{aligned}
p_1(x) &= -\{k_{1,2} + \tfrac{1}{6}\,k_{3,1}\,(x^2-1)\}\,, && (3.24)\\
p_2(x) &= -x\,\{\tfrac{1}{2}\,(k_{2,2}+k_{1,2}^2) + \tfrac{1}{24}\,(k_{4,1}+4\,k_{1,2}\,k_{1,3})(x^2-3)\\
&\qquad + \tfrac{1}{72}k_{3,1}^2\,(x^4-10\,x^2+15)\}\,. && (3.25)
\end{aligned}
$$

Thus, using (3.10) and (3.16)–(3.18) we can express the polynomials p_1 and p_2 in terms of the first four cumulants. Note that as (3.19)–(3.22) show, the coefficients $k_{i,j}$, appearing in (3.24) and (3.25), have a different formula in the case of I_j than they do for D_j. Hence,

$$
\begin{aligned}
P(D_j \le x) = P(I_j \le x) &- n^{-1/2}\,\{a_j + \frac{1}{6}\,c_j\,(x^2-1)\}\phi(x)\\
&- n^{-1}\left\{\frac{1}{2}\,(b_j+2\,a_j\,k_{1,2}+a_j^2) + \frac{1}{24}\,(d_j+4\,a_j\,k_{1,3}+c_j\,k_{1,2})\,(x^2-3)\right.\\
&\left. + \frac{1}{72}\,(2\,c_j\,k_{3,1}+c_j^2)\,(x^4-10\,x^2+15)\right\}x\,\phi(x) + O(n^{-3/2})\,,
\end{aligned}
\tag{3.26}
$$

Substituting (3.23) in equation (3.26) we can obtain an approximation to the distribution of $n^{1/2}\,(\hat{\theta}_j-\theta_j)$ as follows:

$$
\begin{aligned}
P\big(n^{1/2}(\hat{\theta}_j-\theta_j) \le z_\alpha\big) &= \Phi(\sigma_j^{-1}\,z_\alpha) + n^{-1/2}\,p_1^*(\sigma_j^{-1}\,z_\alpha)\,\phi(\sigma_j^{-1}\,z_\alpha)\\
&\quad + n^{-1}\,p_2^*(\sigma_j^{-1}\,z_\alpha)\,\phi(\sigma_j^{-1}\,z_\alpha) + O(n^{-3/2})\,.
\end{aligned}
\tag{3.27}
$$

where:

$$p_1^*(x) = p_1(x) - a_j - \frac{1}{6}\, c_j\,(x^2 - 1)\,, \tag{3.28}$$

$$\begin{aligned} p_2^*(x) = p_2(x) - \Bigg[\frac{1}{2}\,(b_j + 2\,a_j\,k_{1,2} + a_j^2) + \frac{1}{24}\,(d_j + 4\,a_j\,k_{1,3} + c_j\,k_{1,2})\,(x^2 - 3) \\ + \frac{1}{72}\,(2\,c_j\,k_{3,1} + c_j^2)\,(x^4 - 10\,x^2 + 15)\Bigg]\,x\,. \end{aligned} \tag{3.29}$$

The two polynomials p_1^* and p_2^* preserve the parity of p_1 and p_2 according to the index. Consequently,

$$\begin{aligned} 1 - \alpha &= P(n^{-1/2}\,|\hat{\theta}_j - \theta_j| \le z_\alpha) \\ &= P(n^{-1/2}\,(\hat{\theta}_j - \theta_j) \le z_\alpha) - P(n^{-1/2}\,(\hat{\theta}_j - \theta_j) \le -z_\alpha) \\ &= 2\,\Phi(\sigma_j^{-1}\,z_\alpha) - 1 + 2\,n^{-1}\,p_2^*(\sigma_j^{-1}\,z_\alpha)\,\phi(\sigma_j^{-1}\,z_\alpha) + O(n^{-3/2})\,. \end{aligned} \tag{3.30}$$

Inverting the above expansion, we derive a Cornish-Fisher expansion as follows:

$$z_\alpha = \sigma_j\,z_\xi + n^{-1}\,\sigma_j\,p_{13}(z_\xi) + n^{-2}\,\sigma_j\,p_{23}(z_\xi) + \cdots\,, \tag{3.31}$$

where $\xi = 1 - \frac{\alpha}{2}$, and

$$p_{13}(x) = -p_2^*(x) \qquad p_{23}(x) = p_2^*(x)\,p_2^{*\prime}(x) - \frac{1}{2}\,x\,p_2^*(x)^2 - p_4^*(x)\,. \tag{3.32}$$

Note that $p_2^{*\prime}$ denotes the first derivative of p_2^*. The bootstrap estimate of z_α is $\hat{z}_\alpha$, the solution of the equation

$$P(n^{1/2}\,|\hat{\theta}_j^* - \hat{\theta}_j| \le \hat{z}_\alpha|\,\mathcal{X}) = 1 - \alpha\,.$$

By analogy with (3.31), $\hat{z}_\alpha$ admits the obvious bootstrap expansion

$$\hat{z}_\alpha = \hat{\sigma}_j\,z_\xi + n^{-1}\,\hat{\sigma}_j\,\hat{p}_{13}(z_\xi) + n^{-2}\,\hat{\sigma}_j\,\hat{p}_{23}(z_\xi) + \cdots\,, \tag{3.33}$$

where $\hat{\sigma}_j$ is the bootstrap estimate of σ_j.

Therefore, $\hat{z}_\alpha - z_\alpha = O_p(n^{-1/2})$ since $\hat{\sigma}_j - \sigma_j = O_p(n^{-1/2})$. Moreover, using the delta method (see Section 2.7 of Hall, 1992), we have:

$$\begin{aligned} P\big(n^{1/2}\,(\hat{\theta}_j - \theta_j) \le \hat{z}_\alpha\big) & \\ &= P\big((\hat{\theta}_j - \theta_j) \le n^{-1/2}\,\hat{\sigma}_j\, z_\xi + n^{-3/2}\,\hat{\sigma}_j\,\hat{p}_{13}(z_\xi) + n^{-5/2}\,\hat{\sigma}_j\,\hat{p}_{23}(z_\xi) + \cdots\big) \\ &= P\big(S \le z_\xi - n^{-1}\,p_2^*(z_\xi)\big) + O(n^{-2})\,, \qquad (3.34) \end{aligned}$$

where $S = T + n^{-3/2}\,\Delta_n$, in which $T = \frac{n^{1/2}\,(\hat{\theta}_j - \theta_j)}{\hat{\sigma}_j}$ and $\Delta_n = n^{1/2}\,\{p_{13}(z_\xi) - \hat{p}_{13}(z_\xi)\} = n^{1/2}\,\{\hat{p}_2^*(z_\xi) - p_2^*(z_\xi)\}$. It follows that Δ_n is asymptotically Normally distributed, and therefore $\Delta_n = O_p(1)$.

Assume the usual Edgeworth expansion for the Studentized statistic T, i.e.

$$P(T \le x) = \Phi(x) + n^{-1/2}\,q_1(x)\,\phi(x) + n^{-1}\,q_2(x)\,\phi(x) + \cdots\,,$$

where odd/even indexed polynomials q_j are even/odd functions, respectively. Note that we have changed the symbol p to q in the above expansion, because we are now working with Edgeworth expansions for the Studentized statistic T rather than the non-Studentized statistic $n^{1/2}\,(\hat{\theta}_j - \theta_j)$.

The delta method, and (3.34), imply that

$$\begin{aligned} P\big(S \le z_\xi - n^{-1}\,p_2^*(z_\xi)\big) &= P\big(T \le z_\xi - n^{-1}\,p_2^*(z_\xi)\big) + O(n^{-3/2}) \\ &= \Phi\big(z_\xi - n^{-1}\,p_2^*(z_\xi)\big) + n^{-1/2}\,q_1\big(z_\xi - n^{-1}\,p_2^*(z_\xi)\big)\,\phi\big(z_\xi - n^{-1}\,p_2^*(z_\xi)\big) \\ &\quad + n^{-1}\,q_2\big(z_\xi - n^{-1}\,p_2^*(z_\xi)\big)\,\phi\big(z_\xi - n^{-1}\,p_2^*(z_\xi)\big) + O(n^{-3/2})\,. \qquad (3.35) \end{aligned}$$

Taylor expanding the corresponding functions, the following is obtained:

$$P\big(S \leq z_\xi - n^{-1}\, p_2^*(z_\xi)\big) = \Phi(z_\xi) + n^{-1/2}\, r_1(z_\xi)\, \phi(z_\xi) \\ + n^{-1}\, r_2(z_\xi)\, \phi(z_\xi) + n^{-3/2}\, r_3(z_\xi)\, \phi(z_\xi) + \cdots\,, \quad (3.36)$$

where $r_1(x) = q_1(x)$, $r_2(x) = q_2(x) - p_2^*(x)$ and $r_3(x) = x\, p_2^*(x)\, q_1(x) - p_2^*(x)\, q_1'(x)$. Here, as it can be seen, the r_j are odd or even functions according to whether j is even or odd. Therefore,

$$\begin{aligned} P\big(n^{1/2}\, |\hat{\theta}_j - \theta_j| \leq \hat{z}_\alpha\big) &= P\big(n^{1/2}\, (\hat{\theta}_j - \theta_j) \leq \hat{z}_\alpha\big) - P\big(n^{1/2}\, (\hat{\theta}_j - \theta_j) \leq -\hat{z}_\alpha\big) \\ &= 2\, \Phi(z_\xi) - 1 + 2\, n^{-1}\, r_2(z_\xi)\, \phi(z_\xi) + O(n^{-2}) \\ &= 1 - \alpha + O(n^{-1})\,. \end{aligned} \quad (3.37)$$

This means that the coverage error of single-bootstrap confidence interval $(\hat{\theta}_j - \hat{z}_\alpha\, n^{-1/2},\ \hat{\theta}_j + \hat{z}_\alpha\, n^{-1/2})$ is of order $O(n^{-1})$. It should be noted that $\hat{\theta}_j$ used in the confidence interval is not the actual θ_j but an approximation to that. It can be shown that this error can be reduced to $O(n^{-2})$ by using double-bootstrap calibration. See Proposition 1.2 of Hall (1992). We will present the numerical result about bootstrap confidence statements in Chapter 5.

Since the distribution of $\hat{\theta}_j$ is likely to be asymmetric about θ_j, it is better to construct an equal-tailed confidence interval rather than a symmetric one. In numerical work, we can obtain $t_\alpha = \hat{t}_{\alpha/2}$ and $\hat{t}_{1-\alpha/2}$ from

$$\frac{1}{B} \sum_{b=1}^{B} I(\hat{\theta}_b^* - \hat{\theta} \leq t_\alpha) = \alpha\,, \quad (3.38)$$

where θ_b^* is a version of $\hat{\theta}$ which is computed from resample $\chi_b^* = \{X_{1b}^*, \cdots, X_{nb}^*\}$

$(b = 1, \cdots, B)$ rather than the sample χ. Using the fact that

$$P(\hat{\theta}_* - \hat{\theta} \leq x|\chi) \approx P(\hat{\theta} - \theta \leq x)\,,$$

the confidence interval $(\hat{\theta} - \hat{t}_{1-\alpha/2}, \hat{\theta} - \hat{t}_{\alpha/2})$ has nominal coverage $1 - \alpha$. We call it a "single bootstrap" confidence interval. Drawing C resamples from each of our resample, we can obtain $t = \hat{t}^*_{b;\alpha/2}$ and $\hat{t}^*_{b;1-\alpha/2}$ from

$$\frac{1}{C}\sum_{c=1}^{C} I(\hat{\theta}^{**}_{bc} - \hat{\theta}^*_b \leq t^*_{b;\alpha}) = \alpha \tag{3.39}$$

Finally, we can find $t = \tilde{t}_{\alpha/2}$ and $\tilde{t}_{1-\alpha/2}$ so that

$$\frac{1}{B}\sum_{b=1}^{B} I(\hat{\theta}^*_b - \hat{\theta} \leq \hat{t}^*_{b;\alpha/2} + t) = \frac{\alpha}{2}, \text{ and } \frac{1}{B}\sum_{b=1}^{B} I(\hat{\theta}^*_b - \hat{\theta} \leq \hat{t}^*_{b;1-\alpha/2} + t) = 1 - \frac{\alpha}{2} \tag{3.40}$$

respectively. Then our double bootstrap confidence interval is $(\hat{\theta} - \hat{t}_{1-\alpha/2} - \tilde{t}_{1-\alpha/2}, \hat{\theta} - \hat{t}_{\alpha/2} - \tilde{t}_{\alpha/2})$. See Chapter 5 for numerical results.

3.4 Bootstrap Confidence Intervals for eigenfunctions

To construct confidence bands for the eigenfunctions ψ_j we proceed as follows. From the bootstrap dataset $\mathcal{X}^* = \{X_1^*, \ldots, X_n^*\}$, compute the bootstrap versions $\widehat{\psi}_1^*, \widehat{\psi}_2^*, \ldots$ of $\widehat{\psi}_1, \widehat{\psi}_2, \ldots$. Then construct standard two-sided percentile-method confidence regions for $\psi_1, \psi_2, \ldots$ as follows. Define $\hat{z}_{j\alpha}$ by

$$P\left\{\sup_{t\in\mathcal{I}} \left|\widehat{\psi}_j^*(t) - \widehat{\psi}_j(t)\right| \leq \hat{z}_\alpha \;\middle|\; \mathcal{X}\right\} = 1 - \alpha\,,$$

the bivariate region given by

$$\left\{(t,u) : t \in \mathcal{I} \quad \text{and} \quad |\widehat{\psi}(t) - u| \leq \hat{z}_\alpha\right\}$$

is a nominal $(1-\alpha)$-level confidence region for graphs of ψ_j.

Similar approaches such as those employed in the previous Section can be used to show that this band covers the true ψ_j with probability $1-\alpha$, within an error of $O(n^{-1})$, in the sense that

$$P\left\{\sup_{t \in \mathcal{I}} |\widehat{\psi}_j(t) - \psi_j(t)| \leq \hat{z}_\alpha \right\} = 1 - \alpha + O(n^{-1}) .$$

The error can be reduced to $O(n^{-2})$ by using double-bootstrap calibration. Similarly, we can construct confidence bands for the eigenfunctions ψ_j by using the L_2-norm.

3.5 Simultaneous Confidence Bounds for Eigenvalues and Eigenfunctions

If we want the approximation to be valid for a large number of values of j, the above approach for constructing bootstrap confidence intervals is problematic. In theory, we would wish that number to diverge as n increases. The inequalities given in Theorem 1.5 are potentially attractive in this regard, because they can be used as the basis for simultaneous confidence regions with a degree of conservatism. Suppose we have a one-sided prediction interval for $\widehat{\Delta}$, of the form $P(\widehat{\Delta} \leq \widehat{\Delta}_{\text{upp}}) = 1 - \alpha_n$, say, where $\widehat{\Delta}_{\text{upp}}$ is computable from data and the subscript denotes "upper bound." Define

$$\widehat{J}_{\text{upp}} = \inf\left\{j \geq 1 \,:\, \hat{\theta}_j - \hat{\theta}_{j+1} \leq 2\,\widehat{\Delta}_{\text{upp}}\right\}.$$

Then, in view of the theorem, the following is true:

$$\begin{aligned}&\text{with probability at least } 1-\alpha_n,\ \sup_{j\geq 1}|\hat{\theta}_j - \theta_j| \leq \widehat{\Delta}_{\text{upp}},\ \text{and for}\\ &\text{all } 1 \leq j \leq \widehat{J}_{\text{upp}}-1,\ \|\widehat{\psi}_j - \psi_j\| \leq 2^{1/2}\,[1-\{1-4\,(\widehat{\Delta}_{\text{upp}}/\hat{\delta}_j)^2\}^{1/2}]^{1/2}\,.\end{aligned} \tag{3.41}$$

We may readily compute $\widehat{\Delta}_{\text{upp}}$ using bootstrap methods, as follows. Let $\widehat{K}^*$ denote the bootstrap version of $\widehat{K}$, computed from $\mathcal{X}^*$ rather than $\mathcal{X}$, and put $\widehat{\Delta}^* = \|\widehat{K}^* - \widehat{K}\|$. Given $0 < \alpha < 1$, for example $\alpha = 0.05$, take $\widehat{\Delta}_{\text{upp}}$ to be the upper α-level critical point of the distribution of $\widehat{\Delta}^*$, given $\mathcal{X}$.

Note that $P(\widehat{\Delta} \leq \widehat{\Delta}_{\text{upp}})$ converges to $1-\alpha$ as $n \to \infty$. To appreciate why, recall from Section 1.3 that $n^{1/2}\,(\widehat{K}-K)$ converges weakly to a Gaussian process ζ. Analogously, and conditional on $\mathcal{X}$, $n^{1/2}\,(\widehat{K}^*-\widehat{K})$ converges weakly to the same ζ. Therefore, the limit of the distribution of $n^{1/2}\Delta^*$, conditional on $\mathcal{X}$, is identical to the limit of the unconditional distribution of $n^{1/2}\widehat{\Delta}$. It follows that $n^{1/2}\widehat{\Delta}_{\text{upp}}$ converges in probability to the upper α-level critical point of the distribution of $\int\int_{\mathcal{I}^2}\zeta^2$, and hence that $P(\widehat{\Delta} \leq \widehat{\Delta}_{\text{upp}})$ converges to $1-\alpha$.

The simultaneous bootstrap confidence interval for θ_j, suggested by (3.41), is indeed conservative but not especially so. Its numerical properties will be discussed in Section 5.3. The confidence band for ψ_j tends to be quite conservative, however. Therefore, we will not discuss it theoretically any further.

From Chapter 1 we have:

$$n^{1/2}\,(\hat{\theta}_j - \theta_j) = D_j = I_j + n^{-1/2}\,J_j + R_{nj}\,,$$

where $\mathrm{I}_j = \int Z\,\psi_j\,\psi_j$, $J_j = \sum_{\ell:\ell\neq j} w_\ell\, I_{j\ell}^2$, in which $\mathrm{I}_{j\ell} = \int Z\,\psi_j\,\psi_\ell$, $w_\ell = (\theta_j - \theta_\ell)^{-1}$ and R_{nj} can be bounded above by $n^{-1}\,Z_{nj}$.

Assume that from the sequence θ_j, only the first k eigenvalues are non-zero. Then we may write

$$\mathrm{D} = \mathbf{I} + n^{-1/2}\,\mathbf{J} + \mathrm{R}_n\,, \tag{3.42}$$

where $\mathbf{D}$, $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{R}_n$ are k-vectors with elements D_j, I_j, J_j and R_{nj}, respectively, for $1 \le j \le k$.

A Multivariate version of the percentile technique can be used to construct simultaneous confidence regions for $\theta_1, \cdots, \theta_k$. Then, the two k-vectors $\mathbf{D}$ and $\mathbf{I}$ satisfy a vector version of the results (3.19)–(3.22), and for any k-variate sphere $\mathcal{S}$ centered at the origin we have:

$$\begin{aligned} P\big(\mathrm{D} \in \Sigma^{-1/2}\,\mathcal{S}\big) &= P\big(\mathbf{I} + n^{-1/2}\,\mathbf{J} + \mathrm{R}_n \in \Sigma^{-1/2}\,\mathcal{S}\big) \\ &= \int_{\Sigma^{-1/2}\,\mathcal{S}} \{1 + n^{-1}\,r_1(\mathrm{x}) + n^{-2}\,r_2(\mathrm{x}) + \cdots\}\,\phi(\mathrm{x})\,d\mathrm{x}\,, \end{aligned} \tag{3.43}$$

where Σ is the $k \times k$ asymptotic covariance matrix of $n^{1/2}\,\hat\theta_j$ and $n^{1/2}\,\hat\theta_k$ for $1 \le j, \ell \le k$ with the asymptotic variance of $n^{1/2}\,\hat\theta_j$ on its diagonal and the r_j's are polynomials in the components of k-vectors x, and of degree $3j$ and odd/even functions for odd/even j (recall that the odd indexed polynomials were vanished due to the symmetry of the region about the origin).

To construct a bootstrap ellipsoidal region

$$\mathrm{R}_3 = \big\{\hat\theta - n^{-1/2}\,\widehat{\Sigma}^{1/2}\,\mathrm{x} \,:\, \mathrm{x} \in \mathcal{S}_3\big\}\,.$$

Choose $\mathcal{S}_3$ to be the k-variate sphere centered at the origin and of radius $\hat z_\alpha$ such that

$$P(\Delta^* \le \hat z_\alpha | \mathcal{X}) = 1 - \alpha\,.$$

Then, $\theta \in \mathbf{R}_3$ is equivalent to $\widehat{\Sigma}^{-1/2}\,\mathrm{D} \in \mathcal{S}_3$.

Suppose that $\mathrm{S} = \Sigma^{-1/2}\,\mathrm{D}$ and $\mathcal{S}_1$ is a sphere centered at the origin and of radius z_α such that

$$\begin{aligned} 1-\alpha &= P(\mathrm{S} \in \mathcal{S}_1) \\ &= \int_{\mathcal{S}_1} \{1 + n^{-1/2}\, s_1(\mathrm{x}) + n^{-1}\, s_2(\mathrm{x}) + \cdots\}\, \phi(\mathrm{x})\, d\mathrm{x} \\ &= \int_{\mathcal{S}_1} \{1 + n^{-1}\, s_2(\mathrm{x}) + n^{-2}\, s_4(\mathrm{x}) + \cdots\}\, \phi(\mathrm{x})\, d\mathrm{x}\,, \end{aligned} \tag{3.44}$$

where the s_j are polynomials in the components of x, of degree $3j$ and an odd/even function for odd/even j, respectively. Moreover, we have used the fact that for any odd polynomial π and any sphere $\mathcal{S}$ centered at the origin,

$$\int_{\mathcal{S}} \pi(\mathrm{x})\, d\mathrm{x} = 0\,.$$

Therefore, (3.44) implies that

$$z_\alpha = z_0 + n^{-1}\, c_1 + n^{-2}\, c_2 + \cdots\,,$$

where z_0 denotes the radius of the sphere $\mathcal{S}_2$ centered at the origin and such that

$$P(\mathrm{N} \in \mathcal{S}_2) = \int_{\mathcal{S}_2} \phi(\mathrm{x})\, d\mathrm{x} = 1-\alpha\,,$$

in which N is k-variate Standard Normal, and $c_1,\, c_2, \cdots$ are constants depending on population moments. Also, similarly to (3.44), we have:

$$P(\mathrm{S}^* \in \mathcal{S}_3 | \mathcal{X}) = \int_{\mathcal{S}_3} \{1 + n^{-1}\, \hat{r}_1(\mathrm{x}) + n^{-2}\, \hat{r}_2(\mathrm{x}) + \cdots\}\, \phi(\mathrm{x})\, dx\,, \tag{3.45}$$

where, in the k-variate statistic

$$\mathrm{S}^* = n^{1/2}\,\widehat{\Sigma}^{-1/2}\,(\hat{\theta}^* - \hat{\theta}) = \widehat{\Sigma}^{-1/2}\,\mathrm{D}^*\,,$$

the $\hat{r}_j$'s are obtained from r_j on replacing population moments by sample moments, and are of degree $3j$ and odd/even functions for odd/even j.

Applying the argument at (3.44) to (3.45), we deduce that

$$\hat{z}_\alpha = z_0 + n^{-1}\,\hat{c}_1 + n^{-2}\,\hat{c}_2 + \cdots\,, \tag{3.46}$$

where the $\hat{c}_j$ denote the version of the c_j obtained by replacing population moments by sample moments. Therefore, by the delta method,

$$\begin{aligned} P\big(\theta \in \mathcal{R}_3\big) &= P\big(\mathrm{D} \in \Sigma^{-1/2}\,\mathcal{S}_3\big) = P\big(\mathrm{S} \in \mathcal{S}_3\big) \\ &= P\big(\mathrm{S} \in \hat{z}_\alpha\,z_\alpha^{-1}\,\mathcal{S}_1\big) \\ &= P\big(\{1 + (n\,z_0)^{-1}(\hat{c}_1 - c_1) + O_p(n^{-2})\}\;\mathrm{S} \in \mathcal{S}_1\big) \\ &= P\big(\{1 - (n\,z_0)^{-1}(\hat{c}_1 - c_1)\}\;\mathrm{S} \in \mathcal{S}_1\big) + O(n^{-2})\,, \end{aligned} \tag{3.47}$$

where we used the fact that $\hat{c}_j - c_j = O_p(n^{-1/2})$. We know that the distribution of

$$\mathrm{S}_1 = \{1 - (n\,z_0)^{-1}(\hat{c}_1 - c_1)\}\;\mathrm{S}$$

admits an Edgeworth expansion of the form

$$P\big(\mathrm{S}_1 \in \mathcal{S}\big) = \int_{\mathcal{S}} \{1 + n^{-1/2}\,t_1(\mathrm{x}) + n^{-1}\,t_2(\mathrm{x}) + \cdots\}\,\phi(\mathrm{x})\,dx\,, \tag{3.48}$$

where t_j is a polynomial of degree $3j$ and is an odd/even function for odd/even j, respectively, and $\mathcal{S}$ is any k-variate sphere in R^k. More precisely, since S and

$\mathbf{S}_1$ differ only in terms of $O_p(n^{-3/2})$, $t_j = s_j$ for $j = 1, 2$. Furthermore, if $\mathcal{S}$ is a sphere centered at the origin, terms of odd order in particular the term in $n^{-1/2}$ vanish from the expansion (3.48). Thus,

$$\begin{aligned} P(\mathbf{S}_1 \in \mathcal{S}_1) &= \int_{\mathcal{S}_1} \{1 + n^{-1/2}\, t_1(\mathrm{x}) + n^{-1}\, t_2(\mathrm{x}) + \cdots\}\, \phi(\mathrm{x})\, d\mathrm{x} \\ &= \int_{\mathcal{S}_1} \{1 + n^{-1}\, t_2(\mathrm{x})\}\, \phi(\mathrm{x})\, d\mathrm{x} + O(n^{-2})\,, \end{aligned} \tag{3.49}$$

and then

$$P(\theta \in \mathcal{R}_3) = 1 - \alpha + O(n^{-1})\,.$$

However, since we deal with the multivariate distribution of $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ for finding the distribution of $\max_{1\le j\le k}(\hat{\theta}_j - \theta_j)$, and the remainder terms in (3.7) grow up as j increases, we need to restrict $k = k_n$ to increase sufficiently slowly. This condition arises when we obtain the Edgeworth expansion of the distribution function $\mathbf{D}$ in terms of that of $\mathbf{I}$. In particular, the distribution of $\max_{1\le j\le k}(\hat{\theta}_j - \theta_j)$ depends on the moments of the distribution of $(\hat{\theta}_1, \cdots, \hat{\theta}_k)$, needing to be controlled with terms of $n^{-1/2}$ or n^{-1} as k increases. See (3.10) and (3.16). Therefore, we can expect the same results to hold if $k = k_n$ increases sufficiently slowly.

Chapter 4

Properties of Linear Regression Estimators

4.1 Introduction

Suppose that, for $j = 1, \cdots, p$ and $i = 1, \cdots, n$, the $X_i(t_j)$ are discretizations of continuous functions X_i where $t_1, \cdots, t_p \in \mathcal{I}$ are the points at which the continuous functions $X_i(t)$ are digitized. Then, when treating a linear regression model with these predictors, for a function $b(t)$ we may approximate a linear functional like $E(Y_i|X_i) = \int_{\mathcal{I}} X_i(t)\, b(t)\, dt$ by $\sum_{j=1}^{p} X_i(t_j)\, b(t_j)$. In such problems, Frank and Friedman (1993) summarized two methods, Partial Least Square (PLS) and Principal Components Regression (PCR), commonly used by chemometrists in analysing these kinds of data. Then the authors compared the two methods with ridge regression (RR) in a unified situation in which the coefficient vector is constrained to be in some subspace, and the projected sample predictor variables in this subspace have larger spread (variance).

Hastie and Mallows' discussion (1993) of the three techniques (PLS, RR and PCR) as well as the penalized least square method have contributed to the better understanding of the problems arising from analysing functional data. Hastie and Mallows (1993) pointed out that in the presence of a large number of the covariates $X_i(t_j)$, which generally exceed the number of the sample, the model may loose its efficiency for prediction. Furthermore, high correlations of the explanatory variables $X_i(t_j)$ in the model may result in an uninterpretable model. Then, the authors mentioned how to get a better estimation of the coefficient by constraining, via the penalized least square method, the regression slope to be smooth. Hastie and Mallows (1993) emphasized that it is better to use methods creating an order relation among the covariates in terms of their index values. In this way, the problem can be removed by using fewer covariates rather than hundreds of the original covariates.

Hastie and Mallows (1993) also proposed a "Smooth Basis Expansions" method to model the coefficient $b(t)$ smoothly, where the regression coefficient $b(t)$ is expressed in terms of expansion of a sequence of smooth basis functions such as polynomials, cosinusoids and splines. In their response to Hastie and Mallows (1993), Frank and Friedman confirmed that procedures that take an order relation among the covariates indices into account or that constrain the coefficient $b(t)$ to be a smooth function might work better than RR, PLS and PCR in cases in which the predictor curves are not smooth. Indeed, such methods that take the functional nature of the problem into account seem more reasonable than those which do not. See Marx and Eilers (1999), who expressed the advantages of a functional technique by comparing with the nonfunctional ones such as PLS and PCR.

Karhunen (1946) developed a theory of stochastic processes in Hilbert spaces. Using this theory, Grenander (1950) was able to take first steps on FDA by applying the Karhunen-Loève expansions to functional data, including a proposal for functional regression. However, functional regression models were widely used only after recent works such as Ramsay and Dalzell (1991) and Ramsay and Silverman (1997, Chapters 10 and 11).

There are three kinds of functional regression models depending on the nature of predictors and responses:

- Both predictors and response variable are functions.
- The response is a function and predictors are vectors.
- The response is a scalar and predictors are functions.

Models of the first type were introduced by Ramsay and Dalzell (1991). These models can be viewed as extensions of the multivariate linear regression model $E(Y|\mathbf{X}) = \mathbf{B}\,\mathbf{X}$, where $\mathbf{B}$ and $\mathbf{X}$ are matrices. When both response and predictor tend to be a continuum, the model is changing to

$$E[Y(t)|X] = \mu(t) + \int X(s)\,\beta(s,t)\,ds\,,$$

where μ is the mean response function and β is slope of the regression. Then the unknown parameter β can be estimated by applying regularization methods; for example by penalized splines (James, 2002), by basis representations or truncation of series expansions (Ramsay and Silverman 1997, Chapter 11 and Chiou et al., 2004). Also a discussion about models in which the response is a random function and the predictors are scalars or vectors can be founded in Chiou et al. (2004). Accounts of models, in which the predictors are functions and the responses are generalized variables such as binary, count or continuous type, can be founded in work of Cardot, Ferraty and Sarda (1999, 2003), Cardot, Ferraty, Mas and Sarda (2003), James (2002) and Müller and Stadtmüller (2005). This type of data arises in functional prediction problem such as prediction of total annual precipitation for Canadian weather stations from the pattern of temperature variation through the year (Ramsay and Silverman, 1997; Chapter 10) as well as the two examples given in the introduction to Chapter 1. In this research we discuss a linear regression model of the third kind. In Section 4.2 we explain the problem of estimation of the slope function, and also its differences from the prediction problem in functional linear regression. In Section 4.3, using the expansions and their properties given in Chapters 1 and 2, the impact of eigenvalue spacings on properties of linear regression estimators is discussed, and the validity of simple accounts of the performance of functional linear regression is explored. It is observed that those accounts are valid if eigenvalues are reasonably well separated, but not otherwise. We also briefly discuss the prediction problem along with the estimation one in Section 4.4.

4.2 Estimation of Slope Function

The functional simple linear regression model is

$$Y_i = a + \int_{\mathcal{I}} b\, X_i + \epsilon_i\,, \quad 1 \le i \le n\,, \tag{4.1}$$

where b and X_i are square-integrable functions from $\mathcal{I}$ to the real line, a, Y_i and ϵ_i are scalars, a and b are deterministic, the pairs $(X_1, \epsilon_1), (X_2, \epsilon_2), \ldots$ are independent and identically distributed, the random functions X_i are independent of the errors ϵ_i, $\sigma^2 = E(\epsilon^2) < \infty$, $E(\epsilon) = 0$ and $\int_{\mathcal{I}} E(X^2) < \infty$, where ϵ and X are distributed as ϵ_i and X_i, respectively.

Estimation of b is intrinsically an infinite-dimensional problem. Therefore, in functional linear regression, the problem involves using smoothing or regularisation methods which enable us to reduce dimension. Thus, this is an aspect which makes functional regression analysis distinct from classical linear regression. Depending on the purpose for which the estimator b is used, the amount of smoothness needing to be applied is different. When dealing with estimating b, optimal smoothness of $\hat{b}$ will usually result in $\int \hat{b}\, x$, which is used for prediction, being over-smoothed for estimating $\int b X$ given a value x of X. This is due to operation of integration in computing $\int \hat{b}\, x$ from $\hat{b}$, awarding additional smoothness. Therefore, the manner in which $\hat{b}$ is used for prediction is different from that for estimating b.

However, these two problems are not entirely separated from each other, because knowing the behavior of the slope function b, for example in what points $b(t)$ takes large or small values, gives useful information about the role of the functional explanatory variables in the model which in turn provides information about where a future observation x of X will have greatest influence on the value of $\int b\, x$. Thus, it seems that the problem of estimating the slope b is as a prelude to estimating $\int b\, x$, and perhaps because of that, unlike the case of classical linear regression, in this field there is significant interest in estimating b in its own right (Cai and Hall 2004). See for example Ferraty and Vieu (2000), Cuevas et. al (2002), Cardot and Sarda (2003), Hall and Horowitz (2004), in which estimation of b and in particular convergence rates of the estimator $\hat{b}$ to b were discussed. In the present research, our focus is on the problem of estimating the slope function b. Nevertheless, we will discuss shortly the prediction problem along with our results in Section 4.4.

If we express X_i and b in terms of the orthonormal basis $\widehat{\psi}_1, \widehat{\psi}_2, \ldots$, then we have

$$X_i = \sum_{j=1}^{\infty} \xi_{ij}\, \widehat{\psi}_j, \qquad b = \sum_{j=1}^{\infty} \bar{b}_j\, \widehat{\psi}_j\,, \tag{4.2}$$

where $\xi_{ij} = \int X_i\, \widehat{\psi}_j$ and $\bar{b}_j = \int b\, \widehat{\psi}_j$ denote the associated generalised Fourier series of X_i and b, respectively. Then we can write (4.1) equivalently as

$$Y_i = a + \sum_{j=1}^{\infty} \bar{b}_j \, \xi_{ij} + \epsilon_i \,. \tag{4.3}$$

A model equivalent to (4.1) is

$$Y_i - \mu = \int_{\mathcal{I}} b \, (X_i - \eta) + \epsilon_i \,, \quad 1 \le i \le n \,, \tag{4.4}$$

where $\eta = E(X_i)$ is a deterministic function on $\mathcal{I}$ and $\mu = E(Y_i) = a + \int b \, \eta$. Define $g(u) = E\big[(Y - \mu)\{X(u) - \eta(u)\}\big]$, where (X, Y) represents a generic version of (X_i, Y_i). Then $b, g \in L_2(\mathcal{I})$ can be expressed, respectively, as $b = \sum_{j=1}^{\infty} b_j \, \psi_j$ and $g = \sum_{j=1}^{\infty} g_j \, \psi_j$ where $g_j = \int g \, \psi_j$ and $b_j = \int b \, \psi_j$. It can be seen that the slope function $b(t)$ is determined by solving the following equation:

$$K b = g \,, \tag{4.5}$$

where K is the covariance operator. Consequently, $g_j = \langle g, \psi_j \rangle = \theta_j \, \langle b, \psi_j \rangle = \theta_j \, b_j$ and

$$b(t) = \sum_{j=1}^{\infty} b_j \, \psi_j = \sum_{j=1}^{\infty} \theta_j^{-1} \, g_j \, \psi_j \,. \tag{4.6}$$

Below we explain that for estimating the slope function b, dimension reduction is mandatory.

Theorem 15.4 of Kress (1999) points out that for the two normed spaces H_1 and H_2 if $A : H_1 \longrightarrow H_2$ is a compact linear operator, then the equation $A \, x = f$ is ill-posed if H_1 is not of finite dimension; i.e. if H_1 is an infinite dimensional space then the inverse of A may not exist or it may not be bounded. Thus, in problem (4.5) we are facing an ill-posed equation since K is compact and

$H_1 = L_2(\mathcal{I})$, which is of infinite dimension. Indeed, when the θ_j decay rapidly, estimation of b in (4.6) is problematic. To overcome the problems, one way is to project the data on a finite dimensional space. Alternatively, one may treat the problem like a ridge regression where for $\alpha > 0$, the equation (4.4) is replaced by $(\alpha I + K)b_\alpha = g$. The extra parameter $\alpha > 0$ causes the operator $\alpha I + K$ to be injective and its inverse, $(\alpha I + K)^{-1}$, to be bounded (Theorem 13.26 of Kress, 1999). For illustrations of the first approach, see for example, Ramsay and Silverman (1997, Chapter 10) and Cardot, Ferraty and Sarda (2003), and of the second, see Hall and Horowitz (2004).

Ramsay and Silverman (1997, Chapter 10) point out that expressing functional data in terms of a basis such as the Fourier basis or so on might be useful. The underlying idea is that although functional data are usually observed discretely as a large number of points, each being regarded as a recording of the curve $X(t)$ at many points, there is not a large number of important modes of variation there. Therefore, they can be expressed in a finite number of terms, M say, resulting in a smooth presentation of the curves. Similarly, in the case of b, this causes reduction of the degree of freedom in the model (4.3), and providing an interpretable description of the influence of the covariate $X(t)$ on the outcome Y. In particular, Ramsay and Silverman (1997) illustrate the effect of regularization by truncating the basis by choosing a value $r < M$.

Based on the same idea, dimension reduction through functional principal components analysis has become very popular. From previous works such as those of Rao (1958) who used it for growth curves, Besse and Ramsay (1986), Castro, Lawton and Sylvestre (1986), Brekey et al. (1991), Ramsay and Dalzell (1991), Rice and Silverman (1991) and Silverman (1995, 1996) to the recent ones such as Brumback and Rice (1998), Cardot (2000), Girard (2000), James, Hastie and Sugar (2000), Boente and Fraiman (2000), He, Müller and Wang

(2003) and Yao, Müller and Wang (2005) all make significant contributions to PCA. In particular, works on functional linear model, including those by Cardot, Ferraty and Sarda (1999, 2000, 2003), Hall and Horowitz (2004), Cai and Hall (2004) and Hall and Hosseini-Nasab (2006), tackled infinite-dimensionality by employing PCA. Projecting observations onto the space spanned by the first r eigenfunctions of $\widehat{K}$ leads to an optimal linear representation of X_i with respect to the explained variance (Dauxois et al., 1982). In this way, the dominant modes of variation of random curves reduce the functional linear model to a conventional regression model with a finite set of functional principal component scores ξ_{ij}, $j = 1, \cdots, r$ as covariates. Then the unknown parameters can be estimated through conventional methods such as least squares methods.

The true value, (a^0, b^0) say, of (a, b) may be estimated by the least squares method through minimising

$$\sum_{i=1}^{n} (Y_i - a - \sum_{j=1}^{r} \bar{b}_j \, \xi_{ij})^2 \,, \tag{4.7}$$

with respect to $a, \bar{b}_1, \bar{b}_2, \cdots, \bar{b}_r$, and taking $\bar{b}_j = 0$ for $j \geq r+1$. Equation (4.7) can be written as $(\mathbf{Y}-\mathbf{\Lambda}^*\mathbf{b}^*)^T(\mathbf{Y}-\mathbf{\Lambda}^*\mathbf{b}^*)$ where $\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $\mathbf{\Lambda}^* = \begin{pmatrix} 1 & \xi_{11} & \ldots & \xi_{1r} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_{n1} & \ldots & \xi_{nr} \end{pmatrix}$, $\mathbf{b}^* = (a, \bar{b}_1, \cdots, \bar{b}_r)^T$, and T denotes the transpose of the vector. Differentiating with respect to $\mathbf{b}^*$ and equating to zero for the maximum, the results are

$$\hat{a} = \bar{Y} - \sum_{j=1}^{r} \hat{b}_j, \qquad \hat{b}_{(r)} = (\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_r) = \widehat{\Sigma}_{(r)}^{-1} \hat{g}_{(r)} \,, \tag{4.8}$$

where $\bar{Y} = n^{-1} \sum_i Y_i$, $\bar{\xi}_j = n^{-1} \sum_i \xi_{ij}$, $\widehat{\Sigma}_{(r)}$ is the $r \times r$ matrix with (j, k)th component $\hat{\sigma}_{jk}$, $\hat{g}_{(r)} = (\hat{g}_1, \ldots, \hat{g}_r)^{\mathrm{T}}$,

$$\hat{\sigma}_{jk} = \frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\xi_{ik} - \bar{\xi}_k) = \int_{\mathcal{I}} \widehat{K}(u,v)\,\widehat{\psi}_j(u)\,\widehat{\psi}_k(v)\,du\,dv = \hat{\theta}_j\,\delta_{jk}\,,$$
$$\hat{g}_j = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})(\xi_{ij} - \bar{\xi}_j)\,,$$

and δ_{jk} is the Kronecker delta. Therefore, $\widehat{\Sigma}_{(r)} = \mathrm{diag}(\hat{\theta}_1, \cdots, \hat{\theta}_r)$, and by (4.8), for $u \in \mathcal{I}$, our estimator of b is

$$\hat{b}(u) = \sum_{j=1}^{r} \hat{b}_j\,\widehat{\psi}_j(u) = \sum_{j=1}^{r} \hat{\theta}_j^{-1}\,\hat{g}_j\,\widehat{\psi}_j(u)\,.$$

Equivalently, in view of (4.5), $\hat{b}$ can be obtained from a sample version of the equation when the X_i are projected into the space spanned by the eigenfunctions associated with the r greatest eigenvalues $\hat{\theta}_j$, $j = 1, \cdots, r$. Put $\hat{g}(t) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(t) - \bar{X}(t)\}\,(Y_i - \bar{Y})$. Then $\hat{g} = \sum_{j=1}^{\infty} \hat{g}_j\,\widehat{\psi}_j$, where

$$\hat{g}_j = \int \hat{g}\,\widehat{\psi}_j = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})\,(\xi_{ij} - \bar{\xi}_j)\,.$$

Hence, $\hat{b}_j = \hat{\theta}_j^{-1}\,\hat{g}_j$ for $j = 1, \cdots, r$.

The smoothing parameter r can be chosen by cross validation. In the context of functional data analysis, the predictive cross-validation criterion is given by

$$\mathrm{CV}(r) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \hat{a}_{-i;r} - \int_{\mathcal{I}} \hat{b}_{-i;r}\,X_i \right\}^2 . \tag{4.9}$$

Here, $(\hat{a}_{-i;r}, \hat{b}_{-i;r})$ denotes the least-squares estimator of (a, b) that is obtained by confining attention to the set $\mathcal{Z}_i$, say, of all data pairs (X_j, Y_j) excluding the ith, and both $\hat{a}_{-i;r}$ and $\hat{b}_{-i;r}$ use the empirical Karhunen-Loève expansion of length r computed from $\mathcal{Z}_i$. We choose r to minimise $\mathrm{CV}(r)$. Cross validation can be used to select r when computing $\hat{p}(x) = \hat{a} + \int_{\mathcal{I}} \hat{b}x$ (see for example Cardot et al., 2003). Surprisingly, cross-validation also works well when estimating b; the latter problem can be expected to require significantly more smoothing than prediction, as we explained earlier. See numerical results in Chapter 5.

4.3 Asymptotic Mean Integrated Square Error Approximation

Let $b^0 = \sum_{j=1}^{\infty} b_j^0 \,\widehat{\psi}_j$ where $b_j^0 = \int b^0 \,\widehat{\psi}_j$. Assume that the true value, b^0, of the function b is square-integrable. Then,

$$\int (\hat{b} - b^0)^2 = \langle \hat{b} - b^0, \hat{b} - b^0 \rangle = \langle \sum_{j=1}^{r} (\hat{b}_j - b_j^0) - \sum_{j=r+1} b_j^0, \sum_{\ell=1}^{r} (\hat{b}_\ell - b_\ell^0) - \sum_{\ell=r+1} b_\ell^0 \rangle$$
$$= \sum_{j=1}^{r} (\hat{b}_j - b_j^0)^2 + \sum_{j=r+1} {b_j^0}^2 \,. \tag{4.10}$$

Hence,

$$Y_i - \bar{Y} = \sum_{j=1}^{\infty} b_j^0 \,(\xi_{ij} - \bar{\xi}_j) + \epsilon_i - \bar{\epsilon}\,, \quad i = 1, \cdots, n\,.$$

Multiplying both sides of the above equation by $(\xi_{ij} - \bar{\xi}_j)$ and summing up over i, we have

$$\hat{g}_j = \sum_{k=1}^{\infty} \hat{\sigma}_{jk} \, b_k^0 + \hat{\gamma}_j = \sum_{k=1}^{\infty} \hat{\theta}_j \, \delta_{jk} \, b_k^0 + \hat{\gamma}_j = \hat{\theta}_j \, b_j^0 + \hat{\gamma}_j \,,$$

where $\hat{\gamma}_j = \frac{1}{n} \sum_{i=1}^{n} (\epsilon_i - \bar{\epsilon})(\xi_{ij} - \bar{\xi}_j)$. Using the above result and defining $b_{(r)}^0 = (b_1^0, \cdots, b_r^0)^T$, we deduce from the second formula in (4.8) that

$$\hat{b}_{(r)} - b_{(r)}^0 = \widehat{\Sigma}_{(r)}^{-1} \, \hat{g}_{(r)} - b_{(r)}^0 \,.$$

Hence,

$$\hat{g}_{(r)} - \widehat{\Sigma}_{(r)}\, b^0_{(r)} = \begin{pmatrix} \hat{g}_1 \\ \vdots \\ \hat{g}_r \end{pmatrix} - \begin{pmatrix} \hat{\theta}_1\, b^0_1 \\ \vdots \\ \hat{\theta}_r\, b^0_r \end{pmatrix} = \begin{pmatrix} \hat{g}_1 - \hat{\theta}_1\, b^0_1 \\ \vdots \\ \hat{g}_r - \hat{\theta}_r\, b^0_r \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_r \end{pmatrix},$$

and thus $\hat{b}_j - b^0_j = \hat{\theta}_j^{-1}\,\hat{\gamma}_j$ for $j = 1, \cdots, r$. Substituting this result into (4.10) gives us:

$$\int (\hat{b} - b^0)^2 = \sum_{j=1}^{r} \hat{\theta}_j^{-2}\,\hat{\gamma}_j^2 + \sum_{j=r+1} b_j^{0\,2}. \tag{4.11}$$

Since $\hat{\theta}_j$ and $\widehat{\psi}_j$ are functionals of $X_1, \cdots, X_n$, independent of $\epsilon_1, \cdots, \epsilon_n$, and also (X_i, ϵ_i) are independent of each other,

$$\begin{aligned} E(\hat{\gamma}_j^2|\mathcal{X}) &= E\{\big[\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})(\xi_{ij} - \bar{\xi}_j)\big]^2|\mathcal{X}\} = \frac{1}{n^2}\sum_{i=1}^{n} E\{(\epsilon_i - \bar{\epsilon})^2(\xi_{ij} - \bar{\xi}_j)^2|\mathcal{X}\} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}(\xi_{ij} - \bar{\xi}_j)^2 E\big[(\epsilon_i - \bar{\epsilon})^2|\mathcal{X}\big] = \frac{1}{n^2}\sum_{i=1}^{n}(\xi_{ij} - \bar{\xi}_j)^2\,\sigma^2\,(1 - \frac{1}{n}) \\ &= \frac{1}{n}\,(1 - \frac{1}{n})\,\sigma^2\,\hat{\theta}_j\,. \end{aligned}$$

Hence, by (4.11),

$$\int_{\mathcal{I}} E\{(\hat{b} - b^0)^2|\mathcal{X}\} = n^{-1}\,(1 - n^{-1})\,\sigma^2 \sum_{j=1}^{r} \hat{\theta}_j^{-1} + \sum_{j=r+1}^{\infty} \Big(\int b^0\,\widehat{\psi}_j\Big)^2. \tag{4.12}$$

Moreover,

$$\begin{aligned}\sum_{j=r+1}^{\infty}\left(\int b^0\,\widehat{\psi}_j\right)^2 &= \sum_{j=r+1}^{\infty}\left[\int b^0\,(\psi_j+\widehat{\psi}_j-\widehat{\psi}_j)\right]^2\\
&= \sum_{j=r+1}^{\infty}\left[\left(\int b^0\,\psi_j\right)^2+\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}^2+2\left\{\int b^0\,\psi_j\right\}\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}\right]\\
&= \sum_{j=r+1}^{\infty}\left(\int b^0\,\psi_j\right)^2-2\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)\left(\int b^0(\widehat{\psi}_j-\psi_j)\right)\\
&\quad -\sum_{j=1}^{r}\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}^2+\sum_{j=1}^{\infty}\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}^2\\
&\quad +2\sum_{j=1}^{\infty}\left(\int b^0\,\psi_j\right)\left\{\int b^0\,(\widehat{\psi}_j-\psi_j)\right\}.\end{aligned}$$

Because $\|b^0\|^2=\sum_{j=1}^{\infty}\left(\int b^0\,\psi_j\right)^2=\sum_{j=1}^{\infty}\left(\int b^0\,\widehat{\psi}_j\right)^2$, the last two terms above vanish, i.e.

$$\begin{aligned}\sum_{j=r+1}^{\infty}\left(\int b^0\,\widehat{\psi}_j\right)^2 = \sum_{j=r+1}^{\infty}\left(\int b^0\,\psi_j\right)^2-2\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)\left(\int b^0(\widehat{\psi}_j-\psi_j)\right)&\\
-\sum_{j=1}^{r}\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}^2.&\qquad (4.13)\end{aligned}$$

Thus, substituting (4.13) into (4.12) results in

$$\begin{aligned}\int_{\mathcal{I}} E\{(\hat{b}-b^0)^2|\mathcal{X}\} &= n^{-1}\,(1-n^{-1})\,\sigma^2\sum_{j=1}^{r}\hat{\theta}_j^{-1}+\sum_{j=r+1}^{\infty}\left(\int b^0\,\psi_j\right)^2\\
&\quad -2\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)\left(\int b^0(\widehat{\psi}_j-\psi_j)\right)-\sum_{j=1}^{r}\left\{\int b^0(\widehat{\psi}_j-\psi_j)\right\}^2.\end{aligned}\qquad (4.14)$$

Result (4.12) suggests, but of course does not prove, that

$$\int_{\mathcal{I}} E(\hat{b}-b^0)^2 \sim \frac{\sigma^2}{n}\sum_{j=1}^{r}\theta_j^{-1}+\sum_{j=r+1}^{\infty}\left(\int_{\mathcal{I}} b^0\,\psi_j\right)^2, \qquad (4.15)$$

where "$A_n \sim B_n$" means that the ratio of the random variables A_n and B_n converges to 1 as $n \to \infty$. If (4.15) were correct then the first term on the right-hand side would denote the dominant contribution from error-about-the-mean to integrated squared error, and the second term would be the dominant contribution from squared bias. The right-hand side of (4.15) is reminiscent of familiar formulae for the mean integrated squared error of orthogonal series estimators; see, for example, Kronmal and Tarter (1968).

In the following we state a result that implies (4.15). Let $A = X - E(X)$, define $b_j^0 = \int_{\mathcal{I}} b^0\, \psi_j$, and consider the conditions:

$\theta_j = j^{-a}\, L(j)$ and $|b_j^0| = j^{-b}\, M(j)$, where $b > a+\frac{1}{2} > \frac{3}{2}$ and L, M are slowly-varying functions; $\theta_j - \theta_{j+1} \geq \text{const.}\, j^{-a-1}$; the process X has all moments finite; for each integer $r \geq 1$, $\theta_j^{-r}\, E(\int A\, \psi_j)^{2r}$ is bounded uniformly in j; the errors ϵ_i in (4.1) are independent and identically distributed with all moments finite, zero mean and variance σ^2; and the frequency cut-off, r, is in the range $1 \leq r \leq r_0$, where $r_0 = r_0(n)$ satisfies $r_0 = O(n^{(1-\eta)/2(a+1)})$ for some $0 < \eta < 1$. (4.16)

Under these assumptions, r_0 can be chosen so that it is an order of magnitude larger than the value that minimises mean integrated squared error.

Theorem 4.1. To eliminate pathologies arising from too-small values of $\hat{\theta}_j$, replace $\hat{b}_j$ by an arbitrary fixed constant if $|\hat{b}_j| > c_1\, n^{c_2}$, for any given $c_1, c_2 > 0$. Then, if (4.16) holds, so too does (4.15), uniformly in $1 \leq r \leq r_0$.

The assumption $\theta_j - \theta_{j+1} \geq$ const. j^{-a-1} for $j \geq 1$, excludes cases where two or more of the eigenvalues θ_j are close together, in particular where they are tied. In general when using the weaker condition $\frac{\max(\theta_j,\theta_k)}{|\theta_j-\theta_k|} \leq C\, \frac{\max(j,k)}{|j-k|}$ for some $C > 0$

and all $1 \leq j < k < \infty$ (similar to what we assumed in the appendix), it holds, for sufficiently large j, when, for example, $\theta j = C_1 j^{-C_2}$, where $C_1 > 0$ and $C_2 > 2$.

Replacing $\hat{b}_j$ by an arbitrary fixed constant if $|\hat{b}_j| > c_1 n^{c_2}$ guarantees that the expected value on the left-hand side of (4.15) is finite. See the Appendix for a proof and discussion about the constraint.

Versions of Theorem 4.1 are available under more general conditions than (4.16); the coefficients θ_j and b_j need only be "polynomial-like," and in particular need not be regularly varying functions of j. See the Appendix.

Formula (4.15) can be used to construct a plug-in rule for choosing r to optimise performance of $\hat{b}$ as an estimator of b. Assuming that the slowly varying functions L and M in (4.16) are asymptotically constant, the mean integrated squared error of $\hat{b}$ is asymptotic to $C_1 n^{-1} r^{a+1} + C_2 r^{1-2b}$, which is minimised by choosing $r = C_3 n^{1/(a+2b-1)}$, where $C_1, C_2, C_3 > 0$ are constants. The resulting mean-square convergence rate is $n^{-(2b-1)/(a+2b-1)}$.

A valid approximation to the first term on the right-hand side of (4.12) is the first term on the right-hand side of (4.15), provided that

$$\frac{\sum_{j=1}^{r} \hat{\theta}_j^{-1}}{\sum_{j=1}^{r} \theta_j^{-1}} \longrightarrow 1, \qquad \text{in probability}. \tag{4.17}$$

Condition (4.17) can be established in many circumstances, as is explained in the Appendix. However, the second term on the right-hand side of (4.15) is not always appropriate. The properties of expansions of the basis functions $\widehat{\psi}_j$ discussed in Chapters 1 and 2, show that they depend on the spacings of the eigenvalues θ_j, and particularly in that respect we should take care of the spacings, when treating the behaviour of the squared-bias approximation. Indeed, we will show in the following result that the approximation at (4.15) can fail in some circumstances.

Theorem 4.2. Under notation introduced before, if X is a Gaussian process, then, to first order,

$$\sum_{j=r+1}^{\infty} E\Big(\int_{\mathcal{I}} b^0\,\widehat{\psi}_j\Big)^2 = \sum_{j=r+1}^{\infty}\Big(\int_{\mathcal{I}} b^0\,\psi_j\Big)^2 + n^{-1}\sum_{j=1}^{r}\theta_j\sum_{k\,:\,k\neq j}\theta_k\,(\theta_j-\theta_k)^{-2}(\beta_j^2-\beta_k^2)$$
$$+\ (\text{higher-order terms})\,, \qquad (4.18)$$

where $\beta_j = \int_{\mathcal{I}} b^0\,\psi_j$.

Proof: We first need the following Lemma.

lemma 4.1: $E(a_{jk}) = \delta_{jk} + O(n^{-1})$, where a_{jk} introduced in (1.11) and (1.12), and δ_{jk} is the Kronecker delta.

Proof of Lemma: In order to evaluate $E(a_{\ell j})$ we need to compute

$$T(u,v,w,z) = E\{Z(u,v)\,Z(w,z)\} = n\,E\Big[\{\widehat{K}(u,v)-K(u,v)\}\,\{\widehat{K}(w,z)-K(w,z)\}\Big]\,.$$

Put $\eta(u) = E\{X(u)\}$, $Y = X-\eta$, $Y_i = X_i - \eta$ and

$$\widetilde{K}(u,v) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(u)-\eta(u)\}\,\{X_i(v)-\eta(v)\} = \frac{1}{n}\sum_{i=1}^{n} Y_i(u)\,Y_i(v)\,.$$

Note that $\widehat{K}(u,v) = \widetilde{K}(u,v) - \{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}$, whence it follows that

$$\begin{aligned} T(u,v,w,z) = {} & n\,E\Big[\{\widetilde{K}(u,v)-K(u,v)\}\,\{\widetilde{K}(w,z)-K(w,z)\}\Big] \\ & - n\,E\Big[\{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}\,\{\widetilde{K}(w,z)-K(w,z)\}\Big] \\ & - n\,E\Big[\{\widehat{K}(w,z)-K(w,z)\}\,\{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}\Big] \\ & + n\,E\Big[\{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}\,\{\bar{X}(w)-\eta(w)\}\,\{\bar{X}(z)-\eta(z)\}\Big]\,. \end{aligned} \qquad (4.19)$$

The second term on the right-hand side of (4.19) can be bounded as

$$
\begin{aligned}
E\Big[\{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}\,\{\widetilde{K}(w,z)-K(w,z)\}\Big] \\
&= E\bigg[\frac{1}{n^3}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} Y_i(u)\,Y_j(v)\big(Y_k(w)\,Y_k(z)-K(w,z)\big)\bigg] \\
&= \frac{1}{n^3}\sum_{i=1}^{n} E\Big[Y_i(u)\,Y_i(v)\big(Y_i(w)\,Y_i(z)-K(w,z)\big)\Big] = O(n^{-2})\,, \qquad (4.20)
\end{aligned}
$$

where we have used independence of the X_i to obtain the second equality above. This is the case in the third term on the right-hand side of (4.19). Also,

$$
\begin{aligned}
E\Big[\{\bar{X}(u)-\eta(u)\}\,\{\bar{X}(v)-\eta(v)\}\,\{\bar{X}(w)-\eta(w)\}\,\{\bar{X}(z)-\eta(z)\}\Big] \\
&= E\Big[\frac{1}{n^4}\sum_{\ell=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{n} Y_i(u)\,Y_j(v)\,Y_k(w)\,Y_\ell(z)\Big] \\
&= \frac{1}{n^4}\bigg\{\sum_{i=1}^{n} E\big[Y_i(u)\,Y_i(v)\,Y_i(w)\,Y_i(z)\big] + \sum_{i\neq j} E\big[Y_i(u)\,Y_i(v)\,Y_j(w)\,Y_j(z) \\
&\qquad + Y_i(u)\,Y_i(w)\,Y_j(v)\,Y_j(z) + Y_i(u)\,Y_i(z)\,Y_j(v)\,Y_j(w)\big]\bigg\} \\
&= O(n^{-2})\,. \qquad (4.21)
\end{aligned}
$$

Combining (4.19), (4.20) and (4.21) results in

$$
\begin{aligned}
T(u,v,w,z) &= n\,E\Big[\{\widetilde{K}(u,v)-K(u,v)\}\,\{\widetilde{K}(w,z)-K(w,z)\}\Big] + O(n^{-1}) \\
&= n\,E\Big[\frac{1}{n}\sum_{i=1}^{n} Y_i(u)\,Y_i(v)\,\frac{1}{n}\sum_{j=1}^{n} Y_j(w)\,Y_j(z)\Big] - n\,K(u,v)\,K(w,z) + O(n^{-1}) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\Big[Y_i(u)\,Y_i(v)\,Y_i(w)\,Y_i(z)\Big] + \frac{1}{n}\sum_{i\neq j} E\Big[Y_i(u)\,Y_i(v)\Big]\,E\Big[Y_j(w)\,Y_j(z)\Big] \\
&\quad - n\,K(u,v)\,K(w,z) + O(n^{-1})\,.
\end{aligned}
$$

Consequently,

$$T(u,v,w,z) = E\{Y(u)\,Y(v)\,Y(w)\,Y(z)\} - K(u,v)\,K(w,z) + O(n^{-1})\,. \quad (4.22)$$

We can write $Y = \sum_{k\geq 1} \xi_k\,\psi_k$, where the generalised Fourier coefficients ξ_k are uncorrelated, and have zero means. Recall that

$$Z(u,v) = n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n} W_i(u,v) - \bar{Y}(u)\,\bar{Y}(v)\right],$$

where $W_i(u,v) = Y_i(u)\,Y_i(v) - K(u,v)$. Thus,

$$\int Z\,\psi_j\,\psi_k = n^{-1/2}\sum_{i=1}^{n} \xi_{ij}\,\xi_{ik} - n^{1/2}\{\delta_{jk}\,E(\xi_j^2) - \bar{\xi}_j\,\bar{\xi}_k\}\,,$$

where $\bar{\xi}_j = \frac{1}{n}\sum_{i=1}^{n}\xi_{ij}$. Therefore, if $k \neq j$, then $E\big[\int Z\,\psi_j\,\psi_k\big] = 0$. If X is a Gaussian process, the ξ_k's are independent Normal random variables with zero mean and respective variances θ_k. Therefore, by (1.52),

$$\begin{aligned} &E\{Y(u)\,Y(v)\,Y(w)\,Y(z)\} - K(u,v)\,K(w,z) \\ &\quad = \sum_{k=1}^{\infty}\{E(\xi_k^4) - (E\xi_k^2)^2\}\,\psi_k(u)\,\psi_k(v)\,\psi_k(w)\,\psi_k(z) \\ &\qquad + \sum\sum_{k_1\neq k_2} E(\xi_{k_1}^2)\,E(\xi_{k_2}^2)\,\Big\{\psi_{k_1}(u)\,\psi_{k_1}(w)\,\psi_{k_2}(v)\,\psi_{k_2}(z) \\ &\qquad\qquad + \psi_{k_1}(u)\,\psi_{k_1}(z)\,\psi_{k_2}(v)\,\psi_{k_2}(w)\Big\}\,. \end{aligned} \quad (4.23)$$

Combining (4.22) and (4.23) we deduce that

$$
\begin{aligned}
t(j_1, \ldots, j_4) &\equiv E\bigg\{\Big(\int Z\, \psi_{j_1}\, \psi_{j_2}\Big)\Big(\int Z\, \psi_{j_3}\, \psi_{j_4}\Big)\bigg\} \\
&= \int T(u, v, w, z)\, \psi_{j_1}(u)\, \psi_{j_2}(v)\, \psi_{j_3}(w)\, \psi_{j_4}(z)\, du\, dv\, dw\, dz \\
&= \Big\{E\big(\xi_{j_1}^4\big) - \big(E\xi_{j_1}^2\big)^2\Big\}\, I(j_1 = j_2 = j_3 = j_4) \\
&\quad + E\big(\xi_{j_1}^2\big)\, E\big(\xi_{j_2}^2\big)\, \{I(j_1 = j_3 \neq j_2 = j_4) + I(j_1 = j_4 \neq j_2 = j_3)\} + O\big(n^{-1}\big)\,.
\end{aligned}
\tag{4.24}
$$

Result (4.24) implies that if $\ell \neq j$ and $k \neq j$ then $t(j, j, j, k)$, $t(j, \ell, k, \ell)$ equal $O(n^{-1})$, and

$$
t(j, \ell, j, \ell) = E\big(\xi_j^2\big)\, E\big(\xi_\ell^2\big) + O\big(n^{-1}\big) = \theta_j\, \theta_\ell + O\big(n^{-1}\big)\,.
$$

Hence, by (1.11) and (1.12),

$$
\begin{aligned}
&E(a_{jk}) = O(n^{-2})\,, \quad \text{if } k \neq j\,, \\
&E(a_{jj}) = 1 - \tfrac{1}{2}\, n^{-1} \textstyle\sum_{\ell : \ell \neq j} (\theta_j - \theta_\ell)^{-2}\, \theta_j\, \theta_\ell + O\big(n^{-2}\big)\,.
\end{aligned}
\tag{4.25}
$$

This finishes the proof of the lemma. ■

Writing $b^0 = \sum_{k \geq 1} \beta_k\, \psi_k$, by (4.25), we have

$$
\begin{aligned}
t_{1j} &\equiv \Big(\int_{\mathcal{I}} b^0\, \psi_j\Big)\Big\{\int_{\mathcal{I}} b^0\, E(\widehat{\psi}_j - \psi_j)\Big\} = \beta_j \sum_{k=1}^{\infty} E(a_{jk} - \delta_{jk})\, \beta_k \\
&= \beta_j^2\, E(a_{jj} - 1) + O\big(n^{-2}\big) = -n^{-1}\, \tfrac{1}{2}\, \theta_j\, \beta_j^2 \sum_{k : k \neq j} (\theta_j - \theta_k)^{-2}\, \theta_k + O\big(n^{-2}\big)\,,
\end{aligned}
\tag{4.26}
$$

$$\begin{aligned}t_{2j} &\equiv E\bigg\{\int_{\mathcal{I}} b^0\,(\widehat{\psi}_j - \psi_j)\bigg\}^2 = E\bigg\{\sum_{k=1}^{\infty}(a_{jk}-\delta_{jk})\,\beta_k\bigg\}^2 \\ &= E\bigg\{\sum_{k=1}^{\infty}(a_{jk}-Ea_{jk})\,\beta_k\bigg\}^2 + \bigg\{\sum_{k=1}^{\infty}(Ea_{jk}-\delta_{jk})\,\beta_k\bigg\}^2 \\ &= E\bigg\{\sum_{k=1}^{\infty}(a_{jk}-Ea_{jk})\,\beta_k\bigg\}^2 + O\big(n^{-2}\big)\,. \end{aligned} \tag{4.27}$$

Also, by (1.11), (1.12) and (4.25),

$$\begin{aligned} n\,E\bigg\{\sum_{k=1}^{\infty}(a_{jk}-Ea_{jk})\,\beta_k\bigg\}^2 &= E\bigg\{\sum_{k\,:\,k\neq j}(\theta_j-\theta_k)^{-1}\Big(\int Z\,\psi_j\,\psi_k\Big)\,\beta_k\bigg\}^2 + O\big(n^{-1}\big) \\ &= \sum\sum_{k_1,k_2\,:\,k_1\neq j,k_2\neq j} (\theta_j-\theta_{k_1})^{-1}\,(\theta_j-\theta_{k_2})^{-1}\,\beta_{k_1}\,\beta_{k_2}\,t(j,k_1,j,k_2) + O\big(n^{-1}\big) \\ &= \sum\sum_{k_1,k_2\,:\,k_1\neq j,k_2\neq j} (\theta_j-\theta_{k_1})^{-1}\,(\theta_j-\theta_{k_2})^{-1}\,\beta_{k_1}\,\beta_{k_2}\,E\big(\xi_j^2\big)\,E\big(\xi_{k_1}^2\big)\,I(k_1\neq k_2) \\ &\quad + O\big(n^{-1}\big) \\ &= \theta_j\sum_{k\,:\,k\neq j}\theta_k\,(\theta_j-\theta_k)^{-2}\,\beta_k^2 + O\big(n^{-1}\big)\,. \end{aligned} \tag{4.28}$$

Combining (4.27) and (4.28) we deduce that

$$t_{2j} = n^{-1}\,\theta_j\sum_{k\,:\,k\neq j}\theta_k\,(\theta_j-\theta_k)^{-2}\,\beta_k^2 + O\big(n^{-2}\big)\,. \tag{4.29}$$

Results (4.26) and (4.29) imply that

$$\begin{aligned} t_1(r) &\equiv \sum_{j=1}^{r}\Big(\int b^0\,\psi_j\Big)\bigg\{\int_{\mathcal{I}} b^0\,E(\widehat{\psi}_j-\psi_j)\bigg\} \\ &= -n^{-1}\,\tfrac{1}{2}\sum_{j=1}^{r}\theta_j\,\beta_j^2\sum_{k\,:\,k\neq j}(\theta_j-\theta_k)^{-2}\,\theta_k + O\big(n^{-2}\big)\,, \end{aligned}$$

$$\begin{aligned} t_2(r) &\equiv \sum_{j=1}^{r} E\Big\{ \int_{\mathcal{I}} b^0 \, (\widehat{\psi}_j - \psi_j) \Big\}^2 \\ &= n^{-1} \sum_{j=1}^{r} \theta_j \sum_{k\,:\,k \neq j} \theta_k \, (\theta_j - \theta_k)^{-2} \, \beta_k^2 + O\big(n^{-2}\big) \,. \end{aligned}$$

Hence,

$$t(r) \equiv -2\, t_1(r) - t_2(r) = n^{-1}\, t_3(r) + O\big(n^{-2}\big)\,, \tag{4.30}$$

where

$$t_3(r) = \sum_{j=1}^{r} \Big\{ \theta_j \, \beta_j^2 \sum_{k\,:\,k \neq j} (\theta_j - \theta_k)^{-2} \, \theta_k - \theta_j \sum_{k\,:\,k \neq j} \theta_k \, (\theta_j - \theta_k)^{-2} \, \beta_k^2 \Big\} \,. \tag{4.31}$$

We know from (4.13) that the term in n^{-1} on the right-hand side of (4.18) should equal $t(r)$, plus terms of order n^{-2}. By (4.30), the coefficient of n^{-1} must equal $t_3(r)$. Therefore, the form of the coefficient of n^{-1} on the right-hand side of (4.18) follows from (4.31). ■

If along the sequence $\theta_1, \theta_2, \ldots$, there are from time to time very closely spaced eigenvalues, then the term in n^{-1} on the right-hand side of (4.18) can make a non-negligible contribution, and the approximation at (4.15) can fail. However, in other cases (4.15) is valid; see the Appendix. Less generally, if the β_j's decrease to zero very rapidly, and in particular if only a finite number of them are nonzero, then difficulties with spacings will be minor.

Below we derive an asymptotic limit result for the deviation between estimated and true slope function as dimension increases asymptotically; that is, as the number of components in the model increases with sample size.

Theorem 4.3. If $b^0 \in L_2(\mathcal{I})$ and $\int_{\mathcal{I}} E(X^4) < \infty$, where the random function X has the same distribution as the data X_i; and if $r = r(n) \to \infty$ as $n \to \infty$, in such a manner that

$$\frac{1}{n}\sum_{j=1}^{r}\delta_j^{-2}\to 0\,; \tag{4.32}$$

then

$$\int_{\mathcal{I}} E\{(\hat{b}-b^0)^2 \mid \mathcal{X}\}\to 0 \tag{4.33}$$

in probability as $n\to\infty$. In particular, if (4.32) holds then $\int_{\mathcal{I}}(\hat{b}-b^0)^2\to 0$ in probability.

Proof: For the last two terms in (4.14) we have:

$$\sum_{j=1}^{r}\left\{\int b^0\,(\widehat{\psi}_j-\psi_j)\right\}^2\le\left(\int (b^0)^2\right)\sum_{j=1}^{r}\|\widehat{\psi}_j-\psi_j\|^2$$

and

$$\begin{aligned}\left\{\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)\left(\int b^0(\widehat{\psi}_j-\psi_j)\right)\right\}^2&\le\left\{\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)^2\right\}\left\{\sum_{j=1}^{r}\left[\int b^0(\widehat{\psi}_j-\psi_j)\right]^2\right\}\\&\le\left\{\sum_{j=1}^{r}\left(\int b^0\,\psi_j\right)^2\right\}\|b^0\|^2\sum_{j=1}^{r}\|\widehat{\psi}_j-\psi_j\|^2\,.\end{aligned}$$

If $r=r(n)\to\infty$ as $n\to\infty$, these results, assumption $b^0\in L_2(\mathcal{I})$ and (4.14) imply that (4.33) is correct, provided

$$\frac{1}{n}\sum_{j=1}^{r}\hat{\theta}_j^{-1}\to 0\quad\text{and}\quad\sum_{j=1}^{r}\|\widehat{\psi}_j-\psi_j\|^2\to 0 \tag{4.34}$$

in probability.

In the following we prove that a sufficient condition for (4.34) is (4.32).

If (4.32) holds then $n^{-1/2}\,\delta_r^{-1}\to 0$, and so $n^{1/2}\min(\theta_j-\theta_{j+1})\to\infty$. Hence, for each $C>0$ and all sufficiently large n, $\theta_j-\theta_{j+1}>C\,n^{-1/2}$ uniformly in $1\le j\le r$. Since $\widehat{\Delta}=O_p(n^{-1/2})$, the probability that $\theta_j-\theta_{j+1}>2\,\widehat{\Delta}$ for all $1\le j\le r$, converges to 1 as $n\to\infty$. Equivalently, $1\le r\le J$, by definition of J.

Result (2.17) of Theorem 2.4 now implies that

$$i)\ \max_{j\leq r}|\theta_j^{-1}\,\hat{\theta}_j - 1| = \theta_r^{-1}\,\widehat{\Delta} = O_p(n^{-1/2}\,\delta_r^{-1}) = o_p(1)\,,$$

and

$$ii)\ \sum_{j=1}^{r}\|\widehat{\psi}_j - \psi_j\|^2 \leq \sum_{j=1}^{r} C\,\delta_j^{-2}\,\widehat{\Delta}^2 = O_p(n^{-1}\sum_{j=1}^{r}\delta_j^{-2})\,.$$

Furthermore, (i) entails:

$$\sum_{j=1}^{r}\hat{\theta}_j^{-1} = \sum_{j=1}^{r}\theta_j^{-1}\,(\hat{\theta}_j^{-1}\,\theta_j) = \sum_{j=1}^{r}\theta_j^{-1}\{1 + o_p(1)\}$$

Therefore, $\sum_{j=1}^{r}\hat{\theta}_j^{-1} \sim_p \sum_{j=1}^{r}\theta_j^{-1} = O_p(\sum_{j=1}^{r}\delta_j^{-2})$, where we have used the fact that $\theta_j \geq \delta_j$. The first part of (4.34) follows from (4.32). Also, property (ii) and (4.32) imply the second part of (4.34). ■

Condition (4.32) is based on spacings of eigenvalues, and is quite different from constraints imposed by Cardot et al. (1999) in a related problem. When the spacings $\theta_j - \theta_{j+1}$ are decreasing, condition (H_3) of Cardot et al. (1999) assumes the form $(n\,\theta_r^2)^{-1}\,(\sum_{j\leq r}\delta_j^{-1})^2 \to 0$, which is more restrictive than (4.32) above.

4.4 Prediction

The prediction problem is that where, given a candidate value of x of X, we wish to estimate

$$\mu(x) \equiv E(Y|X = x) = \int b\,x\,.$$

Our estimator is $\widehat{\mu}(x) = \int \hat{b}\,x$. An extra ingredient here is the smoothness of x with respect to the orthonormal basis $\psi_1, \psi_2, \cdots$. A major difference in properties, however, is that here, root-n consistency is possible.

For the sake of simplicity we shall discuss this problem in the setting of Theorem 4.1, since then we can be relatively explicit about the rates at which the three Fourier coefficient sequences decrease. However, a more general account, paralleling that given in Theorem I.1 of the Appendix, is possible.

Let us adjoin to condition (4.16) the assumption that $x = \sum_{j=1}^{\infty} x_j \psi_j$, where the real numbers $x_j = \int_{\mathcal{I}} x \psi_j$ satisfy

$$x_j = j^{-c} P(j), \text{ with the function } P \text{ slowly varying.}$$

The relationship among a, b and c needs to be a little different from before, in order to get root-n consistency: we need $a > 1$, $b > a + 2$ and $c > \frac{1}{2}(a + \frac{1}{2})$. In other respects, (4.16) can be left essentially unchanged:

$$\begin{gathered}\theta_j = j^{-a} L(j),\ |b_j^0| = j^{-b} M(j) \text{ and } x_j = j^{-c} P(j), \text{ where } b > a+2 > 3,\ c > \tfrac{1}{2}(a+\tfrac{1}{2}) \text{ and } L, M, P \text{ are slowly-varying functions;} \\ \theta_j - \theta_{j+1} \geq \text{const.}\, j^{-a-1}\text{; the process } X \text{ has all moments finite; for each integer } r \geq 1,\ \theta_j^{-r} E(\textstyle\int A \psi_j)^{2r} \text{ is bounded uniformly in } j\text{; the} \\ \text{errors } \epsilon_i \text{ in (4.1) are independent and identically distributed with all moments finite, zero mean and variance } \sigma^2\text{; and the frequency} \\ \text{cut-off, } r, \text{ is in the range } n^{c_1} \leq r \leq n^{c_2}, \text{ where } \{2(b+c-1)\}^{-1} < c_1 < c_2 < (2a)^{-1}.\end{gathered} \tag{4.35}$$

Under these assumptions, r_0 can be chosen so that it is an order of magnitude larger than the value that minimises mean integrated squared error.

Theorem 4.4. If (4.35) holds, then $\widehat{\mu}(x) = \mu(x) + O_p(n^{-1/2})$ as $n \to \infty$.

A proof of this result is similar to that of Theorem 4.1, although at the same time it has features that are quite different (it is not a nonparametric problem). See Cai and Hall (2004).

Chapter 5

Numerical Properties

5.1 Introduction

The simulation studies reported here provide a numerical assessment of conclusions reached by theoretical work. We first introduce the models used in the simulation studies in Section 5.2. In regard to the results given in Chapter 3, the coverage accuracy of bootstrap confidence intervals for both eigenvalues and eigenfunctions of the covariance operator has been obtained numerically. Section 5.3 presents the numerical results for eigenfunctions, including both simultaneous and individual confidence intervals for θ_j. In Section 5.4 the results corresponding to $\psi_j(t)$, consisting of simultaneous confidence intervals in the argument t, as well as confidence intervals obtained by the L_2-norm, are presented. Section 5.5 contains numerical results for the case where the processes X are observed on a discrete grid of points. In Section 5.6 we explore cross-validation performance in functional regression numerically. Comparison of three cross-validation criteria is discussed in Section 5.7. Then, we present the real data analysis in Section 5.8. In Section 5.9, we compare coverage levels obtained from the single bootstrap method with those resulting from calibrating the bands by the double bootstrap for both the Gaussian and Non-Gaussian processes. Two-sided, equal-tailed coverages of confidence intervals for θ_j and ψ_j are discussed numerically in Sections 5.10 and 5.11, respectively. A comparison of the coverages obtained from Gaussian and Non-Gaussian precesses is summarised in Section 5.12. Finally, numerical results relating to smoothed PCA are presented in Setion 5.13.

5.2 Models Used in Simulation Study

In all the work reported here, each X_i was distributed as $X = \sum_{j\geq 1} \xi_j \psi_j$ and was defined on $\mathcal{I} = [0,1]$, with $\psi_j(t) = 2^{1/2} \cos(j\pi t)$ and the ξ_j's denoting independent variables with zero means and respective variances $\theta_j = j^{-2\ell}$, for $\ell = 1$, 2 or 3. The latter three cases will be referred to as models (i), (ii) and (iii), respectively. The distributions of the ξ_j's were either normal $\mathrm{N}(0, \theta_j)$ or centred exponential with the same variance. When treating the regression problem the errors ϵ_i were normal $\mathrm{N}(0,1)$ and we took $a = 0$ and $b(t) = \pi^2 \left(t^2 - \frac{1}{3}\right) = \sum_j (-1)^j \, 2 \, j^{-2} \, \psi_j(t)$. We know that $\{\psi_0(t) = 1, \, \psi_j(t) = 2^{1/2} \cos(j\pi t); \, j = 1, 2, \cdots\}$ is a complete and orthonormal basis in $L_2(\mathcal{I})$. Therefore, the subspace created by $\psi_0 \equiv 1$ is orthogonal to the subspace spanned by $\{\psi_j(t) = 2^{1/2} \cos(j\pi t); \, j = 1, 2 \cdots\}$. The latter subspace consists of all functions whose integral on the interval $[0,1]$ is zero.

For numerical calculation we truncated the infinite series, defining X and b, at $j = N = 20$. All coverages of confidence regions were computed by averaging over 1000 simulated datasets, i.e. the number of bootstrap replications was $B = 1000$. Also, in the case of the double bootstrap the number of replications drawn from each bootstrap dataset was $C = 500$. However, median values of integrated squared error, discussed in Section 5.6, were calculated from 5000 simulated samples.

To estimate the eigenvalues and eigenfunctions, we need to solve the integral equation $\widehat{K}\,\phi = \lambda\,\phi$ for λ and ϕ, where $\widehat{K}$ is the empirical covariance operator. Therefore, the eigen-solution $(\hat{\theta}, \widehat{\psi})$ satisfies $\widehat{K}\,\widehat{\psi} = \hat{\theta}\,\widehat{\psi}$, and for each $1 \le \ell \le N$,

$$\langle \widehat{K}\,\widehat{\psi}, \psi_\ell \rangle = \hat{\theta}\,\langle \widehat{\psi}, \psi_\ell \rangle\,. \tag{5.1}$$

Assume that the $X_i \in \mathrm{span}\{\psi_1, \cdots, \psi_N\}$. If $\xi_{ij} = \sqrt{\theta_j}\, Z_{ij}$, where the Z_{ij} are iid from the Standard Normal distribution, then

$$\widehat{K}\,\widehat{\psi}(u) = \int_{\mathcal{I}} \widehat{K}(u,v)\,\widehat{\psi}(v)\,dv = \sum_{j=1}^{N}\sum_{k=1}^{N} \sqrt{\theta_j}\; H_{jk} \sqrt{\theta_k}\,\langle \widehat{\psi}, \psi_\ell \rangle\, \psi_j(u)\,, \tag{5.2}$$

where $H_{jk} = \frac{1}{n}\sum_{i=1}^{n}(Z_{ij} - \bar{Z}_j)\,(Z_{ik} - \bar{Z}_k)$.

We know that there is an isometric isomorphism between $L_2(\mathcal{I})$ and l_2 (i.e. $f \to \langle f, e_k \rangle = \varsigma_k$, where the sequence $\{\varsigma_k\}$ satisfies $\sum_{k \ge 1} |\varsigma_k|^2 < \infty$; see Theorem 3.2.15 of Ash, 1972). Combining (5.1) and (5.2) results in, for each $1 \le \ell \le N$,

$$\sum_{j=1}^{N}\sum_{k=1}^{N} \sqrt{\theta_j}\, H_{jk} \sqrt{\theta_k}\, \hat{v}_\ell\, \delta_{jk} = \sum_{k=1}^{N} \sqrt{\theta_\ell}\, \sqrt{\theta_k}\, H_{\ell k}\, \hat{v}_k = \hat{\theta}\, \hat{v}_\ell\,,$$

where $\hat{v}_\ell = \langle \widehat{\psi}, \psi_\ell \rangle$. Equivalently,

$$\mathbf{B}_N\, \hat{\mathbf{v}}_N = \hat{\theta}\, \hat{\mathbf{v}}_N\,, \tag{5.3}$$

where $\mathbf{B}_N = (B_{jk})$ is a N by N matrix with $B_{jk} = \sqrt{\theta_j}\; H_{jk} \sqrt{\theta_k}$ and $\hat{\mathbf{v}}_N = (\hat{v}_1, \cdots, \hat{v}_N)^T$. Therefore, $\widehat{\psi} = \sum_{j=1}^{N} v_j\, \psi_j$. Because $\mathbf{H}_N = (H_{jk})$ is an $N \times N$ positive-definite matrix, $\mathbf{B}_N = \mathbf{D}_N\, \mathbf{H}_N\, \mathbf{D}_N$ is also a positive-definite matrix, where $\mathbf{D}_N = \mathrm{diag}(\sqrt{\theta_1}, \sqrt{\theta_2}, \cdots, \sqrt{\theta_N})$. Suppose that $Y_{ij} = \sqrt{\theta_j}\,(Z_{ij} - \bar{Z}_j)$. Then,

$B_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}\, Y_{ik}$, and if $\frac{1}{n}\sum_{i=1}^{n}(Z_{ij} - \bar{Z}_j)^2 \le C$ a.s.,

$$\langle \hat{\mathbf{v}}, \mathbf{B}\, \hat{\mathbf{v}} \rangle = \frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_i, \mathbf{Y}_i \rangle^2 \le \sum_{j=1}^{\infty} \theta_j \, \frac{1}{n}\sum_{i=1}^{n}(Z_{ij} - \bar{Z}_j)^2 \le C \sum_{j=1}^{\infty} \theta_j \,, \quad \text{a.s.}$$

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \cdots)^T$, $\hat{\mathbf{v}}_i = (\hat{v}_{i1}, \hat{v}_{i2}, \cdots)^T$ and $\mathbf{B}$ is a matrix with elements B_{jk}. Suppose that $\mathbf{v}_{1;i} = (\hat{v}_{i1}, \cdots, \hat{v}_{iN})^T$, $\mathbf{Y}_{1;i} = (Y_{i1}, \cdots, Y_{iN})^T$ and $\mathbf{v}_{2;i} = (\hat{v}_{i;N+1}, \cdots)^T$, $\mathbf{Y}_{2;i} = (Y_{i;N+1}, \cdots)^T$ corresponding to areas (1) and (2) in the matrix

$$\mathbf{B} = \begin{bmatrix} (1) & \vdots & \times \\ \cdots\cdots & \vdots & \cdots\cdots \\ \times & \vdots & (2) \end{bmatrix},$$

respectively. The partition (1) of the matrix is B_N. If we ignore the two areas marked with crosses, then

$$\langle \hat{\mathbf{v}}_i, \mathbf{Y}_i \rangle \approx \langle \hat{\mathbf{v}}_{1;i}, \mathbf{Y}_{1;i} \rangle + \langle \hat{\mathbf{v}}_{2;i}, \mathbf{Y}_{2;i} \rangle \,,$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_i, \mathbf{Y}_i \rangle^2 \approx \frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_{1;i}, \mathbf{Y}_{1;i} \rangle^2 + \frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_{2;i}, \mathbf{Y}_{2;i} \rangle^2 + 2\,\frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_{1;i}, \mathbf{Y}_{1;i} \rangle \, \langle \hat{\mathbf{v}}_{2;i}, \mathbf{Y}_{2;i} \rangle \,. \tag{5.4}$$

We have:

$$\frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_{2;i}, \mathbf{Y}_{2;i} \rangle^2 \le C \sum_{j=N+1}^{\infty} \theta_j \,,$$

$$\frac{1}{n}\sum_{i=1}^{n} \langle \hat{\mathbf{v}}_{1;i}, \mathbf{Y}_{1;i} \rangle \, \langle \hat{\mathbf{v}}_{2;i}, \mathbf{Y}_{2;i} \rangle \le \sqrt{C \sum_{j=1}^{\infty} \theta_j} \sqrt{C \sum_{j=N+1}^{\infty} \theta_j} \,.$$

Let $\mathcal{X}$ be a Banach space with the norm $\|.\|_{\mathcal{X}}$, and let $\|B\| := \|B\|_{L(\mathcal{X})}$ be the operator norm relative to $\|.\|_{\mathcal{X}}$, where $L(\mathcal{X})$ denotes the space of all bounded linear operators from $\mathcal{X}$ to itself. This means that there exists a smallest constant C_1 such that $\|Bx\|_{\mathcal{X}} \leq C_1 \|x\|_{\mathcal{X}}$ for each $x \in \mathcal{X}$. Therefore, by (5.4),

$$\|\Delta \mathrm{B}\| = \|\mathrm{B} - \mathrm{B}_N\| \leq 3\, C\, \text{const.} \sqrt{\sum_{j=N+1}^{\infty} \theta_j} \,.$$

Furthermore,

$$(\mathrm{B} - \Delta\mathrm{B})^{-1} = (I - \mathrm{B}^{-1}\,\Delta\mathrm{B})^{-1}\,\mathrm{B}^{-1} = \sum_{k=0}^{\infty} (\mathrm{B}^{-1}\,\Delta\mathrm{B})^k\,\mathrm{B}^{-1}\,, \tag{5.5}$$

and then,

$$\begin{aligned} \|\mathrm{B}^{-1} - (\mathrm{B} - \Delta\mathrm{B})^{-1}\| &= \|\left(I - (I - \mathrm{B}^{-1}\,\Delta\mathrm{B})^{-1}\right)\mathrm{B}^{-1}\| \\ &= \|\sum_{k=1}^{\infty} (\mathrm{B}^{-1}\,\Delta\mathrm{B})^k\,\mathrm{B}^{-1}\| \\ &\leq \frac{\|\mathrm{B}^{-1}\|^2\,\|\Delta\mathrm{B}\|}{1 - \|\mathrm{B}^{-1}\|\,\|\Delta\mathrm{B}\|}\,. \end{aligned}$$

Let

$$I = \{A : \mathcal{X} \longrightarrow \mathcal{X} \quad \text{such that } A \text{ is invertable}\}.$$

We have $I \subseteq L(\mathcal{X})$. Furthermore, I is open, and the "inverting" mapping $f : B \longrightarrow B^{-1}$ is continuous and differentiable at any point of I. In the following we prove that f is differentiable and its derivative at the point B, $Df(B) \in L(L(\mathcal{X}))$, maps H to $-B^{-1}\,H\,B^{-1}$ ($H = \Delta B$).

We have, by (5.5),

$$f(B-H) - f(B) + Df(B)H = (B-H)^{-1} - B^{-1} - B^{-1}\,H\,B^{-1} = \sum_{i=2}^{\infty} (B^{-1}\,H)^i\,B^{-1}\,.$$

Therefore, for $\|H\| \leq \frac{1-\epsilon}{\|B^{-1}\|}$ (a vicinity of 0 in $L(\mathcal{X})$),

$$\|f(B-H) - f(B) + Df(B)H\| \leq \sum_{i=2}^{\infty} \|B^{-1}\, H\|^i \, \|B^{-1}\| \leq \frac{\|H\|^2 \, \|B^{-1}\|^3}{1 - \|B^{-1}\| \, \|H\|}$$
$$\leq C(\epsilon) \, \|H\|^2 \, .$$

This means that f has the Gatoux derivative at the point B, which is a linear operator on $L(\mathcal{X})$. Moreover, $Df(B)$ maps H to $-B^{-1}\, H\, B^{-1}$. In a computational sense, it shows that computing B^{-1} from B is stable.

5.3 Numerical Results Connected to Eigenvalues

As we have seen in Chapter 3, we can construct confidence statements about the size of $\hat{\theta}_j - \theta_j$. To do that, one can use percentile bootstrap confidence intervals for θ_j, or for a collection of θ_j's if we address several eigenvalues simultaneously. In this section, we first discuss numerical results for simultaneous confidence intervals for eigenvalues, and then treat those for confidence intervals for individual eigenvalues.

5.3.1 Simultaneous Confidence Interval for Eigenvalues

Figure 5.1 presents coverages of the simultaneous, bootstrap, two-sided confidence intervals for θ_j, introduced in Chapter 3, for two different simulated $X(t)$. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. Furthermore, the sample sizes were $n = 20, 50, 100, 200, 500$ and 1000, and we simulated from models (i)–(iii) introduced in Section 5.2.

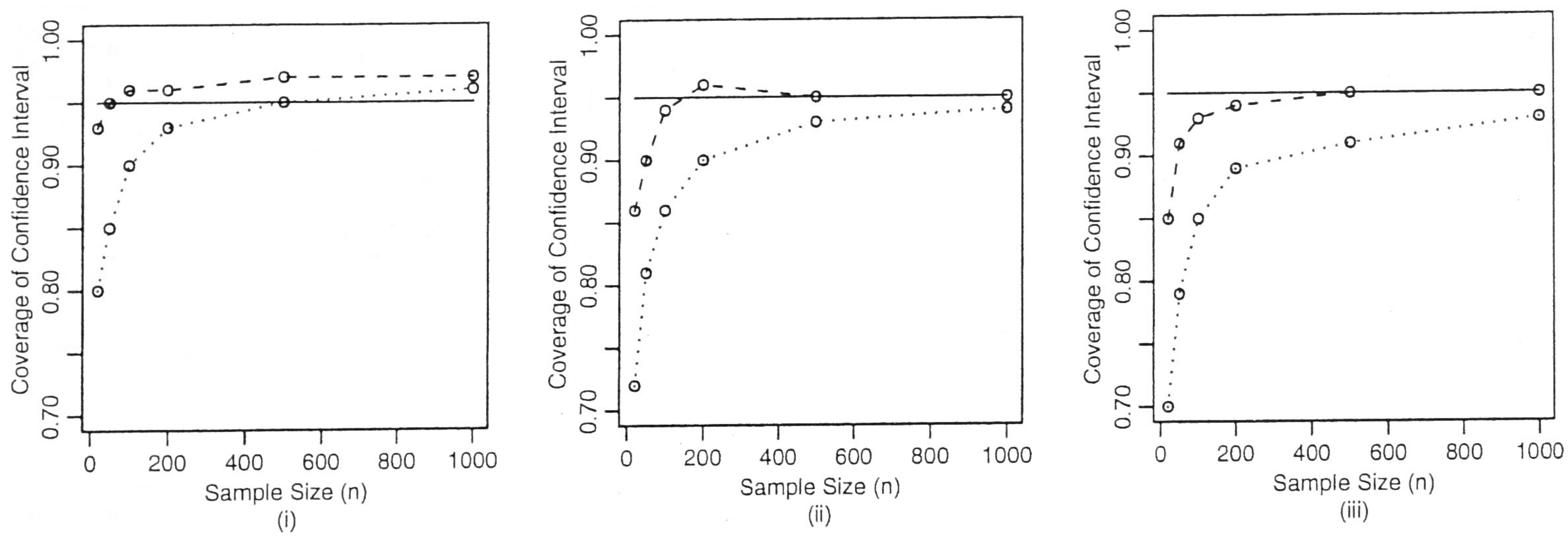

Figure 5.1: Comparison of coverages of two-sided, simultaneous, bootstrap confidence bands for θ_j, when $X(t)$ is generated from a Gaussian or Non-Gaussian process with the nominal coverage $1-\alpha = 0.95$. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The panels from left to right show the results obtained when generating under models (i)-(iii), respectively.

In the case of the Gaussian simulation, coverages tend to be slightly conservative (i.e. have coverage greater than the nominal level) for sample sizes greater than 50 in model (i), and tend to be anti-conservative for small sample sizes in the other two models. In regard to model (i), it can be seen that it has good performance for relatively small samples; especially when $n = 50$, it has actual coverage 0.95. By way of contrast, for the same sample size, coverage is only 0.90 and 0.91 in models (ii) and (iii). Furthermore, when moving from model (i) to (iii), coverage accuracy deteriorates, becoming anti-conservative. However, coverage accuracy under the two models (ii) and (iii) converges exactly to the nominal coverage accuracy, 0.95, for $n \geq 500$.

As regards the Non-Gaussian case, it seems that generally coverage accuracy tends to become worse compared to the Gaussian case. Moreover, here also coverage accuracy is anti-conservative for small sample sizes, and improves as n increases. Furthermore, similarly to the results obtained in the Gaussian case, for all sample sizes, model (i) enjoys better performance compared with the other

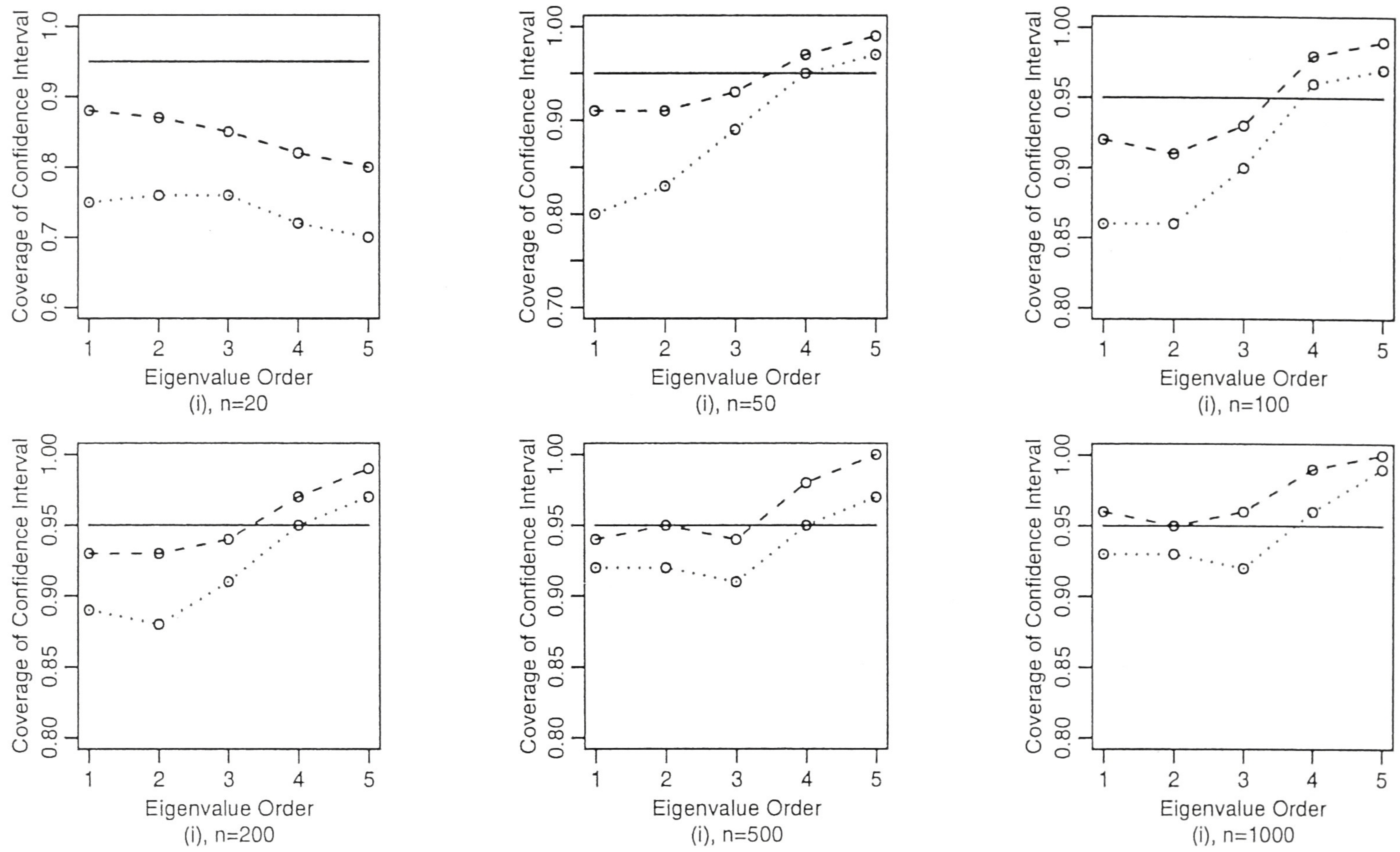

Figure 5.2: Comparison of coverages of two-sided single bootstrap confidence bands for θ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. We simulated from model (i) and the nominal coverage was $1 - \alpha = 0.95$.

two models. Specifically, it has actual coverage for $n = 500$, while coverage is 0.93 and 0.91 in models (ii) and (iii), respectively. Like the Gaussian case, coverage accuracy is reduced when moving from model (i) to (iii).

5.3.2 Confidence Intervals for Individual Eigenvalues

Figures 5.2–5.4 show coverage levels of two-sided, nominal 95 percent, single bootstrap interval for the first five θ_j, for models (i)–(iii) and for both the Gaussian and Non-Gaussian processes. The dashed and dotted lines present the coverage of the confidence intervals obtained by generating from the Gaussian and Non-Gaussian process, respectively.

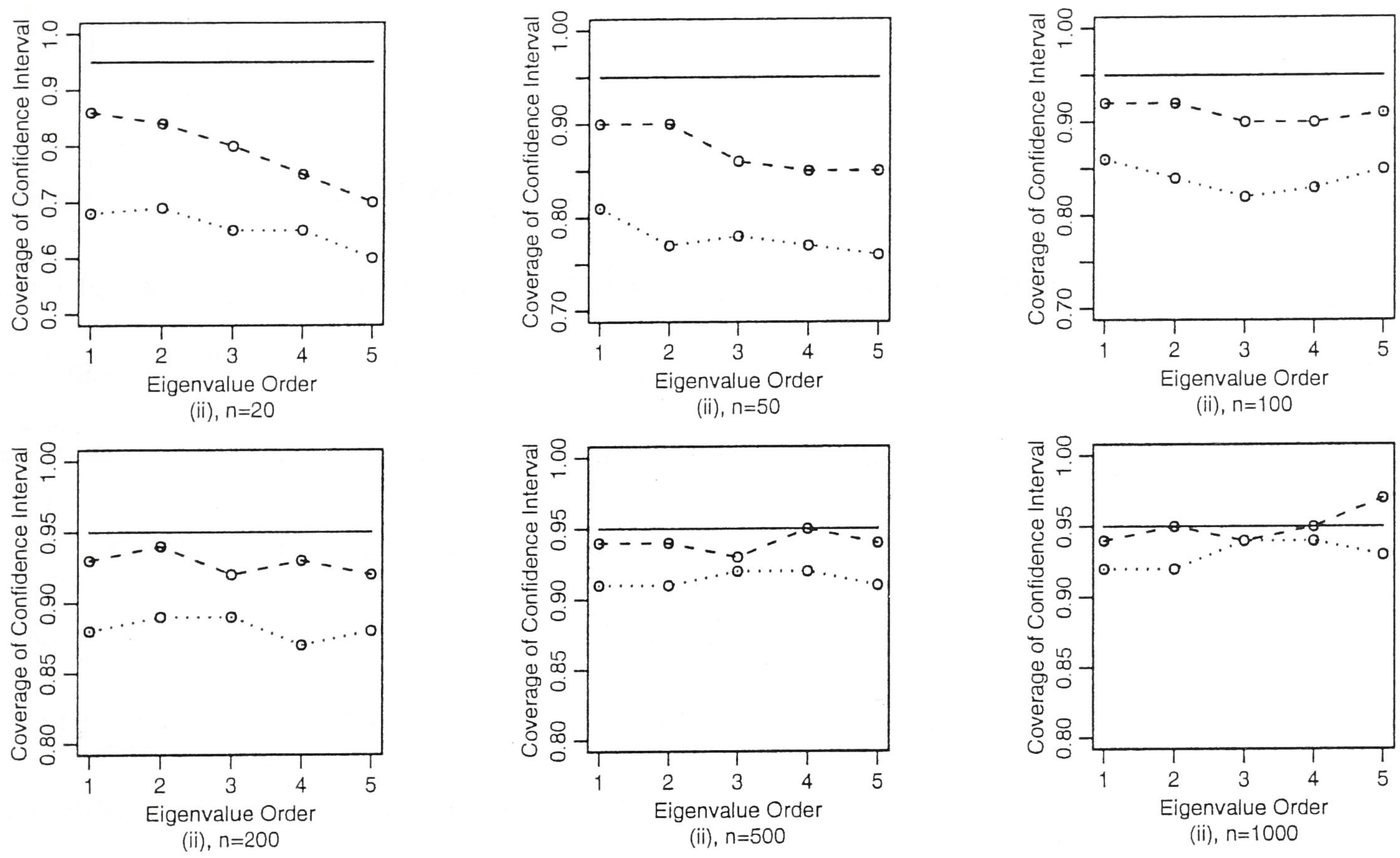

Figure 5.3: Comparison of coverages of two-sided single bootstrap confidence bands for θ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. We simulated from model (ii) and the nominal coverage was $1 - \alpha = 0.95$.

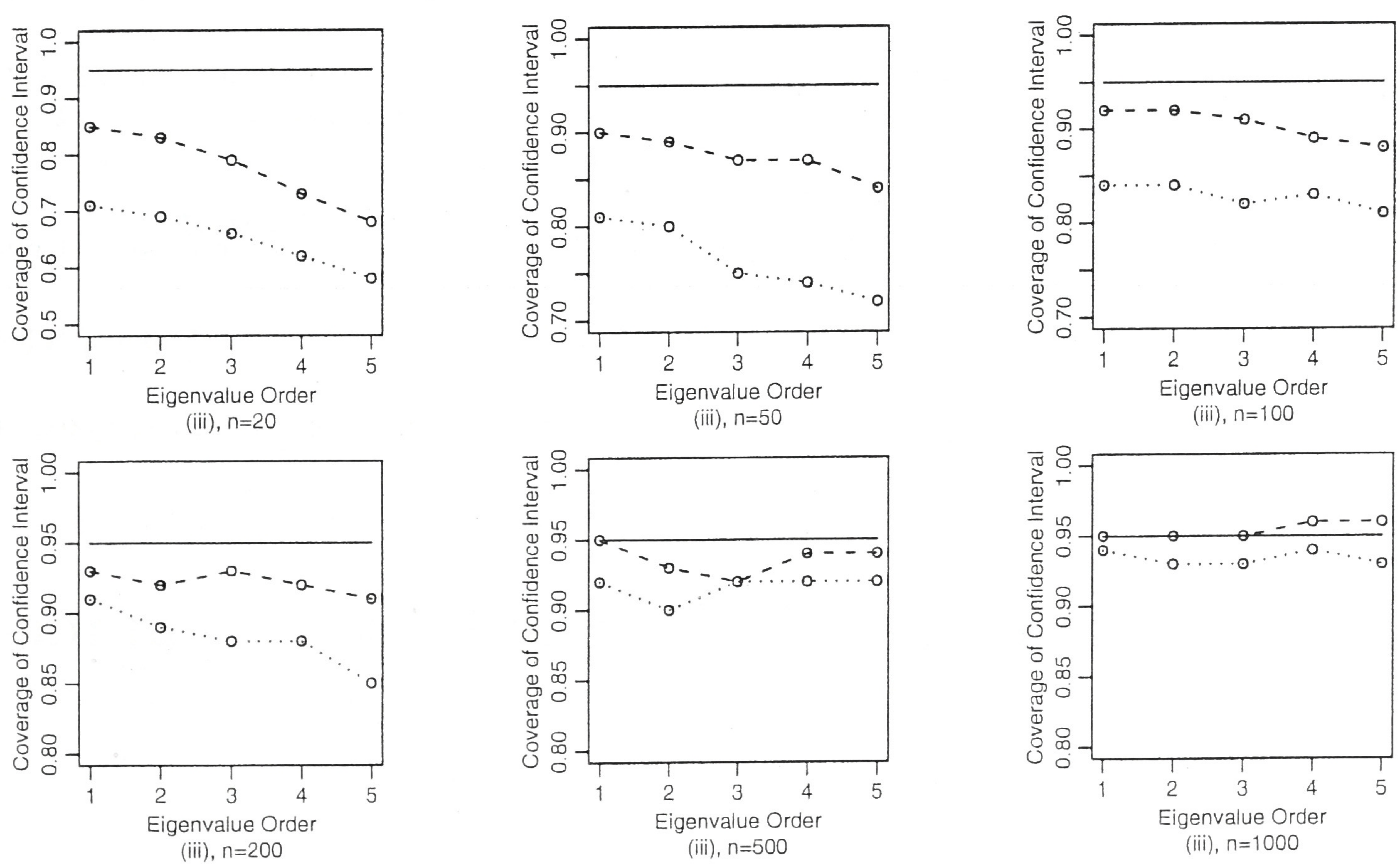

Figure 5.4: Comparison of coverages of two-sided single bootstrap confidence bands for θ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. We simulated from model (iii) and the nominal coverage was $1 - \alpha = 0.95$.

It can be seen that generally in all cases, and for each eigenvalue, coverage accuracy improves as n increases. Moreover, when applying model (i), we obtain slightly higher coverage levels compared with the other two models. Besides, under model (i) there is an increasing trend of coverage as order of eigenvalue increases, except for $n = 20$. The trend tends to fluctuate or decrease from model (i) to (iii). In the case of the Non-Gaussian model, however, as the graphs reveal, coverage accuracy deteriorates. Furthermore, as n increases, coverage accuracy improves and the gap between the coverage resulting from the Gaussian and Non-Gaussian case becomes smaller. For example, coverage levels for θ_1 under the Gaussian process is 0.88, 0.91, 0.92, 0.93, 0.94 and 0.96, respectively for sample sizes $n = 20, 50, 100, 200, 500$ and 1000. However, when altering the distribution of ξ_j to centred exponential, coverages decline to 0.75, 0.82, 0.86, 0.87, 0.93 and 0.93, respectively, for those sample sizes. Moreover, coverage levels for the second eigenvalue is 0.87, 0.91, 0.91, 0.93 and 0.95 for sample sizes $n = 20, 50, 100, 200$ and 500 when simulating from the Gaussian process. However, changing the distribution of ξ_j from normal to centred exponential reduces coverages to 0.77, 0.85, 0.86, 0.90 and 0.92, respectively for those sample sizes.

5.4 Numerical Results Connected to Eigenfunctions

We discuss numerical results related to confidence statements about the sizes of $\sup_t |\widehat{\psi}_j(t) - \psi_j(t)|$ and $\|\widehat{\psi}_j - \psi_j\|$. Using the former we obtain simultaneous bootstrap confidence bands for ψ_j; and using the latter we get confidence intervals for the L_2 distance of $\widehat{\psi}_j$ from ψ_j.

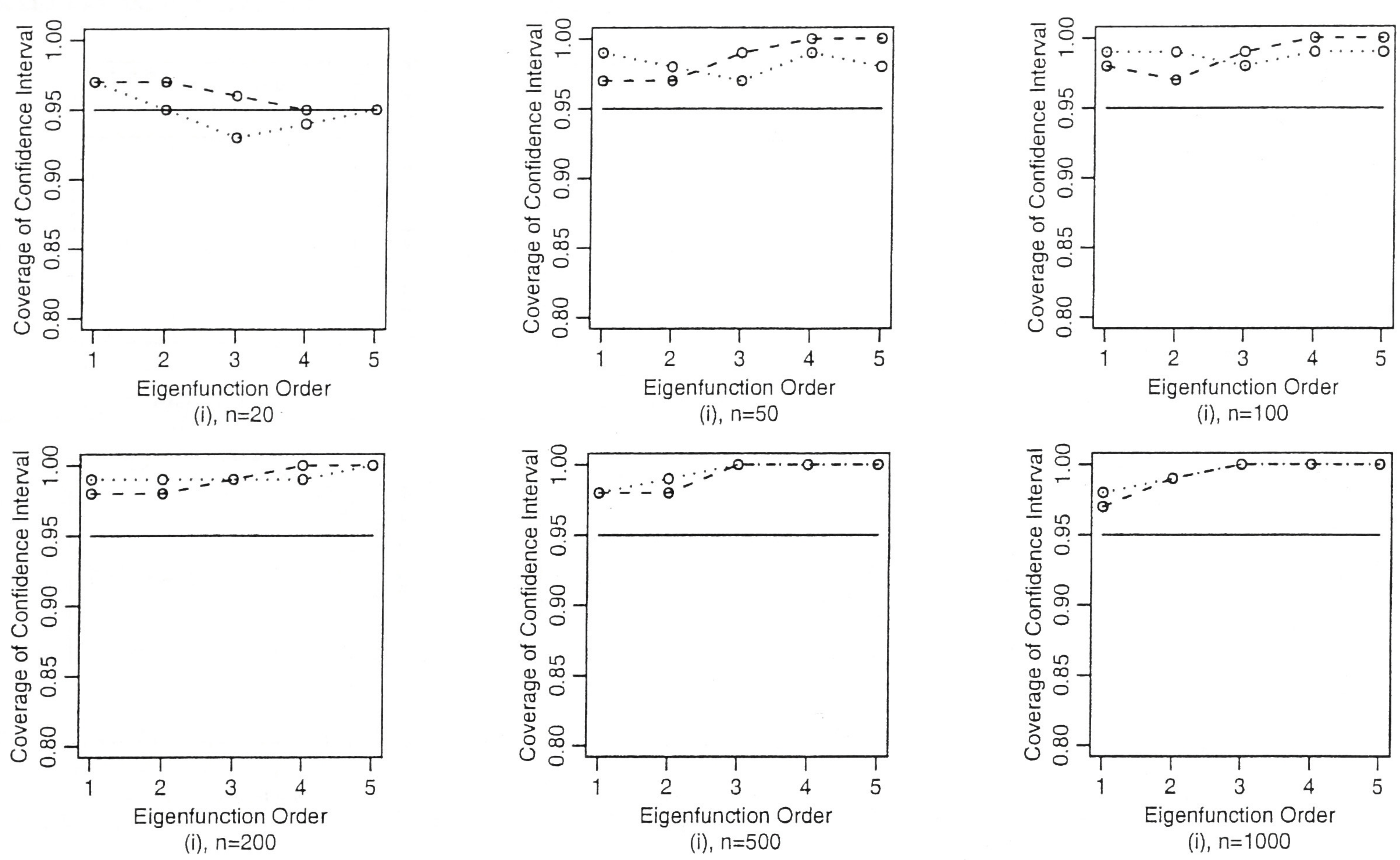

Figure 5.5: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were simultaneous in t, but not in j. We simulated from model (i) and the nominal coverage was $1 - \alpha = 0.95$.

5.4.1 Simultaneous Confidence Intervals for ψ_j Using Sup-norm

Figures 5.5–5.7 present coverage levels of two-sided, single-bootstrap bands for the first five eigenfunctions ψ_j, models (i)–(iii) and for both Gaussian and Non-Gaussian processes. Since each band is interpreted as covering ψ_j for all $t \in \mathcal{I} = [0, 1]$, the bands are called simultaneous. Here also we compare the results obtained from the two different simulated $X(t)$.

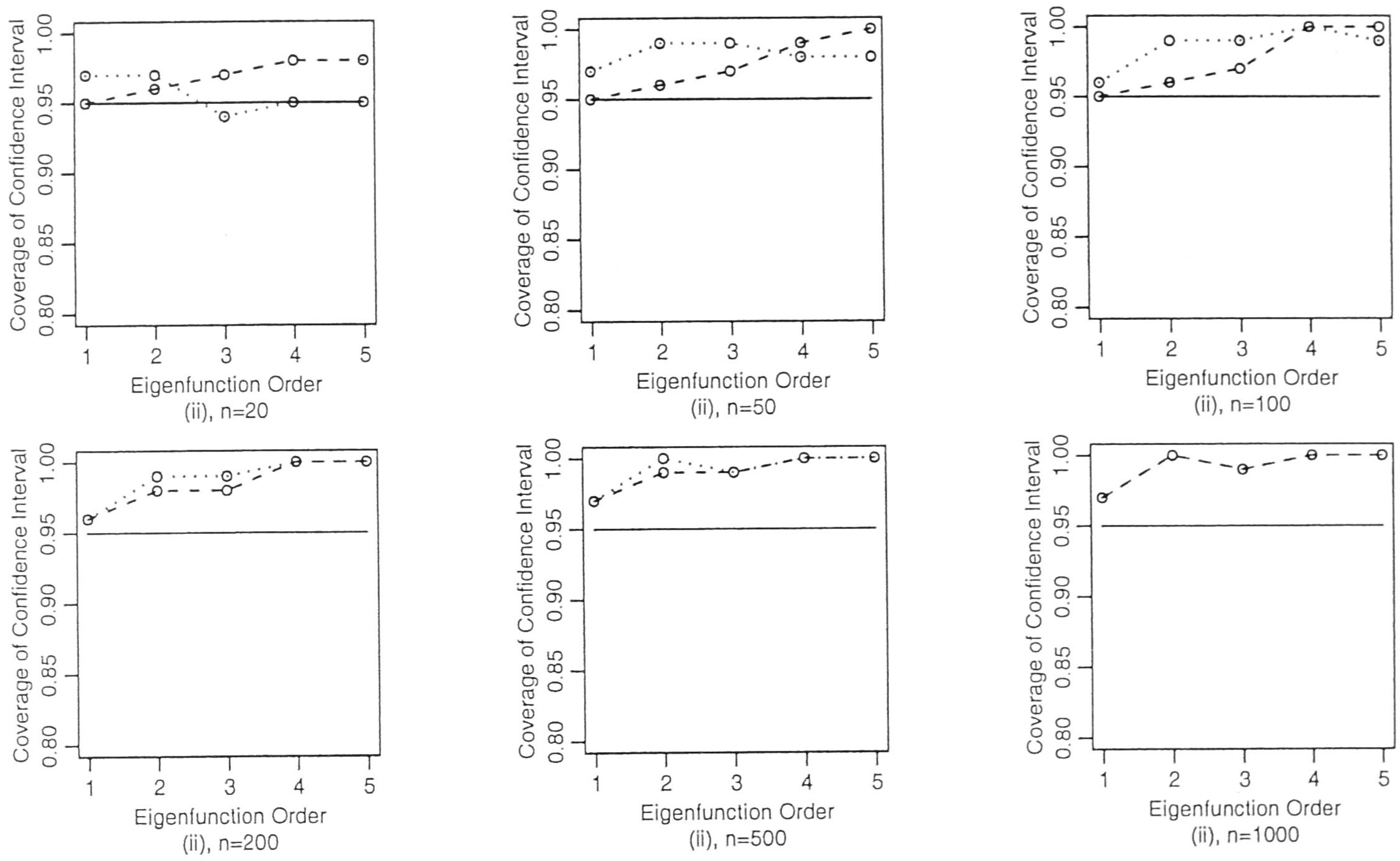

Figure 5.6: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were simultaneous in t, but not in j. We simulated from model (ii) and the nominal coverage was $1 - \alpha = 0.95$.

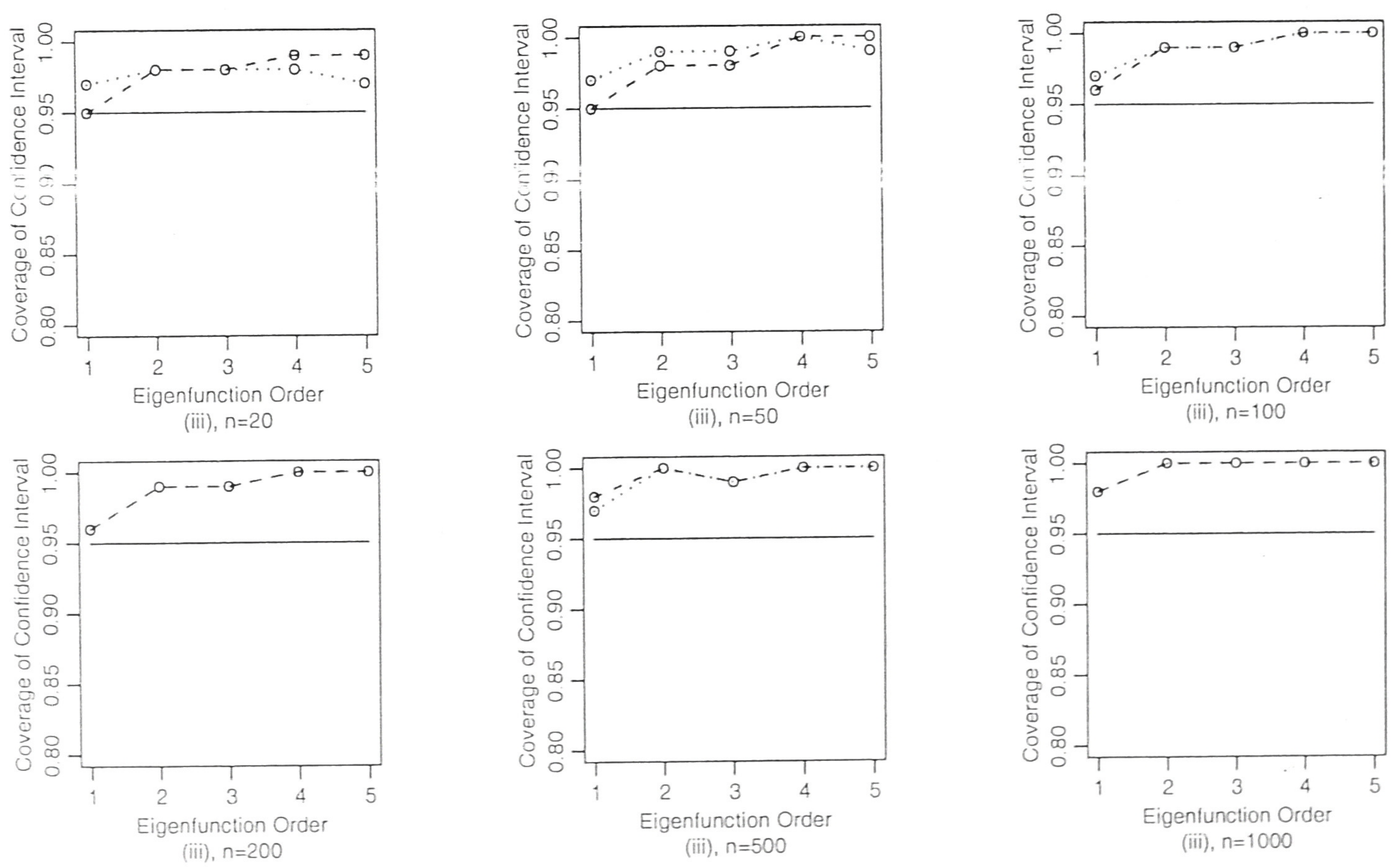

Figure 5.7: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were simultaneous in t, but not in j. We simulated from model (iii) and the nominal coverage was $1 - \alpha = 0.95$.

As the graphs reveal, in the Gaussian case, for each of the three models coverage accuracy almost always increases as the order of the eigenfunction increases, except for model (i) when $n = 20$. This is the case in the Non-Gaussian situation for $n \geq 100$, and it fluctuates for $n < 100$. While in the case of the Gaussian process, model (i) gives higher coverage accuracy compared with the other two models, followed by model (iii), for the Non-Gaussian process only coverage of the first eigenfunction performs similarly. For the eigenfunctions with higher order, however, coverage accuracy becomes more anti-conservative from model (i) to (iii); especially for θ_1. Moreover, when moving from model (i) to (iii), the gap between the coverage obtained in the Gaussian and Non-Gaussian cases becomes smaller as n increases, except for model (ii) when $n \leq 100$. Specifically, there is a perfect match between coverages resulting from the Gaussian and Non-Gaussian situations under model (iii) when $n = 200$.

5.4.2 Confidence Intervals for ψ_j Using L_2-norm

As graphs 5.8–5.10 show, the single bootstrap coverage accuracies obtained by using the L_2 norm, follow almost the same trends as those obtained by using the sup-norm. However, using the L_2-norm, in both the Gaussian and Non-Gaussian cases the coverage accuracy is dominated by its counterpart when using the sup-norm, due to domination of the L_2-norm by the sup-norm. Moreover, the gap between coverages obtained in the Gaussian and Non-Gaussian cases becomes smaller as n increases; especially, it vanishes for all j when $n = 500$ and applying model (ii).

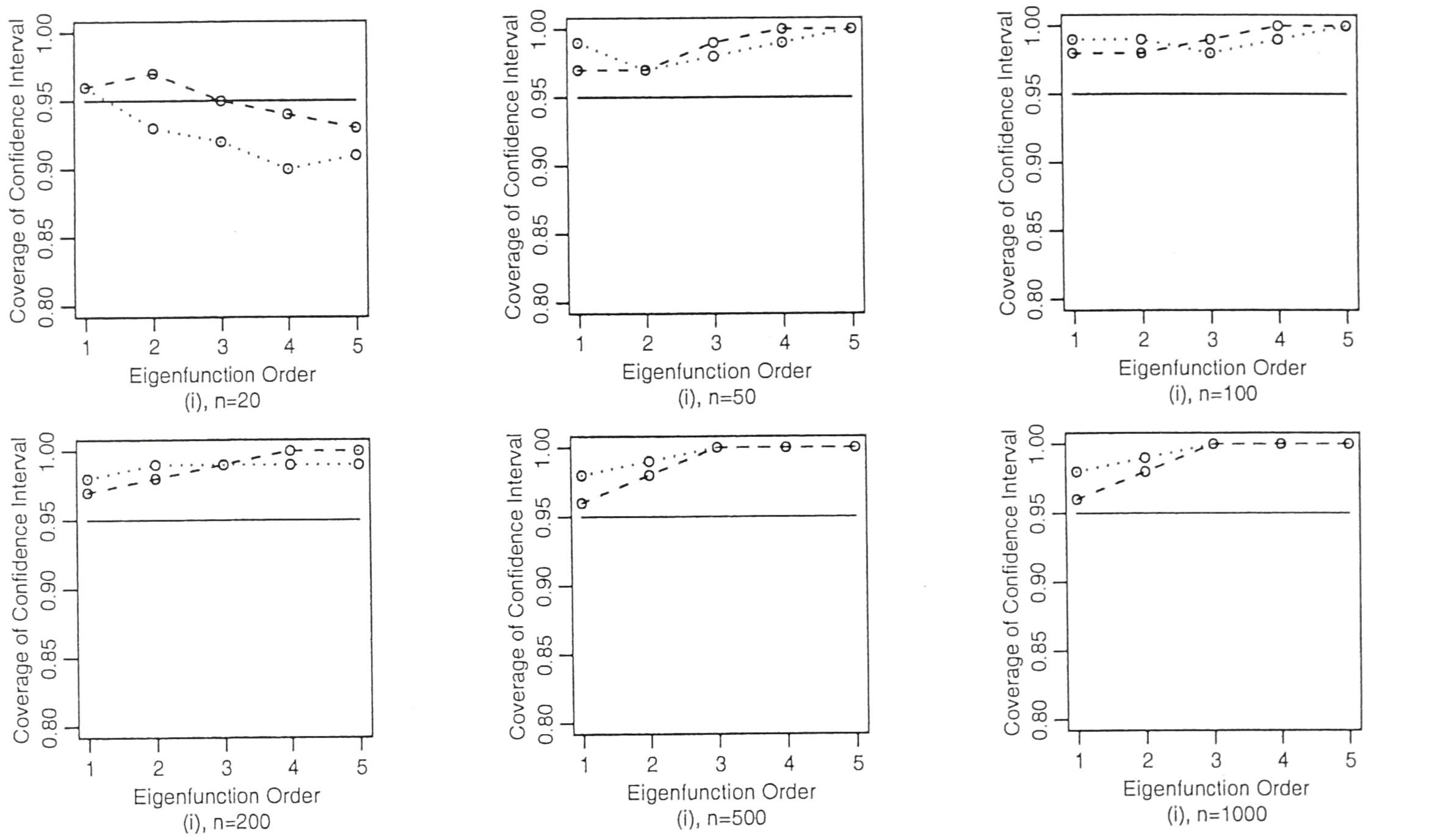

Figure 5.8: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were obtained by using the L_2 distance of $\hat{\psi}_j$ from ψ_j. We simulated from model (i) and the nominal coverage was $1 - \alpha = 0.95$.

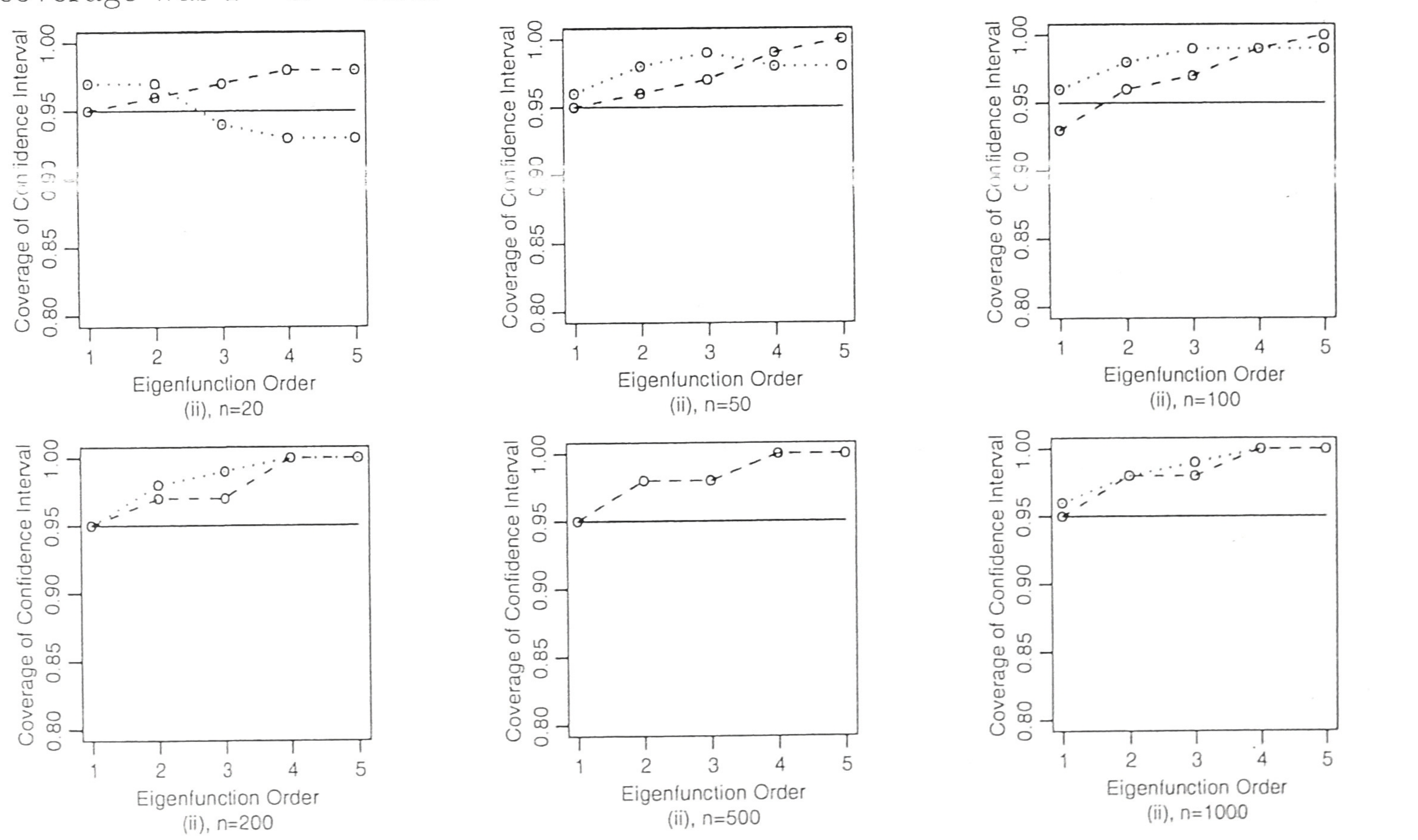

Figure 5.9: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were obtained by using the L_2 distance of $\hat{\psi}_j$ from ψ_j. We simulated from model (ii) and the nominal coverage was $1 - \alpha = 0.95$.

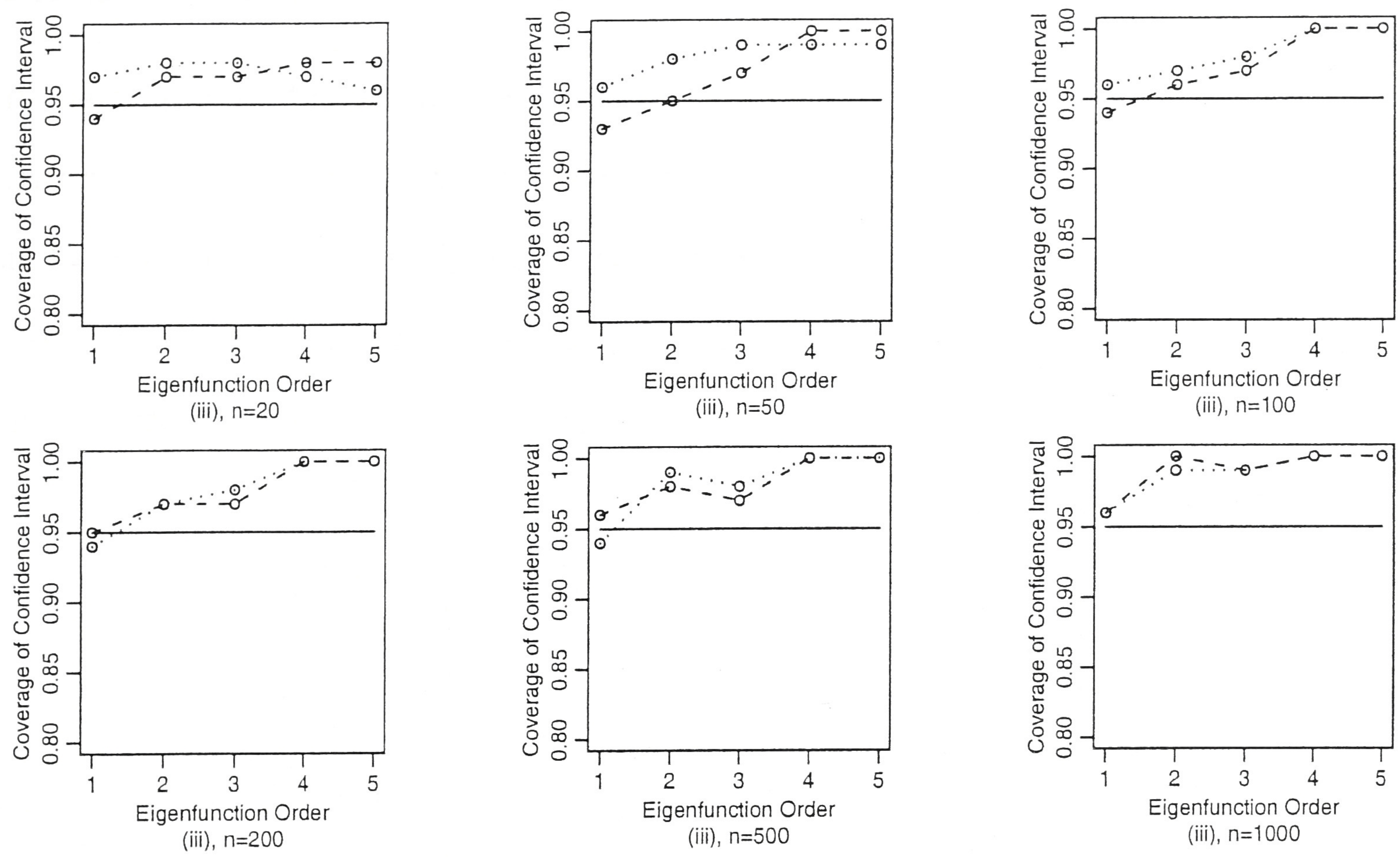

Figure 5.10: Comparison of coverages of two-sided bootstrap confidence bands for ψ_j, when generating $X(t)$ from a Gaussian or Non-Gaussian process. The dashed lines show the coverages when simulating from the Gaussian process, and the dotted lines reveal coverages when simulating from the Non-Gaussian process. The bands were obtained by using the L_2 distance of $\hat{\psi}_j$ from ψ_j. We simulated from model (iii) and the nominal coverage was $1 - \alpha = 0.95$.

5.5 Effect of Discrete Observation on Coverage Accuracy of Confidence Intervals for Eigenvalues and Eigenfunctions

We also explored the case where each X_i was observed discretely on a grid, with additive error, in particular $W_{ij} = X_i(j/20) + \delta_{ij}$, where $1 \leq j \leq 20$ and δ_{ij} was normal $N(0, \sigma^2)$. We passed a conventional local-linear smoother through these data, using the Sheather-Jones method to choose bandwidth, thereby constructing an estimator $\widehat{X}_i$ of X_i; and then we applied our methods as though $\widehat{X}_i$ were X_i. We took $\sigma^2 = 0.025$, 0.05 and 0.1, and calculated the usual empirical approximations to coverage of confidence regions for θ_j and ψ_j, when $n = 20, 50$ and 100. Interestingly, coverage accuracy of confidence intervals for θ_j is almost always improved by discretising and adding noise, the extent of improvement generally decreasing with increasing sample size and with decreasing error variance. The coverage accuracy of confidence bands for ψ_j is hardly affected. (This phenomenon also occurs in related contexts, and so is not unexpected. It arises from the effects of smoothing, including the smoothing associated with adding noise. See Cai and Hall, (2004).

n	σ^2	θ_1	θ_2	θ_3	θ_4	θ_5
20		0.88	0.87	0.85	0.82	0.80
20	0.025	0.88	0.88	0.86	0.83	0.80
20	0.05	0.88	0.88	0.87	0.84	0.83
20	0.1	0.88	0.88	0.88	0.85	0.84
50		0.91	0.91	0.93	0.97	0.99
50	0.025	0.92	0.91	0.94	0.96	0.99
50	0.05	0.91	0.92	0.93	0.96	0.99
50	0.1	0.92	0.92	0.94	0.96	0.98
100		0.92	0.92	0.93	0.97	0.99
100	0.025	0.92	0.93	0.94	0.96	0.99
100	0.05	0.92	0.94	0.95	0.96	0.99
100	0.1	0.93	0.93	0.94	0.96	0.98

Table 5.1: Coverages of bootstrap symmetric confidence bands for θ_j under the Gaussian process and model (i) when the process $X(t)$ is observed at the discrete grid of 20 points. In the table we denote variance of error with σ^2.

5.6 Cross-validation in Regression

We simulated data from the regression models discussed in Section 5.2. For each sample we calculated the value of $\|\hat{b} - b\|^2$, representing integrated squared error (ISE), and analysed, by simulation, the distribution of this quantity. In particular, we calculated median, med(r), say, of the distribution, and found the value, r_0, of r that minimised the median (see Figure 5.11).

n	σ^2	θ_1	θ_2	θ_3	θ_4	θ_5
20		0.86	0.84	0.80	0.75	0.70
20	0.025	0.86	0.85	0.82	0.82	0.80
20	0.05	0.87	0.87	0.86	0.84	0.80
20	0.1	0.87	0.89	0.85	0.83	0.76
50		0.90	0.90	0.86	0.85	0.85
50	0.025	0.90	0.91	0.90	0.91	0.93
50	0.05	0.90	0.91	0.90	0.90	0.85
50	0.1	0.90	0.92	0.92	0.93	0.91
100		0.92	0.92	0.90	0.90	0.91
100	0.025	0.92	0.92	0.93	0.95	0.92
100	0.05	0.92	0.92	0.92	0.94	0.91
100	0.1	0.93	0.93	0.94	0.95	0.92

Table 5.2: Coverages of bootstrap symmetric confidence bands for θ_j under the Gaussian process and model (ii) when the process $X(t)$ is observed at the discrete grid of 20 points. In the table we denote variance of error with σ^2.

n	σ^2	θ_1	θ_2	θ_3	θ_4	θ_5
20		0.86	0.83	0.79	0.73	0.68
20	0.025	0.87	0.83	0.83	0.78	0.73
20	0.05	0.86	0.85	0.87	0.80	0.70
20	0.1	0.86	0.89	0.89	0.82	0.73
50		0.90	0.89	0.87	0.87	0.84
50	0.025	0.90	0.90	0.87	0.88	0.85
50	0.05	0.89	0.92	0.89	0.89	0.84
50	0.1	0.91	0.93	0.90	0.91	0.86
100		0.92	0.92	0.91	0.89	0.88
100	0.025	0.92	0.92	0.92	0.90	0.90
100	0.05	0.92	0.91	0.92	0.90	0.88
100	0.1	0.93	0.91	0.91	0.90	0.89

Table 5.3: Coverages of bootstrap symmetric confidence bands for θ_j under the Gaussian process and model (iii) when the process $X(t)$ is observed at the discrete grid of 20 points. In the table we denote variance of error with σ^2.

Using the values r_0, we obtained the first quartiles (25th percentile), median or second quartile and third quartile (75th percentile) of the distribution ISE. The solid lines in Figure 5.12 indicate the quartiles.

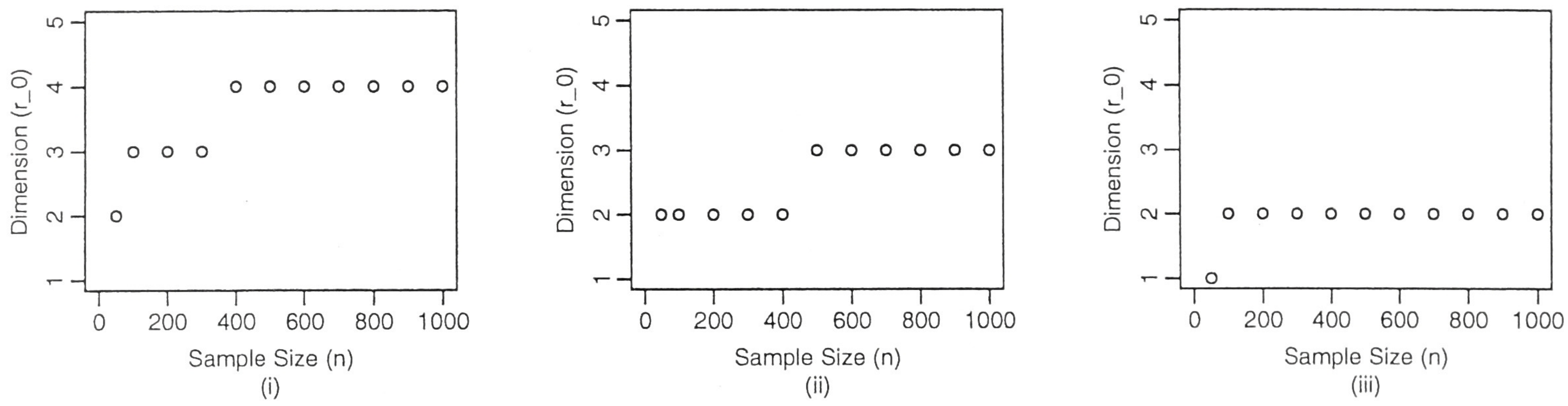

Figure 5.11: The values, r_0, of r that minimised the median of $\|\hat{b}-b\|^2$. We took three different models $\theta_j = j^{-2\ell}$, for $\ell = 1, 2$ and 3, and denote them by (i)–(iii).

We also computed the smallest value of r, $\hat{r}$ say, that produced a local minimum of $\mathrm{CV}(r)$, and calculated the first quartiles (25th percentile), Median or second quartile and third quartile (75th percentile) of the distribution of $\|\hat{b}-b\|^2$ when $\hat{b}$ is computed with $r = \hat{r}$. The dashed lines in Figure 5.12 represent the values of these quartiles.

In some instances there is more than one value of r for which the cross-validation criterion $\mathrm{CV}(r)$, at (4.9), is minimised. Our numerical experience suggests that a good way of selecting among these local minimisers is to choose the smallest. This is in line with recommended practice in more conventional applications of cross-validation for bandwidth choice, where the smoothing parameter which gives a local minimum and also confers most smoothing (i.e. the largest locally-minimising bandwidth, here equivalent to the smallest r) is selected. See, for example, Park and Marron (1990) and Hall and Marron (1991).

It can be seen from the figure that, for model (i), the median performance of the estimator of b, computed using cross-validation to choose r, lies only a little below that when r is selected optimally. This is true even for small sample sizes.

In the case of model (ii), cross-validation experiences difficulty with small samples; note that the dashed and solid lines are well separated there. Nevertheless, for large samples, the performance of cross-validation is not far from optimal.

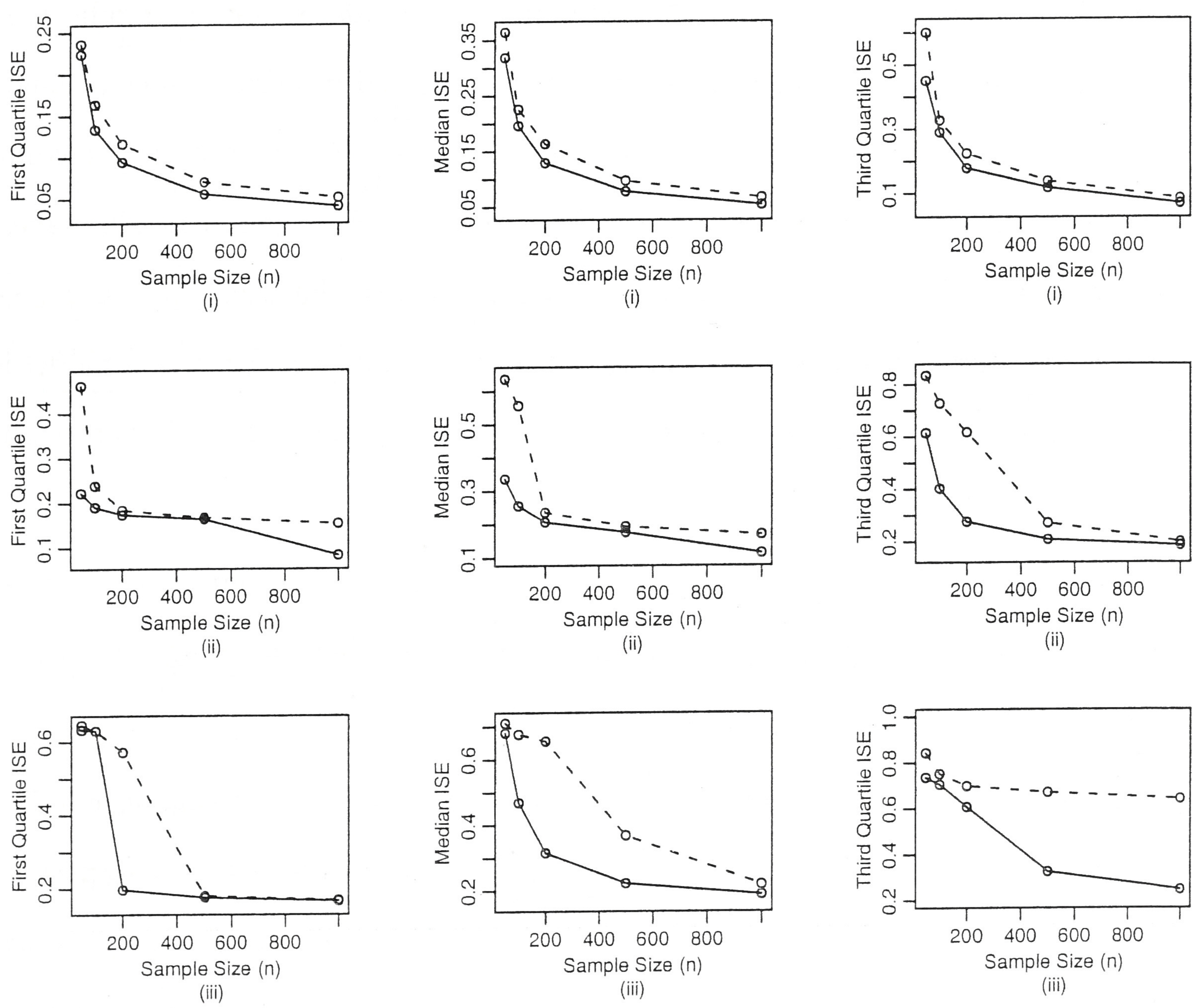

Figure 5.12: Performance of integrated squared error (ISE), when r was chosen optimally or by cross-validation, and $X(t)$ was generated from the Gaussian process. The solid lines graph the three quartiles (25, 50 and 75th percentile) ISE when $r = r_0$ was chosen to minimise the median, and the dashed lines graph those quartiles when r was selected by cross-validation. The value of sample size, n, was graphed on the horizontal axis, and the quartiles integrated squared error were shown on the vertical axis. The first, second and third row represent the results for models (i), (ii) and (iii), respectively.

However, the case of model (iii), which represents a relatively high-dimensional setting, causes greater difficulty, and only when n is approximately 1000 does cross-validation give good median performance there. Moreover, in the third quartile, the dashed and solid lines are distinctly separate under this model even for large samples. These properties are closely reflected in performance at quartiles of the distribution of the $D = \sqrt{||b - \hat{b}||^2 + |a - \hat{a}|^2}$. We obtained median performance of D in the same way as for the ISE (see Figure 5.13). The panels reflect almost similar features as in the case of the ISE. Furthermore, it can be seen that estimating the slope function $b(t)$ in the functional regression model is much more important than estimating the intercept a, due to difficulties brought from its high-dimensionality.

We also found that there is no deterioration in the result connected to cross-validation performance in functional linear regression when changing the distribution of ξ_j from normal to centred exponential.

Following the work on exploring the performance of the cross-validation (CV) criterion, in functional linear regression model (4.1), there are two sources of variability: variability brought to response variable by functional regression coefficient, and variability of noise (ϵ_i). The effect of controlling these two sources of variability with respect to each other in performance of CV was investigated.

In the simulations, we took X considered to be the Gaussian process with $\theta_j = j^{-2}$, and the ϵ_i were generated from $\mathrm{N}(0,1)$. The functional coefficient defined on $[0,1]$ was

$$b(t) = C_1(t^2 - t + \frac{1}{3}) = C_2 \sum_{j=1}^{\infty} \frac{1}{j^2} \psi_j(t) ,$$

where C_1 and C_2 are constant. Furthermore, in the model (4.1), $a = 0$ was chosen.

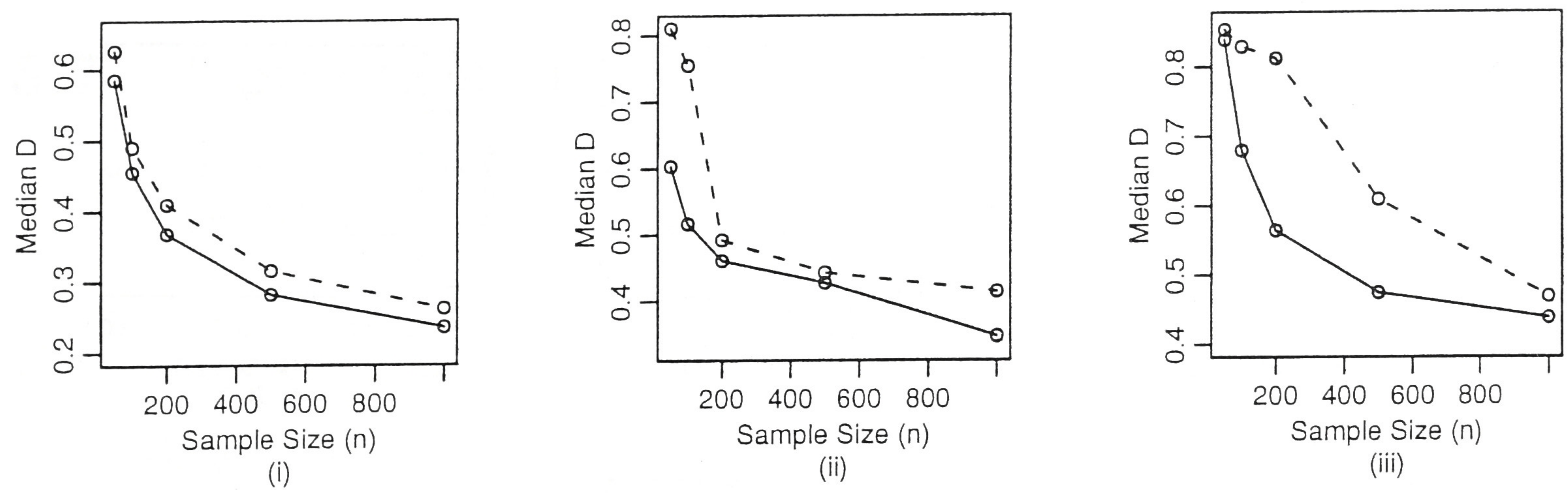

Figure 5.13: Median performance of D, when r was chosen optimally or by cross-validation, and $X(t)$ was generated from the Non-Gaussian process.The solid line graphs the median value of D when $r = r_0$ is chosen to minimise the median, and the dashed line graphs the median when r was selected by cross-validation. The value of sample size, n, was graphed on the horizontal axis, and median D was shown on the vertical axis. The first, second and third panels represent models (i), (ii) and (iii), respectively.

Choosing different values of C_1 affects the amount of variability due to the explanatory variables. We took five different values for the ratio of $R = \frac{\mathrm{var}(<b,X>)}{\mathrm{var}(\epsilon)}$ as follows: $(a)\,0.02$, $(b)\,0.5$, $(c)\,2$, $(d)\,10$, $(e)\,20$. We obtained the value of r, $\hat{r}$ say, that produced a global minimum of $\mathrm{CV}(r)$, and calculated the median of the distribution of $\|\hat{b} - b\|^2$ when $\hat{b}$ is computed with $r = \hat{r}$. The dashed lines in Figure 5.15, created from 500 generated synthetic samples from the model, represent the values of this median.

By simulation, the median of the distribution of $\|\hat{b} - b\|^2$, $\mathrm{med}(r)$ say, was analysed, and the value, r_0, of r that minimised the median was obtained (Figure 5.14). Furthermore, the solid lines in Figure 5.15, created from 5000 synthetic samples from the model, indicate $\mathrm{med}(r_0)$.

It can be seen from Figure 5.15 that, in relation to the performance of the estimation of b, the relative differences between the two lines are improved in all sample sizes, when R increases.

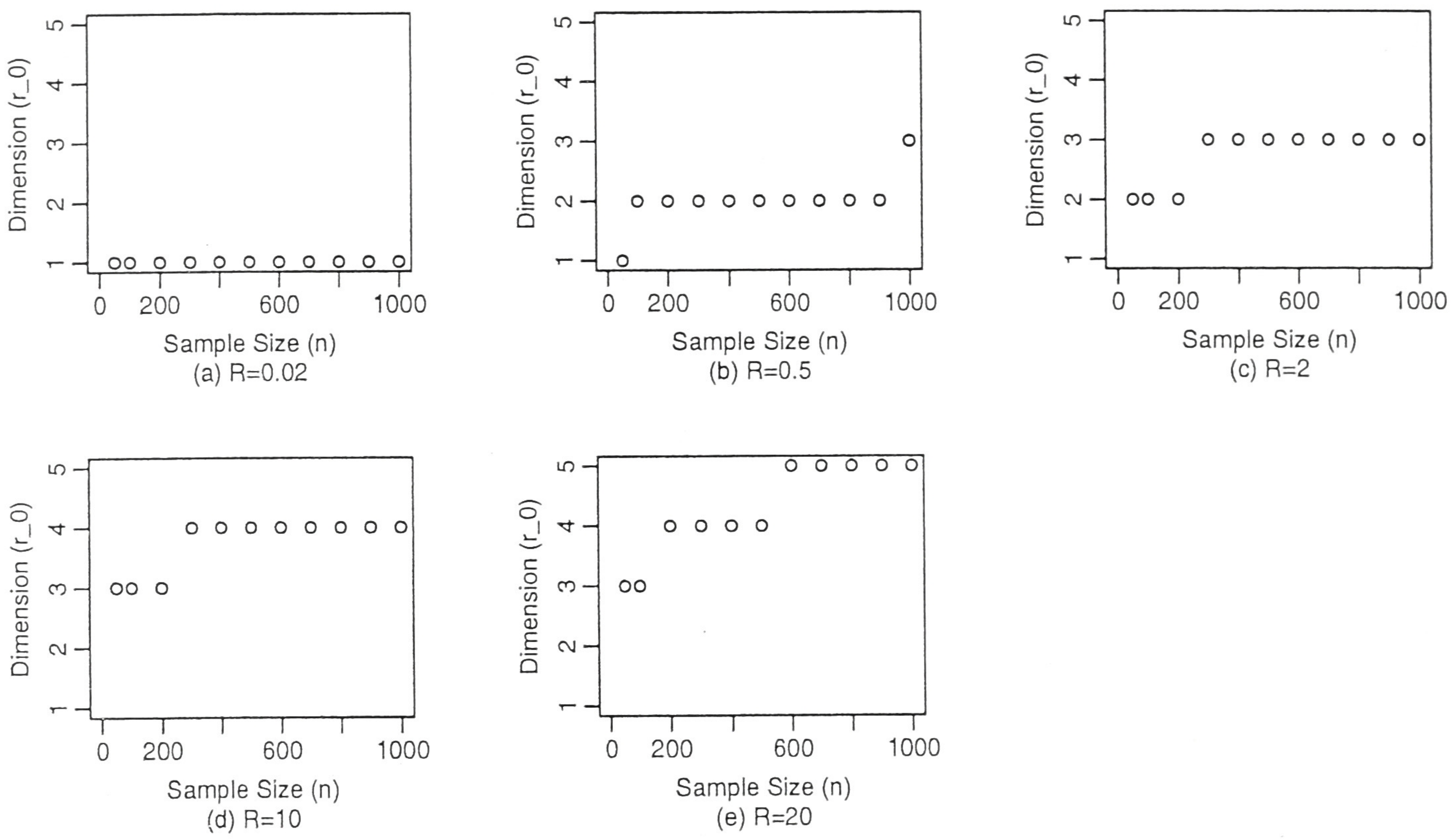

Figure 5.14: The values, r_0, of r that minimised the median of $\|\hat{b} - b\|^2$. We have obtained these values by simulation over 5000 datasets. On the graphs, cases (a)–(e) refer to $R = 0.02$, 0.5, 2, 10 and 20, respectively.

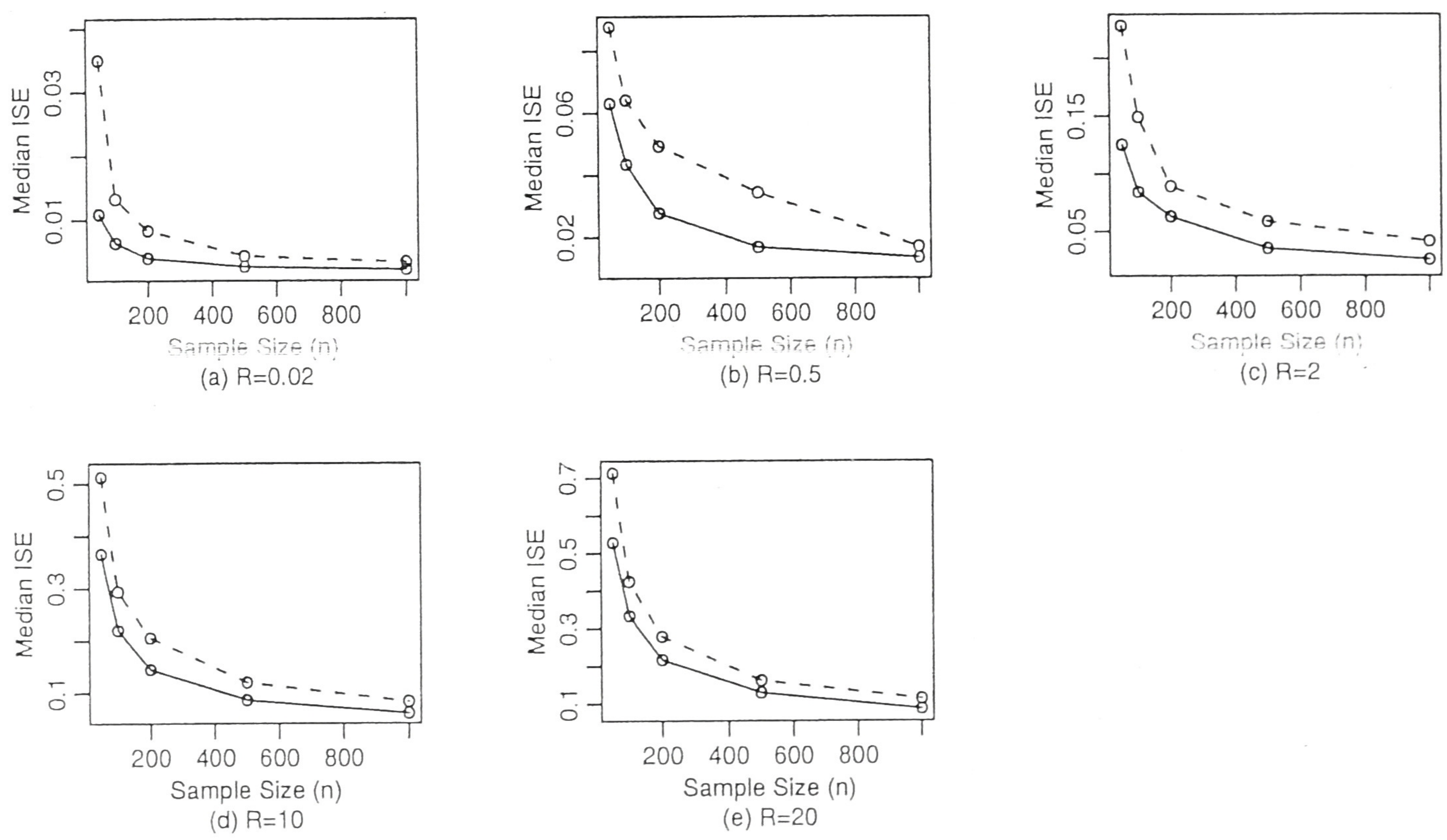

Figure 5.15: Median performance of integrated squared error (ISE), when r was chosen optimally or by cross-validation, and $X(t)$ was generated from the Gaussian process. The solid lines graph the median value of the ISE when $r = r_0$ was chosen to minimise the median, and the dashed line graphs the median of ISE when r was selected by cross-validation. The value of sample size, n, was graphed on the horizontal axis, and median of the ISE on the vertical axis. On the graphs, cases (a)–(e) refer to $R = 0.02$, 0.5, 2, 10 and 20, respectively.

5.7 Comparison of Three Cross-validation Criteria

The purpose of this simulation study is to compare the performances of three cross-validation criteria for choosing the smoothing parameter in functional regression slope estimation, in a functional linear regression model. Assume that the functional linear regression model

$$Y_i = a + \int_{\mathcal{I}} b\, X_i + \epsilon_i = a + \langle b, X_i \rangle + \epsilon_i\,, \quad 1 \le i \le n\,.$$

In the simulation here we chose X to be the Gaussian process introduced in Section 5.2. Furthermore, the errors ϵ_i were Normal $\mathrm{N}(0,1)$ and we took models (i)–(iii), $a = 0$ and $b(t)$ introduced in Section 5.2. All estimated quantities, shown by the graphs later, were computed over 1000 simulated datasets. The three cross-validation criteria were:

1.
$$\mathrm{CV}_1(r) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \hat{a}^{(-i;r)} - \int_{\mathcal{I}} \hat{b}^{(-i;r)}\, X_i\right\}^2. \tag{5.6}$$

   Here, $(\hat{a}^{(-i;r)}, \hat{b}^{(-i;r)})$ denotes the least-squares estimator of (a, b) that is obtained by confining attention to the set $\mathcal{Z}_i$, say, of all data pairs (X_j, Y_j) excluding the ith; and both $\hat{a}^{(-i;r)}$ and $\hat{b}^{(-i;r)}$ use the empirical Karhunen-Loève expansion of length r computed from $\mathcal{Z}_i$. We also show every quantity, obtained by all data pairs except excluding the ith, with index $(-i)$.

2. If we can write $\widehat{\mathbf{Y}}_{n\times 1} = \mathbf{H}\,\mathbf{Y}_{n\times 1}$, where $\widehat{\mathbf{Y}}$ denotes the predictor of $\mathbf{Y}$, then we are able to use
$$\mathrm{CV}_2(r) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \widehat{Y}_i)^2}{(1 - H_{ii})^2}\,, \tag{5.7}$$

where H_{ii} is the ith diagonal element of the $n \times n$ matrix $\mathbf{H}$. This cross-validation criterion is easy to compute since we do not need to calculate the integral in CV_1. It is clear that these two are not the same, since in computing CV_1, when excluding the ith observation, we have to obtain $\widehat{\psi}_j$ with only $n-1$ other observations, and then get the coefficients ξ_{ij} by using the current estimated eigenfunctions. This is not the case for CV_2. However, when computing CV_1, if we neglect that, these two are the same.

If we truncate the two series $b(t)$ and $X_i(t)$ in N, then we have the linear model $\mathbf{Y} = \Lambda^* \theta + \epsilon$, where

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \Lambda^* = \begin{pmatrix} 1 & \xi_{11} & \dots & \xi_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \xi_{n1} & \dots & \xi_{nN} \end{pmatrix}, \theta = \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_N \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

are the vector of response variable, the matrix of covariates, the vector of parameters, the vector of errors, respectively, and $\xi_{ij} = \int x_i \psi_j$. We obtained that $\hat{\theta} = (\Lambda^{*T} \Lambda^*)^{-1} \Lambda^{*T} \mathbf{Y}$, where Λ^{*T} denotes the transpose of matrix Λ^*. To emphasize truncation of the two series in N, we use the index N for the vectors and matrices obtained by those N terms. We assume that

$$\mathbf{Y}^* = (Y_1, \cdots, Y_{i-1}, \xi_i^* \hat{\theta}^{(-i;N)}, Y_{i+1}, \cdots, Y_n)^T,$$

where ${\xi^*}_i = (1, \xi_{i1}, \cdots, \xi_{iN})^T$ and $\hat{\theta}^{(-i;N)} = (\hat{a}^{(-i;N)}, \hat{b}_1^{(-i;N)}, \cdots, \hat{b}_N^{(-i;N)})^T$. Here, $\hat{\theta}^{(-i;N)}$ denotes the estimator of $\theta^{(N)}$ with all data except excluding the ith observation. We have:

$$
\begin{aligned}
(\mathbf{Y}^* - \mathbf{\Lambda}^*\theta)^T(\mathbf{Y}^* - \mathbf{\Lambda}^*\theta) &= \sum_{j=1}^{n}(Y_j^* - \xi_j^{*T}\theta^{(N)})^2 \geq \sum_{j:j\neq i}^{n}(Y_j^* - \xi_j^{*T}\theta^{(N)})^2 \\
&\geq \sum_{j:j\neq i}^{n}(Y_j^* - \xi_j^{*T}\hat{\theta}^{(-i;N)})^2 \geq \sum_{j=1}^{n}(Y_j^* - \xi_j^{*T}\hat{\theta}^{(-i;N)})^2 .
\end{aligned}
\tag{5.8}
$$

Thus, $\hat{\theta}^{(-i;N)}$ is the minimizer of the left hand-side term in (5.8) with respect to θ. On the other hand, differentiating this term with respect to θ results in $\hat{\theta}^{(-i;N)} = (\mathbf{\Lambda}^{*\,T}_{(N)}\mathbf{\Lambda}^*_{(N)})^{-1}\mathbf{\Lambda}^{*\,T}_{(N)}\mathbf{Y}^*$. Therefore, $\widehat{\mathbf{Y}}^{(-i;N)} = \mathbf{H}^{(N)}\mathbf{Y}^*$, where $\mathbf{H}^{(N)} = \mathbf{\Lambda}^*_{(N)}(\mathbf{\Lambda}^{*\,T}_{(N)}\mathbf{\Lambda}^*_{(N)})^{-1}\mathbf{\Lambda}^{*\,T}_{(N)}$. We have:

$$
\begin{aligned}
Y_i - \widehat{Y}_i^{(-i;N)} &= Y_i - \xi_j^{*T}\hat{\theta}^{(-i;N)} = Y_i - \sum_{j\neq i}^{n} H_{ij}^{(N)}Y_j - H_{ii}^{(N)}Y_i^* \\
&= Y_i - \sum_{j\neq i} H_{ij}^{(N)}Y_j - H_{ii}^{(N)}\xi_j^{*T}\hat{\theta}^{(-i;N)} \\
&= Y_i - H_{ii}^{(N)}Y_i - \sum_{j\neq i}^{n} H_{ij}^{(N)}Y_j - H_{ii}^{(N)}\xi_j^{*T}\hat{\theta}^{(-i;N)} + H_{ii}^{(N)}Y_i \\
&= Y_i - \sum_{j=1}^{n} H_{ij}^{(N)}Y_j + H_{ii}^{(N)}(Y_i - \xi_j^{*T}\hat{\theta}^{(-i;N)}) \\
&= Y_i - \widehat{Y}_i^{(N)} + H_{ii}^{(N)}(Y_i - \widehat{Y}_i^{(-i;N)}).
\end{aligned}
\tag{5.9}
$$

Hence, $Y_i - \widehat{Y}_i^{(-i)} = \frac{Y_i - \widehat{Y}_i}{1 - H_{ii}^{(N)}}$, and we can write CV_1 as follows:

$$
\begin{aligned}
\mathrm{CV}_1(r) &= \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \hat{a}^{(-i;r)} - \int_{\mathcal{I}}\hat{b}^{(-i;r)}X_i\right\}^2 \\
&\approx \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{a}^{(-i;r)} - \sum_{j=1}^{r}\xi_{ij}\hat{b}_j^{(-i;r)}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i^{(-i;r)})^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \widehat{Y}_i^{(r)})^2}{(1 - H_{ii}^{(N)})^2} .
\end{aligned}
\tag{5.10}
$$

Note that the approximated quantity in (5.10) was obtained by applying

$X_i = \sum_{j=1}^{r} \xi_{ij}\widehat{\psi}_j$ and $\hat{b}^{(-i;r)} = \sum_{j=1}^{r} b_j^{(-i;r)}\,\widehat{\psi}_j^{(-i)}$, in which the $\widehat{\psi}_j^{(-i)}$ are computed by excluding the ith observation. Moreover, there we have used the approximation $\langle \widehat{\psi}_j^{(-i)}, \widehat{\psi}_k \rangle \approx \delta_{jk}$, where δ_{jk} is the Kronecker delta. If we define $\Lambda^*{}_{(N)} = \big[\mathbf{1}, \Lambda_{(N)}\big]$, where

$$\Lambda_{(N)} = \begin{pmatrix} \xi_{11} & \dots & \xi_{1N} \\ \vdots & & \vdots \\ \xi_{n1} & \dots & \xi_{nN} \end{pmatrix}, \quad \mathbf{1}_{n\times 1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

then we can get $\widehat{\mathbf{Y}}^{(N)}$ as follows:

$$\widehat{\mathrm{Y}}^{(N)} = \Lambda^*_{(N)}\hat{\theta}^{(N)} = \big[1, \Lambda_{(N)}\big] \begin{pmatrix} \bar{\mathrm{Y}} - \frac{1}{n}\mathbf{1}\,\Lambda_{(N)}\,\widehat{\Sigma}^{-1}\big(\Lambda_{(N)}{}^T\mathrm{Y} - \Lambda_{(N)}{}^T\mathbf{1}\,\bar{\mathrm{Y}}\big) \\ \\ \widehat{\Sigma}^{-1}\big(\Lambda_{(N)}{}^T\mathrm{Y} - \Lambda_{(N)}{}^T\,\mathbf{1}\,\bar{\mathrm{Y}}\big) \end{pmatrix}.$$

Therefore,

$$\widehat{\mathrm{Y}}^{(N)} = \big[1, \Lambda_{(N)}\big] \begin{pmatrix} \frac{1}{n}\mathbf{1}^T - \frac{1}{n}\mathbf{1}^T\,\Lambda_{(N)}\,\widehat{\Sigma}^{-1}\Lambda_{(N)}{}^T\big(\mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\big) \\ \\ \widehat{\Sigma}^{-1}\Lambda_{(N)}{}^T\big(\mathrm{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T\big) \end{pmatrix} \mathrm{Y},$$

and finally,

$$\begin{aligned} \widehat{\mathrm{Y}}^{(N)} &= \Big[\frac{1}{n}\mathbf{1}\,\mathbf{1}^T - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T\,\Lambda_{(N)}\,\widehat{\Sigma}^{-1}\Lambda_{(N)}{}^T\,\big(\mathrm{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T\big) + \Lambda\,\widehat{\Sigma}^{-1}\,\Lambda_{(N)}{}^T\big(\mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\big)\Big]\mathrm{Y} \\ &= \Big[\frac{1}{n}\mathbf{1}\,\mathbf{1}^T + \big(\mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\big)\Lambda_{(N)}\,\widehat{\Sigma}^{-1}\Lambda_{(N)}{}^T\big(\mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\big)\Big]\mathrm{Y}, \end{aligned}$$

where

$$\widehat{\Sigma} = \Lambda_{(N)}^T (\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\, \Lambda_{(N)}$$
$$= n\,\mathrm{diag}(\hat{\theta}_1, \cdots, \hat{\theta}_N)\,. \tag{5.11}$$

Consequently, $\mathbf{H}^{(N)} = \left[\frac{1}{n}\mathbf{1}\,\mathbf{1}^T + (\mathbf{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T)\Lambda_{(N)}\,\widehat{\Sigma}^{-1}\Lambda_{(N)}{}^T(\mathbf{I} - \frac{1}{n}\mathbf{1}\,\mathbf{1}^T)\right]$.

3. The third cross-validation criterion is a generalized cross-validation, obtained by substituting the average of the diagonal elements of $\mathbf{H}$ for the denominator of CV_2, instead of the ith diagonal element. It is:

$$\mathrm{CV}_3(r) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \widehat{Y}_i)^2}{(1 - \frac{1}{n}\,\mathrm{tr}(\mathbf{H}))^2}. \tag{5.12}$$

Where $\mathrm{tr}(\mathbf{H})$ in $\mathrm{GCV}_2(r)$ denotes trace of matrix $\mathbf{H}$.

There are three different kinds of lines seen on each graph in the panels. The solid lines graph the first quartile (25th percentile), median or second quartile and third quartile (75th percentile) of the distributions of $||\hat{b} - b||^2$ and $|a - \hat{a}|$ when $\hat{b}$ and $\hat{a}$ are computed with $r = \hat{r}$, producing the minimum of $\mathrm{CV}_1(r)$. The three quartiles (25, 50 and 75th percentile) of the distribution of $||\hat{b} - b||^2$ denoted by ISE, were plotted in the first row of the panels for each case (i), (ii) and (iii) separately, and of the latter were graphed in the second row. For computing the dashed lines, the procedure was the same as the solid lines, except for the value of $r = \hat{r}$ chosen by $\mathrm{CV}_2(r)$. For the dotted lines, however, the minimizer of $\mathrm{GCV}_2(r)$ was considered as the value of $r = \hat{r}$.

When considering model (i) and the first quartile of ISE, as the graphs show, there are no significant differences among the three lines obtained by the three cross-validation criteria. With the second quartile, however, GCV_2 and then, CV_2 have slightly better performance when sample size is small. Regarding the third quartile, we also see a better performance of CV_2 and GCV_2 than CV_1. When $n = 50$, for example, CV_2 works as well as GCV_2 but for $n = 100$, GCV_2 does a better job compared with CV_2. Furthermore, there is no significant difference between the performances of the three cross-validation criteria in estimation of $|a - \hat{a}|$. However, CV_1 in the first quartile of $|a - \hat{a}|$ tends to have slightly better performance when $n \leq 200$.

For model (ii), in the first quartile, GCV_2 works better. Moreover, the solid and dashed lines obtained by CV_1 and CV_2 tend to coincide with each other. With the median, GCV_2 has slightly better performance for all sample sizes. In regard to the third quartile, however, for $n = 50$, GCV_2 and then CV_2 works better than CV_1. When $n = 100$, CV_1 gives a smaller value of ISE compared to the other two criteria. Furthermore, when n is large, GCV_2 is the best. Besides, as the graphs show, there is no significant difference among the performance of the three CV criteria in the case of distribution of $|\hat{a} - a|$.

Under model (iii), there are no significant differences among the performance of the three CV criteria from the viewpoint of the first and second quartile of ISE. But GCV_2 tends to have slightly better performance on the median of ISE when sample size is very large. As regards the third quartile, GCV_2 performed better than the other two for all sample sizes, except for $n = 100$ and 500, where the other two perform better. Moreover, the solid and dashed lines, corresponding to CV_1 and CV_2, tend to coincide with each other. When considering $|a - \hat{a}|$, we do not see any significant differences among the performances of CV_1, CV_2 and GCV_2.

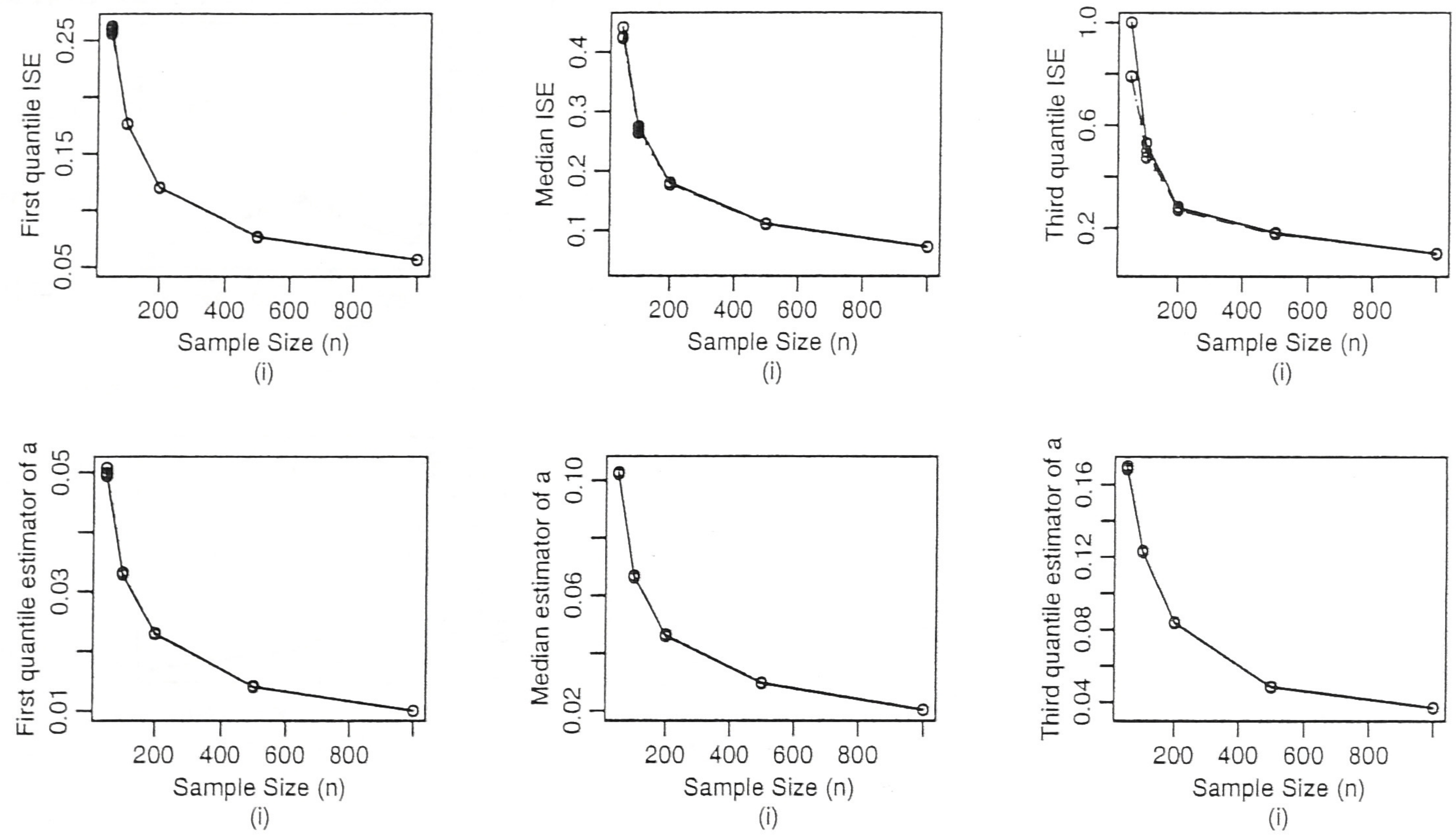

Figure 5.16: Performance of the distribution of integrated squared error (ISE) and of estimators of a, when r is chosen by cross-validation. The solid lines graph the three quartiles (25, 50 and 75th percentile) the distribution of $||\hat{b} - b||^2$ and $|a - \hat{a}|$ when $\hat{b}$ and $\hat{a}$ are computed by $r = \hat{r}$, producing the minimum value of $\mathrm{CV}_1(r)$. The dashed and dotted lines graph the three quartiles when $r = \hat{r}$ is selected by cross-validation $\mathrm{CV}_2(r)$ and $\mathrm{GCV}_2(r)$, respectively. The sample sizes were $n = 50, 100, 200, 500$ and 1000, and the undertaken model was model (i).

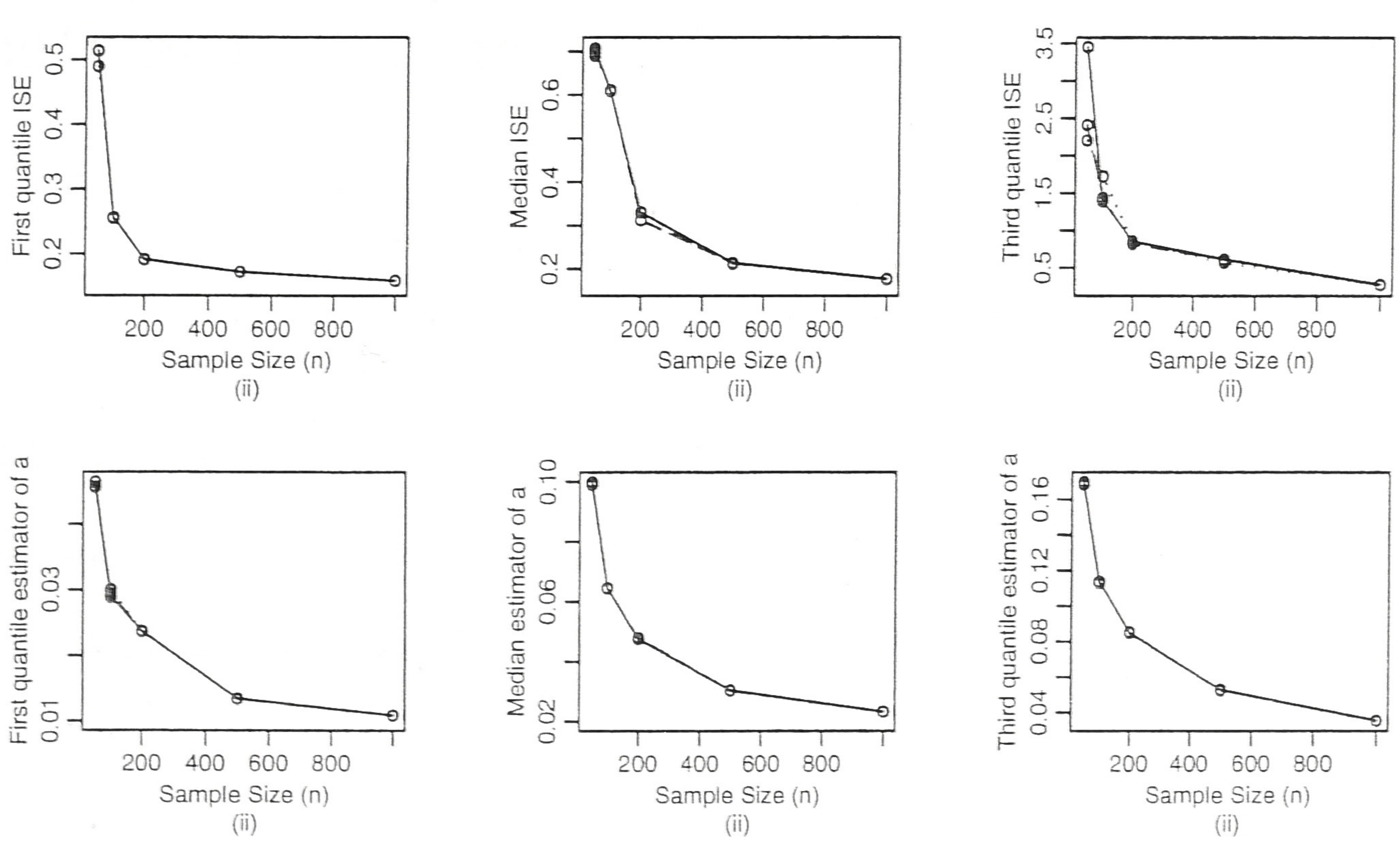

Figure 5.17: Performance of the distribution of integrated squared error (ISE) and of estimators of a, when r is chosen by cross-validation. The solid line graphs the three quartiles (25, 50 and 75th percentile) the distribution of $||\hat{b} - b||^2$ and $|a - \hat{a}|$ when $\hat{b}$ and $\hat{a}$ are computed by $r = \hat{r}$, producing the minimum value of $\mathrm{CV}_1(r)$. The dashed and dotted lines graph the the three quartiles when $r = \hat{r}$ is selected by cross-validation $\mathrm{CV}_2(r)$ and $\mathrm{GCV}_2(r)$, respectively. The sample sizes were $n = 50, 100, 200, 500$ and 1000, and the undertaken model was model (ii).

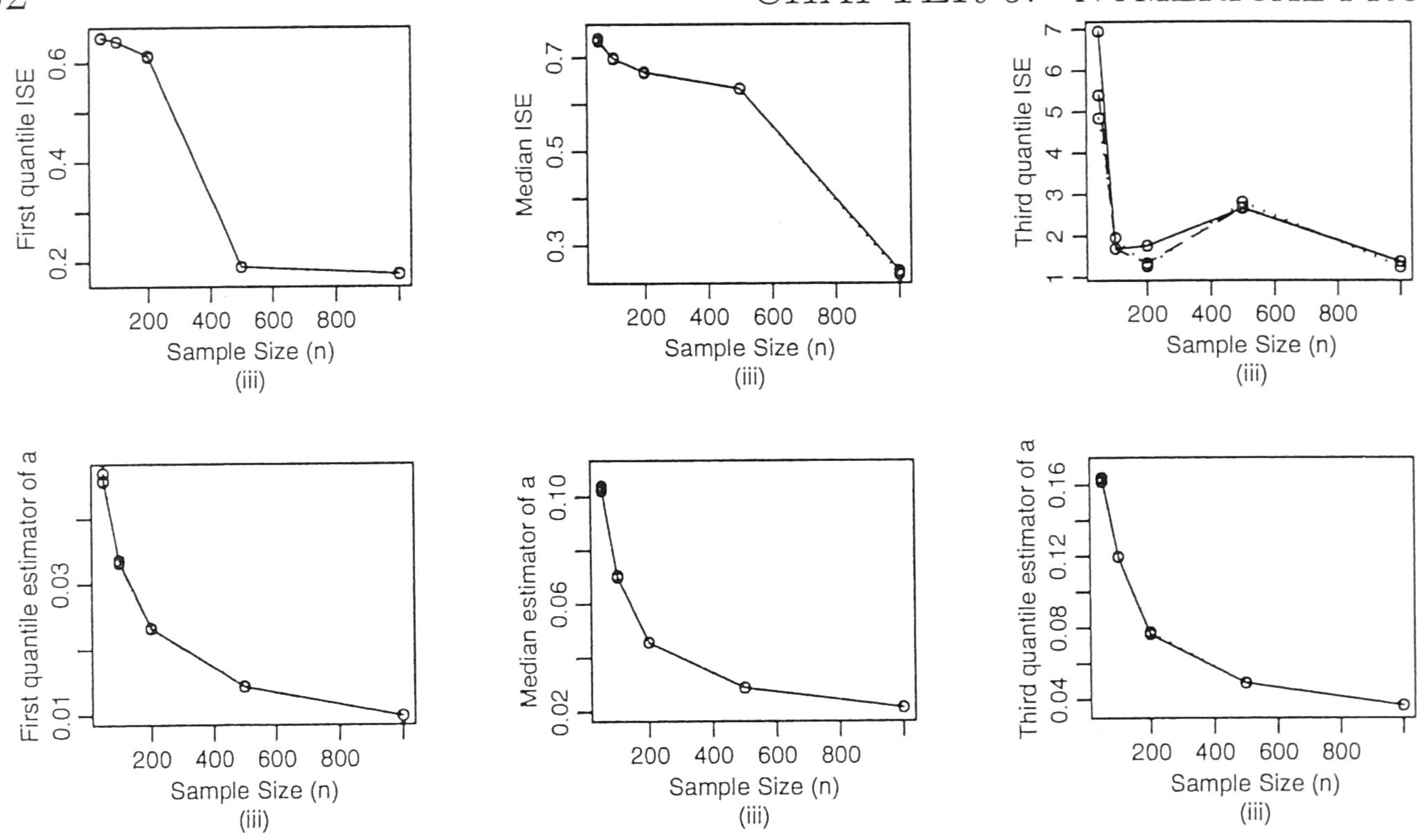

Figure 5.18: Performance of the distribution of integrated squared error (ISE) and of estimators of a, when r is chosen by cross-validation. The solid lines graph the three quartiles (25, 50 and 75th percentile) the distribution of $||\hat{b} - b||^2$ and $|a - \hat{a}|$ when $\hat{b}$ and $\hat{a}$ are computed by $r = \hat{r}$, producing the minimum value of $\mathrm{CV}_1(r)$. The dashed and dotted lines graph the three quartiles when $r = \hat{r}$ is selected by cross-validation $\mathrm{CV}_2(r)$ and $\mathrm{GCV}_2(r)$, respectively. The sample sizes were $n = 50, 100, 200, 500$ and 1000, and the undertaken model was model (iii).

5.8 Numerical Results Connected to the Real Dataset (Canadian Temperature)

Our real data consist of recorded temperatures at 35 Canadian weather-stations for a certain year (1982). The original data were in the form of 12 points for each station, representing monthly averages of Canadian temperature, associated with the middle of each month. Because temperature is periodic, we should use a basis which represents this fact. Hence, we have

$$X(t) = \xi_1 + \xi_2 \sin(\omega t) + \xi_3 \cos(\omega t) + \xi_4 \sin(2\omega t) + \xi_5 \cos(2\omega t) + \cdots, \qquad (5.13)$$

where $\frac{2\pi}{\omega}$ is equal to 12, the length of the interval $\mathcal{I} = [0, 12]$ on which the temperature functions is defined, i.e $\omega = \frac{\pi}{6}$. Because the original data were recorded on a discrete grid (in the middle of every month), they may be contaminated with errors. However, the error in our data, resulting from measuring temperature, is very small, compared to the variation in actual temperature. Nevertheless, it is better to remove the error by taking smoothing into account rather than interpolating, as the error may cause a degree of roughness. Because taking a finite number of the terms in the series in (5.13) causes some degree of smoothness to the functions, we expanded the data by using $N = 13$ terms through "the least squares fitting of basis expansions". Meanwhile, we controlled roughness by adding a small roughness penalty, as follows:

$$\sum_i \{y_i - X(t_i)\}^2 + \mathrm{PEN}(\chi) = \|\mathbf{y} - \Psi\xi\|^2 + \lambda\,\xi^T\,\mathbf{R}\,\xi\,, \qquad (5.14)$$

where PEN denotes the roughness penalty, the 12-vector $\mathbf{y}$ shows the original discrete data, observed in the middle of the 12 months, the $n \times N$ matrix $\boldsymbol{\Psi}$ has elements $\Psi_{ij} = \psi_j(t_i)$, $\psi_j(t_i)$ is the value of the jth term of the basis at point t_i, the $N \times N$ matrix $\mathbf{R}$ has elements $R_{ij} = \int_0^{12} \psi_i''(t)\,\psi_j''(t)\,dt$, in which ψ_j'' denotes the second derivative of ψ_j, and the N-vector ξ denotes the Fourier coefficient. There, the amount of smoothness can be controlled by λ. We chose λ very small and minimised (5.14) with respect to ξ, leading to the optimal solution $\xi = (\Psi^T\Psi + \lambda\,\mathbf{R})^{-1}\,\Psi^T\,\mathbf{y}$. Therefore, the data were registered as 35 functions by expanding the temperature curves in the first 13 terms of the Fourier basis, and setting them up with the discrete temperature data to create smooth curves $X_i(t)$ from the original discrete data. Figure 5.19 presents the 35 registered functions.

After subtracting mean from the data, we obtained the estimated eigenvalues and eigenfunctions. As Figure 5.20 reveals, the value of $\hat{\theta}_1$ shows a strong domination of its associated variation on all other kinds. Furthermore, the first four eigenvalues explained more than 99% of the total variation, in which their contributions individually were $89.3, 8.3, 1.6$ and 0.5 percent, respectively. Contributions from others, however, were less than 0.5 percent. Thus, this drew our attention to the first four kinds of variation rather than the whole. As Figure 5.21 presents, the first estimated principal component (PC) curve shows that the majority of variability (89%) among the data can be attributed to differences between summer and winter temperatures.

The second PC shows regularity of temperature when moving from winter to summer. In other words, it reflects the variation from the average of the difference between the winter and summer temperatures. It contributes positively for winter, and negatively for summer. Therefore, it gives a high positive score to the area for which the difference between winter and summer temperature is small. In contrast, a large negative score is allocated to the areas which are hot in summer and cold in winter. The third PC corresponds to a time shift effect which is accompanied by a slight overall increase in both temperature and range between winter and summer. The fourth PC is due to an effect causing spring to start later and autumn to end earlier.

As a practical illustration we show 95% bootstrap confidence bands, calibrated using the double bootstrap, for the first four principal components in the case of J.O. Ramsay's Canadian weather-station temperature dataset. The extreme closeness of single- and double-bootstrap bands reflects the high degree of accuracy of the uncalibrated bands. This results was confirmed by the simulation study since the coverages from the single and the double bootstrap were almost close to each other.

Table 5.4 reveals a 95% bootstrap confidence intervals for θ_j obtained from Canadian temperature data. We used two different methods of constructions (equal-tailed and symmetric confidence intervals), and for each we applied single- and double-bootstrap methods. The results show that in all cases, as the order of the eigenvalues decreases, the length of the confidence intervals shrinks. The

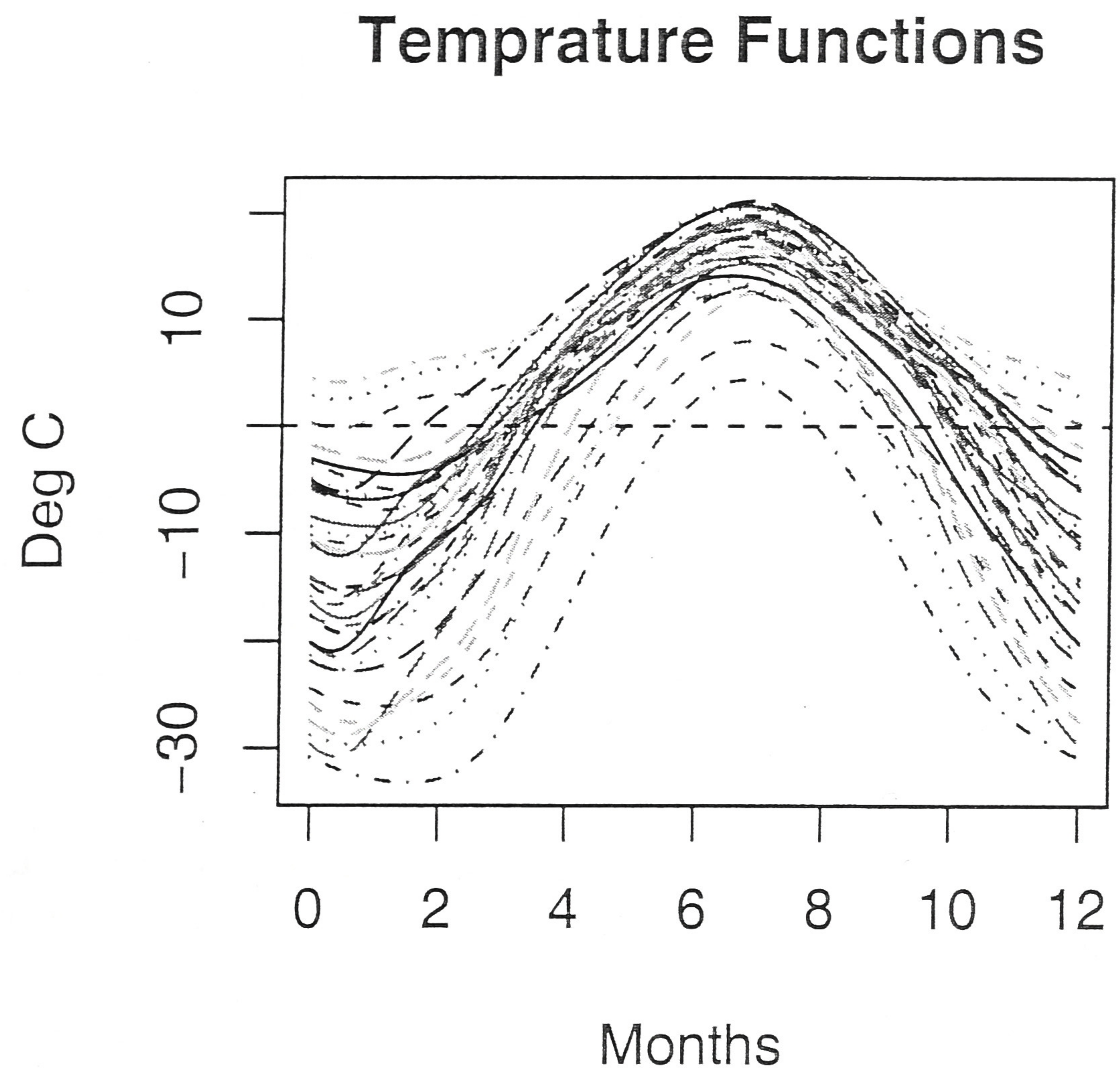

Figure 5.19: Monthly mean of temperature for the Canadian weather stations.

more they are important in view of having a larger proportion of total variation, the more their corresponding confidence intervals are longer.

Regarding the simultaneous bootstrap confidence interval for θ_j, we estimated $\widehat{\Delta}_{\text{upp}} = 259$ from the 95% upper level of the distribution of $\widehat{\Delta}^* = \|\widehat{K}^* - \widehat{K}\|$, i.e. $P(\sup_{j\geq 1} |\hat{\theta}_j - \theta_j| \leq 259) \approx 0.95$, which means that with probability nearly 0.95, for each $1 \leq j \leq 12$ the distance between θ_j and $\hat{\theta}_j$ does not exceed 259.

We were interested in explaining variation in the total annual precipitation by using the temperature variation pattern through the year. Considering the linear regression model, $Y_i = \int_{\mathcal{I}} b(t)\, X_i(t)\, dt + \epsilon_i$, we took Y to be the total annual precipitation and $X(t)$ to be temperature. Because the total annual precipitation was distributed across the four different areas, Atlantic, Continental, Pacific and Arctic, it was highly variable from one weather station to another. Thus, we decided to use its logarithm as the dependent variable. Specifically, we regarded

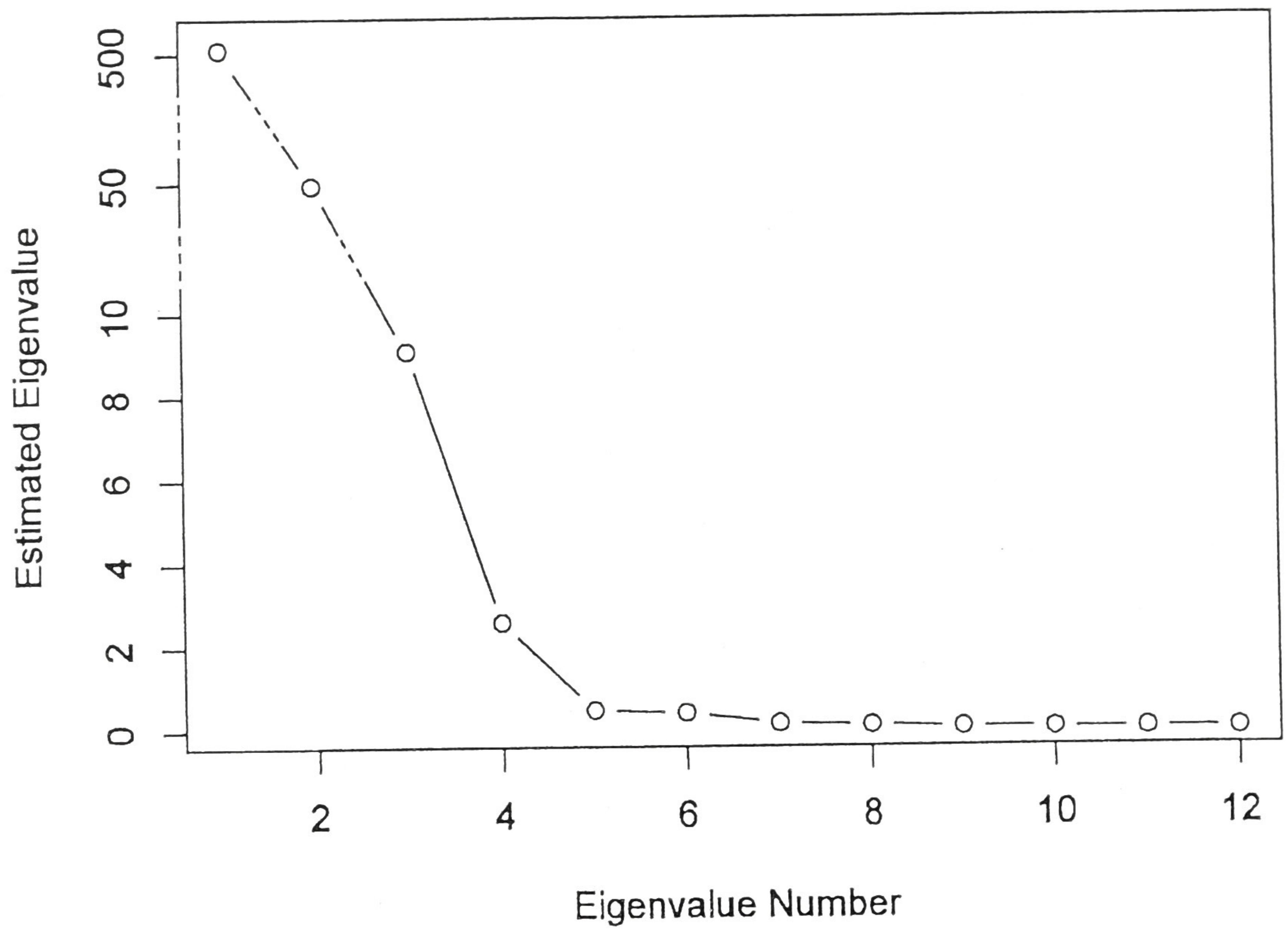

Figure 5.20: The estimated eigenvalues for the Canadian temperature data. The amounts of total variation accounted for by the first four eigenvalues are 89.3, 8.3, 1.6, and 0.5 percent, respectively.

Method	Equal – tailed		Symmetric	
Eigenvalue	SB	DB	SB	DB
θ_1	$(246.26, 737.35)$	$(254.28, 848.07)$	$(260.12, 749.97)$	$(175.24, 834.86)$
θ_2	$(30.019, 71.35)$	$(28.72, 86.13)$	$(24.65, 69.61)$	$(12.05, 82.21)$
θ_3	$(6.57, 12.99)$	$(5.95, 14.71)$	$(5.58, 12.58)$	$(4.35, 13.82)$
θ_4	$(0.99, 4.49)$	$(0.54, 6.15)$	$(0.73, 4.44)$	$(0, 5.93)$
θ_5	$(0.30, 0.65)$	$(0.26, 0.71)$	$(0.30, 0.65)$	$(0.28, 0.67)$
θ_6	$(0.38, 0.66)$	$(0.38, 0.81)$	$(0.19, 0.63)$	$(0.062, 0.76)$
θ_7	$(0.11, 0.25)$	$(0.082, 0.32)$	$(0.059, 0.24)$	$(0, 0.31)$
θ_8	$(0.094, 0.17)$	$(0.089, 0.23)$	$(0.033, 0.17)$	$(0, 0.22)$
θ_9	$(0.049, 0.091)$	$(0.043, 0.12)$	$(0.019, 0.089$	$(0, 0.11)$
θ_{10}	$(0.026, 0.048)$	$(0.022, 0.059)$	$(0.010, 0.047$	$(0, 0.057)$
θ_{11}	$(0.022, 0.037)$	$(0.021, 0.049)$	$(0.0048, 0.036)$	$(0, 0.048)$
θ_{12}	$(0.0093, 0.015)$	$(0.0092, 0.021)$	$(0.0014, 0.015)$	$(0, 0.021)$

Table 5.4: The bootstrap confidence intervals for θ_j obtained from the Canadian temperature data. The nominal coverage was $1 - \alpha = 0.95$, and $B = 5000$ resamples were drawn. Then each of them was sampled $C = 500$ times with replacement. In the table, SB and DB denote single- and double-bootstrap confidence intervals, respectively.

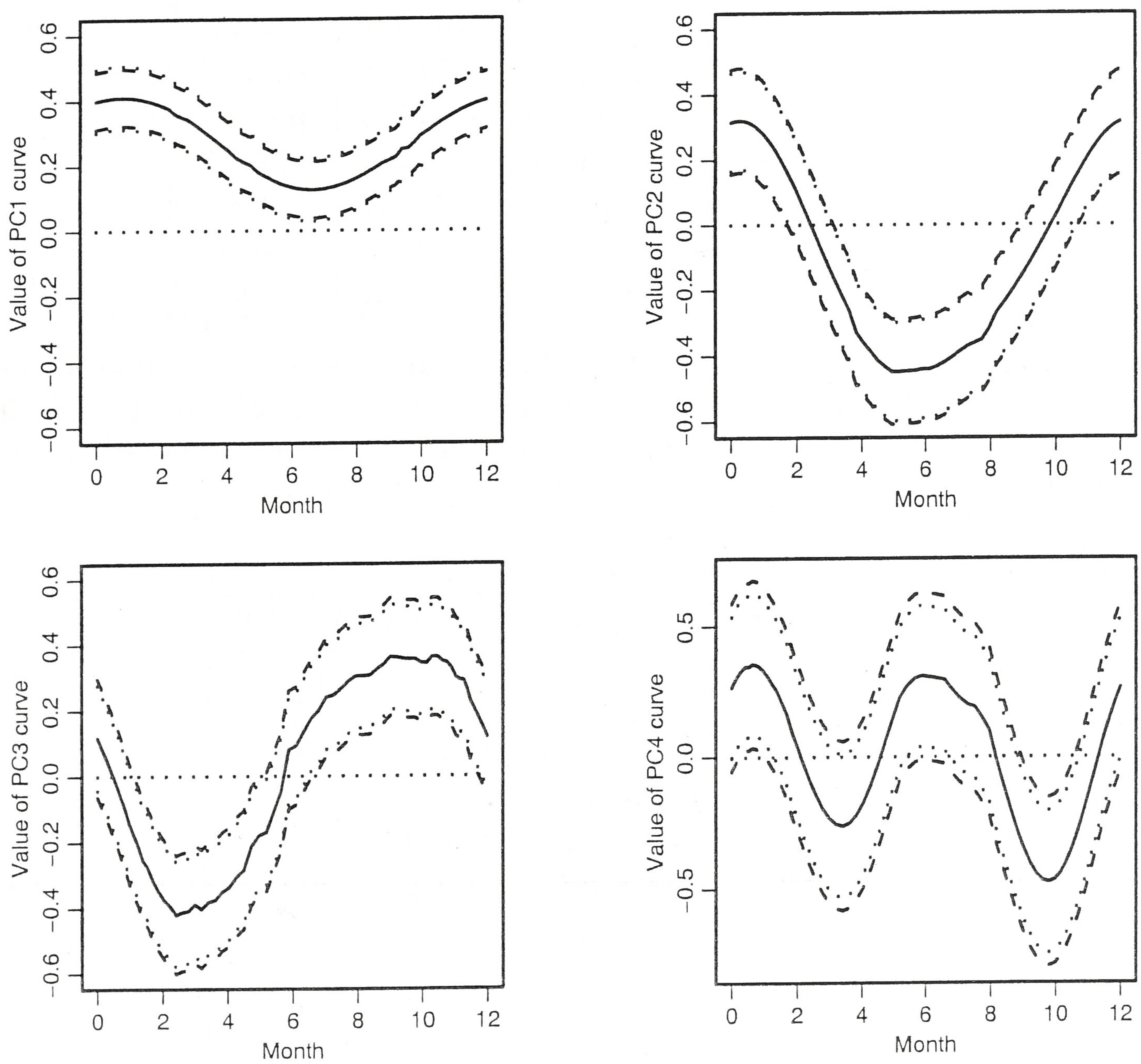

Figure 5.21: First four eigenfunctions with their single and double bootstrap bands, obtained from Ramsay's Canadian weather stations dataset. The dashed line shows single bootstrap bands, the dotted line shows double bootstrap bands, and the unbroken line shows the function estimate $\widehat{\psi}_j$.

Method	Symmetric	Equal − tailed	
Eigenfunction	$\hat{z}_{0.05}$	$\hat{z}_{0.025}$	$\hat{z}_{0.975}$
ψ_1	0.087	−0.0005	0.0210
ψ_2	0.16	−0.0137	0.0115
ψ_3	0.18	−0.0004	−0.0119
ψ_4	0.32	0.00131	−0.0474
ψ_5	0.89	0.0003	0.2540
ψ_6	1.27	−0.4749	0.4872
ψ_7	0.85	−0.0056	−0.0445
ψ_8	1.35	−0.0511	0.6635
ψ_9	1.11	−0.0090	0.3073
ψ_{10}	0.90	−0.0473	0.0289
ψ_{11}	1.25	0.0016	0.4375
ψ_{12}	1.41	−0.0122	0.5961

Table 5.5: The single bootstrap confidence statements for ψ_j obtained from the Canadian temperature data. The nominal coverage was $1-\alpha = 0.95$, and $B = 5000$ resamples were drawn. Here, the confidence statements should be interpreted as regions $\{(t,u) : |\widehat{\psi}(t) - u| \le \hat{z}_{0.05} \text{ and } t \in \mathcal{I} = [0,12]\}$.

Method	Equal − tailed		Symmetric
Eigenfunction	$\hat{z}^*_{0.025}$	$\hat{z}^*_{0.975}$	$\hat{z}^*_{0.05}$
ψ_1	0.0080	0.11	0.097
ψ_2	−0.013	0.15	0.15
ψ_3	0.027	0.17	0.16
ψ_4	0.041	0.29	0.27
ψ_5	0.078	1.19	1.12
ψ_6	−0.36	1.64	1.82
ψ_7	0.11	0.65	0.96
ψ_8	0.037	2.22	1.89
ψ_9	0.14	1.46	1.34
ψ_{10}	0.091	0.92	0.89
ψ_{11}	0.14	1.71	1.65
ψ_{12}	0.071	1.99	2.052

Table 5.6: The double bootstrap confidence statements for ψ_j obtained from the Canadian temperature data. The nominal coverage was $1-\alpha = 0.95$, and $B = 5000$ resamples were drawn. Then each of them was sampled $C = 500$ times with replacement. Here, the confidence statements should be interpreted as regions $\{(t,u) : \hat{z}^*_{0.025} \le \widehat{\psi}(t) - u \le \hat{z}^*_{0.975} \text{ and } t \in \mathcal{I} = [0,12]\}$ and $\{(t,u) : |\widehat{\psi}(t) - u| \le \hat{z}^*_{0.05} \text{ and } t \in \mathcal{I} = [0,12]\}$ for equal-tailed and symmetric construction, respectively.

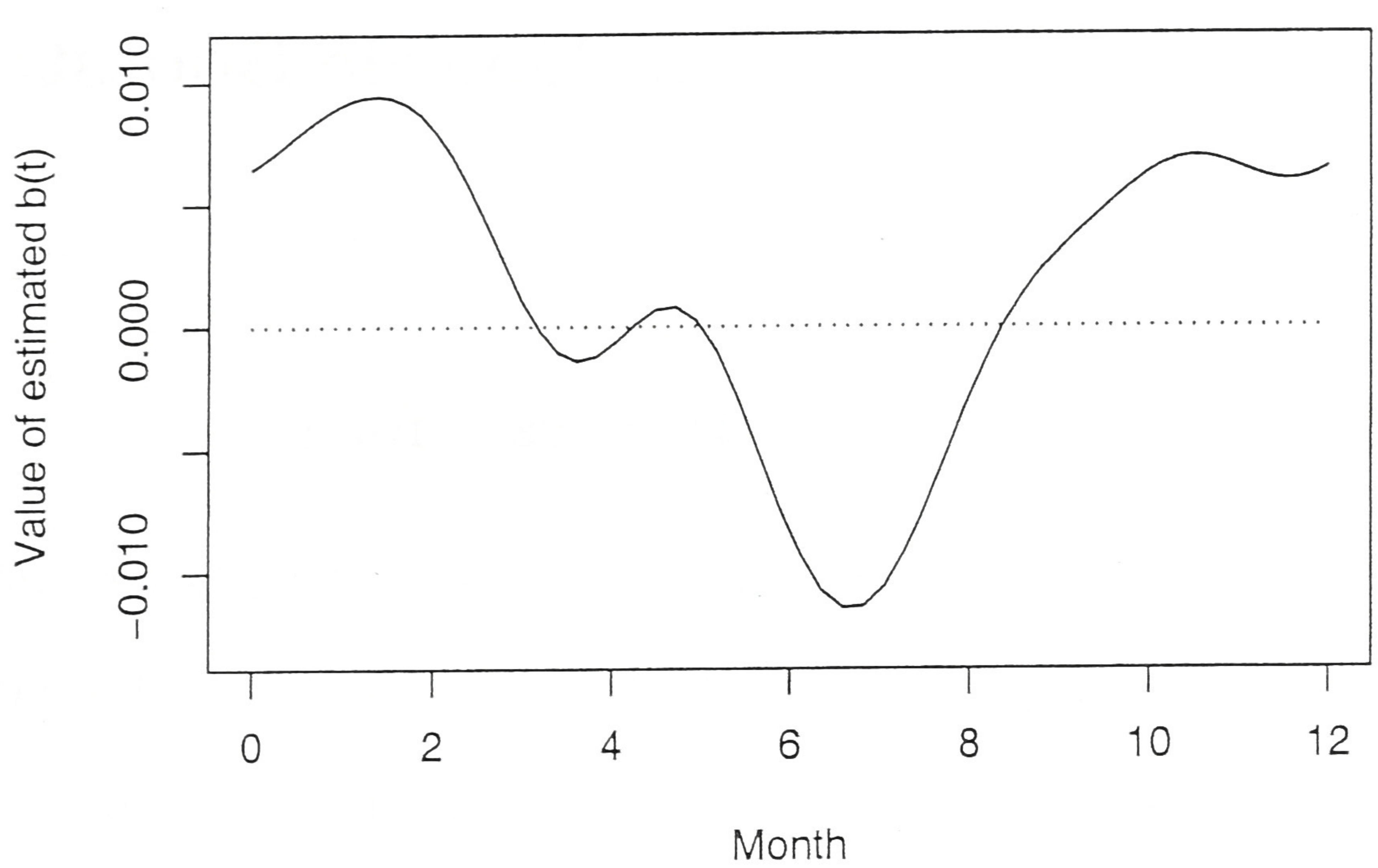

Figure 5.22: The estimated regression slope when taking the log total annual precipitation as the dependent variable. The smoothing parameter, chosen by cross-validation was $\hat{r} = 2$. The estimated intercept was also $\hat{a} = -0.2413092$.

the dependent variable Y as $Y_i = \log_{10}(\sum_{j=1}^{12} \mathrm{prec}_{ij}) - \frac{1}{35}\sum_{i=1}^{35} \log_{10}(\sum_{j=1}^{12} \mathrm{prec}_{ij})$, where prec_{ij} denotes the amount of precipitation reported by station i during month j. Then we estimated the slope b and intercept a from the data using the estimators proposed in (4.8). To estimate the smoothing parameter r, we used the cross-validation criterion

$$\mathrm{CV}_2(r) = \frac{1}{n}\sum_{i=1}^{n} \frac{(Y_i - \widehat{Y}_i)^2}{(1 - H_{ii})^2}, \tag{5.15}$$

where the matrix $\mathbf{H}$ can be determined from $\widehat{\mathbf{Y}} = \mathbf{H}\,\mathbf{Y}$, and H_{ii} is the ith diagonal element of the $n \times n$ matrix $\mathbf{H}$. We obtained $\hat{b}$ and $\hat{a}$ with $\hat{r} = 2$ chosen by CV_2 (see Figure 5.22).

As Figure 5.22 shows, the estimated slope of the functional linear regression gives larger weights to the winter months as well as the second half of autumn. Furthermore, lowest weights corresponds to summer and then at the middle of spring.

5.9 Comparison of Single and Double Bootstrap Coverages for θ_j

5.9.1 Simultaneous Confidence Intervals for θ_j

Figure 5.23 shows coverage of the simultaneous bootstrap confidence interval for θ_j when the nominal coverage level is 95% and X is Gaussian. The dashed and dotted lines show coverage levels when applying single and double bootstrap methods, respectively. For both the single and the double bootstrap, model (i) enjoys good performance compared with the other two models. Furthermore, use of the double bootstrap to calibrate confidence intervals improves coverage accuracy, especially for small sample sizes. It reduces the deficit in the coverage levels from 0.10, resulting from applying the single bootstrap method, to just 0.03. While we have coverage level 0.93 for model (i) and $n = 20$, and under models (ii) and (iii) is 0.90 and 0.91 for $n = 50$, double-bootstrap calibration for these sample sizes provides a good improvement on coverage accuracy with actual coverage 0.95 under each of the three models. As n increases, confidence intervals obtained by applying the double bootstrap method become conservative (having coverages grater than the nominal level).

Figure 5.24 shows coverage of simultaneous bootstrap confidence interval for θ_j when the nominal coverage level is 95% and X is the Non-Gaussian process. Here, as with the Gaussian case, double-bootstrap calibration improves coverage accuracy. In both situations, model (i) has a better performance compared with models (ii) and (iii). For example, when constructing confidence intervals by single-bootstrap method for $n = 20$, the coverages are 0.80, 0.72 and 0.70, respectively under models (i), (ii) and (iii). However, they are 0.84, 0.80 and 0.80

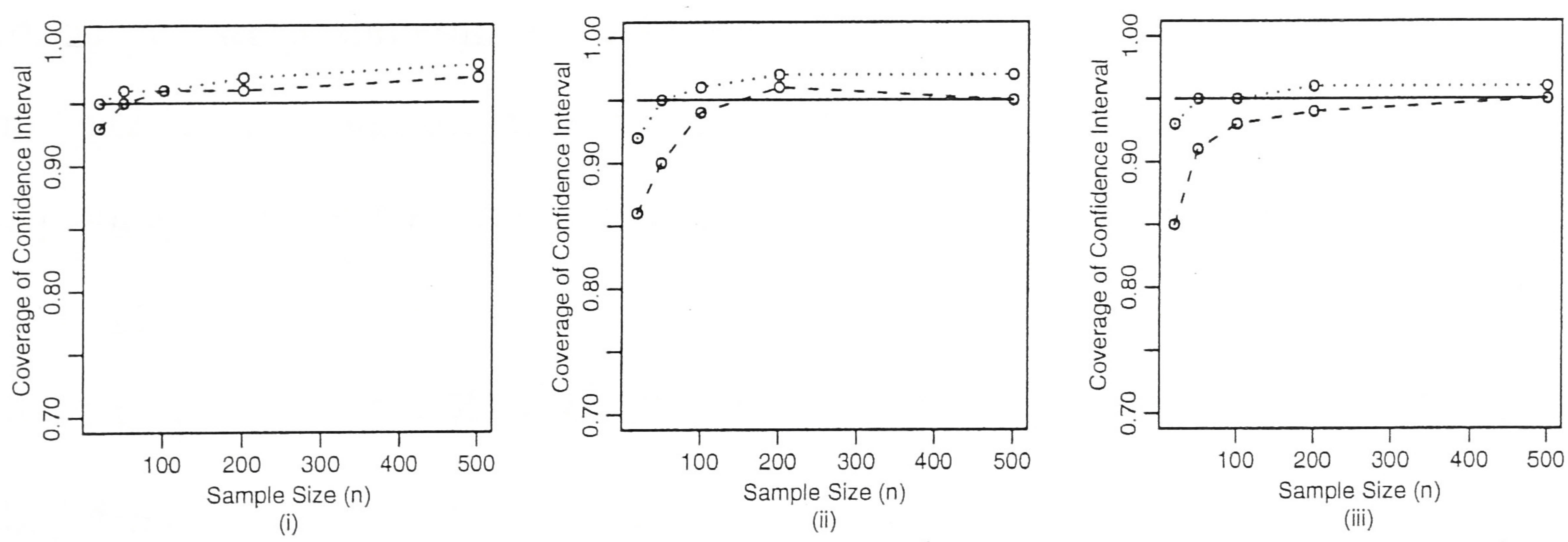

Figure 5.23: Comparison of coverages of two-sided, simultaneous, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process. The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

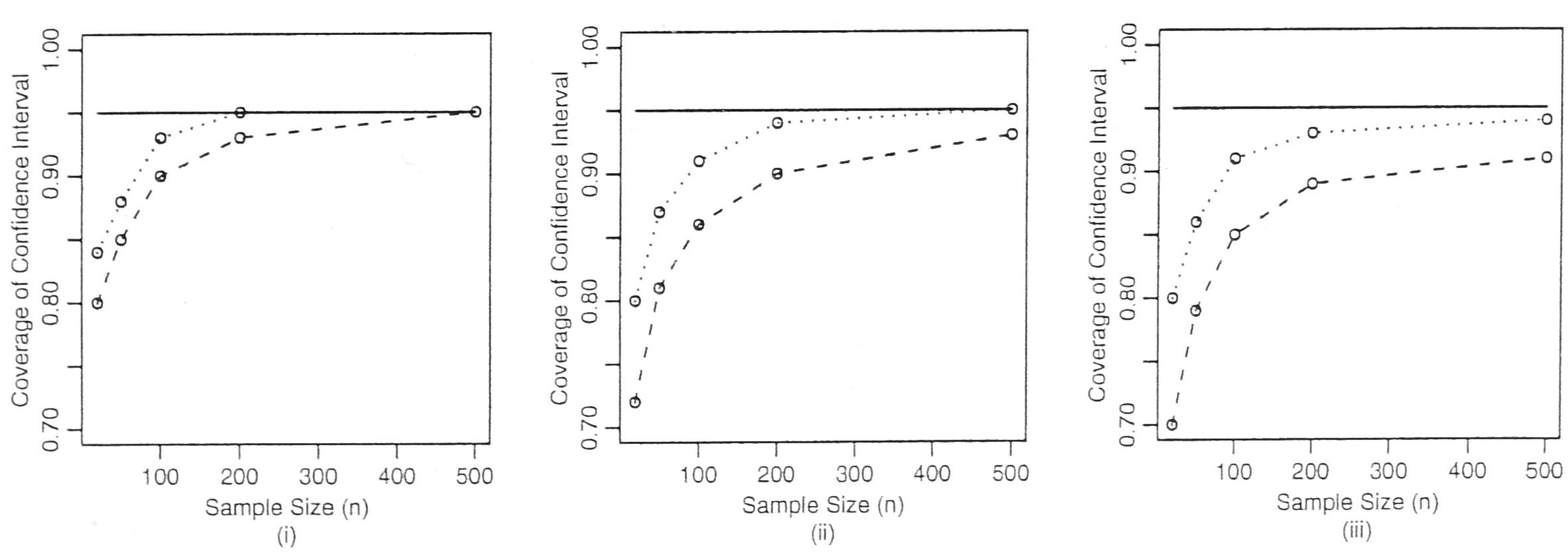

Figure 5.24: Comparison of coverages of two-sided, simultaneous, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process. The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

for those three models when calibrating confidence intervals using the double bootstrap. It can also be seen that the amount of improvement, resulting from this calibration, is lower under model (i) compared with the other two models.

5.9.2 Confidence Intervals for Individual θ_j

Figures 5.25–5.27 show coverage levels of two-sided symmetric, nominal 95% confidence intervals for θ_j individually, when X is simulated from the Gaussian process under models (i)–(iii). The dashed lines show coverages when confidence intervals are constructed by the single bootstrap method, and the dotted lines show coverage levels when calibrating the confidence bounds by the double bootstrap.

Considering model (i), as the panels reveal, generally, using double-bootstrap calibration improves coverage accuracy, especially since it offers better performance for small sample sizes. The amount of improvement decreases as j increases, except for $n = 20$, and this amount is tending to zero for the last eigenvalue. While coverages obtained by applying single-bootstrap are 0.88, 0.91 and 0.92 for sample sizes 20, 50 and 100, respectively, they are 0.91, 0.94 and 0.95 when using double-bootstrap calibration for those sample sizes. Moreover, apart from the last two eigenvalues, which in both situations give conservative coverage levels, use of double-bootstrap calibration reduces the deficit from 0.02-0.15 to 0.00-0.07. Furthermore, the gap between the dotted and dashed lines deceases as n increases, meaning that double-bootstrap calibration has little impact on coverages as sample size increases.

Regarding model (ii), here also coverage levels have the same features as those discussed under model (i). Use of double-bootstrap calibration in constructing confidence intervals, for $n \leq 200$, reduces the deficit 0.01–0.25, resulting from applying the single bootstrap method, to 0.01–0.11. Specifically, we have actual coverage level 95% for the first four eigenvalues when $n = 200$.

Considering model (iii) when $n \leq 200$, use of double-bootstrap calibration reduces the deficit 0.01–0.27, resulting from single-bootstrap method, to 0.00–0.08. This means that double-bootstrap calibration removes about two thirds of the deficit; specifically, under the single bootstrap coverage level for θ_1 is 0.86, 0.90 and 0.92, respectively for $n = 20$, 50 and 100. However, after calibrating confidence bounds by using the double bootstrap, they are 0.92, 0.93 and 0.95 for these sample sizes.

Altering the distribution of ξ_{ij} to the centered exponential causes coverage accuracy of confidence intervals for θ_j decline (see Figures 5.28–5.30). However, a coverage-correction can be achieved by using double-bootstrap method. The coverage level features here, are similar to those in the case of the Gaussian process. The deficit for $n \leq 200$ and the first three eigenvalues caused by applying single-bootstrap, is 0.04–0.20, 0.06–0.30 and 0.04–0.29 for models (i), (ii) and (iii), respectively, and reduce to 0.03–0.16, 0.02–0.20 and 0.02–0.21 for the three models after double-bootstrap calibration. For example, for sample sizes 20, 50 and 100, coverage levels of confidence intervals for the three eigenvalues constructed by single-bootstrap, under model (i) are 0.75, 0.80, 0.86, under model (ii) are 0.68, 0.81, 0.86, and under model (iii) are 0.71, 0.81, 0.84, respectively. However, they increase to 0.79, 0.85, 0.90 under model (i), to 0.78, 0.87, 0.90 under model (ii) and to 0.80, 0.86, 0.88 under model (iii) for these sample sizes.

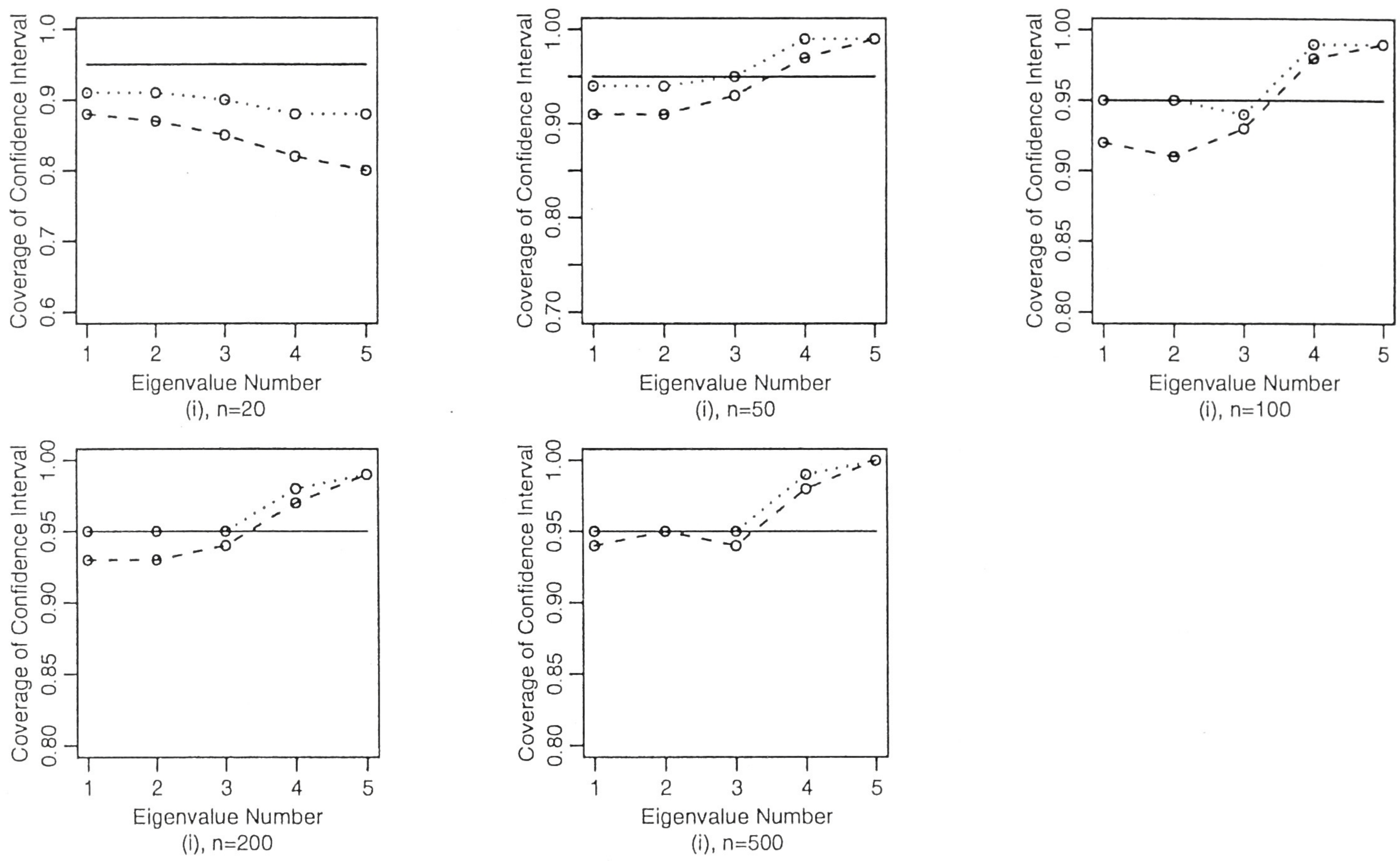

Figure 5.25: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (i). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1-\alpha = 0.95$.

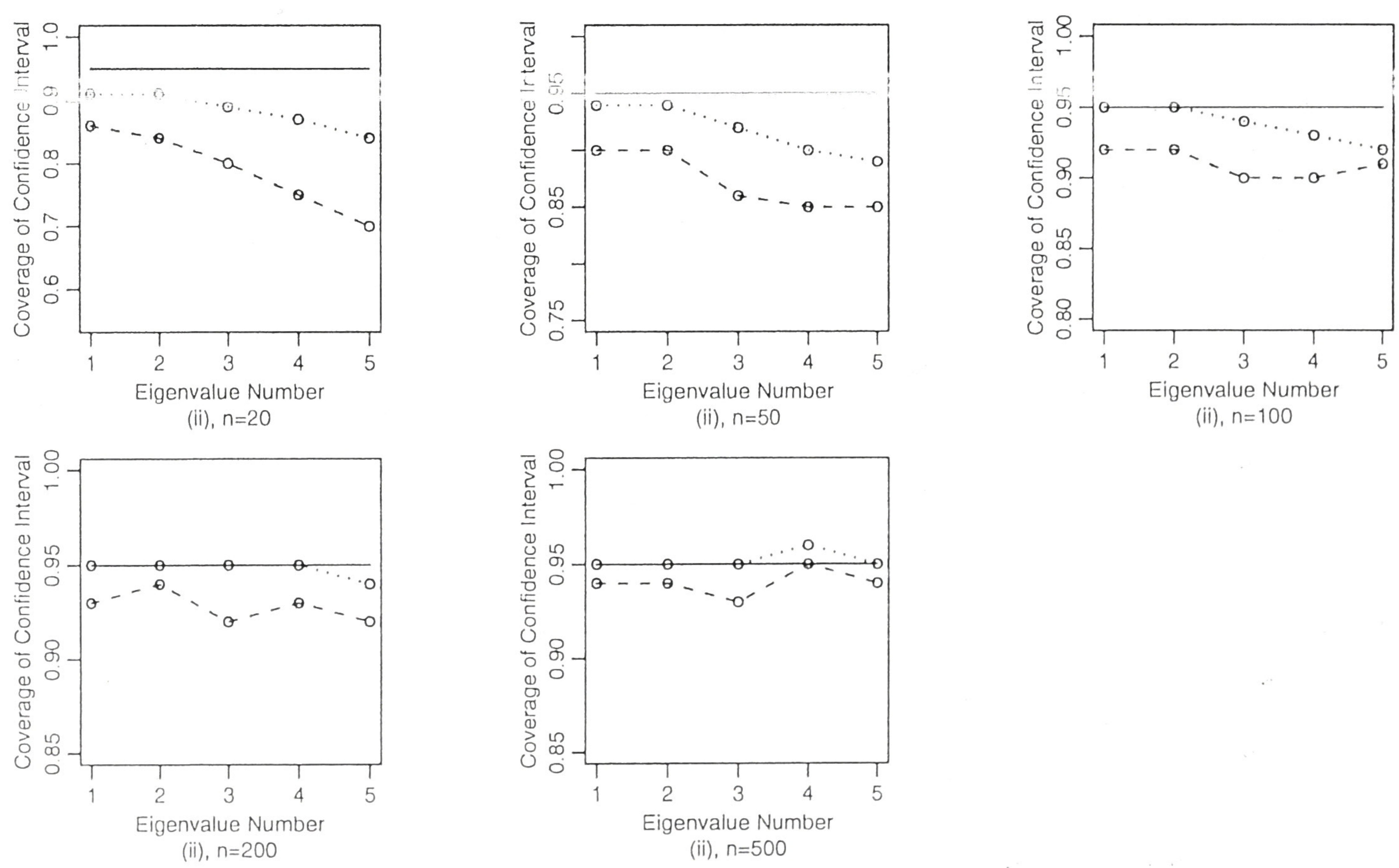

Figure 5.26: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (ii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1-\alpha = 0.95$.

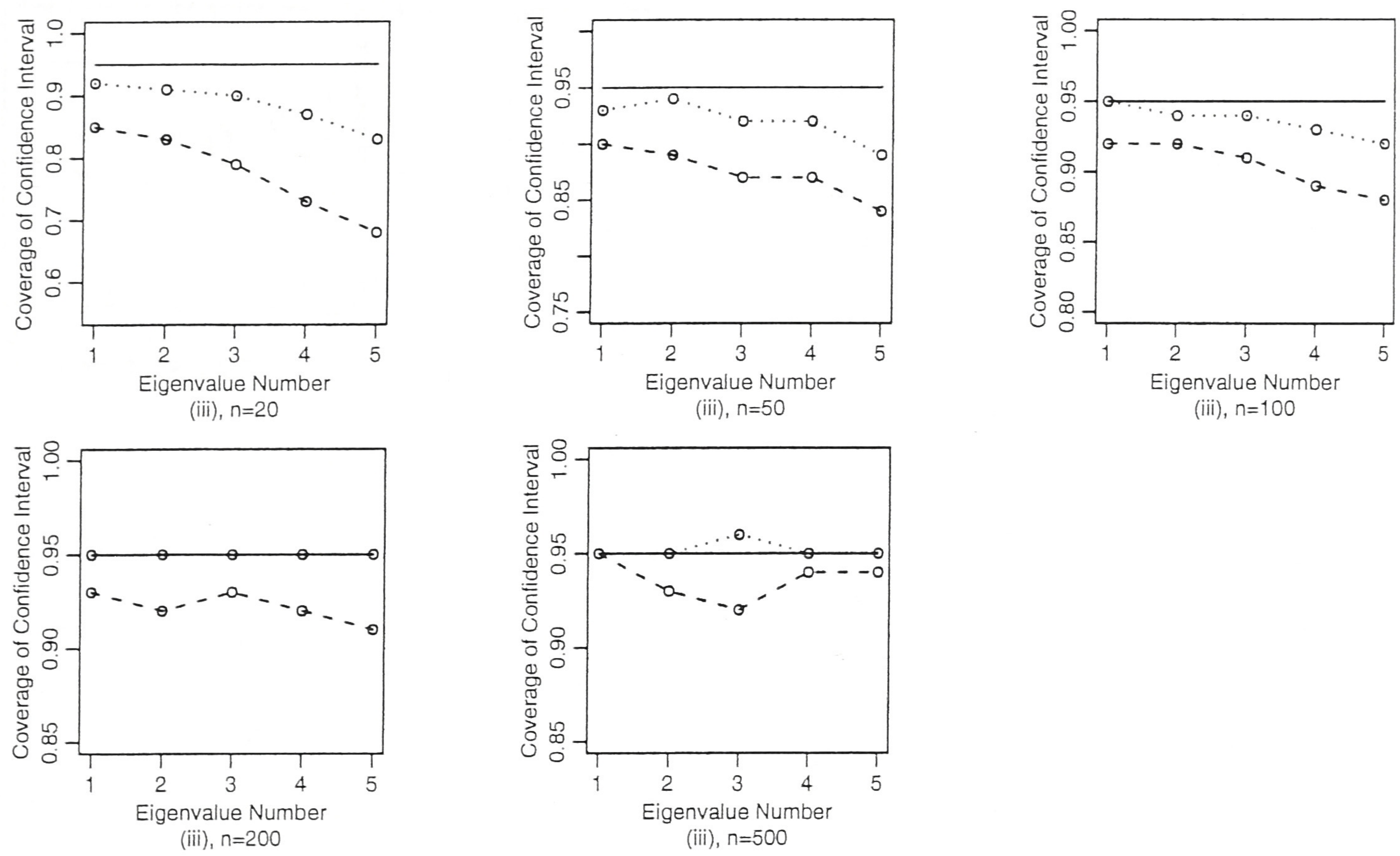

Figure 5.27: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (iii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

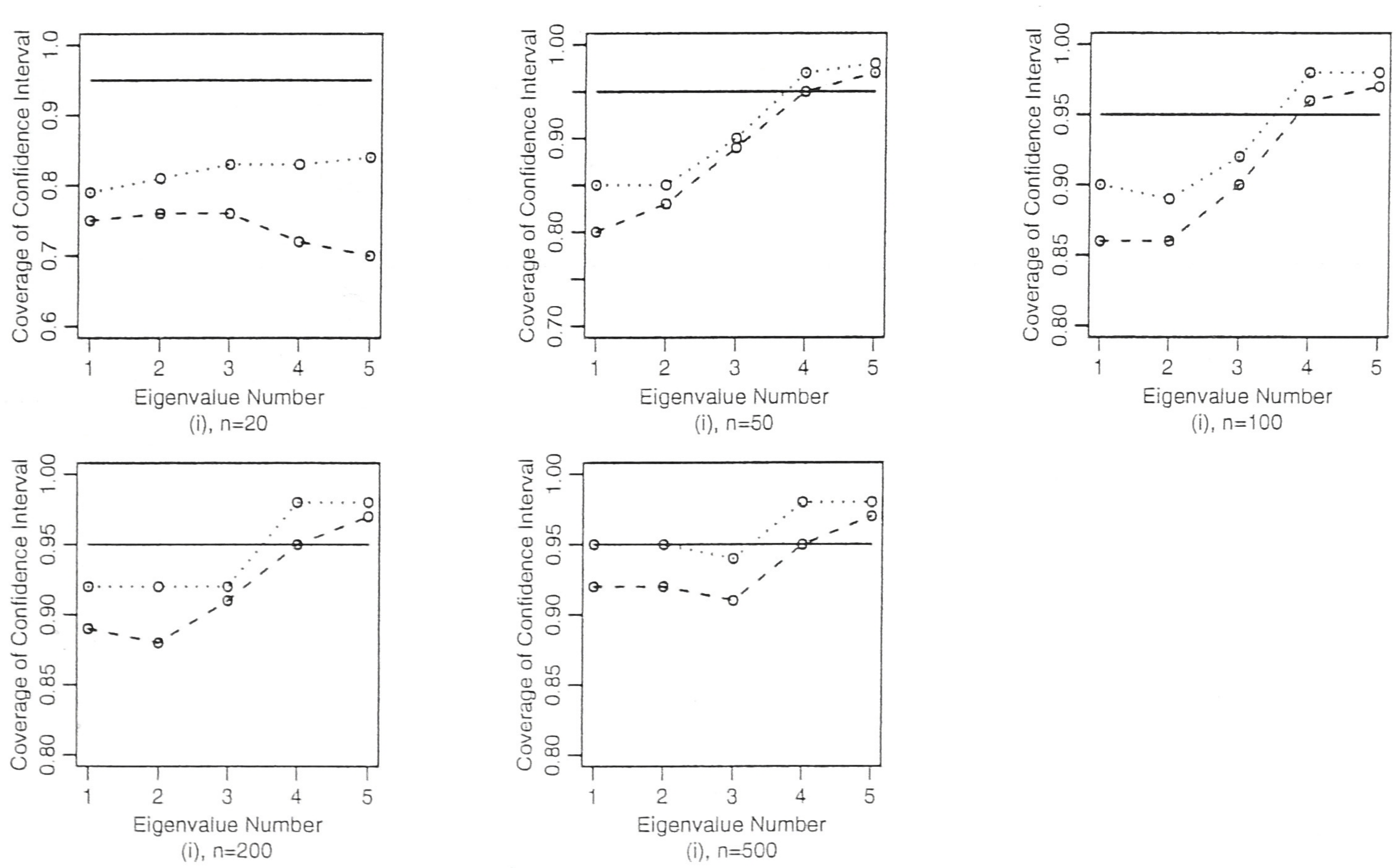

Figure 5.28: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (i). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

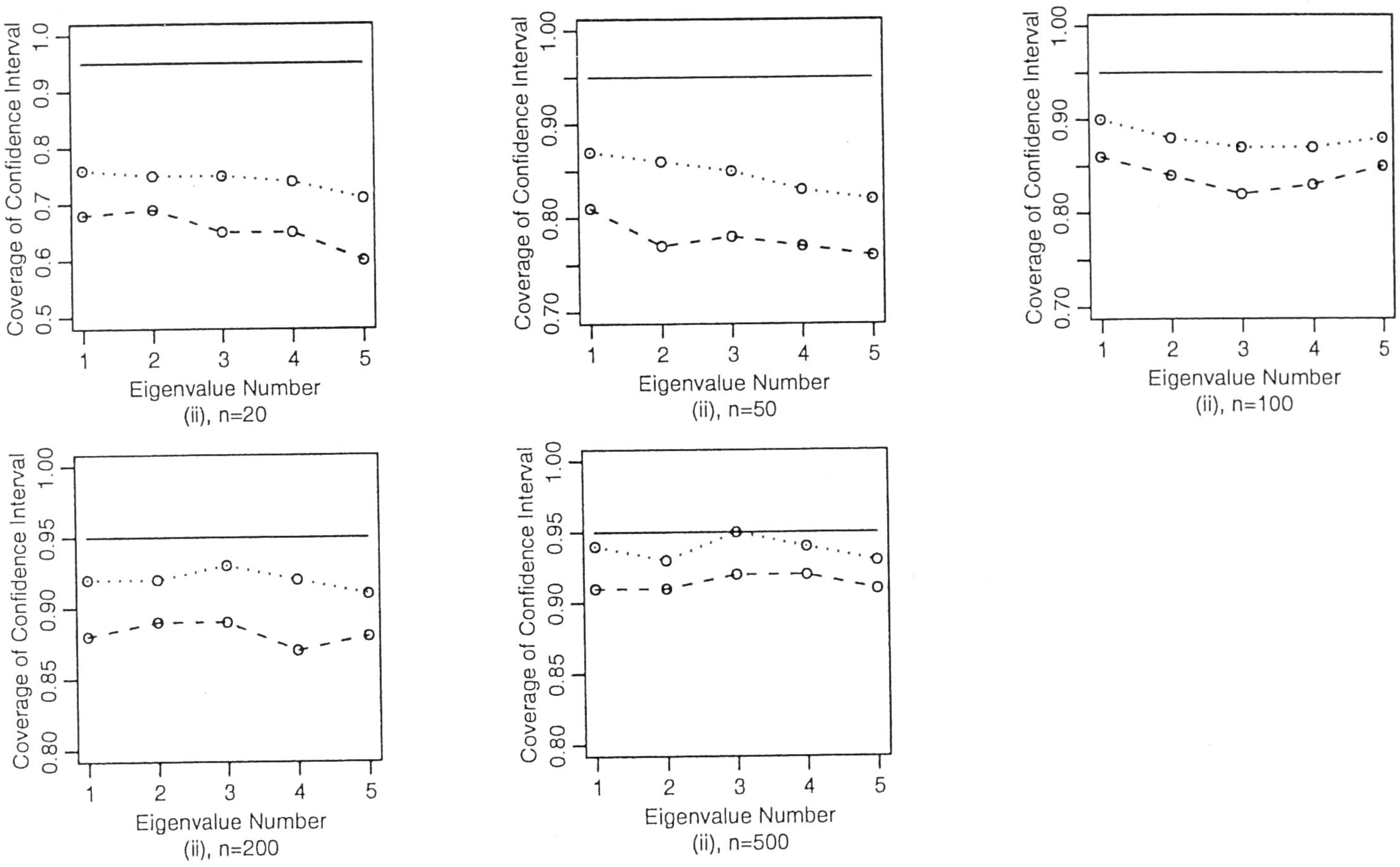

Figure 5.29: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (ii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

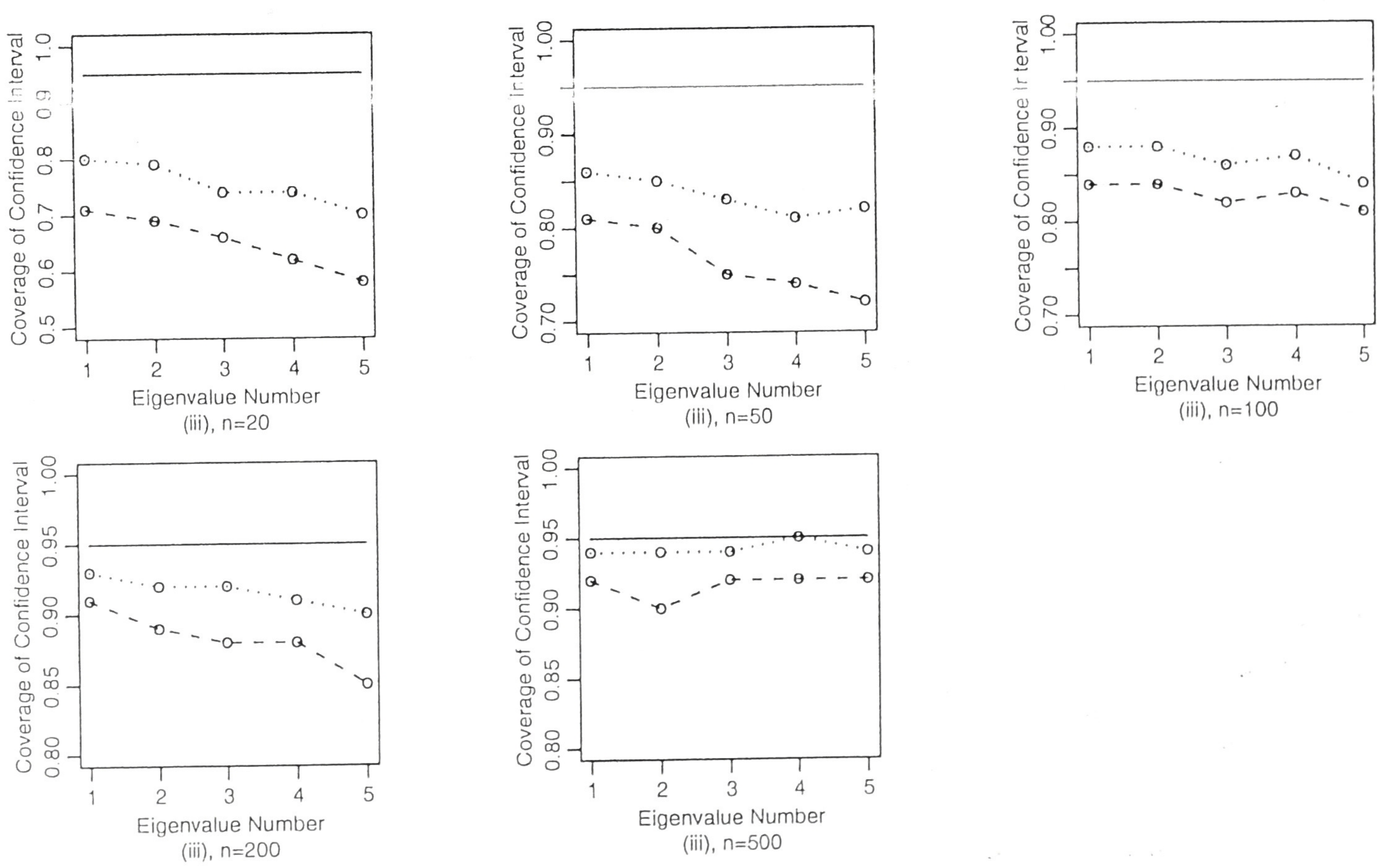

Figure 5.30: Comparison of coverages of two-sided, symmetric bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (iii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

5.10 Two-sided, Equal-tailed Confidence Interval for θ_j, Gaussian

Equal-tailed confidence intervals are appropriate in many circumstances. The asymmetry of an equal-tailed confidence interval can convey important information about our uncertainty as to the location of the true parameter value, and it is not prudent to ignore that information.

Figures 5.31–5.33 show coverage of bootstrap two-sided, equal-tailed confidence intervals for the first three θ_j individually, when the nominal coverage level is 95% and X is Gaussian. The dashed lines show coverage levels when applying the single bootstrap for constructing confidence intervals for the first three eigenvalues and the dotted lines reveal them after considering a coverage-correction on the confidence bands when double-bootstrap calibration is used. In both simulations we took sample sizes $n = 20, 50, 100, 200$ and 500, and obtained coverages for those sample size under the three models (i), (ii) and (iii).

Generally, confidence bands for each θ_j are anti-conservative (have coverage smaller than the nominal level) for small sample sizes, and coverage accuracy of the confidence intervals almost always decreases as j increases. However, it improves as n increases, and enjoy even further improvement after using the double bootstrap. Although the gap between the two lines becomes smaller as n increases, coverage accuracy is improved by calibrating the bands by the double bootstrap.

Under model (i), while the single bootstrap coverage of θ_1 is 0.80, 0.89 and 0.91, respectively for sample sizes 20, 50 and 100, coverage-correction using the double bootstrap substantially increases them to 0.90, 0.93 and 0.95 for those sample sizes. Moreover, one can see that coverages of the three eigenvalues is 0.95 when $n = 200$, something which could not be obtained from the single bootstrap even for $n = 500$. Using the single bootstrap for $n \leq 200$ causes coverage error 0.01–0.18; however more than half of the deficit in the coverage is removed by double-bootstrap calibration, decreasing the error to 0.00–0.08.

Considering model (ii), double-bootstrap calibration also causes coverages to move towards the nominal level, 0.95. It removes about two thirds of the error 0.00–0.17, obtained by applying single-bootstrap, reducing the error to 0.00–0.06. For example, coverage of θ_2 interval for sample sizes 20, 50 and 100 are 0.80, 0.90 and 0.91, respectively, in the case of single-bootstrap application. However, they increase to 0.89, 0.95 and 0.95 after being calibrated by the double bootstrap.

With model (iii), when $n \leq 200$, constructing confidence bands of θ_j using the single bootstrap causes a coverage error 0.01–0.16, but it decreases to 0.00–0.05 after double-bootstrap calibration on the bands. For example, while single-bootstrap coverage levels of θ_1 and θ_2 intervals are 0.89 for $n = 50$ and 0.91 for $n = 100$, they are 0.93 and 0.95 after double-bootstrap calibration of the confidence bands.

5.11 Two sided Equal-tailed Confidence Interval for θ_j, Non-Gaussian

Graphs 5.34–5.36 show coverage of bootstrap two sided, equal-tailed confidence intervals for the first three θ_j when the nominal coverage level is $1-\alpha = 0.95$, and X is generated from the Non-Gaussian process. The dashed lines show coverages of confidence intervals constructed by the single bootstrap, and dotted lines reveal those calibrated by the double bootstrap methods. General features of the results are as follows:

Coverages are anti-conservative, i.e. they fall below 0.95, the nominal level, but move towards it as n increases. Furthermore, coverages obtained from the single bootstrap are dominated by their counterparts, resulting from the double bootstrap calibration of bands; the gap between the two, however, becomes smaller as n increases; and both usually decline as j increases.

Simulating from the Non-Gaussian process under model (i) causes a coverage error 0.04–0.35 when constructing confidence intervals by the single bootstrap. However, coverage-correction using the double bootstrap removes about half of the deficit in the coverage, decreasing the error to 0.00–0.20. This means that using double-bootstrap calibration substantially improves performance. For example, coverage level of the single bootstrap confidence intervals for θ_1 when $n = 50$ is 0.79, whereas it is 0.87 after double-bootstrap calibration. It is also 0.90 when $n = 100$ for the double bootstrap construction of confidence intervals for θ_1, but only 0.82 for its single-bootstrap one.

In the model (ii), using double-bootstrap calibration decreases the coverage error from 0.02–0.35 to 0.00–0.21. Specifically, coverage of single-bootstrap confidence intervals for θ_2 is 0.80 when $n = 50$, but improves to 0.89 after calibrating bands by double-bootstrap. This is the case when $n = 100$, increasing it from 0.85 to 0.92.

Regarding model (iii), coverage error from single-bootstrap usage is 0.03–0.31 but declines to 0.00–0.21 by using double-bootstrap calibration of the confidence intervals. While coverage of single-bootstrap bands for $\theta_1, \theta_2, \theta_3$ are, respectively, 0.82, 0.83, 0.82 when $n = 100$, they are 0.90 for the three eigenvalues after calibrating the confidence intervals by the double bootstrap.

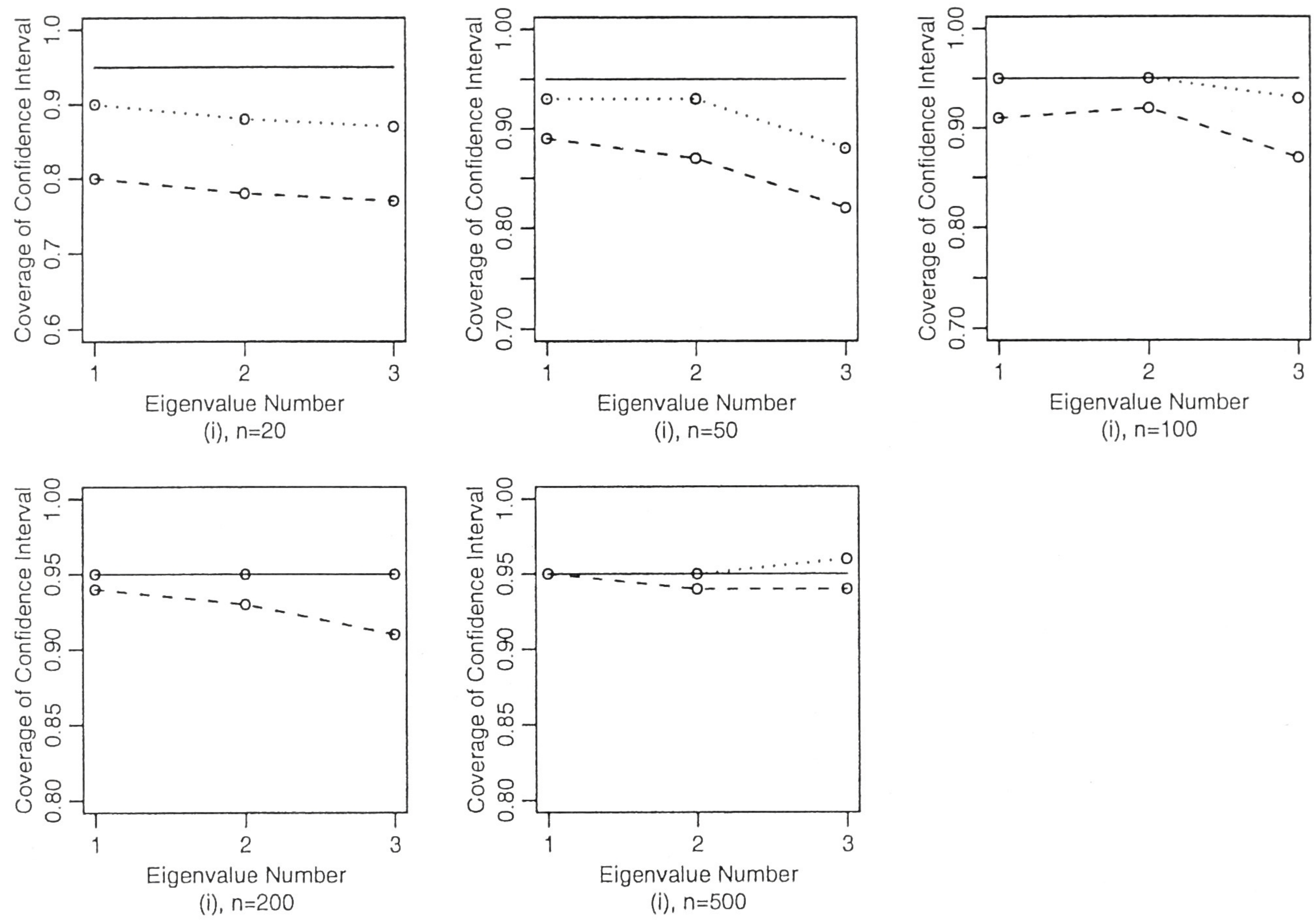

Figure 5.31: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (i). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1-\alpha=0.95$.

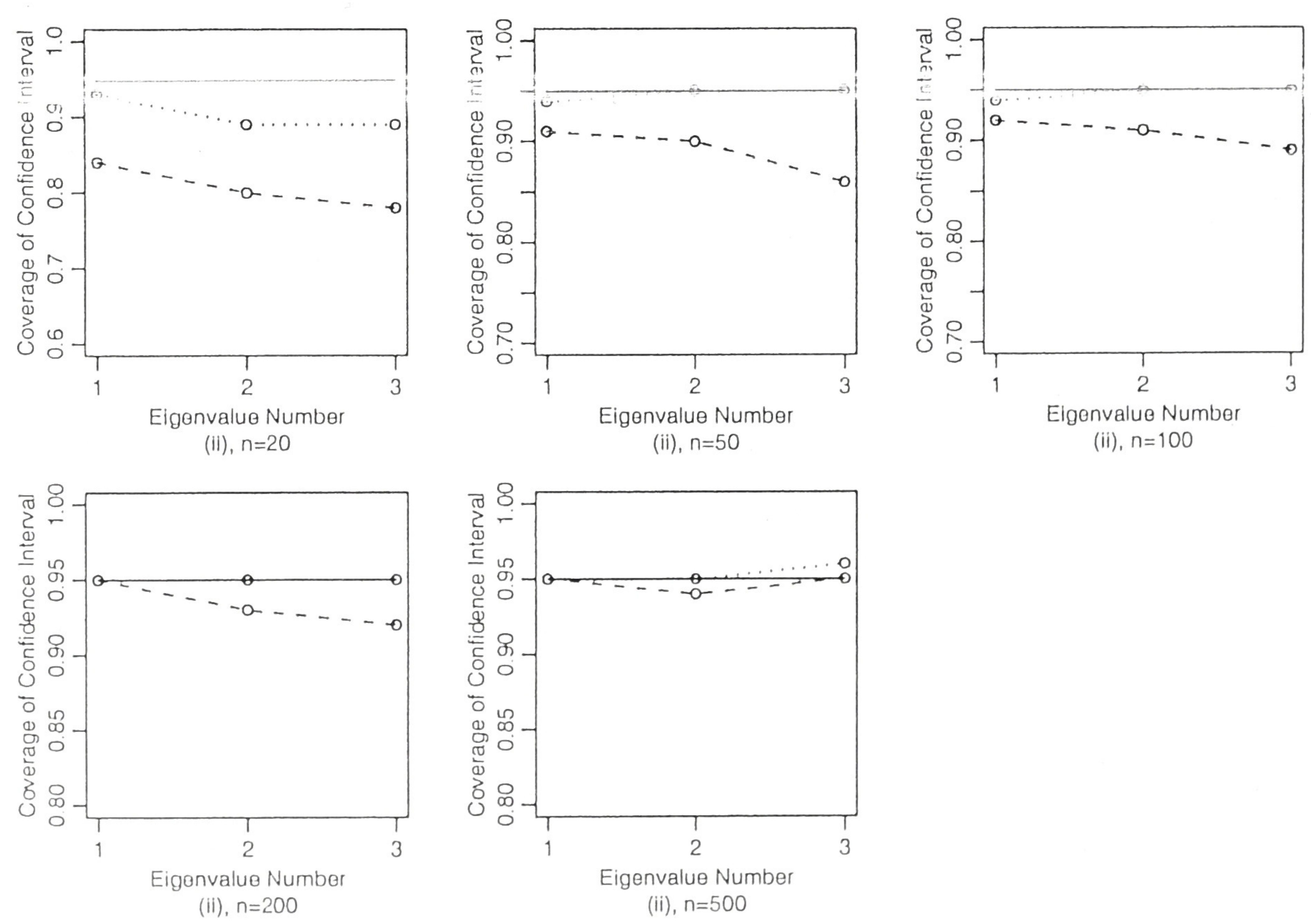

Figure 5.32: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (ii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1-\alpha=0.95$.

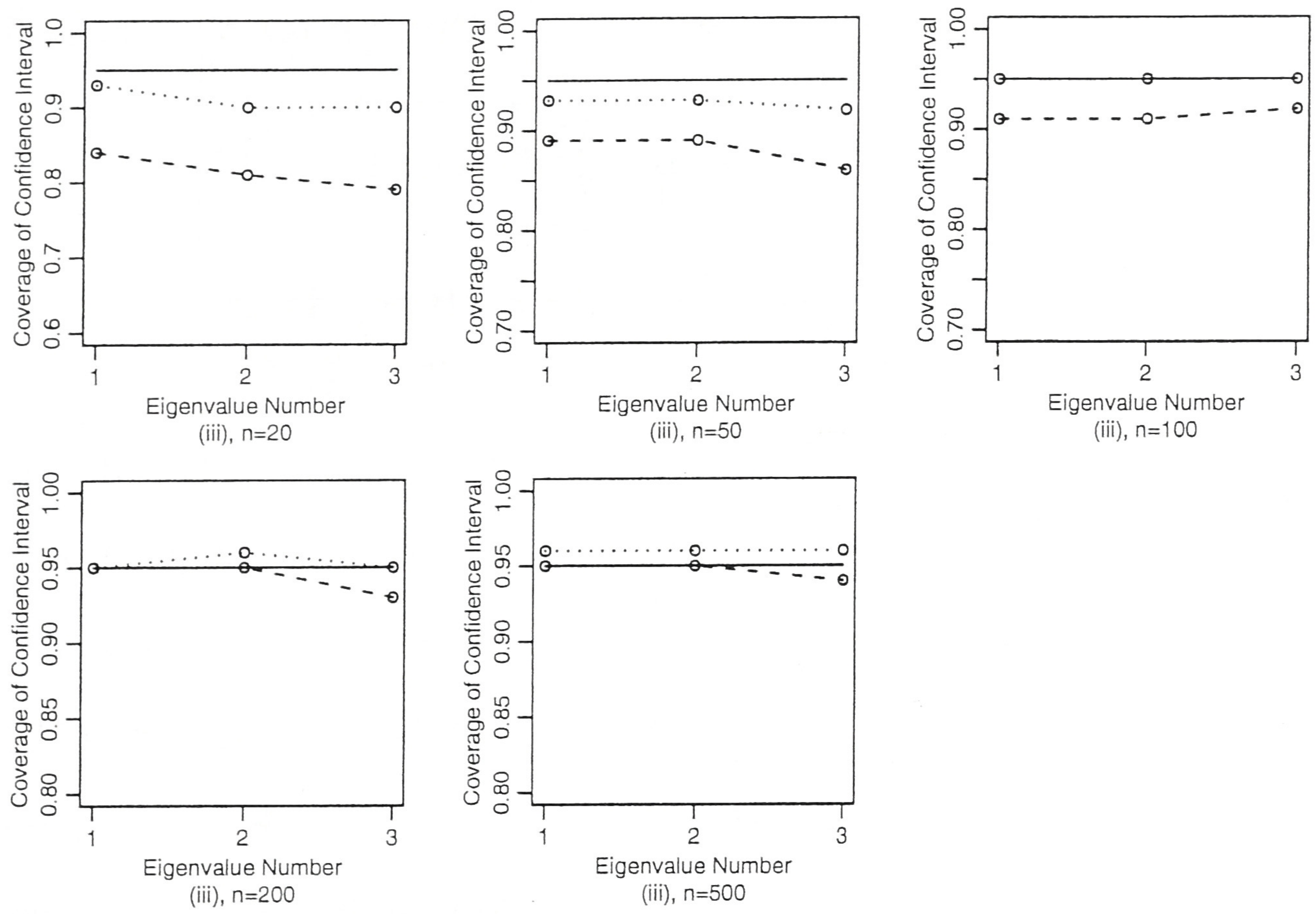

Figure 5.33: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Gaussian process under model (iii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

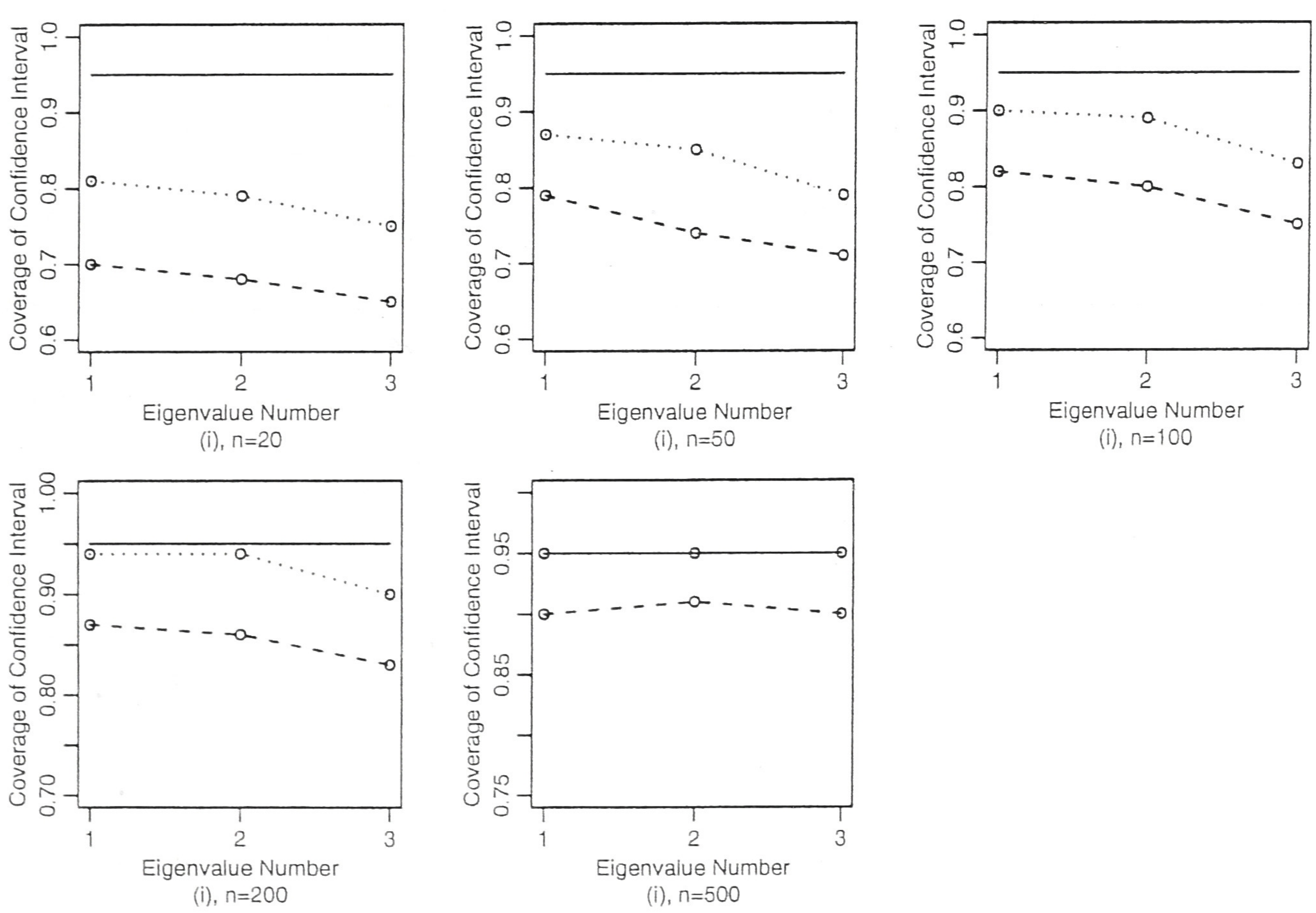

Figure 5.34: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (i). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

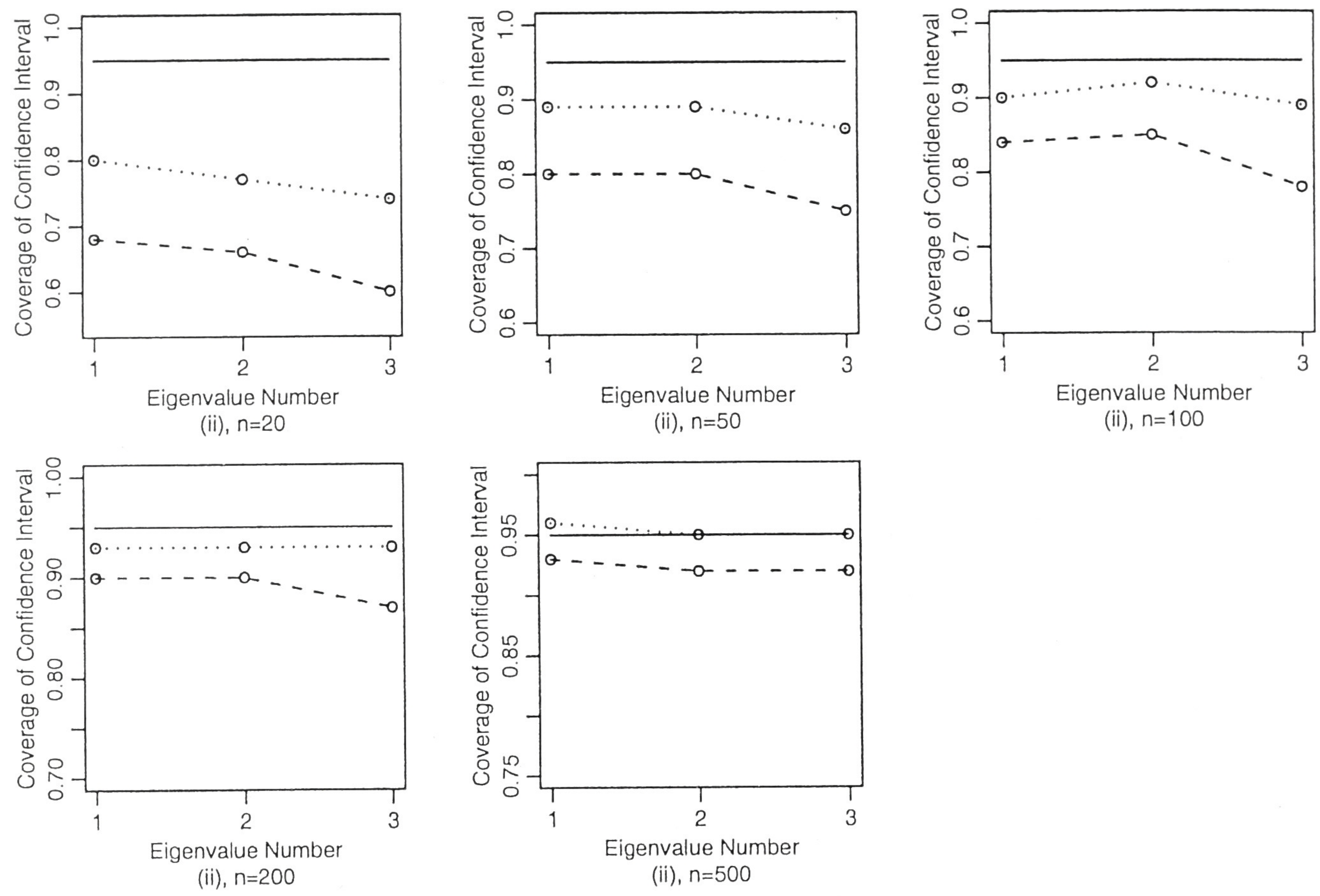

Figure 5.35: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (ii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

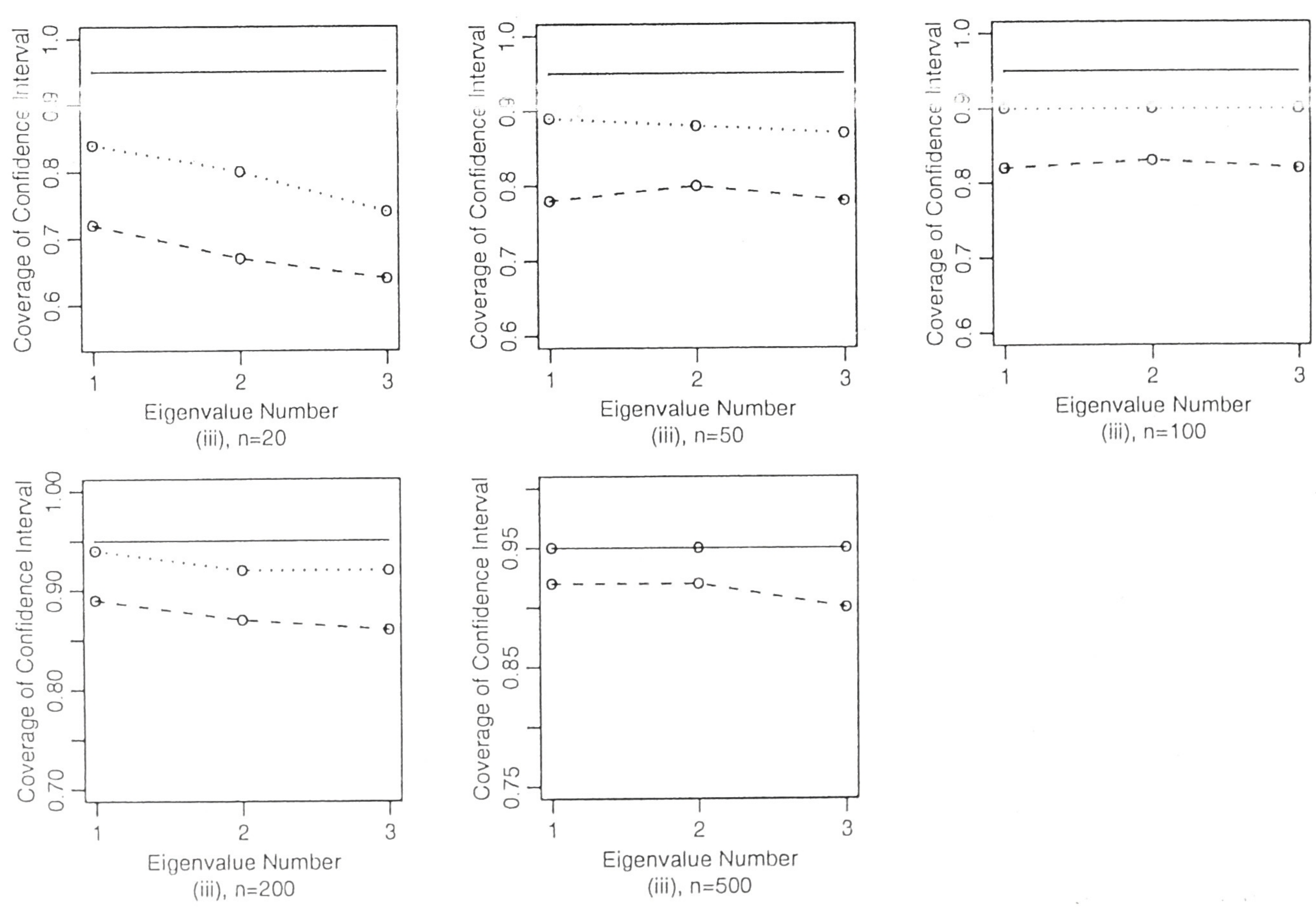

Figure 5.36: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j, when generating $X(t)$ from the Non-Gaussian process under model (iii). The dashed lines show the coverages obtained from the single bootstrap, and the dotted lines reveal coverages from the double bootstrap. The nominal coverage was $1 - \alpha = 0.95$.

5.12 Comparison of coverages of two sided equal-tailed confidence interval for θ_j, for the Gaussian and Non-Gaussian process

Graphs 5.37–5.39 show coverage levels of two sided, equal tailed confidence intervals for θ_j when the nominal level is 95%. These were obtained for the first three eigenvalues, under models (i), (ii) and (iii), and for sample sizes 20,50,100,200 and 500, when X is simulated from a Gaussian or Non-Gaussian process. For each model, each θ_j and sample size, the dashed lines show coverage levels obtained from the Non-Gaussian, and the dotted lines reveal coverages when simulating from the Gaussian process.

As the panels show, generally, the coverages for small sample sizes fall below the nominal coverage level, 95%, with both Gaussian and Non-Gaussian. For each sample size, coverages obtained by the Non-Gaussian process are dominated by their Gaussian counterparts although the gap between the two becomes smaller as n increases. Furthermore, although in both situations, coverage accuracy improves as n increases, coverages resulting from the Non-Gaussian process, are still below 95% (the nominal level) even for large sample sizes. It can also been seen that for both situations, coverage accuracy almost always declines as j increases.

Under model (i) and the Gaussian process, coverage level of confidence interval for θ_1 is 0.80, 0.89 and 0.91, respectively, for $n = 20, 50$ and 100, but if we alter the distribution of ξ_j to the centered exponential, then its coverage level declines to 0.70, 0.79 and 0.82 for those sample sizes.

Regarding model (ii), it can be seen that coverage level for θ_1 is 0.84, 0.91 and 0.92, respectively, for sample sizes 20, 50 and 100 when the Gaussian process is considered. However, they decrease under the Non-Gaussian process to 0.68, 0.80

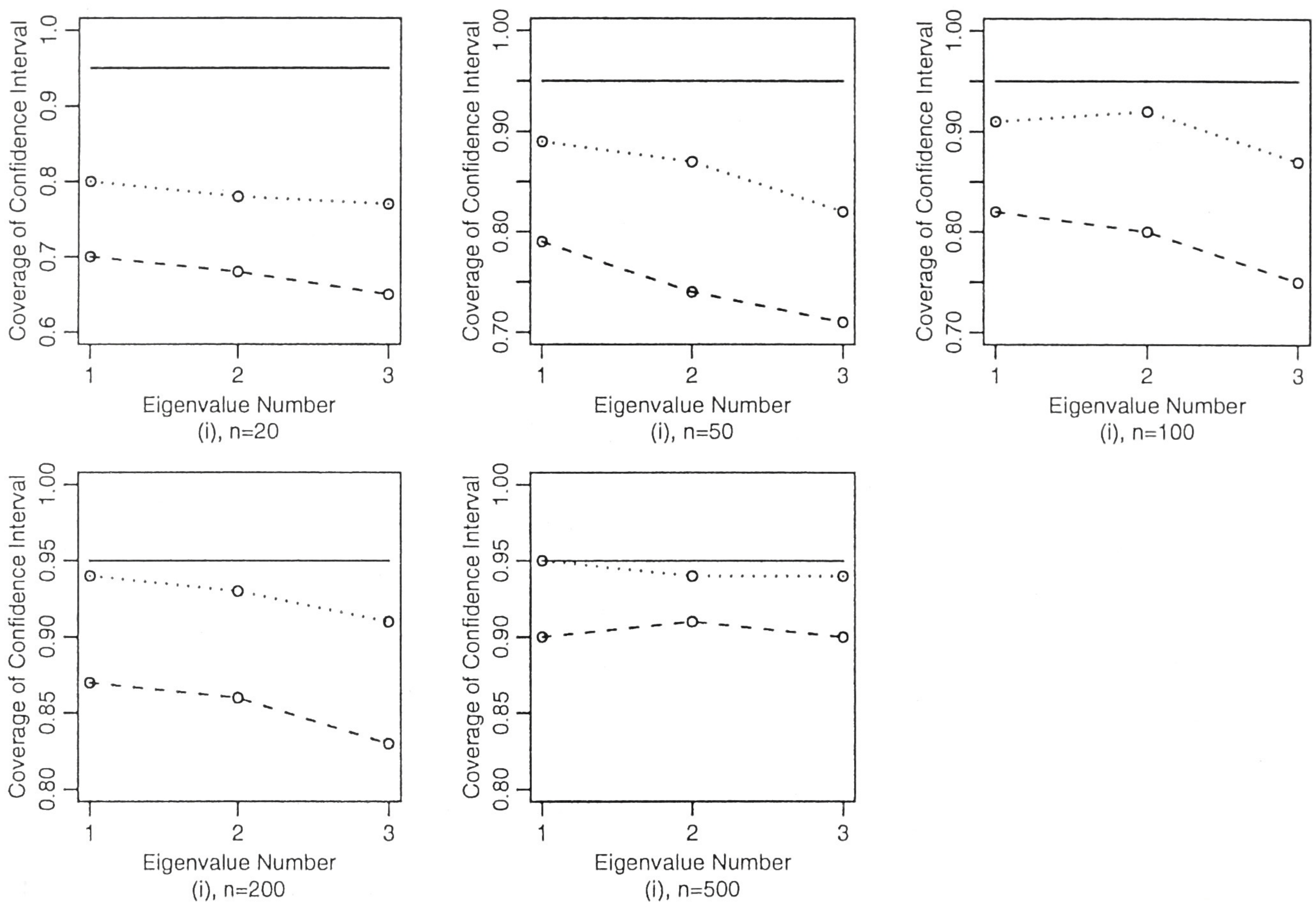

Figure 5.37: Comparison of coverages of two sided, equal-tailed, bootstrap confidence bands for θ_j when generating $X(t)$ from a Gaussian or Non-Gaussian processes under model (i). The dashed lines show the coverages for the Gaussian process, and the dotted lines reveal coverages for the Non-Gaussian process. The nominal coverage was $1 - \alpha = 0.95$.

and 0.84 for those sample sizes. For the Non-Gaussian situation, the coverages for θ_2 are 0.66, 0.80 and 0.85, respectively, for $n = 20$, 50 and 100, but they increase to 0.80, 0.90 and 0.91 for these sample sizes when altering the distribution to the Gaussian.

Considering model (iii), we see that under the Gaussian situation, coverages for θ_1 are 0.84, 0.89, 0.91 and 0.95 for $n = 20, 50, 100$ and 200, respectively. However, like the two models (i) and (ii), it declines under the Non-Gaussian process to 0.72, 0.78, 0.82 and 0.89 for these sample sizes. For θ_2 also coverages obtained from the Gaussian situation, continue to dominate their counterpart, resulting from the Non-Gaussian process. For example, for the former they are 0.81, 0.89 and 0.91 for $n = 20, 50, 100$ respectively, and for the latter only 0.67, 0.80 and 0.83 for these sample sizes, respectively.

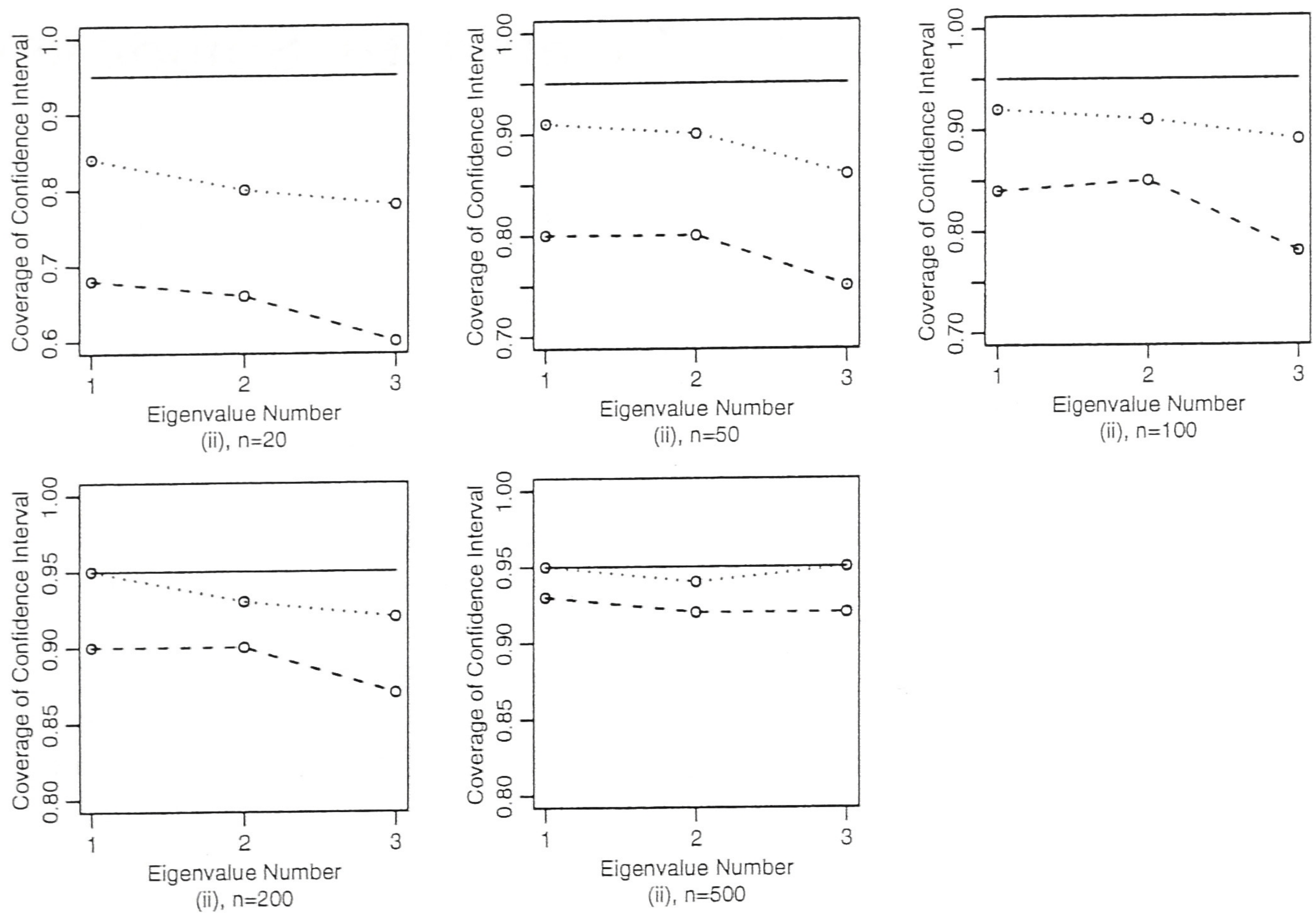

Figure 5.38: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j when generating $X(t)$ from a Gaussian or Non-Gaussian processes under model (ii). The dashed lines show the coverages for the Gaussian process, and the dotted lines reveal coverages for the Non-Gaussian process. The nominal coverage was $1 - \alpha = 0.95$.

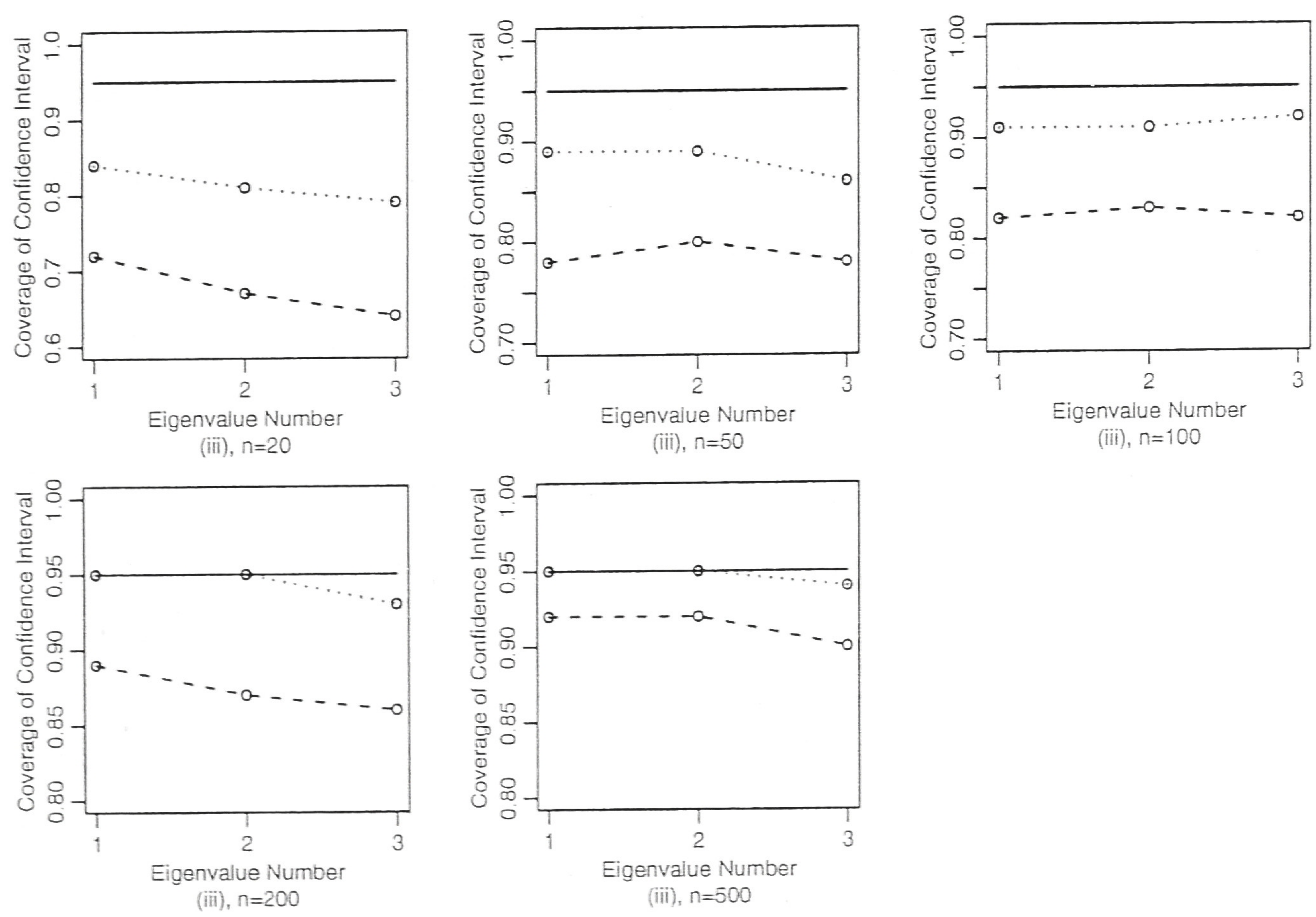

Figure 5.39: Comparison of coverages of two-sided, equal-tailed, bootstrap confidence bands for θ_j when generating $X(t)$ from a Gaussian or Non-Gaussian processes under model (iii). The dashed lines show the coverages for the Gaussian process, and the dotted lines reveal coverages for the Non-Gaussian process. The nominal coverage was $1 - \alpha = 0.95$.

5.13 Numerical Results Related to Smoothed FPCA

5.13.1 Introduction

In this Section we investigate the performance of functional principal component analysis (FPCA) when considering smoothing. We propose a new method of smoothing and study its consistency under suitable conditions. We also compare the effectiveness of the new method for estimating eigenfunctions with that of the non-smoothing method as well as another existing method of smoothing in terms of the mean integrated squared error (MISE). The new idea is based on considering another parameter with the smoothing parameter proposed by Silverman (1996). The numerical results show that adding the new parameter improves the performance of the estimator considerably towards having smaller error, compared with the situation in which we only consider the smoothing parameter itself.

Silverman (1996) proposed an approach, in which the roughness penalty is incorporated in the orthonormality constraints imposed by the Sobolev norm. Let $\mathcal{S}$ be the space of functions with square-integrable second derivative on $\mathcal{I}$, and D^2 be the differential operator of order two on $\mathcal{S}$. We define the bilinear form $[f, g] = \langle D^2 f, D^2 g\rangle = \int_{\mathcal{I}} f'' g''$ for each f, $g \in \mathcal{S}$ (recall that $\langle .,. \rangle$ is the usual L_2 norm). Here, the roughness penalty is incorporated in the orthonormality constraint, instead of penalizing the sample variance of a principal component. This can be done through defining the new inner product

$$\langle f, g\rangle_\alpha = \langle f, g\rangle + \alpha\,[f, g]\,, \tag{5.16}$$

where $\alpha \geq 0$. The inner product $\langle .,. \rangle_\alpha$ is a "slight generalisation" of standard Sobolov inner products (Adams, 1975) with corresponding squared norm $||.||_\alpha^2 = \langle .,. \rangle_\alpha$. Then the problem is to find a maximum of:

$$\frac{\iint_{\mathcal{I}^2} \widehat{K}(u,v)\, \phi(u)\, \phi(v)\, du\, dv}{||\phi||^2 + \alpha||D^2\phi||^2}. \tag{5.17}$$

Silverman (1996) showed that under appropriate conditions the estimates $\hat{\theta}_j$ and $\widehat{\psi}_j$ are consistent. Using the asymptotic expansions for $\hat{\theta}_j$ and $\widehat{\psi}_j$, the author investigated the advantages of the smoothed estimates in terms of the MISE. Furthermore, the choice of smoothing parameter by cross-validation was discussed by Silverman (1996).

5.13.2 Smoothing by Two Parameters; a Generalisation of Silverman's Method

In this Section we first introduce our new idea, and then investigate some properties of smoothed FPCA obtained by using this method. Our new idea is to maximize:

$$\frac{\iint_{\mathcal{I}^2} \widehat{K}(u,v)\, \phi(u)\, \phi(v)\, du\, dv \,+ \alpha_0 \,\langle \phi, \phi \rangle}{||\phi||^2 + \alpha_1||D^2\phi||^2}, \tag{5.18}$$

where α_1 is the smoothing parameter introduced in the previous Section, and α_0 is a positive real number.

5.13.3 Consistency of eigenvalues and eigenfunctions estimators

Define $\psi_\ell^* = \frac{\tilde{\psi}_\ell}{||\tilde{\psi}_\ell||}$ for each $\ell \geq 1$, and $P_\psi x = \langle \psi, \psi \rangle^{-1} \langle \psi, x \rangle\, \psi$ as the projection onto the subspace generated by ψ.

Assume that:

- $\theta_1 \geq \theta_2 \geq \cdots > 0$, i.e. the covariance operator K is strictly positive-definite, and $\int K(t,t) = \sum_{i=1} \theta_i < \infty$. Without loss of generality, we assume that the eigenvalues θ_ℓ have multiplicity one, which means, $\theta_1 > \theta_2 > \cdots > 0$.

- Each of the eigenfunctions ψ_ℓ belongs to $\mathcal{S}$, the space of smooth functions, in which $\int \left(\psi_\ell''\right)^2 < \infty$.

- α_0 and $\alpha_1 \to 0$ as $n \to \infty$

Theorem 5.1. Under the above assumptions, for each $\ell \geq 1$ and with probability 1 (almost surely),

$$\tilde{\theta}_\ell \to \theta_\ell \quad \text{as } n \to \infty \tag{5.19}$$

and

$$\langle \psi_\ell^*, \psi_\ell \rangle^2 \to 1 \text{ as } n \to \infty \tag{5.20}$$

Equivalently, we can write result (5.20) as

$$\|\psi_\ell^* - \psi_\ell\|^2 = 2\left(1 - \langle \psi_\ell^*, \psi_\ell \rangle\right) \to 0 \text{ as } n \to \infty\,, \tag{5.21}$$

if we choose the sign of the ψ^* properly. This also results in

$$\|P_{\tilde{\psi}_\ell} - P_{\psi_\ell}\|^2 \longrightarrow 0\,, \text{with probability 1 as } n \to \infty\,, \tag{5.22}$$

Proof: The proof is similar to the proof given by Theorem 1 of Silverman (1996), and it is done by induction. All we need is to replace K and $\widehat{K}$ by $\alpha_0 I + K$ and $\alpha_0 I + \widehat{K}$, respectively, in the Silverman's proof.

5.13.4 Advantages of the New Idea for Smoothing

Now we are going to investigate the effect of using α_0 with α_1. Let R be the differential operator of order two defined on the space $\mathcal{S}$, i.e $R = D^2$. The operator R is self-adjoint subject to some periodic boundary conditions. Let $\mathcal{V}$ be the space of functions g such that g and its first three derivatives are absolutely continuous on $\mathcal{I}$, and its fourth derivative belongs to $L_2(\mathcal{I})$. We call $\mathcal{V}$ the space of "very smooth function". Define the fourth-derivative operator $Q = R^2$ in the space $\mathcal{V}$, then we have $[f, g] = \langle R\,f, R\,g\rangle = \langle f, Q\,g\rangle$, for each $f \in \mathcal{S}$ and $g \in \mathcal{V}$. The periodic boundary conditions needed are those by which we can have $\int_{\mathcal{I}} f''\,g'' = \int_{\mathcal{I}} f\,g''''$, where g'''' denotes the fourth derivative of g. It should be mentioned that in the case of non-periodic functions, the space $\mathcal{V}$ can be defined as the space of functions g such that g has square-integrable fourth derivative and its second and third derivatives are zero at the boundaries, by which we still have $\int_{\mathcal{I}} f''\,g'' = \int_{\mathcal{I}} f\,g''''$. We assume that all eigenfunctions of covariance operator K fall in the space of "very smooth" functions $\mathcal{V}$. Therefore, (5.18) is equivalent to the equation:

$$\int_{\mathcal{I}} \left(\widehat{K}(u,v) + \alpha_0\, I(u,v)\right) \tilde{\psi}(u)\, du = \tilde{\theta}\left(I + \alpha_1\, D^4\right)\tilde{\psi}\,, \tag{5.23}$$

where $\tilde{\theta}$ is the maximum value of the ratio in (5.18) obtained when $\phi = \tilde{\psi}$. The problem is reduced to seeking the extremum of the Lagrangian function:

$$\begin{aligned}\int\!\!\int_{\mathcal{I}^2} \left(\widehat{K}(u,v) + \alpha_0\, I(u,v)\right) \phi_\ell(u)\, \phi_\ell(v)\, du\, dv \,+\, \lambda_0\big(\langle \phi_\ell, \phi_\ell\rangle + \alpha_1\,[\phi_\ell, \phi_\ell] - 1\big)\\ + 2\sum_{r=1}^{\ell}\sum_{s=1}^{r-1} \lambda_{rs}\left\{\int \phi_r\,\phi_s + \alpha_1\,[\phi_r, \phi_s]\right\},\end{aligned} \tag{5.24}$$

where λ_0 and λ_{rs} are Lagrange multipliers. Inserting $\phi_\ell = \sum_{j=1}^{\infty} a_{\ell j}\, \psi_j$ into the above equation results in:

$$\sum_{j=1}^{\infty} \left(a_{\ell j}^2\,(\theta_j + \alpha_0) + \lambda_0 \right) - \lambda_0 + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \left[\lambda_0\, \alpha_1 \int \psi_j'' \psi_k'' + n^{-1/2} \left(\int Z\, \psi_j\, \psi_k \right) \right]$$
$$+ 2 \sum_{r=1}^{\ell} \sum_{s=1}^{r-1} \lambda_{rs} \left[\sum_{j=1}^{\infty} a_{rj}\, a_{sj} + \alpha_1 \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} a_{rj}\, a_{sj} \left(\int \psi_j'' \psi_k'' \right) \right], \qquad (5.25)$$

where $Z = n^{-1/2}\,(\widehat{K} - K)$ and ψ_j'' is the second derivative of ψ_j. Differentiating with respect to $a_{\ell j}$ and equating to zero for an extremum, we obtain that

$$a_{\ell j}\,(\theta_j + \alpha_0 + \lambda_0) + \sum_{k=1}^{\infty} a_{\ell k} \left[\lambda_0\, \alpha_1 \int \psi_j'' \psi_k'' + n^{-1/2} \left(\int Z\, \psi_j\, \psi_k \right) \right]$$
$$+ \sum_{s=1}^{\ell-1} \lambda_{\ell s} \left[a_{sj} + \alpha_1 \sum_{k=1}^{\infty} a_{sj} \left(\int \psi_j'' \psi_k'' \right) \right] = 0\,. \qquad (5.26)$$

Also,

$$a_{rj} = \delta_{rj} + n^{-1/2}\,(1 - \delta_{rj})\, c_{rj}^{(1)} + \alpha_1 \left((1 - \delta_{rj})\, c_{rj}^{(2)} - \delta_{rj}\,(\tfrac{1}{2}) \int (\psi_\ell'')^2 \right)$$
$$+ n^{-1/2} \alpha_0\, a_{rj}^{(01)} + n^{-1/2}\, \alpha_1\, a_{rj}^{(12)} + \alpha_0\, \alpha_1\, a_{rj}^{(02)} + n^{-1}\, a_{rj}^{(11)} + \alpha_1^2\, a_{rj}^{(22)}$$
$$+ O_p(n^{-3/2},\, \alpha_0^2\, n^{-1/2},\, \alpha_1^2\, n^{-1/2},\, \alpha_0^2\, \alpha_1,\, \alpha_0\, \alpha_1^2,\, \alpha_0^3)\,. \qquad (5.27)$$

After doing the algebraic calculations, the results are:

$$c_{\ell j}^{(1)} = (\theta_\ell - \theta_j)^{-1} \left(\int Z\, \psi_j\, \psi_\ell \right), \qquad c_{\ell j}^{(2)} = (\theta_j - \theta_\ell)^{-1}\, \theta_\ell \left(\int \psi_j'' \psi_\ell'' \right),$$
$$a_{\ell j}^{(01)} = (\theta_\ell - \theta_j)^{-2} \left[-\, \theta_\ell \left(\int \psi_j'' \psi_\ell'' \right) - \left(\int Z\, \psi_j\, \psi_\ell \right) \right],$$

$$
\begin{aligned}
a_{\ell j}^{(12)} &= (\theta_\ell - \theta_j)^{-1} \biggl\{ \Bigl(- \int Z\, \psi_\ell \psi_\ell \Bigr) \Bigl(\int \psi_j'' \psi_\ell'' \Bigr) \\
&\quad + \theta_\ell \sum_{k\,:\,k\neq\ell} (\theta_k - \theta_\ell)^{-1} \Bigl(\int Z\, \psi_k\, \psi_\ell \Bigr) \Bigl(\int \psi_j'' \psi_k'' \Bigr) \\
&\quad + \theta_\ell \sum_{k\,:\,k\neq\ell} (\theta_k - \theta_\ell)^{-1} \Bigl(\int Z\, \psi_k\, \psi_\ell \Bigr) \Bigl(\int \psi_\ell'' \psi_k'' \Bigr) - \tfrac{1}{2} \Bigl(\int Z\, \psi_j\, \psi_\ell \Bigr) \Bigl(\int (\psi_\ell'')^2 \Bigr) \\
&\quad + (\theta_\ell - \theta_j)^{-1}\, \theta_\ell \biggl[\Bigl(\int Z\, \psi_\ell \psi_\ell \Bigr) \Bigl(\int \psi_j'' \psi_\ell'' \Bigr) + \Bigl(\int Z\, \psi_j\, \psi_\ell \Bigr) \Bigl(\int (\psi_\ell'')^2 \Bigr) \biggr\},
\end{aligned}
$$

$$
\begin{aligned}
a_{\ell j}^{(02)} &= -(\theta_\ell - \theta_j)^{-1} \Bigl(\int \psi_j'' \psi_\ell'' \Bigr), \\
a_{\ell j}^{(11)} &= (\theta_\ell - \theta_j)^{-1} \biggl\{ \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_k)^{-1} \Bigl(\int Z\, \psi_k\, \psi_\ell \Bigr) \Bigl(\int Z\, \psi_j\, \psi_k \Bigr) \\
&\qquad - (\theta_\ell - \theta_j)^{-1} \Bigl(\int Z\, \psi_j\, \psi_\ell \Bigr) \Bigl(\int Z\, \psi_\ell \psi_\ell \Bigr) \biggr\}, \\
a_{\ell j}^{(22)} &= (\theta_\ell - \theta_j)^{-1} \biggl\{ \theta_\ell^2 \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_k)^{-1} \Bigl(\int \psi_k'' \psi_\ell'' \Bigr) \Bigl(\int \psi_j'' \psi_k'' \Bigr) \\
&\qquad + \tfrac{3}{2}\, \theta_\ell \Bigl(\int (\psi_\ell'')^2 \Bigr) \Bigl(\int \psi_j'' \psi_\ell'' \Bigr) \biggr\} - (\theta_\ell - \theta_j)^{-1}\, \theta_\ell^2 \Bigl(\int (\psi_\ell'')^2 \Bigr) \Bigl(\int \psi_j'' \psi_\ell'' \Bigr) \biggr\}.
\end{aligned}
$$

Therefore, $a_{\ell j}$ can be obtained by substituting the above terms into (5.27).

As a result, we have:

$$
\begin{aligned}
\tilde{\psi}_\ell - \psi_\ell &= n^{-1/2}\, \varphi^{(1)} + \alpha_1\, \varphi^{(2)} + n^{-1/2}\, \alpha_0\, \varphi^{(01)} + n^{-1/2}\, \alpha_1\, \varphi^{(12)} + \alpha_0\, \alpha_1\, \varphi^{(02)} \\
&\quad + n^{-1}\, \varphi^{(11)} + \alpha_1^2\, \varphi^{(22)} + O_p(n^{-3/2},\, \alpha_0^2\, n^{-1/2},\, \alpha_1^2\, n^{-1/2},\, \alpha_0^2\, \alpha_1,\, \alpha_0\, \alpha_1^2,\, \alpha_0^3)\,,
\end{aligned}
\tag{5.28}
$$

where where,

$$
\begin{aligned}
\varphi^{(1)} &= \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-1} \Bigl(\int Z\, \psi_j\, \psi_\ell \Bigr)\, \psi_j\,, \\
\varphi^{(2)} &= \biggl\{ \theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_j - \theta_\ell)^{-1} \Bigl(\int \psi_j'' \psi_\ell'' \Bigr)\, \psi_j - \tfrac{1}{2} \Bigl(\int (\psi_\ell'')^2 \Bigr)\, \psi_\ell \biggr\},
\end{aligned}
$$

$$
\begin{aligned}
\varphi^{(11)} = \Bigg\{ & \sum_{j\,:\,j\neq\ell}\ \sum_{k\,:\,k\neq\ell} (\theta_\ell-\theta_j)^{-1}\,(\theta_\ell-\theta_k)^{-1}\psi_j\,\left(\int Z\,\psi_k\,\psi_\ell\right)\left(\int Z\,\psi_k\,\psi_j\right) \\
& - \tfrac{1}{2}\,\psi_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\left(\int Z\,\psi_j\,\psi_\ell\right)^2 \\
& - \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\psi_j\left(\int Z\,\psi_j\,\psi_\ell\right)\left(\int Z\,\psi_\ell\,\psi_\ell\right)\Bigg\},
\end{aligned}
\tag{5.29}
$$

$$
\begin{aligned}
\varphi^{(01)} &= \left\{ -\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_j-\theta_\ell)^{-2}\left(\int \psi_j''\psi_\ell''\right)\psi_j - \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-2}\left(\int Z\,\psi_j\,\psi_\ell\right)\psi_j \right\}, \\
\varphi^{(02)} &= \left\{ \left(\int (\psi_\ell'')^2\right)\psi_\ell - \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-1}\left(\int \psi_j''\psi_\ell''\right)\psi_j \right\},
\end{aligned}
\tag{5.30}
$$

$$
\begin{aligned}
\varphi^{(12)} = \Bigg\{ & \tfrac{1}{2}\left(\int(\psi_\ell'')^2\right)\textstyle\sum_{j\,:\,j\neq\ell}(\theta_j-\theta_\ell)^{-1}\left(\int Z\,\psi_j\,\psi_\ell\right)\psi_j \\
& + \theta_\ell \sum_{j\,:\,j\neq\ell}\ \sum_{k\,:\,k\neq\ell}(\theta_\ell-\theta_j)^{-1}\,(\theta_k-\theta_\ell)^{-1}\psi_k\Bigg[\left(\int Z\,\psi_j\,\psi_k\right)\left(\int \psi_\ell''\psi_k''\right) \\
& \qquad\qquad + \left(\int Z\,\psi_k\,\psi_j\right)\left(\int \psi_j''\psi_\ell''\right)\Bigg] \\
& + \theta_\ell \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-2}\,\psi_j\left(\int Z\,\psi_\ell\,\psi_\ell\right)\left(\int \psi_\ell''\psi_j''\right) \\
& - \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-1}\,\psi_j\left(\int Z\,\psi_\ell\,\psi_\ell\right)\left(\int \psi_\ell''\psi_j''\right) \\
& - \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-1}\,\psi_\ell\left(\int Z\,\psi_j\,\psi_\ell\right)\left(\int \psi_\ell''\psi_j''\right) \\
& + \theta_\ell \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-2}\,\psi_j\left(\int Z\,\psi_j\,\psi_\ell\right)\left(\int (\psi_\ell'')^2\right) \\
& + \psi_\ell\,\theta_\ell \sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-2}\left(\int Z\,\psi_j\,\psi_\ell\right)\left(\int \psi_\ell''\psi_j''\right)\Bigg\},
\end{aligned}
\tag{5.31}
$$

$$\varphi^{(22)} = \Bigg\{ \tfrac{3}{2}\,\theta_\ell \left(\int (\psi_\ell'')^2 \right) \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-1}\,\psi_j \left(\int \psi_\ell''\psi_j'' \right)$$
$$+ \theta_\ell^2 \sum_{j\,:\,j\neq\ell}\ \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_j)^{-1}\,(\theta_\ell - \theta_k)^{-1}\psi_j \left(\int \psi_\ell''\psi_k'' \right) \left(\int \psi_k''\psi_j'' \right)$$
$$+ \tfrac{3}{8}\,\psi_\ell \left(\int (\psi_\ell'')^2 \right)^2 + \theta_\ell\,\psi_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-1} \left(\int \psi_\ell''\psi_j'' \right)^2$$
$$- \theta_\ell^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-2}\psi_j \left(\int (\psi_\ell'')^2 \right) \left(\int \psi_\ell''\psi_j'' \right)$$
$$- \tfrac{1}{2}\,\theta_\ell^2\,\psi_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-2} \left(\int \psi_\ell''\psi_j'' \right)^2 \Bigg\}. \tag{5.32}$$

It can be seen that

$$\tilde\theta_\ell - \theta_\ell = n^{-1/2} \left(\int Z\,\psi_\ell\,\psi_\ell \right) + \alpha_0 + \alpha_1 \left(-\,\theta_\ell \int (\psi_\ell'')^2 \right) - n^{-1/2}\alpha_0 \left\{ \tfrac{1}{2} \int (\psi_\ell'')^2 \right\}$$
$$- \alpha_0\,\alpha_1 \left(\int (\psi_\ell'')^2 \right) - n^{-1/2}\alpha_1 \Bigg\{ \left(\int (\psi_\ell'')^2 \right) \left(\int Z\,\psi_\ell\,\psi_\ell \right)$$
$$+ 2\,\theta_\ell \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_k)^{-1} \left(\int Z\,\psi_k\,\psi_\ell \right) \left(\int \psi_\ell''\psi_k'' \right) \Bigg\}$$
$$+ \alpha_1^2 \left\{ \theta_\ell \left(\int (\psi_\ell'')^2 \right)^2 - \theta_\ell^2 \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_k)^{-1} \left(\int \psi_k''\psi_\ell'' \right)^2 \right\}$$
$$+ n^{-1} \sum_{k\,:\,k\neq\ell} (\theta_\ell - \theta_k)^{-1} \left(\int Z\,\psi_k\,\psi_\ell \right)^2$$
$$+ O_p(n^{-3/2},\ \alpha_0^2\,n^{-1/2},\ \alpha_1^2\,n^{-1/2},\ \alpha_0^2\,\alpha_1,\ \alpha_0\,\alpha_1^2,\ \alpha_0^3)\,. \tag{5.33}$$

Define $S \equiv \left(I + \alpha_1 D^4\right)^{-1/2}$. If we take, for example, any of the following orthonormal bases

$$\left\{1, \cos(\pi t), \cos(2\pi t), \cdots\right\},\ \left\{\sin(\pi t), \sin(2\pi t), \cdots\right\} \text{ and}$$
$$\left\{1, \sin(\pi t), \cos(\pi t), \sin(2\pi t), \cos(2\pi t), \cdots\right\} \tag{5.34}$$

as a basis, the operator D^4 is not bounded, since $\|D^4\| \geq \sup_{i\geq 1} d_i = \infty$, where the

d_i are the D^4's eigenvalues. However, S is a compact and bounded operator. If we substitute S into the above equation, the result is $S\big(\widehat{K} + \alpha_0\, I\big)S\, S^{-1}\tilde{\psi} = \tilde{\theta}\, S^{-1}\tilde{\psi}$. Now let $\tilde{u} = S^{-1}\tilde{\psi}$, and substitute this term into the previous equation. We have $S\big(\widehat{K} + \alpha_0\, I\big)S\,\tilde{u} = \tilde{\theta}\,\tilde{u}$. Consequently, it is now clear that $\tilde{\theta}$ and $\tilde{u}$ are the eigenvalue and eigenfunction of the operator $S\big(\widehat{K} + \alpha_0\, I\big)S$. As a result, $\tilde{\psi}$ is the maximizer of (5.18), and $\tilde{\theta}$ is the maximum value. Furthermore, $\tilde{\psi}$ can be computed from $\tilde{\psi} = S\tilde{u}$.

Our idea has two advantages:

- In solving equations, say $Ax = 0$, computation is done in such a way that A is changed to a stepwise matrix, called the Reduced Echelon form of A. If the elements of A satisfy

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|\,, \tag{5.35}$$

  then computation towards obtaining its Reduced Echelon form will be faster. Since in numerical work we have to truncate the series involved in calculating $\widehat{K}$, we will have $S(V + \alpha_0 I)S$ if we take α_0 into account. For a positive α_0 the diagonal elements of $S(V + \alpha_0 I)S$ are greater than those of SVS, which is obtained without considering α_0. This will satisfy condition (5.35). Apart from this computational point, adding an adequate α_0 causes $\widehat{K}$ to be more stable. So, the results might be improved.

- In maximization of the ratio $\frac{\int\int_{\mathcal{I}^2} \widehat{K}(u,v)\, \phi(u)\, \phi(v)\, du\, dv}{\|\phi\|^2 + \alpha_1\, \|D^2\phi\|^2}$, adding a multiple of term $\langle \phi, \phi \rangle$, which is well controlled computationally, to the numerator has very good effect in computation, especially when $\widehat{K}$ is not a good operator.

We have

$$
\begin{aligned}
E\tilde{\psi}_\ell = \psi_\ell + \alpha_1\,\varphi^{(2)} + n^{-1/2}\alpha_0 \Big\{ -\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_j-\theta_\ell)^{-2} \Big(\int \psi_j''\psi_\ell'' \Big)\,\psi_j \Big\} + \alpha_0\,\alpha_1\,\varphi^{(02)} \\
+\, n^{-1}\Big\{ -\tfrac{1}{2}\,\theta_\ell\,\psi_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\,\theta_j \Big\} + \alpha_1^2\,\varphi^{(22)} \\
+\, O_p(n^{-3/2},\, \alpha_0^2\,n^{-1/2},\, \alpha_1^2\,n^{-1/2},\, \alpha_0^2\,\alpha_1,\, \alpha_0\,\alpha_1^2,\, \alpha_0^3)\,. \qquad (5.36)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{Bias}(\alpha_0,\alpha_1) \equiv \|E\tilde{\psi}_\ell - \psi_\ell\|^2 \approx \alpha_1^2 \Big\{ \tfrac{1}{4}\Big(\int (\psi_\ell'')^2\Big)^2 + \theta_\ell^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \Big(\int \psi_\ell''\psi_j''\Big)^2 \Big\} \\
+\, \alpha_0\alpha_1^2 \Big\{ -\Big(\int (\psi_\ell'')^2\Big)^2 + 2\,\theta_\ell \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \Big(\int \psi_\ell''\psi_j''\Big)^2 \Big\} \\
+\, n^{-1/2}\alpha_0\alpha_1 \Big\{ 2\,\theta_\ell^2 \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3} \Big(\int \psi_\ell''\psi_j''\Big)^2 \Big\} \\
+\, n^{-1}\alpha_1 \Big\{ \tfrac{1}{2}\,\theta_\ell \Big(\int (\psi_\ell'')^2\Big) \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\,\theta_j \Big\} \\
+\, n^{-1}\,\alpha_0^2 \Big\{ \theta_\ell^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4} \Big(\int \psi_\ell''\psi_j''\Big)^2 \Big\} \\
+\, \alpha_0^2\,\alpha_1^2 \Big\{ \Big(\int (\psi_\ell'')^2\Big)^2 + \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \Big(\int \psi_\ell''\psi_j''\Big)^2 \Big\}. \qquad (5.37)
\end{aligned}
$$

Taking either of the bases in (5.34) as a basis, causes $\left(\int \psi_\ell''\psi_j''\right) = 0$ for j $\neq \ell$. In this case, differentiation of the above bias approximation with respect to α_0, and equating to zero for the minimum, results in $\mathrm{Bias}(\frac{1}{2},\alpha_1) \leq \mathrm{Bias}(0,\alpha_1)$, for all values of α_1.

As regards variance of the estimator we have that

$$\begin{aligned}\mathrm{Var}(\alpha_0,\alpha_1) \equiv E\|\tilde{\psi}_\ell - E\tilde{\psi}_\ell\|^2 \approx E\Big\| n^{-1/2} \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-1}\left(\int Z\,\psi_j\,\psi_\ell\right)\psi_j \\ - n^{-1/2}\alpha_0 \left\{ \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\left(\int Z\,\psi_j\,\psi_\ell\right)\psi_j \right\}\Big\|^2 \\ = n^{-1}\,\theta_\ell \left\{ \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\theta_j - 2\,\alpha_0 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\theta_j \right. \\ \left. + \alpha_0^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4}\theta_j \right\}. \end{aligned} \tag{5.38}$$

This approximation does not depend on the amount of smoothing (α_1) applied. This effect on the variance can be seen in the next terms. Differentiating the above variance approximation with respect to α_0, we obtain the optimal values of α_0, $\alpha_0^* = \max(\frac{\sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-3}\theta_j}{\sum_{j\,:\,j\neq\ell}(\theta_\ell-\theta_j)^{-4}\theta_j}, 0)$. Consequently, $\mathrm{Var}(\alpha_0^*, \alpha_1) \leq \mathrm{Var}(0,\alpha_1)$. This means that applying some degree of α_0 results in smaller variance than when we consider only some degree of α_1. To see the effect of α_0 on the MISE of the estimator, we have:

$$\begin{aligned}\mathrm{MISE}(\alpha_0,\alpha_1) &\equiv E\|\tilde{\psi}_\ell - \psi_\ell\|^2 \\ &\approx E\|n^{-1/2}\,\varphi^{(1)} + \alpha_1\,\varphi^{(2)} + n^{-1/2}\,\alpha_0\,\varphi^{(01)} + n^{-1/2}\alpha_1\,\varphi^{(12)} + \alpha_0\,\alpha_1\,\varphi^{(02)}\|^2\,,\end{aligned} \tag{5.39}$$

where $\varphi^{(1)}, \varphi^{(2)}, \varphi^{(01)}, \varphi^{(12)}, \varphi^{(02)}$ introduced in (5.29)–(5.32) and (5.30). We can write the MISE as follows:

$$\mathrm{MISE}(\alpha_0,\alpha_1) = \mathrm{MISE}(0,\alpha_1) + \mathrm{f}(\alpha_0,\alpha_1)\,, \tag{5.40}$$

where

$$\begin{aligned}
\mathrm{MISE}(0,\alpha_1) \approx n^{-1}\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\,\theta_j \;-\, n^{-1}\alpha_1 \Bigg\{ \bigg(\int (\psi_\ell'')^2 \bigg)\,\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2}\,\theta_j \\
+\, 2\,\theta_\ell^2 \textstyle\sum_{k\neq\ell} \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_k)^{-2}\,(\theta_\ell-\theta_j)^{-1}\,\theta_j \left(\int \psi_j''\psi_k'' \right) \\
-\, 2\,\theta_\ell^2 \left(\int (\psi_\ell'')^2 \right) \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\,\theta_j \Bigg\} \\
+\, \alpha_1^2 \left\{ \tfrac{1}{4} \left(\int (\psi_\ell'')^2 \right)^2 + \theta_\ell^2 \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \left(\int \psi_\ell''\psi_j'' \right)^2 \right\},
\end{aligned} \tag{5.41}$$

and,

$$\begin{aligned}
\mathrm{f}(\alpha_0,\alpha_1) = -2\,n^{-1}\,\alpha_0\,\theta_\ell \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\,\theta_j \\
+\, 2\,n^{-1/2}\,\alpha_0\,\alpha_1\,\theta_\ell^2 \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3} \left(\int \psi_j''\psi_\ell'' \right)^2 \\
+\, \alpha_0\,\alpha_1^2 \left\{ 2\,\theta_\ell \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \left(\int \psi_j''\psi_\ell'' \right)^2 - \left(\int (\psi_\ell'')^2 \right)^2 \right\} \\
+\, n^{-1}\,\alpha_0^2 \left\{ \theta_\ell^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4} \left(\int \psi_j''\psi_\ell'' \right)^2 + \theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4}\,\theta_j \right\} \\
+\, 2\,n^{-1}\,\alpha_0\,\alpha_1 \Bigg\{ \tfrac{1}{2} \left(\int (\psi_\ell'')^2 \right) \theta_\ell \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\,\theta_j \\
+\, \theta_\ell^2 \sum_{k\neq\ell} \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_k)^{-3}\,(\theta_\ell-\theta_j)^{-1}\,\theta_j \left(\int \psi_j''\psi_k'' \right) \\
+\, \theta_\ell^2 \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_k)^{-4}\,\theta_j \left(\int \psi_j''\psi_\ell'' \right) + \theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\,\theta_j \left(\int \psi_j''\psi_\ell'' \right) \\
+\, \theta_\ell^2 \left(\int (\psi_\ell'')^2 \right) \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4}\,\theta_j \Bigg\} \\
+\, 2\,n^{-1/2}\,\alpha_0^2\,\alpha_1 \left\{ \theta_\ell \textstyle\sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3} \left(\int \psi_j''\psi_\ell'' \right)^2 \right\} \\
+\, \alpha_0^2\,\alpha_1^2 \left\{ \left(\int (\psi_\ell'')^2 \right)^2 + \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-2} \left(\int \psi_j''\psi_\ell'' \right)^2 \right\}.
\end{aligned} \tag{5.42}$$

If we ignore the term of $n^{-1}\,\alpha_0\,\alpha_1$ in (5.40), then in our case in which $\left(\int \psi_\ell''\,\psi_j''\right) = 0$ for $j \neq \ell$, minimising with respect to α_0, results in,

$$\alpha_0^* = \frac{n^{-1}\,\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-3}\,\theta_j + \frac{1}{2}\,\alpha_1^2 \left(\int (\psi_\ell'')^2\right)^2}{\alpha_1^2 \left(\int (\psi_\ell'')^2\right)^2 + n^{-1}\,\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell - \theta_j)^{-4}\,\theta_j}\,. \tag{5.43}$$

$f(\alpha_0, \alpha_1)$ is decreasing with respect to α_0, when it is between 0 and α_0^*. Thus, it is beneficial to apply some degree of α_0 if $\alpha_0^* \geq 0$. Only in this case do we have $\mathrm{MISE}(\alpha_0^*, \alpha_1) \leq \mathrm{MISE}(0, \alpha_1)$. It can be seen that α_0^* tends to $\frac{1}{2}$ when n is large enough (compared with the two sums in the ratio) for fixed α_1 or when $\alpha_1 \to \infty$ for fixed n. However, in (5.40) if we take the term $n^{-1}\,\alpha_0\,\alpha_1$ into account, then minimise it with respect to α_0, we obtain,

$$\alpha_0^* = \frac{n^{-1}\theta_\ell \left\{ \left(1 - \frac{1}{2}\,\alpha_1 \int (\psi_\ell'')^2\right) \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-3}\,\theta_j + \alpha_1 \left(\int (\psi_\ell'')^2\right) \theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4}\,\theta_j \right\} + \frac{1}{2}\,\alpha_1^2 \left(\int (\psi_\ell'')^2\right)^2}{\alpha_1^2 \left(\int (\psi_\ell'')^2\right)^2 + n^{-1}\theta_\ell \sum_{j\,:\,j\neq\ell} (\theta_\ell-\theta_j)^{-4}\,\theta_j} \tag{5.44}$$

Here also it can be seen that α_0^* tends to $\frac{1}{2}$ when n is large enough (compared with the two sums, multiplied to it) for fixed α_1 or when $\alpha_1 \to \infty$ for fixed n. Numerical work on (5.44) showed that for a small positive choice of α_1, which depends on the rate of decay of the θ_j, α_0 quickly converges to $\frac{1}{2}$ for all $\ell \geq 1$. Hence, these all support that applying some value of α_0 might be useful when some degree of smoothing (α_1) is used. Our simulation result also confirms this result. However, if we have $\alpha_1 = 0$, applying any degree of α_0 to (5.18) is useless, because the ratio is changed to maximizing of

$$\frac{\int\int_{\mathcal{I}^2} \widehat{K}(u,v)\,\phi(u)\,\phi(v)\,du\,dv}{\|\phi\|^2} + \alpha_0\,. \tag{5.45}$$

This shows that any amount of α_0 does not affect the maximization problem. See the numerical results at the end of this Section.

It should be noted that without any problem, the above calculations can be generalised to the case in which $\rho_{jk} = [\psi_j, \psi_k] = \langle R\psi_j, R\psi_k\rangle = \langle \psi_j, Q\psi_k\rangle$ is replaced by $\int \psi_j'' \psi_k''$, where $Q = R^*R$ and R^* denotes the adjoint of R.

5.13.5 How to Choose Appropriate Values of the Smoothing Parameters?

Without considering α_0 in the problem, Ramsay and Silverman (1997) pointed out that the smoothing parameter α_1 in many circumstances can be adequately chosen subjectively. However, one may use a usual cross-validation paradigm for an automatic choice of the smoothing parameter. In our case, in which we have the two smoothing parameters α_0 and α_1, we may also use the cross-validation criterion proposed by Silverman (1996) with the two parameters α_0 and α_1.

In each case, Gaussian and Non-Gaussian, and each model, the quartiles of the integrated squared error (ISE$\equiv \|\widehat{\psi} - \psi\|^2$) were computed by using 500 simulated datasets. The series related to X was truncated at $j = N = 20$, and the first 20 eigenfunctions were estimated. We computed the three quartiles of the ISE for the original estimations of the eigenfunctions (without applying α_0 and α_1), using only α_1, and applying both α_0 and α_1. On the graphs, these are shown with "o" lines, "+" lines, and "$*$" lines, respectively.

Furthermore, the horizontal axes show the order of the eigenfunctions, enumerated from 1 to 20. As "o" lines on the graphs show, the ISE is increasing as the order of the eigenfunction increases. This error might be due to increased roughness, which is increasing as the order of the eigenfunction increases. This situation has been reported by others, such as Pezzulli and Silverman (1993). Although applying some degree of smoothing (α_1) improves the estimations, the ISE of the smoothed eigenfunctions increases when the order of the eigenfunction increases. This error, however, can be decreased considerably(even for small sample sizes) by applying some degree of α_0 with α_1.

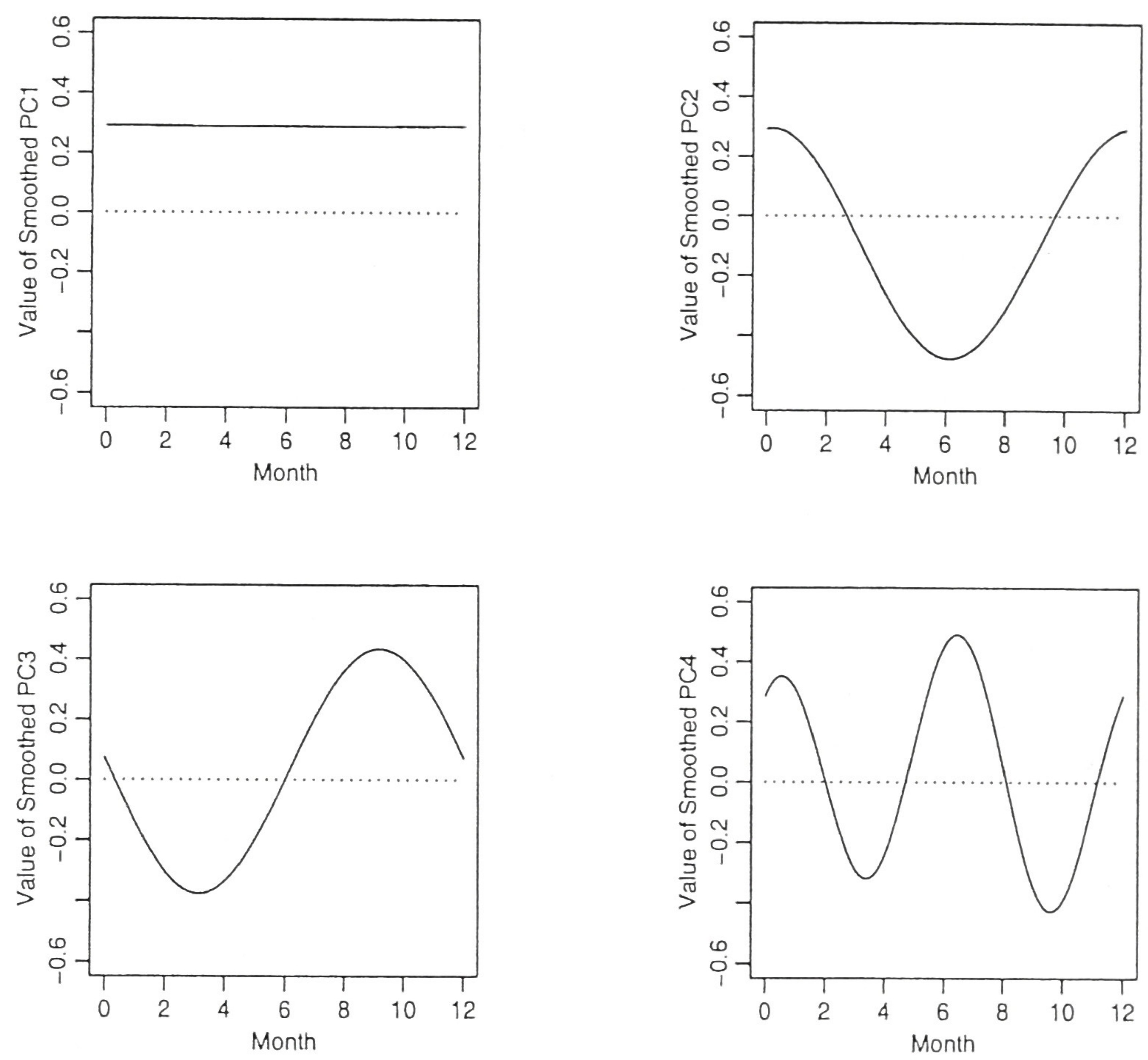

Figure 5.40: First four smoothed eigenfunctions obtained from Ramsay's Canadian weather stations dataset. The smoothing parameters α_0 and α_1 were chosen by cross-validation.

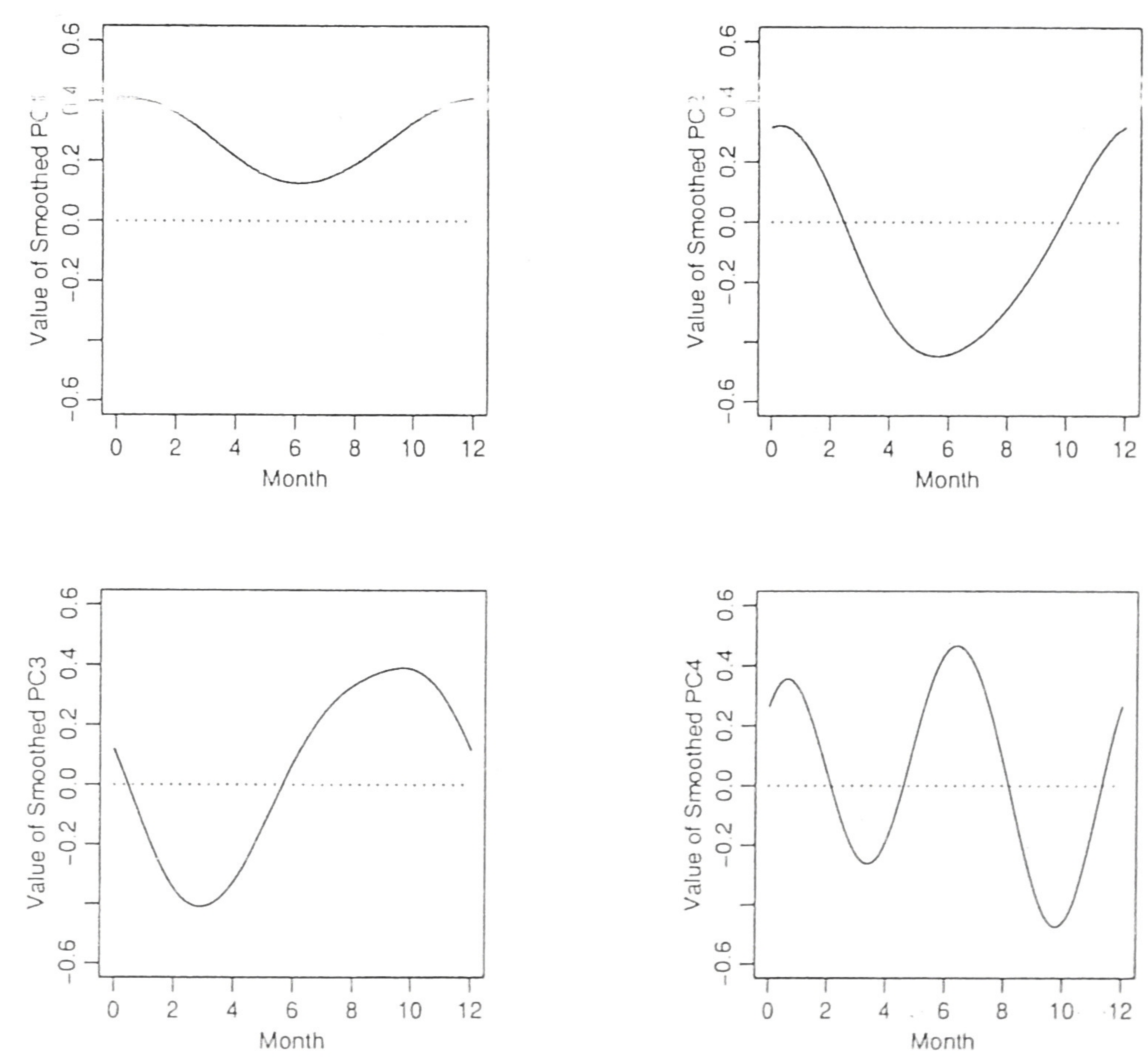

Figure 5.41: First four smoothed eigenfunctions obtained from Ramsay's Canadian weather stations dataset. We chose the smoothing parameters $\alpha_0 = 0.01$ and $\alpha_1 = 0.001$.

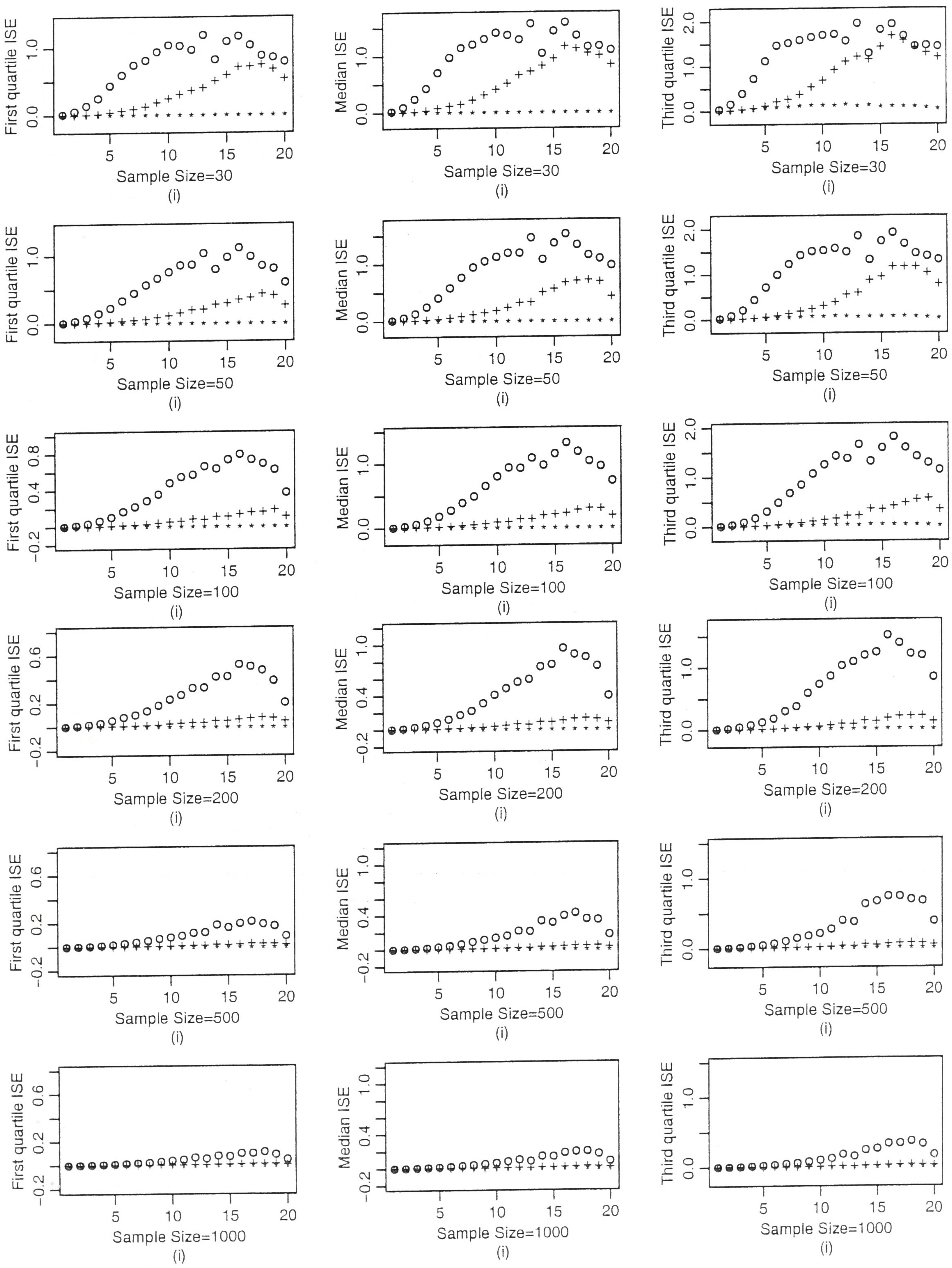

Figure 5.42: Performance of integrated squared error (ISE), when applying α_0 and α_1), using only α_1, and without applying α_0 and α_1. On the graphs, the latter cases are shown with "$*$" lines, "$+$" lines, and "o" lines, respectively. We generated from the Non-Gaussian process under model (i), and for the sample sizes $n = 30, 50, 100$. The first column shows the first quartile, the second and third columns reveal median and third quartiles, respectively.

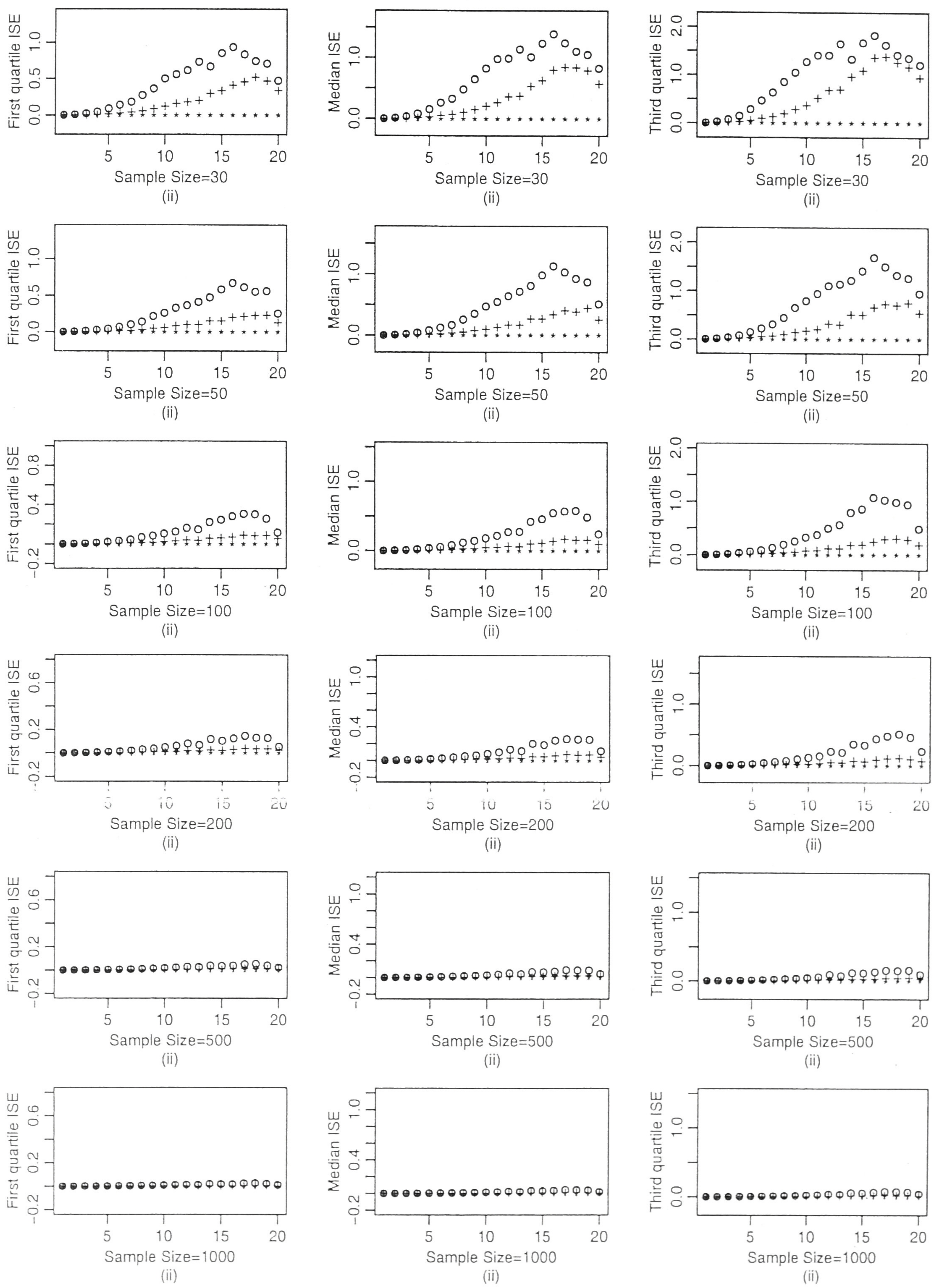

Figure 5.43: Performance of integrated squared error (ISE), when applying α_0 and α_1), using only α_1, and without applying α_0 and α_1. On the graphs, the latter cases are shown with "$*$" lines, "+" lines, and "o" lines, respectively. We generated from the Non-Gaussian process under model (ii), and for the sample sizes $n = 30, 50, 100, 200, 500, 1000$. The first column shows the first quartile, the second and third columns reveal median and third quartiles, respectively.

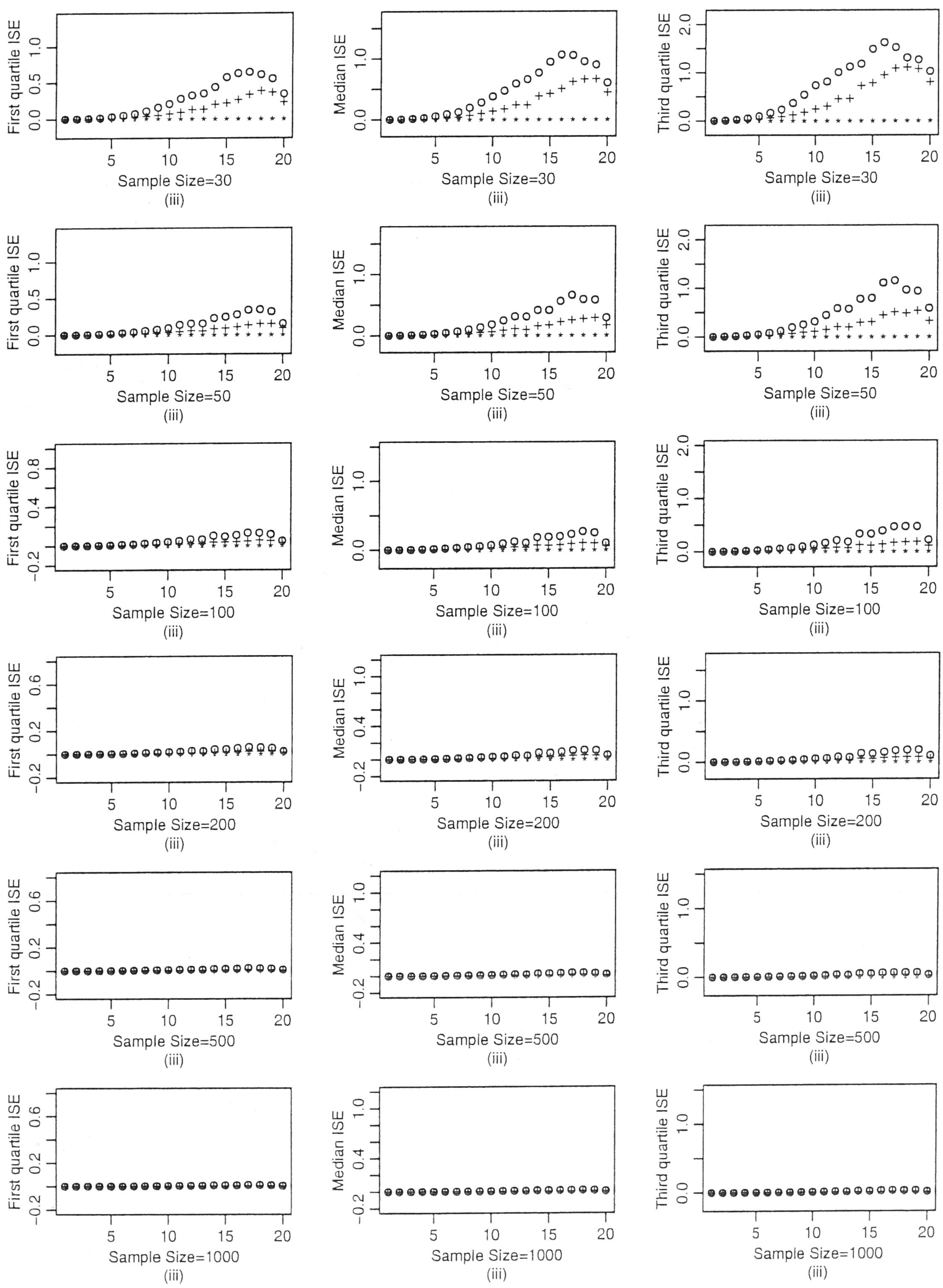

Figure 5.44: Performance of integrated squared error (ISE), when applying α_0 and α_1), using only α_1, and without applying α_0 and α_1. On the graphs, the latter cases are shown with "$*$" lines, "$+$" lines, and "o" lines, respectively. We generated from the Non-Gaussian process under model (iii), and for the sample sizes $n = 30, 50, 100, 200, 500, 1000$. The first column shows the first quartile, the second and third columns reveal median and third quartiles, respectively.

6

Conclusions

We have shown how to develop stochastic "Taylor expansions" of estimators of eigenvalues and eigenfunctions in functional principal components analysis, and provided a theoretical account of the accuracy of those expansions. We have seen that eigenvalue spacings have only a second-order effect on properties of eigenvalue estimators, but a first-order effect on properties of eigenfunction estimators. We have also shown that the stochastic expansions not only are valid for any finite number of principal components, but they are available uniformly in increasingly many components. The expansions themselves, or the methods used to derive them, can be used as the basis for an extensive theory of properties of functional principal components analysis, including the bootstrap. It should be mentioned that first order stochastic expansions of estimators of eigenvalues and eigenfunctions were given by Bosq (2000). We have addressed higher order stochastic expansions of the estimators, in particular those of order $n^{-3/2}$.

We have shown how stochastic, bootstrapped versions of uniform bounds that are obtainable via the mathematical theory of infinite-dimensional operators can be used to construct simultaneous confidence regions for literally all eigenvalue estimates, and for increasing numbers of eigenfunction estimates. The nature of these regions means that their coverage accuracy errs on the side of conservatism. In the case of simultaneous confidence regions for eigenvalues the degree of conservatism is small, and the bootstrap confidence regions are attractive practical tools. We have also addressed the problem of bootstrap confidence regions for individual, or relatively small numbers of, eigenvalues and eigenfunctions. These regions have been described theoretically and numerically.

In regard to smoothing in principal component analysis (PCA) for Functional Data, the effect of the smoothing parameter has been discussed theoretically and numerically.

Also, the impact of eigenvalue spacings on properties of functional linear regression has been addressed. We have explored the validity of simple accounts of the performance of functional linear regression. It has been observed that those accounts are valid if eigenvalues are reasonably well separated, but not necessarily otherwise. However, the expansions have implications beyond that setting. For example, they can be used to describe properties of functional-data methods applied to classification and clustering.

Appendix I

Generalisation and Proof of Theorem 4.1

I.1. *Generalisation.* Theorem 4.1 as a rigorous version of (4.15) can be generalised. A version of Theorem 4.1 is available under more general conditions than (4.16). In the present Section we develop and state the general theorem under explicit regularity conditions, and in Section (I.2) we derive the theorem in that general form.

For brevity, write simply b for b^0. One way in which we generalise (4.16) is through assuming that θ_j and b_j satisfy only Césaro-summability conditions, rather than being constrained to equal regularly varying functions:

for each p, q and r with $p \geq 0$, it is true that for all $j \geq 1$ and all $\eta > 0$,

$$\sum_{k=1}^{j} |b_k|^p \theta_k^q k^r \leq \text{const.}\,(1 + |b_j|^p \theta_j^q\, j^{r+1+\eta}),$$

where the constant depends on p, q, r and η; and, if the infinite series converges, for all $j \geq 1$ and all $\eta > 0$, (I.1)

$$\sum_{k=j}^{\infty} |b_k|^p\, \theta_k^q\, k^r \leq \text{const.}\ |b_j|^p \theta_j^q\, j^{r+1+\eta}.$$

Also, we work with weaker conditions on spacings than are given in (4.16):

for all $1 \leq j < k < \infty$, $\frac{\max(\theta_j,\theta_k)}{|\theta_j - \theta_k|} \leq \text{const.}\ \frac{\max(j,k)}{|j-k|}$. (I.2)

We also ask that:

for some $\eta > 0$, $b_j^2\, \theta_j^{-2}\, j^{1+\eta} \to 0$ as $j \to \infty$; and $b_j^2\, \theta_j^{-2}\, j$ is ultimately decreasing in j; (I.3)

the process X has all moments finite, and defining $A = X - E(X)$, we have $E(\int A\, \psi_j)^{2r} \leq C_r\, \{E(\int A\psi_j)^2\}^r$ for all integers $r \geq 1$, where C_r depends only on r, not on j. (I.4)

$E(|\epsilon|^r) < \infty$ for all $r > 0$, $E(\epsilon) = 0$, and $E(\epsilon^2) = \sigma^2$; (I.5)

for some $\eta > 0$, $n^{-1+\eta}\, r_0^2\, \theta_{r_0}^{-2} \to 0$. (I.6)

Define

$$W(t) = \frac{1}{n} \sum_{i=1}^{n} \{X_i(t) - \bar{X}(t)\}(\epsilon_i - \bar{\epsilon}), \tag{I.7}$$

where $\bar{\epsilon} = n^{-1} \sum_{i=1}^{n} \epsilon_i$, and put

$$S(r) = \sum_{j=1}^{r} \theta_j^{-2} \left\{ \int_{\mathcal{I}} W(t)\phi_j(t)\, dt \right\}^2. \tag{I.8}$$

Assumption (4.16) implies (I.1)-(I.6). Moreover, we have the following result, of which Theorem 4.1 is a corollary:

Theorem I.1. Assume (I.1)-(I.6). To eliminate pathologies arising from too small values of $\hat{\theta}_j$, replace $\hat{b}_j$ by an arbitrary fixed constant if $|\hat{b}_j| > c_1 n^{c_2}$, for any given $c_1, c_2 > 0$. Then for some $\eta > 0$ we have, with probability 1 for all $1 \le r \le r_0$,

$$\left| \int_{\mathcal{I}} (\hat{b} - b)^2 - S(r) - \sum_{j=r+1}^{\infty} b_j^2 \right| \le \text{const.}\, n^{-\eta}\, r\, (n^{-1}\, \theta_r^{-1} + b_r^2)\, \hat{A}(n),$$

where $\hat{A}(n)$ denotes a positive random variables which satisfies $\sup_{n \ge 1} E\{\hat{A}(n)^s\} < \infty$ for each integer $s \ge 1$, and the constant does not depend on n and r. Additionally, for each $r \ge 1$,

$$E\{S(r)\} = \frac{1}{n}\,(1 - \frac{1}{n})\, \sigma^2 \sum_{j=1}^{r} \theta_j^{-1}.$$

I.2. *Proof of Theorem I.1.* The proof has similarities to the arguments in Chapter 2, in that it proceeds by developing "Taylor approximations" (analogues of those at (1.11), (1.12), (1.32) and (1.33)). However, in order to make the regularity conditions relatively weak it is advantageous to go back to the methods used to derive Theorem 2.1, rather than use Theorem 2.1 itself as the basis for a proof of Theorem I.1.

I.2.1. *Preliminaries.* Note that by (I.2),

$$\rho_j^{-1} \equiv \max_{k\,:\,k\neq j} |\theta_j - \theta_k|^{-1} \leq \text{const.} \max_{k\,:\,k\neq j} \left\{ \frac{\max(j,k)}{|j-k| \max(\theta_j, \theta_k)} \right\} \leq \text{const.}\ \theta_j^{-1}\, j. \quad \text{(I.9)}$$

Define $\bar{b}_j = \int b\, \widehat{\psi}_j$. Then $b = \sum_{j=1}^{\infty} \bar{b}_j\, \widehat{\psi}_j$, and also $\hat{b} = \sum_{j=1}^{r} \hat{b}_j\, \widehat{\psi}_j$, where $\hat{b}_j = \hat{\theta}_j^{-1} \hat{g}_j$ and $\hat{g}_j = \int \hat{g}\, \widehat{\psi}_j$. Therefore,

$$\int (\hat{b} - b)^2 = \sum_{j=1}^{r} (\hat{b}_j - \bar{b}_j)^2 + \sum_{j=r+1}^{\infty} \bar{b}_j^2. \quad \text{(I.10)}$$

In the proof below, if Ψ is a symmetric function of two variables and ψ_1 and ψ_2 are both functions of a single variable, then we shall write $\int \Psi\, \psi_1\, \psi_2$ and $\int \Psi\, \psi_1$ to denote, respectively, the scalar $\int\int_{\mathcal{I}^2} \Psi(u,v)\, \psi_1(u)\, \psi_2(v)\, du\, dv$ and the function of which the value at u is $\int_{\mathcal{I}} \Psi(u,v)\, \psi_1(v)\, dv$.

Lemma I.1. Let $S(r_0)$ be as introduced in (I.8), and $\hat{b}_j$ be replaced by an arbitrary fixed constant if its absolute value exceeds $c_1 n^{c_2}$, for some $c_1, c_2 > 0$. Then, there exists $d > 0$ such that for each $s \geq 1$,

$$E\big(\max_{r \leq r_0} \|\hat{b}\|^s\big) + E\{S(r_0)^s\} \leq \text{const.}\, n^{ds},$$

where the constant depends on r but not on n.

(Note that this is the only part of the proof of Theorem I.1 where we use this assumption.)

Proof of Lemma:

$$
\begin{aligned}
E\|\hat{b}\|^s = E\|\sum_{j=1}^{r} \hat{b}_j\, \widehat{\psi}_j\|^s &\le r^s\, E\big[\max_{1\le j\le r} (\hat{b}_j\, \|\widehat{\psi}_j\|)\big]^s \le r^s \sum_{j=1}^{r} E|\hat{b}_j|^s \\
&\le r^s \sum_{j=1}^{r} E\Big[|\hat{b}_j|^s\, I(|\hat{b}_j| \le c_1\, n^{c_2}) + C^s\, I(|\hat{b}_j| > c_1\, n^{c_2})\Big] \\
&= r^s \sum_{j=1}^{r} \Big[E\{|\hat{b}_j|^s\, I(|\hat{b}_j| \le c_1\, n^{c_2})\} + C^s\, \mathrm{P}(|\hat{b}_j| \ge c_1\, n^{c_2})\Big] \\
&= r^s \sum_{j=1}^{r} \Big[E\big[|\hat{\theta}_j^{-1}\, \hat{g}_j|^s\, \{I(\hat{\theta}_j \le \tfrac{1}{2}\, \theta_j) + I(\hat{\theta}_j > \tfrac{1}{2}\, \theta_j)\}\, I(|\hat{b}_j| \le c_1\, n^{c_2})\big] \\
&\qquad + C^s\, \mathrm{P}(|\hat{b}_j| \ge c_1\, n^{c_2})\Big]\,.
\end{aligned}
$$

So,

$$
\begin{aligned}
E\|\hat{b}\|^s &\le r^s \sum_{j=1}^{r} \Big[E\big[(\tfrac{1}{2}\,\theta_j)^{-1}\, \hat{g}_j\big]^s + (c_1\, n^{c_2})^s\, \mathrm{P}(\hat{\theta}_j \le \tfrac{1}{2}\,\theta_j) + C^s\Big] \\
&\le r^s \sum_{j=1}^{r} \Big[(2/\theta_j)^s\, 2^{s-1}\{|g_j|^s + E|\hat{g}_j - g_j|^s\} + (c_1\, n^{c_2})^s\, \mathrm{P}(\hat{\theta}_j \le \tfrac{1}{2}\,\theta_j) + C^s\Big] \\
&\le r^s \sum_{j=1}^{r} \Big[(2/\theta_j)^s\, 2^{s-1}\{|g_j|^s + E|\hat{g}_j - g_j|^s\} + (c_1\, n^{c_2})^s\, \mathrm{P}(|\hat{\theta}_j - \theta_j| \ge \tfrac{1}{2}\,\theta_j) + C^s\Big] \\
&\le r^s \sum_{j=1}^{r} \Big[(2/\theta_j)^s\, 2^{s-1}\{|g_j|^s + E|\hat{g}_j - g_j|^s\} + (c_1\, n^{c_2})^s\, (2/\theta_j)^k\, E|\hat{\theta}_j - \theta_j|^k + C^s\Big] \\
&\le r^s \sum_{j=1}^{r} \Big[(2/\theta_j)^s\, 2^{s-1}\{|g_j|^s + E|\hat{g}_j - g_j|^s\} + (c_1\, n^{c_2})^s\, (2/\theta_j)^k\, E(\widehat{\Delta}^k) + C^s\Big]\,,
\end{aligned}
\tag{I.11}
$$

where we have used Markov's inequality for some $k > 0$, and the fact that $|\hat{\theta}_j - \theta_j| \le \widehat{\Delta}$ to obtain the last two inequalities above, respectively.

We also have:

$$
|\hat{g}_j - g_j| = |\int \hat{g}\, \widehat{\psi}_j - \int g\, \psi_j| = \int (\hat{g} - g)\, \psi_j + \int (\hat{g} - g)\, (\widehat{\psi}_j - \psi_j) + \int g\, (\widehat{\psi}_j - \psi_j)\,,
$$

Hence,

$$
\begin{aligned}
|\hat{g}_j - g_j|^s &= \left\{ \int (\hat{g} - g)\,\psi_j + \int (\hat{g} - g)\,(\widehat{\psi}_j - \psi_j) + \int g\,(\widehat{\psi}_j - \psi_j) \right\}^s \\
&\le C_s \left\{ \|\hat{g} - g\|^s + \|\hat{g} - g\|^s\,\|\widehat{\psi}_j - \psi_j\|^s + \|g\|^s\,\|\widehat{\psi}_j - \psi_j\|^s \right\},
\end{aligned}
$$

and then,

$$
E|\hat{g}_j - g_j|^s \le C_s \left\{ E\|\hat{g} - g\|^s + E\big[\|\hat{g} - g\|^s\,\|\widehat{\psi}_j - \psi_j\|^s\big] + \|g\|^s\,E\|\widehat{\psi}_j - \psi_j\|^s \right\}. \tag{I.12}
$$

Also, we shall prove in (I.42) that $E\left\{ \int (\hat{g} - g)\,\psi_j \right\}^{2s} = O(\{n^{-1}\,\theta_j\}^s)$. Thus, this result leads us to:

$$
\begin{aligned}
E\|\hat{g} - g\|^{2s} = E\big(\|\hat{g} - g\|^2\big)^s &= E\left(\sum_{j=1}^{\infty} \left\{ \int (\hat{g} - g)\,\psi_j \right\}^2 \right)^s \\
&\le \left(\sum_{j=1}^{\infty} \left[E\left\{ \int (\hat{g} - g)\,\psi_j \right\}^{2s} \right]^{1/s} \right)^s \\
&\le \left(\sum_{j=1}^{\infty} \left[(\theta_j/n)^s\right]^{1/s} \right)^s \\
&= n^{-s} \left(\sum_{j=1}^{\infty} \theta_j \right)^s = O(n^{-s}).
\end{aligned} \tag{I.13}
$$

Also, in Section (I.2.8) we shall prove that $\|\widehat{\psi}_j - \psi_j\|^2 \le n^{-1}\,j^2\,\max_{1\le j\le r_0} \tilde{a}_j$, where $E\big(\max_{1\le j\le r_0} \tilde{a}_j\big)^s \le r_0\,\max_{1\le j\le r_0} E(\tilde{a}_j^s) \le \text{const.}\, r_0$, in which the constant does not depend on n or r_0. Therefore, for $1 \le j \le r_0$,

$$
E\|\widehat{\psi}_j - \psi_j\|^{2s} \le n^{-s}\,j^{2s}\,E\big[\max_{1\le j\le r_0} \tilde{a}_j\big]^s \le n^{-s}\,j^{2s}\text{const.}\, r_0 \le \text{const.}\; n^{-s}\,r_0^{2s+1},
$$

and as a result,

$$E\left[\left(1+\|\widehat{\psi}_j-\psi_j\|^s\right)\|\hat{g}-g\|^s\right] \le \left(E\left[1+\|\widehat{\psi}_j-\psi_j\|^s\right]^2\right)^{1/2}\left(E\|\hat{g}-g\|^{2s}\right)^{1/2}$$

$$\le \text{const.}\,(1+n^{-s}\,r_0^{2s+1})^{1/2}\,n^{-s/2}$$

$$\le \text{const.}\; n^{-s}\,r_0^{s+\frac{1}{2}}\,. \tag{I.14}$$

Substituting results (I.13) and (I.14) into (I.12) gives:

$$E|\hat{g}_j-g_j|^s \le C_s\left\{n^{-s/2}+n^{-s}\,r_0^{s+\frac{1}{2}}+n^{-s/2}\,r_0^{s+1}\right\}$$

$$\le C_s\,n^{-s/2}\left\{1+n^{-s/2}\,r_0^{s+\frac{1}{2}}+r_0^{s+1}\right\}$$

$$\le C_s\,n^{-s/2}\,r_0^{s+1}\,.$$

Therefore, we have from (I.11) that for $s \ge 1$,

$$E\|\hat{b}\|^s \le \text{const.}\,r^s\{r+(\theta_r^{-1}\,r^2\,n^{-1})^{s/2}\,r^2$$

$$+\,c_1\,n^{c_2\,s}\,(r\,\theta_r^{-1}\,n^{-1/2})^k+C^s\}\,.$$

This also implies that

$$E\big(\max_{r\le r_0}\|\hat{b}\|^s\big) \le \sum_{r=1}^{r_0} E\|\hat{b}\|^s$$

$$\le \text{const.}\; r_0^{s+1}\{r_0+(\theta_{r_0}^{-1}\,r_0^2\,n^{-1})^{s/2}\,r_0^2+c_1\,n^{c_2\,s}\,(r_0\,\theta_{r_0}^{-1}\,n^{-1/2})^k+C^s\}$$

$$\le c_1^*\,n^{d\,s}\,, \tag{I.15}$$

where c_1^* depends on r but not n, and d is positive. Furthermore, we have used (I.6) to obtain the last inequality. We see that

$$
\begin{aligned}
E\{S(r_0)^s\} &= E\Bigg[\sum_{j=1}^{r_0} \theta_j^{-2} \bigg\{ \int_{\mathcal{I}} W(t)\,\psi_j\,dt \bigg\}^2\Bigg]^s \\
&= E\Bigg[\sum_{j=1}^{r_0} \theta_j^{-2} \bigg\{ \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{I}} (X_i(t) - \bar{X}(t))\,\psi_j(t)\,dt\,(\epsilon_i - \bar{\epsilon}) \bigg\}^2\Bigg]^s \\
&= E\Bigg[\sum_{j=1}^{r_0} \theta_j^{-2} \bigg\{ \frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\epsilon_i - \bar{\epsilon}) \bigg\}^2\Bigg]^s \\
&\le r_0^{s-1} \sum_{j=1}^{r_0} \theta_j^{-2s}\, E\Bigg[\frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\epsilon_i - \bar{\epsilon})\Bigg]^{2s}. \qquad \text{(I.16)}
\end{aligned}
$$

By using independence of the (X_i, ϵ_i) and of X_i from ϵ_i, and then applying the Rosenthal's inequality, we have:

$$
\begin{aligned}
E\Bigg[\frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\epsilon_i - \bar{\epsilon})\Bigg]^{2s} &\le n^{-2s}\, C_s \Bigg[\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)^{2s}\,(\epsilon_i - \bar{\epsilon})^{2s}\} \\
&\qquad + \bigg(\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)^2\,(\epsilon_i - \bar{\epsilon})^2\}\bigg)^s\Bigg] \\
&= n^{-2s}\, C_s \Bigg[\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)\}^{2s}\, E\{(\epsilon_i - \bar{\epsilon})\}^{2s}\} \\
&\qquad + \bigg(\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)^2\}\, E\{(\epsilon_i - \bar{\epsilon})^2\}\bigg)^s\Bigg].
\end{aligned}
$$

Thus, by (I.4),

$$
\begin{aligned}
E\Bigg[\frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\epsilon_i - \bar{\epsilon})\Bigg]^{2s} &\le n^{-2s}\, C_s \Bigg[\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)^{2s}\}\, E\{(\epsilon_i - \bar{\epsilon})^{2s}\} \\
&\qquad + \bigg(\sum_{i=1}^{n} E\{(\xi_{ij} - \bar{\xi}_j)^2\}\, E\{(\epsilon_i - \bar{\epsilon})^2\}\bigg)^s\Bigg] \\
&\le n^{-2s}\, C_s \Bigg[\sum_{i=1}^{n} \theta_j^s\, E\{(\epsilon_i - \bar{\epsilon})^{2s}\} + \bigg(\theta_j \sum_{i=1}^{n} E\{(\epsilon_i - \bar{\epsilon})^2\}\bigg)^s\Bigg] \\
&\le C_1(s)\,(n^{-1}\,\theta_j)^s.
\end{aligned}
$$

Therefore, substituting the above result into (I.16) we deduce that

$$E\{S(r_0)^s\} \leq r_0^{s-1} \sum_{j=1}^{r_0} \theta_j^{-s} n^{-s} \leq r_0^s n^{-s} \theta_{r_0}^{-s} . \tag{I.17}$$

Combining (I.15) and (I.17) finishes the proof. ■

In view of Lemma I.1, noting particulary that C in (I.18) below is arbitrarily large, Theorem I.1 will follow if we derive the following result. Recall that $\widehat{\Delta} = \|\widehat{K} - K\|$.

Proposition I.1. Assume (I.1)-(I.6). Given $c > 0$, let $\mathcal{G}$ denote the event that $\widehat{\Delta} \leq c r_0^{-1} \theta_{r_0}$. Then, for each choice of c,

$$1 - P(\mathcal{G}) = O(n^{-C}) \quad \text{as} \quad n \to \infty, \qquad \text{for each } C > 0. \tag{I.18}$$

Moreover, if c is sufficiently small then for some $\eta > 0$ we have, for all $1 \leq r \leq r_0$, provided $\mathcal{G}$ holds,

$$\left| \int_{\mathcal{I}} (\hat{b} - b)^2 - S(r) - \sum_{j=r+1}^{\infty} b_j^2 \right| \leq \text{const.}\, n^{-\eta}\, r\, (n^{-1}\, \theta_r^{-1} + b_r^2)\, \hat{A}(n) , \tag{I.19}$$

where $\hat{A}(n)$ denotes a positive random variable which satisfies $\sup_{n \geq 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \geq 1$, and the constant does not depend on n and r. Additionally, for each $r \geq 1$,

$$E\{S(r)\} = \frac{1}{n}\,(1 - \frac{1}{n})\, \sigma^2 \sum_{j=1}^{r} \theta_j^{-1} . \tag{I.20}$$

I.2.2. *Approximation to first term on right-hand side of* (I.10). Write

$$
\begin{aligned}
\hat{b}_j - \bar{b}_j &= \hat{\theta}_j^{-1} \int \hat{g}\, \widehat{\psi}_j - \int b\, \widehat{\psi}_j \\
&= \hat{\theta}_j^{-1} \int (\hat{g} - g)(\widehat{\psi}_j - \psi_j) - \int b\, (\widehat{\psi}_j - \psi_j) \\
&\quad + \hat{\theta}_j^{-1} \int g\, (\widehat{\psi}_j - \psi_j) + \hat{\theta}_j^{-1} \int \hat{g}\, \psi_j - \int b\, \psi_j \\
&= \hat{\theta}_j^{-1} \int (\hat{g} - g)\, (\widehat{\psi}_j - \psi_j) - \int b\, (\widehat{\psi}_j - \psi_j) + (\hat{\theta}_j^{-1} - \theta_j^{-1}) \int g\, (\widehat{\psi}_j - \psi_j) \\
&\quad + \hat{\theta}_j^{-1} \int (\hat{g} - g)\, \psi_j - \int b\, \psi_j + \theta_j^{-1} \int g\, (\widehat{\psi}_j - \psi_j) + \hat{\theta}_j^{-1} \int g\, \psi_j \\
&= \hat{\theta}_j^{-1} \int (\hat{g} - g)\, (\widehat{\psi}_j - \psi_j) - \int b\, (\widehat{\psi}_j - \psi_j) + \theta_j^{-1} \int (\hat{g} - g)\, \psi_j \\
&\quad + (\hat{\theta}_j^{-1} - \theta_j^{-1}) \Big\{ \int (\hat{g} - g)\, \psi_j + \int g\, (\widehat{\psi}_j - \psi_j) \Big\} \\
&\quad - \int b\, \psi_j + \hat{\theta}_j^{-1} \int g\, \psi_j + \theta_j^{-1} \int g\, (\widehat{\psi}_j - \psi_j)\,.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\hat{b}_j - \bar{b}_j &= \theta_j^{-1} \int (\hat{g} - g)\, \psi_j + \int (\theta_j^{-1}\, g - b)\, (\widehat{\psi}_j - \psi_j) - b_j\, \theta_j^{-1}\, (\hat{\theta}_j - \theta_j) \\
&\quad + \hat{\theta}_j^{-1} \int (\hat{g} - g)\, (\widehat{\psi}_j - \psi_j) + (\hat{\theta}_j^{-1} - \theta_j^{-1}) \Big\{ \int (\hat{g} - g)\, \psi_j + \int g\, (\widehat{\psi}_j - \psi_j) \Big\} \\
&\quad - \int b\, \psi_j + \hat{\theta}_j^{-1} \int g\, \psi_j + b_j\, \theta_j^{-1} (\hat{\theta}_j - \theta_j) \\
&= T_{1j} + T_{2j}
\end{aligned}
$$

where

$$
\begin{aligned}
T_{1j} =& \theta_j^{-1} \int (\hat{g} - g)\, \psi_j + \int (\theta_j^{-1} g - b)\, (\widehat{\psi}_j - \psi_j) - b_j\, \theta_j^{-1}\, (\hat{\theta}_j - \theta_j), \\
T_{2j} =& \hat{\theta}_j^{-1} \int (\hat{g} - g)\, (\widehat{\psi}_j - \psi_j) + (\hat{\theta}_j^{-1} - \theta_j^{-1}) \Big\{ \int (\hat{g} - g)\, \psi_j + \int (g\, (\widehat{\psi}_j - \psi_j) \Big\} \\
& + \theta_j^{-1}\, \hat{\theta}_j^{-1}\, b_j\, (\hat{\theta}_j - \theta_j)^2\,.
\end{aligned}
$$

Here we have used the fact that $g_j = \theta_j\, b_j$. Note too that

$$\left|\left\{\sum_{j=1}^{r}(\hat{b}_j - \bar{b}_j)^2\right\}^{1/2} - \left(\sum_{j=1}^{r} T_{1j}^2\right)^{1/2}\right| \leq \left(\sum_{j=1}^{r} T_{2j}^2\right)^{1/2}, \tag{I.21}$$

$$\begin{aligned}\tfrac{1}{4}\sum_{j=1}^{r} T_{2j}^2 \leq \sum_{j=1}^{r} \hat{\theta}_j^{-2}\left\{\int (\hat{g}-g)\,(\widehat{\psi}_j - \psi_j)\right\}^2 + \sum_{j=1}^{r}(\hat{\theta}_j^{-1} - \theta_j^{-1})^2\left\{\int (\hat{g}-g)\,\psi_j\right\}^2 \\ + \sum_{j=1}^{r}(\hat{\theta}_j^{-1} - \theta_j^{-1})^2\left\{\int g\,(\widehat{\psi}_j - \psi_j)\right\}^2 + \sum_{j=1}^{r} b_j^2\,\theta_j^{-2}\,\hat{\theta}_j^{-2}\,(\hat{\theta}_j - \theta_j)^4 .\end{aligned} \tag{I.22}$$

By (2.17) we have,

$$\sup_{j\geq 1} |\hat{\theta}_j - \theta_j| \leq \widehat{\Delta}\,. \tag{I.23}$$

Provided

$$\widehat{\Delta} \leq \tfrac{1}{2}\,\rho_{r_0}\,, \tag{I.24}$$

it follows from (I.22), (I.23) and the bound $\rho_{r_0} \leq \theta_{r_0}$, that for all $1 \leq r \leq r_0$,

$$\begin{aligned}\tfrac{1}{16}\textstyle\sum_{j=1}^{r} T_{2j}^2 \leq \|\hat{g}-g\|^2 \displaystyle\sum_{j=1}^{r} \theta_j^{-2}\,\|\widehat{\psi}_j - \psi_j\|^2 + \widehat{\Delta}^2 \sum_{j=1}^{r} \theta_j^{-4}\left\{\int (\hat{g}-g)\,\psi_j\right\}^2 \\ + \widehat{\Delta}^2 \sum_{j=1}^{r} \theta_j^{-4}\left\{\int g\,(\widehat{\psi}_j - \psi_j)\right\}^2 + \widehat{\Delta}^4 \sum_{j=1}^{r} b_j^2\,\theta_j^{-4}\,.\end{aligned} \tag{I.25}$$

Let $\mathcal{G}$ denote the event that $\widehat{\Delta} \leq c\,r_0^{-1}\,\theta_{r_0}$, where $c > 0$ is chosen so small that $c\,r_0^{-1}\,\theta_{r_0} \leq \frac{1}{2}\,\delta_{r_0}$ for all n. Result (I.9) implies that such a c exists. From Theorem 2.5 we have that for each $s \geq 1$, $E(\widehat{\Delta}^{2s}) = O(n^{-s})$. Then,

$$\begin{aligned}1 - P(\mathcal{G}) = P(\widehat{\Delta} \leq c\,r_0^{-1}\,\theta_{r_0}) \leq \frac{E(\widehat{\Delta}^{2s})}{(c\,r_0^{-1}\,\theta_{r_0})^s} \leq \text{const.}\,\frac{n^{-s}}{(c r_0^{-1}\theta_{r_0})^s} \\ \leq \text{const.}\,(n^{-1+\eta}\,r_0\,\theta_{r_0}^{-1})^s\, n^{-s\eta} \leq \text{const.}\, n^{-C}\,,\end{aligned}$$

where $C = s\,\eta$, and we have used Markov's inequality to obtain the first inequality and (I.6) to get the last one. So, (I.18) has proved. Moreover, if $\mathcal{G}$ holds, then in view of our choice of c, so too does (I.24), and hence (I.25) is valid for $1 \le r \le r_0$.

I.2.3. *Approximation to* $\sum_{j=1}^r T_{1j}^2$. Write $\widehat{\Delta}_g = \hat{g} - g$ and define $\widehat{\Delta}_j(t)$ and $\hat{\delta}_j$ by

$$\widehat{\psi}_j(t) = \psi_j(t) + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1} \psi_k(t) \int (\widehat{K} - K)\, \psi_j\, \psi_k + \widehat{\Delta}_j(t)\,, \tag{I.26}$$
$$\hat{\theta}_j = \theta_j + \int (\widehat{K} - K)\, \psi_j\, \psi_j + \hat{\delta}_j\,.$$

Note that $W(t) = \widehat{\Delta}_g(t) - \int (\widehat{K} - K)(t, v)\, b(v)\, dv$. Put

$$\begin{aligned} T_{3j} =& \theta_j^{-1} \int \widehat{\Delta}_g\, \psi_j + \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, (\theta_j^{-1}\, g_k - b_k) \int (\widehat{K} - K)\, \psi_j\, \psi_k \\ &- \theta_j^{-1}\, b_j \int (\widehat{K} - K)\, \psi_j\, \psi_j \\ =& \theta_j^{-1} \int \left(\widehat{\Delta}_g - \int (\widehat{K} - K)\, b \right) \psi_j = \theta_j^{-1} \int W\, \psi_j\,, \end{aligned} \tag{I.27}$$
$$T_{4j} = T_{1j} - T_{3j} = \int (\theta_j^{-1}\, g - b)\, \widehat{\Delta}_j - \theta_j^{-1}\, b_j\, \hat{\delta}_j$$

Then,

$$\sum_{j=1}^r T_{3j}^2 = \sum_{j=1}^r \theta_j^{-2} \left(\int W \psi_j \right)^2,$$
$$\sum_{j=1}^r T_{4j}^2 \le 2 \sum_{j=1}^r \theta_j^{-2} \left[\left\{ \int (g - \theta_j b)\, \widehat{\Delta}_j \right\}^2 + (b_j\, \hat{\delta}_j)^2 \right], \tag{I.28}$$
$$\left| \left(\sum_{j=1}^r T_{1j}^2 \right)^{1/2} - \left(\sum_{j=1}^r T_{3j}^2 \right)^{1/2} \right| \le \left(\sum_{j=1}^r T_{4j}^2 \right)^{1/2}. \tag{I.29}$$

I.2.4. *Bounds for* $\sum_{j=1}^r T_{4j}^2$.

Lemma I.2. The following results hold:

$$\widehat{\psi}_j(t) - \psi_j(t) = \sum_{k\,:\,k\neq j} (\hat{\theta}_j - \theta_k)^{-1}\, \psi_k(t) \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j\,, \tag{I.30}$$

$$\left| \hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j \right| \leq \|\widehat{\psi}_j - \psi_j\| \left(|\hat{\theta}_j - \theta_j| + \Delta_{(j)} \right), \tag{I.31}$$

where $\Delta_{(j)} = \| \int (\widehat{K} - K)\, \psi_j \|$.

Proof of Lemma: Result (I.30) follows from (2.9) on using $\hat{\theta}_j$ and $\widehat{\psi}_j$ instead of λ_j and ϕ_j. To prove (I.31), by (2.20), we have

$$\begin{aligned} \left| \hat{\theta}_j - \theta_j - \int (\widehat{K} - K)\, \psi_j\, \psi_j \right| &\leq \left| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_j \right| + |\hat{\theta}_j - \theta_j| \left| \int (\widehat{\psi}_j - \psi_j)\, \psi_j \right| \\ &\leq \left\| \int (\widehat{K} - K)\, \psi_j \right\| \|\widehat{\psi}_j - \psi_j\| + |\hat{\theta}_j - \theta_j|\, \|\widehat{\psi}_j - \psi_j\| \\ &= \|\widehat{\psi}_j - \psi_j\| \left(|\hat{\theta}_j - \theta_j| + \Delta_{(j)} \right). \blacksquare \end{aligned}$$

From (I.30) it follows that

$$\begin{aligned} \widehat{\Delta}_j = &\sum_{k\,:\,k\neq j} \left\{ (\hat{\theta}_j - \theta_k)^{-1} - (\theta_j - \theta_k)^{-1} \right\} \psi_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \\ &+ \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, \psi_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k + \psi_j \int (\widehat{\psi}_j - \psi_j)\, \psi_j\,. \end{aligned}$$

Therefore, using (I.23),

$$\begin{aligned}
\left|\int (\theta_j\, b - g_j)\, \widehat{\Delta}_j\right|^2 &= \Bigg| \sum_{k\,:\,k\neq j} \left\{ \left(1 + \frac{\hat\theta_j - \theta_j}{\theta_j - \theta_k}\right)^{-1} - 1 \right\} b_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \\
&\quad + \sum_{k\,:\,k\neq j} b_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \Bigg|^2 \\
&\le 8\, \widehat{\Delta}^2 \left\{ \sum_{k\,:\,k\neq j} \left| (\theta_j - \theta_k)^{-1}\, b_k \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k \right| \right\}^2 \\
&\quad + 2 \left\{ \sum_{k\,:\,k\neq j} \left| b_k \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \right| \right\}^2 \\
&\le 8\, \widehat{\Delta}^2 \left\{ \sum_{k\,:\,k\neq j} b_k^2\, (\theta_j - \theta_k)^{-2} \right\} \left\| \int (\widehat{K} - K)\, \widehat{\psi}_j \right\|^2 \\
&\quad + 2 \left(\int b^2 \right) \left\| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j) \right\|^2 \\
&\le 8\, \widehat{\Delta}^4 \sum_{k\,:\,k\neq j} b_k^2\, (\theta_j - \theta_k)^{-2} + 2\, \widehat{\Delta}^2\, \|\widehat{\psi}_j - \psi_j\|^2 \int b^2\,, \qquad \text{(I.32)}
\end{aligned}$$

where the first identity always holds, but the three inequalities are predicated on $\mathcal{G}$ obtaining and $1 \le j \le r_0$.

In view of (I.2) we have, for a constant not depending on j,

$$s(j) \equiv \sum_{k\,:\,k\neq j} b_k^2\, (\theta_j - \theta_k)^{-2} \le \text{const.} \sum_{k\,:\,k\neq j} b_k^2 \left\{ \frac{\max(j,k)}{|j-k|\, \max(\theta_j, \theta_k)} \right\}^2 \equiv \text{const.}\, t(j)\,,$$

say. Using the first part of (I.3) to bound $b_k^2\, \theta_k^{-2}$ we can show that for some $\eta > 0$,

$$\begin{aligned}
t(j) &\le \text{const.} \left\{ \sum_{k \le j/2,\, k > 2j} b_k^2\, \theta_k^{-2} + j^2 \sum_{j/2 < k \le 2j} b_k^2\, \theta_k^{-2}\, (j-k)^{-2} \right\} \\
&\le \text{const.} \left\{ 1 + j^{1-\eta} \sum_{j/2 < k \le 2j} (j-k)^{-2} \right\} \le \text{const.}\, j^{1-\eta}.
\end{aligned}$$

Therefore,

$$s(j) \equiv \sum_{k\,:\,k\neq j} b_k^2\, (\theta_j - \theta_k)^{-2} \le \text{const.}\, j^{1-\eta}\,. \qquad \text{(I.33)}$$

Combining (I.32) and (I.33) we deduce that if $\mathcal{G}$ holds and $1 \le j \le r_0$, we have for some $\eta > 0$:

$$\left|\int (\theta_j b - g)\,\widehat{\Delta}_j\right|^2 \leq \text{const.}\,\widehat{\Delta}^4\, j^{1-\eta} + 2\,\widehat{\Delta}^2\,\|\widehat{\psi}_j - \psi_j\|^2 \int b^2\,. \tag{I.34}$$

More simply, by (I.23) and (I.31), $|\hat{\delta}_j| \leq 2\,\widehat{\Delta}\,\|\widehat{\psi}_j - \psi_j\|$, whence by (I.1), (I.28) and (I.34), if $\mathcal{G}$ holds and $1 \leq r \leq r_0$, then

$$\begin{aligned}\sum_{j=1}^{r} T_{4j}^2 &\leq \text{const.}\,\widehat{\Delta}^2 \sum_{j=1}^{r} \theta_j^{-2}\,(\widehat{\Delta}^2\, j^{1-\eta} + \|\widehat{\psi}_j - \psi_j\|^2)\\ &\leq \text{const.}\,\widehat{\Delta}^4\,\theta_r^{-2}\,r^2 + \text{const.}\,\widehat{\Delta}^2 \sum_{j=1}^{r} \theta_j^{-2}\,\|\widehat{\psi}_j - \psi_j\|^2\,.\end{aligned} \tag{I.35}$$

We shall prove in Section (I.2.8) that for each $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \leq r \leq r_0$, $\sum_{j=1}^{r} \theta_j^{-2}\,\|\widehat{\psi}_j - \psi_j\|^2 \leq$ const. $n^{-1+\eta}\,r^3\,\theta_r^{-2}\,\hat{A}_1(n,\eta)$, where the random variable $\hat{A}_1(n,\eta)$ satisfies $\sup_{n\geq 1} E\{\hat{A}_1(n,\eta)^s\} < \infty$ for each $s \geq 1$. (I.36)

It therefore follows from (I.35) that for each $\eta > 0$,

if $\mathcal{G}$ holds, then for all $1 \leq r \leq r_0$, $\sum_{j=1}^{r} T_{4j}^2 \leq$ const. $n^{-2+\eta}\,r^3\,\theta_r^{-2}$ $\times\ \hat{A}(n,\eta)$, where $\hat{A}(n,\eta)$ satisfies $\sup_{n\geq 1} E\{\hat{A}(n,\eta)^s\} < \infty$ for each $s \geq 1$. (I.37)

The factor $n^{-2+\eta}\,r^3\,\theta_r^{-2}$ on the right-hand side of (I.37) can equivalently be written as $n^{-1}\,r\,\theta_r^{-1}\,.\,n^{-1+\eta}\,r^2\theta_r^{-1}$. By (I.6), for some $\eta' > 0$, we have $n^{-1+\eta'}\,r_0^2\,\theta_{r_0}^{-1} \to 0$. Put $\hat{A}_r(n) = n^{-1+2\eta}\,r^2\,\theta_r^{-1}\,\hat{A}(n,\eta)$. Then, for $\eta = \frac{1}{2}\eta'$, we have $\hat{A}_r(n) \leq$ const. $\hat{A}(n,\eta)$, for each $1 \leq r \leq r_0$. Therefore, for some $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \leq r \leq r_0$, $\sum_{j=1}^{r} T_{4j}^2 \leq$ const. $(n^{-1}r\theta_r^{-1}) \times (n^{-1+\eta}r^2\theta_r^{-1})\,\hat{A}(n,\eta) \leq$ const. $n^{-1-\eta}\,r\,\theta_r^{-1}\,\hat{A}(n)$, where $\hat{A}(n) = \max_{1\leq r\leq r_0} \hat{A}_r(n)$ satisfies $\sup_{n\geq 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \geq 1$. (I.38)

I.2.5. *Bound for* $\sum_{j=1}^{r} T_{2j}^2$. We shall bound successive terms on the right-hand side of (I.25). Denote the series components of those terms by

$$S_1 = \sum_{j=1}^{r} \theta_j^{-2}\, \|\widehat{\psi}_j - \psi_j\|^2, \; S_2 = \sum_{j=1}^{r} \theta_j^{-4} \Big\{ \int (\hat{g} - g)\psi_j \Big\}^2,$$
$$S_3 = \sum_{j=1}^{r} \theta_j^{-4} \Big\{ \int g\, (\widehat{\psi}_j - \psi_j) \Big\}^2, \; S_4 = \sum_{j=1}^{r} \theta_j^{-4}\, b_j^2,$$

so that, by (I.25), if $\mathcal{G}$ holds and $1 \le r \le r_0$,

$$\tfrac{1}{16} \textstyle\sum_{j=1}^{r} T_{2j}^2 \le \|\hat{g} - g\|^2\, S_1 + \widehat{\Delta}^2\, (S_2 + S_3) + \widehat{\Delta}^4\, S_4 \,. \tag{I.39}$$

A bound to S_1 is given at (I.36), which, in view of (I.6), implies that for some $\eta > 0$,

$$\begin{aligned}&\text{if } \mathcal{G} \text{ holds for all } 1 \le r \le r_0,\ S_1 \le \text{const.}\, (n^{-1+\eta}\, r^2\, \theta_r^{-1}) \times \\ &(\theta_r^{-1}\, r) \hat{A}_1(n, \eta) \le \text{const.}\ n^{-\eta}\, r\, \theta_r^{-1}\, \hat{A}(n), \text{ where } \hat{A}(n) \text{ satisfies} \\ &\sup_{n\ge1} E\{\hat{A}(n)^s\} < \infty \text{ for each } s \ge 1.\end{aligned} \tag{I.40}$$

A bound for S_4 is available via (I.1): for all $\eta > 0$, $S_4 = \sum_{j=1}^{r} \theta_j^{-4}\, b_j^2 \le \text{const.}\, (1 + b_r^2\, \theta_r^{-4}\, r^{1+\eta}) \le \text{const.}\, n^{\eta}\, (1 + b_r^2\, \theta_r^{-4}\, r)$. Hence, by (I.6), we have for some $\eta > 0$,

$$\begin{aligned}\text{for all} \quad 1 \le r \le r_0,\ n^{-2} S_4 &\le \text{const.}\, (n^{-2+\eta} + n^{-2+\eta}\, \theta_r^{-4}\, b_r^2\, r) \\ &\le \text{const.} \Big(n^{-1} + (n^{-1+\eta}\, \theta_r^{-2})^2\, n^{-\eta}\, b_r^2\, r \Big) \\ &\le \text{const.}\, (n^{-1} + n^{-\eta}\, b_r^2\, r)\,.\end{aligned} \tag{I.41}$$

Next we bound S_2. Assume, without loss of generality, that $\eta = E(X) = 0$. Put $\xi_{ij} = \int X_i\, \psi_j$ and $\bar{\xi}_j = n^{-1} \sum_{i=1}^{n} \xi_{ij}$, and recall the notation from Chapter 4. In this notation,

$$\int (\hat{g} - g)\, \psi_j = \int \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}) \, \{X_i(u) - \bar{X}(u)\} - E\big[(Y - \mu)\, X(u)\big] \right\} \psi_j$$
$$= n^{-1} \sum_{i=1}^{n} \left\{ \xi_{ij} \int bX_i - \bar{\xi}_j \int b\,\bar{X} - E\Big(\xi_{ij} \int b\, X_i \Big) \right\}$$
$$+\, n^{-1} \sum_{i=1}^{n} (\xi_{ij} \epsilon_i - \bar{\xi}_j \bar{\epsilon}) \,,$$

from which, using (I.4) and Rosenthal's inequality, for each integer $s \geq 1$, we deduce that

$$E\bigg\{ \int (\hat{g} - g)\psi_j \bigg\}^{2s} \leq \text{const.} \bigg[n^{-s} \bigg\{ \text{var}\Big(\xi_{1j} \int b\, X_1 \Big) + \text{var}(\xi_{1j}) \bigg\}^{s}$$
$$+\, n^{1-2s} \bigg(E\bigg| \xi_{1j} \int bX_1 \bigg|^{2s} + E|\xi_{1j}|^{2s} \bigg) \bigg]$$
$$\leq \text{const.}\, (\theta_j / n)^s \,, \tag{I.42}$$

where the constants depend on s but not on j or n. Define

$$\hat{U} = \hat{U}(n) = \max_{1 \leq j \leq r_0} n\, \theta_j^{-1} \bigg\{ \int (\hat{g} - g)\psi_j \bigg\}^2 .$$

Result (I.42) implies that, for each $s \geq 1$,

$$E(\hat{U}^s) \leq \sum_{j=1}^{r_0} E\bigg[n\, \theta_j^{-1} \bigg\{ \int (\hat{g} - g)\, \psi_j \bigg\}^2 \bigg]^s \leq \text{const.}\, r_0 \,, \tag{I.43}$$

where the constants depends on s but not on n. Hence, for each $\eta > 0$ and all

$1 \le r \le r_0$,

$$S_2 = \sum_{j=1}^{r} \theta_j^{-4} \Big\{ \int (\hat{g} - g)\, \psi_j \Big\}^2 = \sum_{j=1}^{r} \theta_j^{-3}\, n^{-1} \Big[n\, \theta_j^{-1} \Big\{ \int (\hat{g} - g)\, \psi_j \Big\}^2 \Big]$$
$$\le n^{-1}\, \hat{U}(n) \sum_{j=1}^{r} \theta_j^{-3} \le \text{const.}\ n^{-1+\eta}\, r\, \theta_r^{-3}\, \hat{A}(n, \eta)\,, \tag{I.44}$$

where $\hat{A}(n, \eta) = n^{-\eta}\, \hat{U}(n)$ satisfies, for $t > s \ge 1$,

$$E\{\hat{A}(n,\eta)^s\} = n^{-\eta s}\, E\big[\hat{U}(n)^s\big] \le n^{-\eta s} \Big(E\big[\hat{U}(n)^t\big] \Big)^{s/t} < C_{s,t}\, n^{-\eta s}\, (r_0)^{s/t}\,.$$

Furthermore, we have $\frac{s\,(1-\eta)}{2\,t\,(a+1)} \le \eta\, s$, for large values of t, and then $(r_0)^{s/t} \le n^{\eta\, s}$. Thus, $\sup_{n \ge 1} E\{\hat{A}(n,\eta)^s\} < \infty$ for $s \ge 1$, and we deduce from (I.44) that for each $\eta > 0$,

for all $1 \le r \le r_0$, $S_2 \le \text{const.}\ n^{-1+\eta}\, r\, \theta_r^{-3}\, \hat{A}(n,\eta)$, where $\hat{A}(n,\eta)$ satisfies $\sup_{n\ge 1} E\{\hat{A}(n,\eta)^s\} < \infty$ for each $s \ge 1$. (I.45)

By (I.6) we have $n^{-1+\eta'}\, \theta_{r_0}^{-2} \to 0$. Put $\hat{A}_r(n) = n^{-1+2\eta}\, \theta_r^{-2}\, \hat{A}(n,\eta)$. Then, for $\eta = \frac{1}{2}\eta'$, we have $\hat{A}_r(n) \le \text{const.}\ \hat{A}(n,\eta)$, for each $1 \le r \le r_0$. Thus, (I.45) implies that for some $\eta > 0$,

for all $1 \le r \le r_0$, $S_2 \le \text{const.}\ n^{-\eta}\, r\, \theta_r^{-1}\, \hat{A}(n)$, where $\hat{A}(n) = \max_{1 \le r \le r_0} \hat{A}_r(n)$ satisfies $\sup_{n \ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$. (I.46)

Next we bound S_3, noting first that

$$\Big| \int g(\widehat{\psi}_j - \psi_j) \Big| \le U_{1j} + \theta_j\, U_{2j} + U_{3j} + \theta_j\, U_{4j}\,, \tag{I.47}$$

where, using (I.26),

$$U_{1j} = \Big| \sum_{k\,:\,k\neq j} b_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big|,\ U_{2j} = \Big| \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-1}\, b_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big|,$$

$$U_{3j} = \Big| \int (g - \theta_j\, b)\, \widehat{\Delta}_j \Big| \text{ and } U_{4j} = \Big| \int b\, \widehat{\Delta}_j \Big|.$$

Now,

$$U_{1j}^2 \leq 2 \Big| \int (\widehat{K} - K)\, b\, \psi_j \Big|^2 + 2\, b_j^2 \Big| \int (\widehat{K} - K)\, \psi_j\, \psi_k \Big|^2 \leq 4 \left(\int b^2 \right) \Delta_{(j)}^2 . \quad \text{(I.48)}$$

Recall that

$$\Delta_{(j)}^2 = \| \int (\widehat{K} - K)\, \psi_j \|^2 = \sum_{k=1}^{\infty} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2$$

By Lemma 2.5 we have $E\left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^{2s} = O\big((n^{-1}\, \theta_j\, \theta_k)^s \big)$. Therefore,

$$\begin{aligned} E[\Delta_{(j)}]^{2s} &\leq \left(\sum_{k=1}^{\infty} \left[E\left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^{2s} \right]^{1/s} \right)^s \\ &\leq \left(C_s\, n^{-1}\, \theta_j \sum_{k=1}^{\infty} \theta_k \right)^s = O\big((n^{-1}\, \theta_j)^s \big) . \end{aligned}$$

An argument similar to that used to derive (I.43) can be employed to prove that, defining $\hat{V} = \max_{1 \leq j \leq r_0} n\, \theta_j^{-1} \Delta_{(j)}^2$, we have

$$E(\hat{V}^s) \leq \text{const.}\, r_0 . \quad \text{(I.49)}$$

for each $s \geq 1$, the constant not depending on n. Moreover, using (I.48),

$$\begin{aligned}
\sum_{j=1}^{r} \theta_j^{-4} U_{1j}^2 &\leq \text{const.}\, n^{-1} \sum_{j=1}^{r} \theta_j^{-3} \left(n\, \theta_j^{-1}\, \Delta_{(j)}^2\right) \\
&\leq \text{const.}\, n^{-1} \left(\max_{1 \leq j \leq r_0} n\, \theta_j^{-1}\, \Delta_{(j)}^2\right) \sum_{j=1}^{r} \theta_j^{-3} \\
&\leq \text{const.}\, n^{-1+\eta} \sum_{j=1}^{r} \theta_r^{-3}\, \hat{A}(n, \eta),
\end{aligned}$$

where $\hat{A}(n,\eta) = n^{-\eta}\, \hat{V}(n)$ satisfies $\sup_{n \geq 1} E\{\hat{A}(n,\eta)^s\} < \infty$ for each $s \geq 1$. This leads to the following analogue of (I.45): for each $\eta > 0$,

$$\begin{aligned}
&\text{for all } 1 \leq r \leq r_0,\ \textstyle\sum_{j=1}^{r} \theta_j^{-4} U_{1j}^2 \leq \text{const.}\, n^{-1+\eta} \sum_{j=1}^{r} \theta_r^{-3}\, \hat{A}(n,\eta), \\
&\text{where } \hat{A}(n,\eta) \text{ satisfies } \sup_{n \geq 1} E\{\hat{A}(n,\eta)^s\} < \infty, \text{ for each } s \geq 1.
\end{aligned} \tag{I.50}$$

With $s(j)$ as at (I.33) we have, for some $\eta > 0$,

$$\begin{aligned}
U_{2j}^2 &= \left| \sum_{k\,:\,k \neq j} (\theta_j - \theta_k)^{-1}\, b_k \int (\widehat{K} - K)\, \psi_j\, \psi_k \right|^2 \\
&\leq \left(\sum_{k\,:\,k \neq j} (\theta_j - \theta_k)^{-2}\, b_k^2 \right) \left(\sum_{k\,:\,k \neq j} \left\{ \int (\widehat{K} - K)\, \psi_j\, \psi_k \right\}^2 \right) \\
&\leq s(j) \sum_{k=1}^{\infty} \left(\int (\widehat{K} - K)\, \psi_j\, \psi_k \right)^2 \\
&= s(j)\, \Delta_{(j)}^2 \leq \text{const.}\, j^{1-\eta}\, \Delta_{(j)}^2,
\end{aligned}$$

where the last inequality used (I.33). The argument employed to obtain (I.47) now implies that for each $\eta > 0$,

$$\text{for all } 1 \le r \le r_0,\ \sum_{j=1}^{r} \theta_j^{-2} U_{2j}^2 \le \text{const.}\ n^{-1} \sum_{j=1}^{r} \theta_j^{-1} \{ n\, \theta_j^{-1} \times \Delta_{(j)}^2 \} j^{1-\eta} \ \le\ \text{const.}\, n^{-1} \max_{1\le j\le r_0} \{ n\, \theta_j^{-1}\, \Delta_{(j)}^2 \} \sum_{j=1}^{r} \theta_j^{-1}\, j^{1-\eta} \ \le\ \text{const.}\ n^{-1+\eta}\, n^{-\eta}\, \hat V(n) (r^{1-\eta}\, \theta_r^{-1}\, r) \ \le\ \text{const.}\ n^{-1+\eta}\, r^2\, \theta_r^{-1} \hat A(n,\eta), \tag{I.51}$$

where $\hat A(n,\eta) = n^{-\eta}\, \hat V(n)$ satisfies $\sup_{n\ge1} E\{\hat A(n,\eta)^s\} < \infty$ for each $s \ge 1$.

Using the bound (I.34) and the argument leading to (I.36) we have

$$\sum_{j=1}^{r} \theta_j^{-4}\, U_{3j}^2 \le \text{const.} \sum_{j=1}^{r} \theta_j^{-4} \left(\widehat\Delta^4 \sum_{k\,:\,k\ne j} b_k^2\, (\theta_j - \theta_k)^{-2} + \widehat\Delta^2\, \|\widehat\psi_j - \psi_j\|^2 \int b^2 \right)$$
$$\le \text{const.}\, n^{-2+\eta}\, r^3\, \theta_r^{-4}\, \hat A(n,\eta)\,,$$

where $\hat A(n,\eta)$ satisfies $\sup_{n\ge1} E\{\hat A(n,\eta)^s\} < \infty$ for each $s \ge 1$. Therefore, for $\eta > 0$,

$$\text{if } \mathcal{G} \text{ holds for all } 1 \le r \le r_0,\ \sum_{j=1}^{r} \theta_j^{-4} U_{3j}^2 \le \text{const.}\, n^{-2+\eta}\, r^3 \theta_r^{-4} \times \hat A(n,\eta), \text{ where } \hat A(n,\eta) \text{ satisfies } \sup_{n\ge1} E\{\hat A(n,\eta)^s\} < \infty \text{ for each } s \ge 1. \tag{I.52}$$

Combining (I.50)-(I.52), we deduce that for each $\eta > 0$,

$$\sum_{j=1}^{r} \{ \theta_j^{-4} (U_{1j}^2 + U_{3j}^2) + \theta_j^{-2} U_{2j}^2 \} \le \text{const.}\ (n^{-1+\eta}\, r\, \theta_r^{-3} + n^{-2+\eta}\, r^3\, \theta_r^{-4} + n^{-1+\eta}\, r^2\, \theta_r^{-1}) \hat A(n,\eta)$$
$$\le \text{const.}\ \left[(n^{-1+\eta}\, \theta_r^{-2})(r\, \theta_r^{-1}) + (n^{-1+\eta}\, r^2\, \theta_r^{-2})\, (n^{-1}\, \theta_r^{-1})\, r\, \theta_r^{-1} + (n^{-1+\eta}\, r)\, r\, \theta_r^{-1} \right] \hat A(n,\eta)\,.$$

Thus, by (I.6), for some $\eta > 0$,

$$\text{if } \mathcal{G} \text{ holds for all } 1 \le r \le r_0,\ \sum_{j=1}^{r}\{\theta_j^{-4}(U_{1j}^2 + U_{3j}^2) + \theta_j^{-2}U_{2j}^2\} \le \text{const. } n^{-\eta}\, r\, \theta_r^{-1}\, \hat{A}(n) \text{ where } \hat{A}(n) \text{ satisfies } \sup_{n\ge1} E\{\hat{A}(n)^s\} < \infty \text{ for each } s \ge 1. \tag{I.53}$$

To bound U_{4j} we first use the argument leading to (I.33) to show that, for some $\eta > 0$,

$$\begin{aligned}
u(j) \equiv \sum_{k\,:\,k\ne j} b_k^2(\theta_j - \theta_k)^{-4} &\le \text{const.} \sum_{k\,:\,k\ne j} b_k^2 \left\{ \frac{\max(j,k)}{|j-k|\max(\theta_j,\theta_k)} \right\}^4 \\
&\le \text{const. } \theta_j^{-2} \sum_{k\,:\,k\ne j} b_k^2 \{\max(\theta_j,\theta_k)\}^{-2} \left(\frac{\max(j,k)}{|j-k|} \right)^4 \\
&\le \text{const. } \theta_j^{-2} \left\{ \sum_{k\le j/2,\,k>2j} b_k^2\,\theta_k^{-2} + j^4 \sum_{j/2<k\le 2j} b_k^2\,\theta_k^{-2}\,(j-k)^{-4} \right\} \\
&\le \theta_j^{-2}\,(1 + b_{[j/2]}^2\,\theta_{[j/2]}^{-2}\,j^4)\,,
\end{aligned} \tag{I.54}$$

where $[j/2]$ denotes the integer part of $j/2$. Next we argue as in Section (I.2.4), obtaining, provided $\mathcal{G}$ holds and $1 \le j \le r_0$:

$$\begin{aligned}
U_{4j}^2 = \Bigg| &\sum_{k\,:\,k\ne j} \left\{ \left(1 + \frac{\hat{\theta}_j - \theta_j}{\theta_j - \theta_k}\right)^{-1} - 1 \right\} \frac{b_k}{\theta_j - \theta_k} \int (\widehat{K} - K)\,\widehat{\psi}_j\,\psi_k \\
&+ \sum_{k\,:\,k\ne j} \frac{b_k}{\theta_j - \theta_k} \int (\widehat{K} - K)\,(\widehat{\psi}_j - \psi_j)\,\psi_k + b_j \int (\widehat{\psi}_j - \psi_j)\,\psi_j \Bigg|^2,
\end{aligned}$$

and then,

$$\begin{aligned}
U_{4j}^2 &\le 12 \left\{ \sum_{k\,:\,k\ne j} b_k^2\,(\theta_j - \theta_k)^{-4} \right\} \widehat{\Delta}^2 \left\| \int (\widehat{K} - K)\,\widehat{\psi}_j \right\|^2 \\
&\quad + 3 \left\{ \sum_{k\,:\,k\ne j} b_k^2\,(\theta_j - \theta_k)^{-2} \right\} \left\| \int (\widehat{K} - K)(\widehat{\psi}_j - \psi_j) \right\|^2 + 3\,b_j^2\,\|\widehat{\psi}_j - \psi_j\|^2 \\
&\le 24\,u(j)\,\widehat{\Delta}^2\,(\widehat{\Delta}^2\,\|\widehat{\psi}_j - \psi_j\|^2 + \Delta_{(j)}^2) + 3\,\{s(j)\,\widehat{\Delta}^2 + b_j^2\}\,\|\widehat{\psi}_j - \psi_j\|^2\,.
\end{aligned} \tag{I.55}$$

In Section (I.2.8) we shall show that (I.55) implies that for each $\eta > 0$,

$$\sum_{j=1}^{r} \theta_j^{-2} U_{4j}^2 \le \text{const. } n^\eta \hat{A}_1(n,\eta) \sum_{j=1}^{r} \theta_j^{-2} \{ \theta_j^{-2} (1 + b_{[j/2]}^2 \theta_{[j/2]}^{-2} j^4)(n^{-3} j^2 + n^{-2} \theta_j)$$

$$+ n^{-2} j^3 + n^{-1} b_j^2 j^2 \}$$

$$\le \text{const. } n^\eta \hat{A}_1(n,\eta) \{ \theta_r^{-2} (n^{-3} r^2 + n^{-2} \theta_r) \sum_{j=1}^{r} \theta_j^{-2} (1 + b_{[j/2]}^2 \theta_{[j/2]}^{-2} j^4)$$

$$+ r (n^{-2} r^3 + n^{-1} b_r^2 r^2) \}$$

$$\le \text{const. } n^\eta \hat{A}_1(n,\eta) r^{1+\eta} \{ \theta_r^{-4} (1 + b_r^2 \theta_r^{-2} r^4)(n^{-3} r^2 + n^{-2} \theta_r)$$

$$+ r (n^{-2} r^3 + n^{-1} b_r^2 r^2) \}$$

$$\le \text{const. } n^{2\eta} \hat{A}_1(n,\eta) r \{ \theta_r^{-4} (1 + b_r^2 \theta_r^{-2} r^4)(n^{-3} r^2 + n^{-2} \theta_r)$$

$$+ \theta_r^{-2} (n^{-2} r^3 + n^{-1} b_r^2 r^2) \},$$

where the random variable $\hat{A}_1(n,\eta)$ satisfies $\sup_{n \ge 1} E\{\hat{A}_1(n,\eta)^s\} < \infty$ for each $s \ge 1$. Hence,

$$n^{-1} \sum_{j=1}^{r} \theta_j^{-2} U_{4j}^2 \le \text{const. } n^{-1+\eta} \hat{A}_1(n,\eta) r \theta_r^{-1} \Big\{ n^{-3+\eta} r^2 \theta_r^{-3} + n^{-2+\eta} \theta_r^{-2}$$

$$+ n^{-3+\eta} r^6 \theta_r^{-5} b_r^2 + \theta_r^{-4} n^{-2+\eta} b_r^2 r^4 + \theta_r^{-1} n^{-2+\eta} r^3 + \theta_r^{-1} n^{-1+\eta} b_r^2 r^2 \Big\}.$$

Therefore,

$$n^{-1} \sum_{j=1}^{r} \theta_j^{-2} U_{4j}^2 \le \text{const. } n^{-1-\eta} \hat{A}_1(n,\eta) r \theta_r^{-1} \Big\{ (n^{-1+2\eta} r^2 \theta_r^{-2})(n^{-2+\eta} \theta_r^{-1}) + n^{-2+3\eta} \theta_r^{-2}$$

$$+ (n^{-1+\eta} r \theta_r^{-1})(r b_r^2 \theta_r^{-2})(n^{-1+2\eta} r^2 \theta_r^{-1})^2 + (r b^2 \theta_r^{-2})(n^{-1+2\eta} \theta_r^{-2} r^2)(n^{-1+\eta} r)$$

$$+ (n^{-1+2\eta} r^2 \theta_r^{-1})(n^{-1+\eta} r) + (n^{-1+3\eta} r \theta_r^{-1})(b_r^2 r) \Big\}. \tag{I.56}$$

Using (I.3) and (I.6) in conjunction with (I.56) we deduce that for some $\eta > 0$, for all $1 \le r \le r_0$, $n^{-1} \sum_{j=1}^{r} \theta_j^{-4} U_{4j}^2 \le \text{const. } n^{-1-\eta} r \theta_r^{-1} \hat{A}(n)$, where $\hat{A}(n)$ satisfies $\sup_{n \ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$.

Combining this result with (I.47) and (I.53), we conclude that for some $\eta > 0$,

$$
\begin{aligned}
S_3 = \sum_{j=1}^{r} \theta_j^{-4} \left\{ \int g(\widehat{\psi}_j - \psi_j) \right\}^2 &\le \sum_{j=1}^{r} \theta_j^{-4} \{U_{1j} + \theta_j U_{2j} + U_{3j} + \theta_j U_{4j}\}^2 \\
&\le 2 \sum_{j=1}^{r} \theta_j^{-4} \{U_{1j}^2 + \theta_j^2 U_{2j}^2 + U_{3j}^2 + \theta_j^2 U_{4j}^2\} \\
&= 2 \left\{ \sum_{j=1}^{r} \theta_j^{-4} U_{1j}^2 + \sum_{j=1}^{r} \theta_j^{-2} U_{2j}^2 + \sum_{j=1}^{r} \theta_j^{-4} U_{3j}^2 + \sum_{j=1}^{r} \theta_j^{-2} U_{4j}^2 \right\} \\
&\le \text{const. } \left\{ n^{-\eta} r \theta_r^{-1} \hat{A}(n) + n^{-1-\eta} r \theta_r^{-1} \hat{A}(n) \right\}.
\end{aligned}
$$

Consequently, for some $\eta > 0$,

if $\mathcal{G}$ holds for all $1 \le r \le r_0$, $S_3 \le \text{const. } n^{-\eta} r \theta_r^{-1} \hat{A}(n)$, where $\hat{A}(n)$ satisfies $\sup_{n \ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$. (I.57)

Define $\|.\|_s = \left[E(.)^s\right]^{1/s}$ for $s \ge 1$. Then from (I.42) we obtain:

$$
\begin{aligned}
E\|\hat{g} - g\|^{2s} = E\left(\|\hat{g} - g\|^2\right)^s &= E\left[\sum_{j=1}^{\infty} \left\{ \int (\hat{g} - g)\, \psi_j \right\}^2 \right]^s \\
&\le \left(\sum_{j=1}^{\infty} \left\| \left\{ \int (\hat{g} - g)\, \psi_j \right\}^2 \right\|_s \right)^s \\
&= \left(\sum_{j=1}^{\infty} \left[E\left\{ \int (\hat{g} - g)\, \psi_j \right\}^{2s} \right]^{1/s} \right)^s \\
&\le \text{const. } \left(\sum_{j=1}^{\infty} \left[(\theta_j/n)^s\right]^{1/s} \right)^s \\
&= \text{const. } n^{-s} \left(\sum_{j=1}^{\infty} \theta_j \right)^s = O(n^{-s}).
\end{aligned}
$$

Since by Theorem 2.5, we have $E\widehat{\Delta}^{2s} = O(n^{-s})$, now we conclude that $E\|\hat{g} -$

$g\|^{2s} + E\widehat{\Delta}^{2s} = O(n^{-s})$. Hence, noting (I.39)–(I.41), (I.46) and (I.57) lead to

$$\begin{aligned}\sum_{j=1}^{r} T_{2j}^2 &\le \text{const.}\left(\|\hat{g} - g\|^2 S_1 + \widehat{\Delta}^2 (S_2 + S_3) + \widehat{\Delta}^4 S_4\right)\\ &\le \text{const.}\left\{\left(n^{-\eta} r\, \theta_r^{-1} \hat{A}(n)\right)\left(\|\hat{g} - g\|^2 + \widehat{\Delta}^2\right) + (n^{-2} S_4)\, n^2\, \widehat{\Delta}^4\right\}\\ &\le \text{const.}\left\{\left(n^{-\eta} r\, \theta_r^{-1} \hat{A}(n)\right)\left(\|\hat{g} - g\|^2 + \widehat{\Delta}^2\right) + (n^{-\eta} r\, b_r^2)\, n^2\, \widehat{\Delta}^4\right\}\\ &\le \text{const.}\, n^{-\eta} r \{n^{-1} \theta_r^{-1} + b_r^2\} \hat{A}_1(n)\,,\end{aligned}$$

where $\hat{A}_1(n)$ satisfies $\sup_{n\ge 1} E\{\hat{A}_1(n)^s\} < \infty$, for each $s \ge 1$, and the last inequality was obtained by using the fact that $E\|\hat{g} - g\|^{2s} + E\widehat{\Delta}^{2s} = O(n^{-s})$. Therefore,

if $\mathcal{G}$ holds for all $1 \le r \le r_0$, $\sum_{j=1}^{r} T_{2j}^2 \le$ const. $n^{-\eta}\, r\, (n^{-1}\, \theta_r^{-1} +$ $b_r^2)\, \hat{A}(n)$, where $\hat{A}(n)$ satisfies $\sup_{n\ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$. (I.58)

I.2.6. *Approximation to second term on right-hand side of* (I.10). Define

$$U_j = \int b\, (\widehat{\psi}_j - \psi_j),$$

in which notation $\bar{b}_j = b_j + U_j$. Therefore,

$$\sum_{j=r+1}^{\infty} b_j^2 - \sum_{j=r+1}^{\infty} \bar{b}_j^2 = \sum_{j=1}^{r} \bar{b}_j^2 - \sum_{j=1}^{r} b_j^2 = \sum_{j=1}^{r} U_j^2 + 2 \sum_{j=1}^{r} b_j\, U_j\,. \tag{I.59}$$

Note that, by (I.30), $U_j = U_{1j} + \cdots + U_{4j}$, where, redefining the U_{jk}'s from

Section (I.2.5),

$$U_{1j} = \sum_{k\,:\,k\neq j} b_k\,(\theta_j-\theta_k)^{-1}\int(\widehat{K}-K)\,\psi_j\,\psi_k\,,$$
$$U_{2j} = \sum_{k\,:\,k\neq j} b_k\,\{(\hat\theta_j-\theta_k)^{-1}-(\theta_j-\theta_k)^{-1}\}\int(\widehat{K}-K)\,\psi_j\,\psi_k\,,$$
$$U_{1j} = \sum_{k\,:\,k\neq j} b_k\,(\hat\theta_j-\theta_k)^{-1}\int(\widehat{K}-K)\,(\widehat{\psi}_j-\psi_j)\,\psi_k\,,$$
$$U_{4j} = b_j\int(\widehat{\psi}_j-\psi_j)\,\psi_j\,.$$

We shall develop bounds for $V_{\ell m} = \sum_{j=1}^r U_{\ell j}^2$ and $W_{\ell m} = |\sum_{j=1}^r b_j U_{\ell j}|$. In view of (I.59),

$$\left|\sum_{j=r+1}^{\infty} b_j^2 - \sum_{j=r+1}^{\infty}\bar b_j^2\right| \le 4\sum_{\ell=1}^{4}(V_{\ell j}+W_{\ell j})\,. \tag{I.60}$$

First we bound terms involving U_{1j}. With $s(j)$ defined as at (I.33), we have:

$$U_{1j}^2 \le s(j)\sum_{j=1}^{\infty}\left(\int(\widehat{K}-K)\,\psi_j\,\psi_k\right)^2 = s(j)\int\left\{\int(\widehat{K}-K)(u,v)\,\psi_j(v)\,dv\right\}^2 du.$$

Therefore, using (I.33), we obtain for some $\eta>0$:

$$\begin{aligned} V_{1r} &\le \text{const.}\sum_{j=1}^{r} j^{1-\eta}\int\left\{\int(\widehat{K}-K)(u,v)\,\psi_j(v)\,dv\right\}^2 du \\ &\le r^{1-\eta}\sum_{j=1}^{\infty}\int\left\{\int(\widehat{K}-K)(u,v)\,\psi_j(v)\,dv\right\}^2 du \\ &= \text{const.}\,r^{1-\eta}\,\widehat{\Delta}^2\,. \end{aligned} \tag{I.61}$$

Note that by Parseval's identity $\widehat{\Delta}^2 = \sum_{j=1}^{\infty}\widehat{\Delta}_j^2$, by which we have obtained the last identity above. Also, if $A_{jk} = b_j\,b_k\,(\theta_j-\theta_k)^{-1}\int(\widehat{K}-K)\,\psi_j\,\psi_k$, then because

$A_{jk} = -A_{kj}$, by antisymmetry and thence (I.5), for some integer $m > 0$ we have

$$\begin{aligned}
W_{1r} &= \left| \sum_{j=1}^{r} \sum_{k=m+1}^{\infty} b_j\, b_k\, (\theta_j - \theta_k)^{-1} \int (\widehat{K} - K)\, \psi_j\, \psi_k \right| \\
&\leq \left| \sum_{j=1}^{r} \sum_{k=m+1}^{\infty} b_j\, b_k\, \{(\theta_j - \theta_k)^{-1} - \theta_j^{-1}\} \int (\widehat{K} - K)\, \psi_j\, \psi_k \right| \\
&\quad + \left| \sum_{j=1}^{r} \sum_{k=m+1}^{\infty} b_j\, b_k\, \theta_j^{-1} \int (\widehat{K} - K)\, \psi_j\, \psi_k \right| \\
&\leq \text{const.}\, \{\widehat{\Delta}\, W_1(r) + W_2(r)\},
\end{aligned}$$

where, using (I.1),

$$\begin{aligned}
W_1(r) &= \sum_{j=1}^{r} \sum_{k=m+1}^{\infty} |b_j\, b_k|\, \theta_j^{-2}\, \theta_k = \left(\left| \sum_{j=1}^{r} |b_j|\, \theta_j^{-2} \right) \left(\sum_{k=m+1}^{\infty} |b_k|\, \theta_k \right) \\
&\leq \text{const.}\, (1 + |b_r|\, \theta_r^{-2}\, r)\, |b_r|\, \theta_r\, r^{\eta},
\end{aligned}$$

$$\begin{aligned}
W_2(r) &= \left| \int (\widehat{K} - K) \left(\sum_{j=1}^{r} b_j\, \theta_j^{-1}\, \psi_j \right) \left(\sum_{k=m+1}^{\infty} b_k\, \psi_k \right) \right| \\
&\leq \left(\left[\int \left\{ \int (\widehat{K} - K)(u, v) \sum_{j=1}^{r} b_j\, \theta_j^{-1}\, \psi_j(u)\, du \right\}^2 dv \right] \right)^{1/2} \left(\sum_{j=r+1}^{\infty} b_k^2 \right)^{1/2} \\
&\leq \widehat{\Delta} \left(\sum_{j=1}^{r} b_j^2\, \theta_j^{-2} \right)^{1/2} \left(\sum_{j=r+1}^{\infty} b_k^2 \right)^{1/2} \leq \text{const.}\, \widehat{\Delta} \left(\sum_{j=r+1}^{\infty} b_k^2 \right)^{1/2} \\
&\leq \text{const.}\, \widehat{\Delta}\, |b_r|\, r^{1/2+\eta}\, .
\end{aligned}$$

where we have used (I.3) to obtain the second-last inequality. Therefore, for all $\eta > 0$,

$$W_{1r} \leq \text{const.}\, \widehat{\Delta} \left(|b_r|\, r^{1/2} + |b_r|\, \theta_r + b_r^2\, \theta_r^{-1}\, r \right) r^{\eta}\, . \tag{I.62}$$

To bound terms involving U_{2j}, observe that if $\mathcal{G}$ holds and $r \le r_0$ then

$$U_{2j}^2 \le 4 \left\{ \sum_{k\,:\,k \ne j} b_k^2 \frac{(\hat\theta_j - \theta_j)^2}{(\theta_j - \theta_k)^4} \sum_{k=1}^{\infty} \left(\int (\widehat{K} - K)\, \psi_j\, \psi_k \right)^2 \le 4\, u(j)\, \widehat{\Delta}^2\, \Delta_{(j)}\,, \quad \text{(I.63)}$$

where $u(j)$ is as at (I.54). In view of (I.1) and (I.54),

$$\begin{aligned}
\sum_{j=1}^{r} u(j)\, \theta_j &\le \text{const.} \sum_{j=1}^{r} \theta_j^{-1} \,(1 + b_{[j/2]}^2\, \theta_{[j/2]}^{-2}\, j^4) \\
&\le \text{const.}\, \theta_r^{-1} \big(r + \sum_{j=1}^{2r} b_{[j/2]}^2\, \theta_{[j/2]}^{-2}\, j^4\big) \le \text{const.}\, r^{1+\eta}\, \theta_r^{-1}\, \big(1 + b_r^2\, \theta_r^{-2}\, r^4\big) \\
&\le \text{const.}\, n^{\eta}\, \theta_r^{-1}\, (1 + b_r^2\, \theta_r^{-2}\, r^4)\, r,
\end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^{r} |b_j|\, u(j)^{1/2}\, \theta_j^{1/2} &\le \text{const.} \sum_{j=1}^{r} |b_j|\, \theta_j^{-1/2} (1 + |b_{[j/2]}|\, \theta_{[j/2]}^{-1}\, j^2) \\
&\le \text{const.} \sum_{j=1}^{r} |b_j|\, \theta_j^{-1/2}\, j\, (j^{-1} + |b_{[j/2]}|\, \theta_{[j/2]}^{-1}\, j)\,,
\end{aligned}$$

and then,

$$\begin{aligned}
\sum_{j=1}^{r} |b_j|\, u(j)^{1/2}\, \theta_j^{1/2} &\le \text{const.} \left(\sum_{j=1}^{r} b_j^2\, \theta_j^{-1}\, j^2 \right)^{1/2} \left(\sum_{j=1}^{r} (j^{-1} + |b_{[j/2]}|\, \theta_{[j/2]}^{-1}\, j)^2 \right)^{1/2} \\
&\le \text{const.} \left(\sum_{j=1}^{r} b_j^2\, \theta_j^{-1}\, j^2 \right)^{1/2} \left(2 \sum_{j=1}^{r} (j^{-2} + b_{[j/2]}^2\, \theta_{[j/2]}^{-2}\, j^2) \right)^{1/2} \\
&\le \text{const.} \left(\sum_{j=1}^{r} b_j^2\, \theta_j^{-1}\, j^2 \right)^{1/2} \left(1 + \sum_{j=1}^{2r} b_{[j/2]}^2\, \theta_{[j/2]}^{-2}\, j^2 \right)^{1/2}.
\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{j=1}^{r} |b_j|\, u(j)^{1/2}\, \theta_j^{1/2} &\le \text{const.}\, (1 + b_r^2\, \theta_r^{-1}\, r^{2+1+\eta})^{1/2}\, (1 + b_r^2\, \theta_r^{-2}\, r^{2+1+\eta})^{1/2} \\ &\le \text{const.}\, r^{\eta}\, (1 + |b_r|\, \theta_r^{-1/2}\, r)(1 + |b_r|\, \theta_r^{-1}\, r)\, r \\ &\le \text{const.}\, n^{\eta}\, (1 + |b_r|\, \theta_r^{-1/2}\, r + |b_r|\, \theta_r^{-1}\, r + b_r^2\, \theta_r^{-3/2}\, r^2)\, r \\ &\le \text{const.}\, n^{\eta}\, (1 + |b_r|\, \theta_r^{-1}\, r + |b_r|\, \theta_r^{-1}\, r + b_r^2\, \theta_r^{-3/2}\, r^2)\, r \\ &\le \text{const.}\, n^{\eta}(1 + |b_r|\, \theta_r^{-1}\, r + b_r^2\, \theta_r^{-3/2}\, r^2)\, r\,,\end{aligned}$$

where $\eta > 0$ is arbitrary and, in each string of inequalities, the last constant depends on η. From these results, (I.63) and (I.49), we deduce that for each $\eta > 0$,

$$\begin{aligned}&\text{for all } 1 \le r \le r_0,\ V_{2m} \le \text{const.}\, n^{-2+\eta}\, \theta_r^{-1}\, (1+b_r^2\, \theta_r^{-2}\, r^4)\, r\, \hat{A}(n,\eta), \\ &W_{2m} \le \text{const.}\, n^{-1+\eta}(1 + |b_r|\, \theta_r^{-1}\, r + b_r^2\, \theta_r^{-3/2}\, r^2)\, r\, \hat{A}_2(n,\eta), \text{ where} \\ &\text{the random variable } \hat{A}_j(n,\eta) \text{ satisfies } \sup_{n\ge1} E\{\hat{A}_j(n,\eta)^s\} < \infty \\ &\text{for each } s \ge 1.\end{aligned} \tag{I.64}$$

Next we bound terms involving U_{3j}. If $\mathcal{G}$ holds then, noting that $s(j)$ is defined at (I.33), we have:

$$\begin{aligned}|U_{3j}| &\le \text{const.} \sum_{k\,:\,k\ne j} |b_k|\, |\theta_j - \theta_k|^{-1} \left| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k \right| \\ &\le \text{const.}\, s(j)^{1/2} \left\| \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j) \right\| \le \text{const.}\, \widehat{\Delta}\, \|\widehat{\psi}_j - \psi_j\|\, s(j)^{1/2}\,.\end{aligned}$$

Hence, using (I.33), we obtain:

$$|U_{3j}| \le \text{const.}\, j^{1/2}\, \widehat{\Delta}\, \|\widehat{\psi}_j - \psi_j\|\,. \tag{I.65}$$

We shall prove in Section (I.2.8) that this result implies that for each $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \le r \le r_0$, $V_{3r} \le \text{const.}\, n^{-2+\eta}\, r^4\, \hat{A}_1(n, \eta)$, and $W_{3r} \le \text{const.}\, n^{-1+\eta}\, (1+|b_r|\, r^{5/2})\, \hat{A}_2(n, \eta)$, where random variable $\hat{A}_j(n, \eta)$ satisfies $\sup_{n\ge1} E\{\hat{A}_j(n, \eta)^s\} < \infty$ for each $s \ge 1$. (I.66)

Also in Section (I.2.8) we shall show that for each $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \le r \le r_0$, $V_{4j} \le \text{const.}\, n^{-2+\eta}\, (1 + b_r^2\, r^5)\, \hat{A}_1(n, \eta)$, and $W_{4j} \le \text{const.}\, n^{-1+\eta}(1+|b_r^2|\, r^3)\, \hat{A}_2(n, \eta)$, where random variable $\hat{A}_j(n, \eta)$ satisfies $\sup_{n\ge1} E\{\hat{A}_j(n, \eta)^s\} < \infty$ for each $s \ge 1$. (I.67)

Combining (I.60)–(I.62), (I.64), (I.66) and (I.67), and using (I.3) and (I.6) to simplify the right-hand sides of these formulae, we conclude that, for some $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \le r \le r_0$, $|\sum_{j>r}(b_j^2 - \bar{b}_j^2)| \le \text{const.}\, n^{-\eta}\, r\, (n^{-1}\, \theta_r^{-1} + b_r^2)\, \hat{A}(n)$, where $\sup_{n\ge1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$. (I.68)

I.2.7. *Combining earlier results.* First we develop an approximation to $\sum_{j=1}^{r} T_{3j}^2$. Note, from the definition of W at (I.7) and from (I.27) that $T_{3j} = \theta_j^{-1}\, Z_j$, where

$$Z_j = \frac{1}{n} \sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)(\epsilon_i - \bar{\epsilon})$$

and $\xi_{ij}, \epsilon_i, \bar{\xi}_j$ and $\bar{\epsilon}$ are as in Section (I.2.5). Defining $\widehat{W} = \max_{1\le j\le r_0} n\, \theta_j^{-1}\, Z_j^2$, we have:

$$\sum_{j=1}^{r} T_{3j}^2 \le n^{-1}\, \widehat{W} \sum_{j=1}^{r} \theta_j^{-1} \le \text{const.}\, n^{-1}\, r\, \theta_r^{-1}\, \widehat{W}\,. \tag{I.69}$$

Using an argument similar to that leading to (I.43), it can be shown that for each $s \ge 1$, $E(\widehat{W}^s) \le \text{const.}\, r_0$, the constant not depending on n. Nevertheless, we prove it as follows. We have $\widehat{W}^s = \big(\max_{1\le j\le r_0} n\,\theta_j^{-1}\,Z_j^2\big)^s \le \sum_{j=1}^{r_0} n^s\,\theta_j^{-s}\,Z_j^{2s}$, and then

$$E\big[\widehat{W}^s\big] \le \sum_{j=1}^{r_0} n^s\,\theta_j^{-s}\,E\big[Z_j^{2s}\big]\,.$$

Also,

$$\begin{aligned}
E\big[Z_j^{2s}\big] &= E\Big[\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\,\epsilon_i - \bar{\xi}_j\,\bar{\epsilon}\Big]^{2s} \\
&\le C_s\Big\{E\Big[\Big|\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\,\epsilon_i\Big|^{2s}\Big] + E\big|\bar{\xi}_j\,\bar{\epsilon}\big|^{2s}\Big\}\,. \qquad\qquad (\text{I.70})
\end{aligned}$$

Using Rosenthal's inequality for the first term on the right-hand side of (I.70), and noting (I.4), give:

$$\begin{aligned}
E\Big[\Big|\frac{1}{n}\sum_{i=1}^{n}\xi_{ij}\,\epsilon_i\Big|^{2s}\Big] &\le n^{-2s}\,C_{1s}\Big\{\sum_{i=1}^{n}E|\xi_{ij}\,\epsilon_i|^{2s} + \Big(\sum_{i=1}^{n}E|\xi_{ij}\,\epsilon_i|^2\Big)^s\Big\} \\
&= n^{-2s}\,C_{1s}\Big\{\sum_{i=1}^{n}E|\xi_{ij}|^{2s}\,E|\epsilon_i|^{2s} + \Big(\sum_{i=1}^{n}E\big[\xi_{ij}^2\big]\,E\big[\epsilon_i^2\big]\Big)^s\Big\} \\
&\le C_{2s}\,n^{-s}\,\theta_j^s\,, \qquad\qquad (\text{I.71})
\end{aligned}$$

where C_{1s} and C_{2s} are constants which depend only on s. The second term on the right-hand side of (I.70) leads to:

$$\begin{aligned}
E\big|\bar{\xi}_j\,\bar{\epsilon}\big|^{2s} = E\big|\bar{\xi}_j\big|^{2s}\,E|\bar{\epsilon}|^{2s} &\le \Big[n^{-2s}\,C_{1s}\Big\{\sum_{i=1}^{n}E|\xi_{ij}|^{2s} + \Big(\sum_{i=1}^{n}E|\xi_{ij}|^2\Big)^s\Big\}\Big]\times \\
&\qquad \Big[n^{-2s}\,C_{2s}\Big\{\sum_{i=1}^{n}E|\epsilon_i|^{2s} + \Big(\sum_{i=1}^{n}E|\epsilon_i|^2\Big)^s\Big\}\Big] \\
&\le C_{3s}\,n^{-s}\,\theta_j^s \times C_{4s}\,n^{-s} \\
&\le C_{5s}\,n^{-2s}\,\theta_j^s\,, \qquad\qquad (\text{I.72})
\end{aligned}$$

where C_{1s}, C_{2s}, C_{3s}, C_{4s} and C_{5s} are constants which depend only on s. If we combine the results (I.70)–(I.72), we deduce that for each $s \geq 1$, $E[Z_j^{2s}] \leq$ const. $(\theta_j n^{-1})^s$. Therefore, for each $s \geq 1$, $E(\hat{W}^s) \leq$ const. r_0, where the constant does not depend on n. Hence, by (I.69), for each $\eta > 0$,

for all $1 \leq r \leq r_0$, $\sum_{j=1}^{r} T_{3j}^2 \leq$ const. $n^{-1+\eta}\, r\, \theta_r^{-1}\, \hat{A}(n,\eta)$, where random variable $\hat{A}(n,\eta)$ satisfies $\sup_{n\geq 1} E\{\hat{A}(n,\eta)^s\} < \infty$ for each $s \geq 1$. (I.73)

From (I.29) we have:

$$|\sum_{j=1}^{r} T_{1j}^2|^{1/2} - |\sum_{j=1}^{r} T_{3j}^2|^{1/2} \leq |\big(\sum_{j=1}^{r} T_{1j}^2\big)^{1/2} - \big(\sum_{j=1}^{r} T_{3j}^2\big)^{1/2}| \leq |\sum_{j=1}^{r} T_{4j}^2|^{1/2},$$

which by (I.37) and (I.73) implies that, for each $\eta > 0$,

$$\sum_{j=1}^{r} T_{1j}^2 = \Big(|\sum_{j=1}^{r} T_{1j}^2|^{1/2}\Big)^2 \leq \Big(|\sum_{j=1}^{r} T_{3j}^2|^{1/2} + |\sum_{j=1}^{r} T_{4j}^2|^{1/2}\Big)^2 \leq 2\Big(\sum_{j=1}^{r} T_{3j}^2 + \sum_{j=1}^{r} T_{4j}^2\Big)$$
$$\leq \text{const.}\, n^{-1+\eta}\, r\, \theta_r^{-1}\, \hat{A}(n,\eta) + \text{const.}\, n^{-2+\eta}\, r^3\, \theta_r^{-2}\, \hat{A}(n,\eta)\,. \qquad (I.74)$$

Thus, combining (I.73) with the above result, and using (I.6), we deduce that for some $\eta > 0$ we have:

for all $1 \leq r \leq r_0$, $\sum_{j=1}^{r}(T_{1j}^2 - T_{3j}^2) \leq$ const. $n^{-1-\eta}\, r\, \theta_r^{-1}\, \hat{A}(n)$, where the random variable $\hat{A}(n)$ satisfies $\sup_{n\geq 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \geq 1$. (I.75)

Combining (I.21), (I.58), (I.74) and the above result we see that, for some $\eta > 0$,

$$
\begin{aligned}
|\sum_{j=1}^{r}(\hat{b}_j-\bar{b}_j)^2-\sum_{j=1}^{r}T_{1j}^2| &\le |\sum_{j=1}^{r}(\hat{b}_j-\bar{b}_j)^2|+\sum_{j=1}^{r}T_{1j}^2 \\
&\le \text{const. } n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)+\text{const. } n^{-1-\eta}\, r\,\theta_r^{-1}\,\hat{A}(n) \\
&\le \text{const.}\, n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)\,,
\end{aligned}
$$

then, by (I.75),

$$
\begin{aligned}
|\sum_{j=1}^{r}(\hat{b}_j-\bar{b}_j)^2-\sum_{j=1}^{r}T_{3j}^2| &\le |\sum_{j=1}^{r}(\hat{b}_j-\bar{b}_j)^2-\sum_{j=1}^{r}T_{1j}^2|+|\sum_{j=1}^{r}(T_{1j}^2-T_{3j}^2)| \\
&\le \text{const. } n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)+\text{const. } n^{-1-\eta}\, r\,\theta_r^{-1}\,\hat{A}(n) \\
&\le \text{const.}\, n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)\,.
\end{aligned}
$$

Thus, from the above result we see that for some $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \le r \le r_0$, $|\sum_{j\le r}\{(\hat{b}_j-\bar{b}_j)^2-T_{3j}^2\}| \le$ const. $n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)$, where $\sup_{n\ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$. (I.76)

Combining (I.10), (I.68) and (I.76), and noting that dependence of the function $\hat{b}$ on m has been suppressed, we conclude that, for some $\eta > 0$,

if $\mathcal{G}$ holds then for all $1 \le r \le r_0$, $|\int(\hat{b}-b)^2-\sum_{j=1}^{r}T_{3j}^2-\sum_{j\ge r}b_j^2| \le$ const. $n^{-\eta}\, r\,(n^{-1}\,\theta_r^{-1}+b_r^2)\,\hat{A}(n)$, where $\sup_{n\ge 1} E\{\hat{A}(n)^s\} < \infty$ for each $s \ge 1$.

This property implies (I.19). To get (I.20), we have:

$$
\begin{aligned}
E\{S(r)\} &= E\bigg[\sum_{j=1}^{r} \theta_j^{-2} \bigg\{\frac{1}{n}\sum_{i=1}^{n} \int (X_i(t) - \bar{X}(t))\, \psi_j\bigg\}^2\bigg] \\
&= \sum_{j=1}^{r} \theta_j^{-2} E\bigg[\frac{1}{n}\sum_{i=1}^{n} (\xi_{ij} - \bar{\xi}_j)\,(\epsilon_i - \bar{\epsilon})\bigg]^2 \\
&= \sum_{j=1}^{r} \theta_j^{-2}\, n^{-2} \sum_{i=1}^{n} E[\xi_{ij} - \bar{\xi}_j]^2\, E[\epsilon_i - \bar{\epsilon}]^2 \\
&= n^{-2} \sum_{j=1}^{r} \theta_j^{-2}\, n\,(1 - \frac{1}{n})\,\theta_j \times (1 - \frac{1}{n})\,\sigma^2 \\
&= n^{-1}\,(1 - n^{-1})^2\,\sigma^2 \sum_{j=1}^{r} \theta_j^{-1}\,. \blacksquare
\end{aligned}
$$

I.2.8. *Bounds relating to* $\widehat{\psi}_j - \psi_j$. It can be proved from (I.30) that $\|\widehat{\psi}_j - \psi_j\|^2 = \hat{u}_j^2 + \hat{v}_j^2$, where

$$
\hat{u}_j^2 = \sum_{k\,:\,k \neq j} (\theta_j - \theta_k)^{-2}\, \hat{w}_{jk}^2, \quad \hat{v}_j^2 = \bigg\{ \int (\widehat{\psi}_j - \psi_j)\, \psi_j \bigg\}^2
$$

and $\hat{w}_{jk} = \int (\widehat{K} - K)\, \widehat{\psi}_j\, \psi_k$. Since both ψ_j and $\widehat{\psi}_j$ are of unit length, then $\hat{v}_j^2 = 2\{1 - (1 - \hat{u}_j^2)^{1/2}\} - \hat{u}_j^2$, which implies that

$$
\text{for all } j \geq 1, \qquad \|\widehat{\psi}_j - \psi_j\|^2 \leq 2\,\hat{u}_j^2, \qquad \hat{v}_j^2 \leq \hat{u}_j^4\,. \tag{I.77}
$$

If the event $\mathcal{G}$ obtains, then so too does (I.24), and therefore,

$$
|\theta_j - \theta_k| \leq |\hat{\theta}_j - \theta_j| + |\hat{\theta}_j - \theta_k| \leq \widehat{\Delta} + |\hat{\theta}_j - \theta_k| \leq \tfrac{1}{2}\,\rho_j + |\hat{\theta}_j - \theta_k|\,.
$$

Thus, $|\hat{\theta}_j - \theta_k| \geq \frac{1}{2}\,|\theta_j - \theta_k|$, which means $|\hat{\theta}_j - \theta_k|^{-1} \leq 2|\theta_j - \theta_k|^{-1}$ for all j, k such that $j \neq k$ and $1 \leq j \leq r_0$. For the same range of values of j and k, $|\theta_j - \theta_k|^{-1} \leq C\,\theta_{r_0}^{-1}\, r_0$; see (I.9). Here, C denotes a positive constant not depending on c in the definition of $\mathcal{G}$. Defining $\hat{x}_{ij} = \int (\widehat{K} - K)\, \psi_j\, \psi_k$ and $\hat{y}_{jk} = \int (\widehat{K} - K)\, (\widehat{\psi}_j - \psi_j)\, \psi_k$, we have $\hat{w}_{jk}^2 \leq 2\,(\hat{x}_{jk}^2 + \hat{y}_{jk}^2)$. Hence, assuming $\mathcal{G}$ holds, we have for $1 \leq j \leq r_0$,

$$\begin{aligned} \hat{u}_j^2 &\leq 8 \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\,(\hat{x}_{jk}^2 + \hat{y}_{jk}^2) \leq 8\,\hat{a}_j + 8\,C^2\,\theta_{r_0}^{-2}\,r_0^2\,\hat{c}_j \\ &\leq 8\hat{a}_j + 8\,C^2\,\theta_{r_0}^{-2}\,r_0^2\,\widehat{\Delta}^2\,\|\widehat{\psi}_j - \psi_j\|^2\,, \end{aligned} \tag{I.78}$$

where $\hat{a}_j = \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\,\hat{x}_{jk}^2$ and $\hat{c}_j = \sum_{k\,:\,k\neq j} \hat{y}_{jk}^2 \leq \widehat{\Delta}^2\,\|\widehat{\psi}_j - \psi_j\|^2$. Condition (I.4) and Lemma 2.5 imply that $n\,E(\hat{x}_{jk}^2) \leq \text{const.}\,\theta_j\,\theta_k$, where the constant does not depend on j, k or n. Moreover, by (I.1) and (I.2),

$$\begin{aligned} \sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\theta_j\,\theta_k &\leq \text{const.} \sum_{k\,:\,k\neq j} \left\{\frac{\max(j,k)}{\max(\theta_j,\theta_k)\,|j-k|}\right\}^2 \theta_j\,\theta_k \\ &\leq \text{const.}\left\{\theta_j \sum_{k\leq j/2} \theta_k^{-1} + j^2 \sum_{j/2<k\leq 2j} (j-k)^{-2} + \theta_j^{-1} \sum_{k>2j} \theta_k\right\} \leq \text{const.}\,j^2\,. \end{aligned} \tag{I.79}$$

Therefore, $E(\hat{a}_j) \leq \text{const.}\,n^{-1}\,j^2$. Also, for $s \geq 1$,

$$E(\hat{a}_j^s) = E\left[\left(\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\,\hat{x}_{jk}^2\right)^s\right] \leq \left(\sum_{k\,:\,k\neq j} (\theta_j - \theta_k)^{-2}\,\left(E[\hat{x}_{jk}^{2s}]\right)^{1/s}\right)^s. \tag{I.80}$$

Using Lemma 2.5, we see that

$$E(\hat{x}_{jk}^{2s}) \leq C_s\,n^{-r}\,\theta_j^s\,\theta_k^s \quad \text{for all } k \neq j \quad \text{and } s \geq 1\,, \tag{I.81}$$

where C_r is a constant depending only on r. So, (I.79)–(I.81) imply that for each $s \geq 1$,

$$E(\hat{a}_j^s) \leq C_s(n^{-1}\,j^2)^s, \quad \text{uniformly in } j\,, \tag{I.82}$$

where C_s does not depend on j. Moreover, defining $\tilde{a}_j = n\, j^{-2}\, \hat{a}_j$, we have,

$$E\big(\max_{1\le j\le r_0} \tilde{a}_j\big)^s \le E\Big(\sum_{j=1}^{r_0} \tilde{a}_j^s\Big) \le \sum_{j=1}^{r_0} E(\tilde{a}_j^s) \le r_0 \max_{1\le j\le r_0} E(\tilde{a}_j^s) \le \text{const.}\, r_0,$$

where the constant does not depend on n or r_0. Therefore, similarly to the argument used to obtain (I.45), with $\hat{U} = \max_{1\le j\le r_0} \tilde{a}_j$ we have, for any sequence $u_1, \cdots, u_{r_0}$ and all $r \le r_0$,

$$\sum_{j=1}^{r} u_j\, \tilde{a}_j \le \hat{U} \sum_{j=1}^{r} u_j, \quad \text{where} \quad \sup_{n\ge 1} E(n^{-\eta}\, \hat{U})^r < \infty \quad \text{for all } r \ge 1 \text{ and } \eta > 0\,. \tag{I.83}$$

Combining (I.78) with the first part of (I.77) we deduce that if $\mathcal{G}$ holds,

$$\|\widehat{\psi}_j - \psi_j\|^2 \le 16\, \hat{a}_j + 16\, C^2\, \theta_{r_0}^{-2}\, r_0^2\, \widehat{\Delta}^2\, \|\widehat{\psi}_j - \psi_j\|^2\,, \tag{I.84}$$

for $1 \le j \le r_0$. However, by definition of $\mathcal{G}$, if that event holds then $\widehat{\Delta} \le c\, r_0^{-1}\, \theta_{r_0}$. Hence, by (I.84), if $\mathcal{G}$ holds then for $1 \le j \le r_0$,

$$(1 - 16\, C^2\, c^2)\, \|\widehat{\psi}_j - \psi_j\|^2 \le 16\, \hat{a}_j\,.$$

Choosing c, in the definition of $\mathcal{G}$, so small that $16\, C^2\, c^2 \le \frac{1}{2}$, we deduce that if $\mathcal{G}$ holds then for $1 \le j \le r_0$, $\|\widehat{\psi}_j - \psi_j\|^2 \le 32\, \hat{a}_j$. Combining this result with (I.78), and noting the choice of c, we deduce that if $\mathcal{G}$ holds then for $1 \le j \le r_0$, $\hat{u}_j^2 \le 16\, \hat{a}_j$. From this property and the second part of (I.77) we conclude that

$$\text{if } \mathcal{G} \text{ holds then for } 1 \le j \le r_0,\ \|\widehat{\psi}_j - \psi_j\|^2 \le 32\, \hat{a}_j \text{ and}$$
$$\Big|\int (\widehat{\psi}_j - \psi_j)\, \psi_j\Big| \le 16\, \hat{a}_j \tag{I.85}$$

Results (I.37), (I.66) and (I.67) follow from (I.1), (I.65), (I.82), (I.83) and (I.85). Using (I.55), together with the bounds at (I.33) and (I.54) on $s(j)$ and $u(j)$ respectively, plus also the bounds at (I.49) on the maximum of $\Delta_{(j)}$, and the bounds (I.83) and (I.85) on the maximum of $\|\widehat{\psi}_j - \psi_j\|$, we can similarly derive (I.56) from (I.55).

I.1 On Validity of the Asymptotic MISE Approximation

Conditions under which (4.17) holds can be quickly deduced from Theorem 2.4, as follows. Under condition (2.1), we have $E\left\{\int(\widehat{K}-K)\,\psi_j\,\psi_j\right\}^2 \leq \text{const.}\; n^{-1}\,\theta_j^2$ (see Lemma 2.5), and also from (2.28) and (2.29) we conclude that $E(\|\widehat{K} - K\|\,\|\widehat{K}-K\|_{\sup}) = O(n^{-1})$. From these results and (2.18) it is deduced that, for each $1 \leq j \leq r$,

$$E\left|\,\hat{\theta}_j - \theta_j\right| = O\left(n^{-1/2}\,\theta_j + n^{-1}\,\delta_j^{-1}\right). \tag{I.86}$$

Recall that $\mathcal{E}$ denote the event that $\widehat{\Delta} \leq c\,r_0\,\theta_{r_0}$, where $c > 0$ is chosen so small that $c r_0^{-1}\,\theta_{r_0} \leq \frac{1}{2}\,\delta_{r_0}$, for all n. We proved that $1 - P(\mathcal{E}) = O(n^{-C})$, for each $C > 0$. If c is chosen sufficiently small, then we have $\widehat{\Delta} \leq \frac{1}{2}\,\delta_{r_0}$. Since $\delta_{r_0} \leq \theta_{r_0}$, then $\widehat{\Delta} \leq \frac{1}{2}\,\theta_{r_0}$ which implies $|\hat{\theta}_j - \theta_j| \leq \frac{1}{2}\,\theta_j$ for $1 \leq j \leq r \leq r_0$. Hence, for sufficiently small choice of c, we have $1 - P(|\hat{\theta}_j - \theta_j| \leq \frac{1}{2}\,\theta_j) \leq 1 - P(\mathcal{E}) = O(n^{-C})$. So, denoting $\mathcal{H}$ as the event $|\hat{\theta}_j - \theta_j| \leq \frac{1}{2}\,\theta_j$, provided the event $\mathcal{H}$ obtains, for each $1 \leq j \leq r$ we have:

$$E\left|\,\hat{\theta}_j^{-1} - \theta_j^{-1}\right| \leq 2\,\theta_j^{-2}\,E\left[|\hat{\theta}_j - \theta_j|\right]. \tag{I.87}$$

Combining (I.86) and (I.87) and summing up over j, and choosing C large enough, we have:

$$\sum_{j=1}^{r} E\big|\,\hat{\theta}_j^{-1} - \theta_j^{-1}\big| = O\Big(n^{-1/2}\sum_{j=1}^{r}\theta_j^{-1} + n^{-1}\sum_{j=1}^{r}\theta_j^{-2}\,\delta_j^{-1}\Big)\,. \tag{I.88}$$

Consequently,

$$E\bigg[\bigg|\Big(\sum_{j=1}^{r}\hat{\theta}_j^{-1}\Big)\Big/\Big(\sum_{j=1}^{r}\theta_j^{-1}\Big) - 1\bigg|\bigg] \leq \frac{1}{\Big(\sum_{j=1}^{r}\theta_j^{-1}\Big)}\sum_{j=1}^{r} E\big|\hat{\theta}_j^{-1} - \theta_j^{-1}\big|$$
$$= O\bigg(n^{-1/2} + \frac{(\sum_{j\leq r}\theta_j^{-2}\,\delta_j^{-1})}{(n\sum_{j\leq r}\theta_j^{-1})}\bigg)\,,$$

from which it follows that (4.17) holds, provided r increases sufficiently slowly, a sufficient condition for the rate being: $(\sum_{j\leq r}\theta_j^{-2}\,\delta_j^{-1})/(n\sum_{j\leq r}\theta_j^{-1}) \to 0$.

Bibliography

[1] Adams, R. A. (1975). *Sobolev Spaces.* Academic Press, New York.

[2] Aguilera, A.M., Ocana, F.A. and Valderrama, M.J. (1999a). Forecasting time series by functional PCA. Discussion of several weighted approaches. *Comput. Statist.* **14**, 443–467.

[3] Aguilera, A.M., Ocana, F.A. and Valderrama, M.J. (1999b). Forecasting with unequally spaced data by a functional principal component approach. *Test* **8**, 233–253.

[4] Ash, R. B. (1972). *Real Analysis and Probability.* Academic Press, Inc.

[5] Berkey, C.S., Laird, N.M., Valadian, I. and Gardner, J. (1991), Modeling Adolescent Blood Pressure Patterns and Their Prediction of Adult Pressure. *Biometrics.* **47**, 1005–1018.

[6] Besse, P. (1992). PCA stability and choice of dimensionality. *Statist. Probab. Lett.* **13**, 405–410.

[7] Besse, P.C. and Cardot, H. (1996). Spline approximation of the forecast of a first-order autoregressive functional process *Canad. J. Statist.* **24**, 467–487.

[8] Besse, P. and Ramsay J.O. (1986). Principal components-analysis of sampled functions. *Psychometrika* **51**, 285–311.

[9] Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley and Sons Inc.

[10] Bhatia, R., Davis, C. and McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.* **52/53**, 45–67.

[11] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.* **48**, 335–345.

[12] Bosq, D. (1989). Propriétés des opérateurs de covariance empiriques d'un processus stationnaire hilbertien. *C. R. Acad. Sci. Paris* Sér. I **309**, 873–875.

[13] Bosq, D. (2000). *Linear Processes in Function Spaces. Lect. Notes Statist.*, **149**.

[14] Brumback, B.A. and Rice, J.A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.* **93**, 961–976.

[15] Cai, T.T. and Hall, P. (2004). Prediction in functional linear regression. Manuscript.

[16] Capra, B. and Müller, H.-G. (1997). An accelerated-time model for response curves. *J. Amer. Statist. Assoc.* **92**, 72–83.

[17] Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparam. Statist.* **12**, 503–538.

[18] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11–22.

[19] Cardot, H., Ferraty, F. and Sarda, P. (2000). Étude asymptotique d'un estimateur spline hybride pour le modèle linéaire fonctionnel. *C.R. Acad. Sci. Paris* Sér. I **330**, 501–504.

[20] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13**, 571–591.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 241–255.

[21] Cardot and Sarda, P. (2003). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Anal.*, to appear.

[22] Castro, P. E., Lawton W. H. and Sylvestre, E.A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics.* **28**, 329–337.

[23] Cardot and Sarda, P. (2003b). Linear regression models for functional data. Manuscript.

[24] Chiou, J.M., Müller, H.-G. and Wang, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc.* Ser. B **65**, 405–423.

[25] Chiou, J.M., Müller, H.-G. and Wang, J.-L. (2004). Functional Response Models. *Statistica Sinica.* **14**, 675–693.

[26] Chung, K.L. (1974).*A Course in Probability Theory, Second Edition.* Academic Press, Inc.

[27] Conway J. B. (1985). *A Course in Functional Analysis.* Springer-Verlag New York Inc.

[28] Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *Canad. J. Statist.* **30**, 285–300.

[29] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multivariate Anal.* **12**, 136–154.

[30] Dunford, N. and Schwartz, J. T. (1963). *Linear Operators: Part II.* Interscience Publishers; A Division of John Wiley& Sons.

[31] Fan, J. and Zhang, J.T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *J. Roy. Statist. Soc.* Ser. B **62**, 303–322.

[32] Faraway, J.J. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254–261.

[33] Ferraty, F., Goia, A. and Vieu, P. (2002). Functional nonparametric model for time series: a fractal approach for dimension reduction. *Test* **11**, 317–344.

[34] Ferraty, F. and Vieu, P. (2000). Fractal dimensionality and regression estimation in semi-normed vectorial spaces. *C.R. Acad. Sci. Paris* Sér. I **330**, 139–142.

[35] Girard, S. (2000). A nonlinear PCA based on manifold approximation. *Comput. Statist.* **15**, 145–167.

[36] Grambsch, P.M., Randall, B.L., Bostick, R.M., Potter, J.D. and Louis, T.A. (1995). Modeling the labeling index distribution: An application of functional data analysis. *J. Amer. Statist. Assoc.* **90**, 813–821.

[37] Grenander, U. (1950). Stochastic processes and ststistics inference. *Arkiv för Mstematik*, 196–276.

[38] Hall, P. and Heckman, N.E. (2002). Estimating and depicting the structure of a distribution of random functions. *Biometrika* **89**, 145–158.

[39] Hall P. and Heyde C. C. (1980). *Martingale limit theory and its application.* Academic Press: New York.

[40] Hall, P. and Horowitz, J.L. (2004). Methodology and convergence rates for functional linear regression. Manuscript.

[41] Hall, P. and Hosseini-Nasab, M.. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **68** , pp 109-126.

[42] Hall, P. and Marron, J.S. (1991). Local minima in cross-validation functions. *J. Roy. Statist. Soc.* Ser. B **53**, 245–252.

[43] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

[44] Hastie, T. and Mallows, C. (1993). A discussion of "A statistical view of some chemometrics regression tools" by I. E. Frank and J. H. Friedman. *Technometrics.* **35**, 140–143.

[45] He, G.Z., Müller, H.-G. and Wang, J.-L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivar. Anal.* **85**, 54–77.

[46] Huang, J.H.Z., Wu C.O. and Zhou L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.

[47] James, G.M., (2002). Generalized linear models with functional predictors. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **64**, 411–432.

[48] James, G.M., Hastie, T.J. and Sugar, C.A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587–602.

[49] Jolliffe, I.T. (2002). *Principal Component Analysis, 2nd ed.* Springer-Verlag New York, Inc.

[50] Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse.*Ann. Acad. Sci. Fennicae.* **A I 37**.

[51] Kneip, A. and Utikal, K.J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96**, 519–532.

[52] Kress, R. (1999). *Linear Integral Equations, 2nd ed.* Springer, New York.

[53] Kronmal, R. and Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **63**, 925–952.

[54] Lee, S.Y., Zhang, W.Y. and Song, X.Y. (2002). Estimating the covariance function with functional data. *Brit. J. Math. Statist. Psych.* **55**, 247–261.

[55] Marx, B. D. and Eilers P.H. (1996). Generalized linear regression on sampled signals with penalized likelihood. In *Statistical Modelling. Proceedings of the 11th International workshop on Statistical modelling, Orvietto.* (Edited by A.Forcina, G. M. Marchetti, R. Hatzinger, G. Galmacci).

[56] Marx, B. D. and Eilers P.H. (1999). Generalized linear regression on sampled signals and curves: a p-spline approch. *Technometrics.* **41**, 1-13

[57] Mas, A. (2002). Weak convergence for the covariance operators of a Hilbertian linear process. *Stochastic Process. Appl.* **99**, 117–135.

[58] Müller, H. -G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.***33**, 774–805.

[59] Ocana, F.A., Aguilera, A.M. and Valderrama, M.J. (1999). Functional principal components analysis by choice of norm. *J. Multivariate Anal.* **71**, 262–276.

[60] Park, B.U. and Marron, J.S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85**, 66–72.

[61] Pezzulli, S. and Silverman, B.W. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics* **8**, 1–16.

[62] Pfeiffer, R.M., Bura, E., Smith, A. and Rutter, J.L. Two approaches to mutation detection based on functional data. *Statist. Med.* **21**, 3447–3464.

[63] Ramsay, J.O. (1982). When the data are functions. *Psychometrika* **47** (4), 379–396.

[64] Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis. (With discussion.) *J. Roy. Statist. Soc.* Ser. B **53**, 539–572.

[65] Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*. Springer, New York.

[66] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.

[67] Rao, C.R. (1958). Some Statistical Methods for the Comparison of growth curves. *Biometrics* **14**, 1–17.

[68] Rice, J.A. and Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc.* Ser. B **53**, 233–243.

[69] Riesz, F. and SZ.-Nagy, B. (1955). *Functional Analysis.* Frederick Ungar Publishing Co., New York.

[70] Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Satistics: An Introduction With Applications.* Chapman & Hall, Inc.

[71] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser.* **B**, **53**, 683–690.

[72] Silverman, B.W. (1995). Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser.* **B**, **57**, 673–689.

[73] Silverman, B.W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24**, 1–24.

[74] Staniswalis, J.G. and Lee, J.J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1403–1418.

[75] Wang, K.M. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25**, 1251–1276.

[76] Wang, K.M. and Gasser, T. (1999). Synchronizing sample curves nonparametrically. *Ann. Statist.* **27**, 439–460.

[77] Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Ann. Statist.*, in press.

[78] Yu, Y. and Lambert, D. (1999). Fitting trees to functional data, with an application to time-of-day patterns. *J. Comput. Graph. Statist.* **8**, 749–762.