Smoothing Based on Stretched Interpolated Moving Average Approach

Övgü Çıdar İyikal

Submitted to the
Institute of Graduate Studies and Research
in partial fulfilments of the requirements for the Degree of

Doctor of Philosophy
in
Mathematics

Eastern Mediterranean University
August 2013
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Mathematics.

Prof. Dr. Nazım İ. Mahmudov
Chair, Department of Mathematics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Mathematics.

Asst. Prof. Dr. Yücel Tandoğdu
Supervisor

Examining Committee

1. Prof. Dr. Agamirza Bashirov
2. Prof. Dr. İnci Batmaz
3. Prof. Dr. Memmedağa Memmedli
4. Assoc. Prof. Dr. Sonuç Zorlu Oğurlu
5. Asst. Prof. Dr. Yücel Tandoğdu

ABSTRACT

In this thesis, some smoothing techniques in multivariate and functional data analysis such as, kernel smoothing, local linear regression (LLR), spline smoothing and smoothing together with principal components analysis through conditional expectation (PACE) methods are considered. Their details are studied and a new smoothing method benefiting from moving average concept and applicable under certain conditions is proposed. Due to the steps involved in its logic, the proposed method is named Strecthed Interpolated Moving Average (SIMA). Its application to different data sets produced better results in terms of involved error, compared with LLR and similar results when compared with PACE.

Keywords: Karhunen–Loève Expansion, Stretched Interpolated Moving Average, Principal Component Scores, Lag Interval, Weight Function.

ÖZ

Bu tezde, çok değişkenli ve fonksiyonel veri analizinin; çekirdek pürüzsüzleştirme, yerel lineer regresyon (LLR), spline pürüzsüzleştirme, ve koşullu beklenti ile temel bileşenler analizi (PACE) gibi bazı pürüzsüzleştirme tekniklerine yer verilmiştir. Bunların ayrıntıları incelenmiş ve belirli koşullar altında hareketli-ortalamadan yararlanılarak yeni bir pürüzsüzleştirme tekniği önerilmiştir. Kendi mantığı içinde yer alan adımları nedeniyle önerilen yöntem Gerilmiş İnterpolasyonlu Hareketli-Ortalama (SIMA) diye adlandırılır. SIMA'nın farklı verilerde yapılan uygulamasında LLR uygulamasına kıyasla daha iyi sonuçlar elde edilmiş, PACE ile kıyaslandığında ise benzer sonuçlar elde edilmiştir.

Anahtar Kelimeler: Karhunen–Loève Açılımı, Gerilmiş Interpolasyonlu Hareketli-Ortalama, Temel Bileşenler Skorları, Gecikme Aralığı, Ağırlık Fonksiyonu.

To My Father

ACKNOWLEDGMENTS

I would like to thank Asst. Prof. Dr. Yücel Tandoğdu for his continuous support, help and guidance in the preparation of this thesis. Without his invaluable supervision, all my efforts could have been short-sighted.

Prof. Dr. Nazım Mahmudov, Chairman of the Department of Mathematics, Eastern Mediterranean University and Prof. Dr. Agamirza Bashirov helped me with various issues during the study and I am grateful to them. I am also obliged to Assoc. Prof. Dr. Sonuç Zorlu Oğurlu for her helps. Besides, a number of friends had always been around to support me morally. I would like to thank them as well.

I want to thank a lot to my family who provided financial and endless moral support towards my studies. I would like to dedicate this study to my husband Ramadan İyikal as an indication of his significance in this study as well as in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\mathbf{b}$	Local linear regression coefficient vector.
$G(t,t^{+})$	Covariance function.
$\mathbf{G}$	Covariance matrix of the functional representation of the data set.
h	Band width or the local neighbourhood of smoothing parameter.
ℓ_k	Eigenvalue of k^{th} eigenvector w_k.
n	Number of objects or trajectories.
m	Number of data values used for moving averaging.
M	Total number of the moving averaged data values.
p	Number of variables.
p_i	Number of observations on the *i*th trajectory.
p_{x_i}	Projected coordinate of data.
S	Sample covariance operator.
$\mathbf{S}$	Sample covariance matrix.
$\mathbf{u}_i$	i^{th} unit eigenvector of $\mathbf{X}^T\mathbf{X}$ / weight function.
$\mathbf{v}_j$	j^{th} unit eigenvector of $\mathbf{X}\mathbf{X}^T$.
x_{ij}	i^{th} observation on the j^{th} object.
$\mathbf{x}_{(j)}$	The column vector.
$\mathbf{x}_i$	The row vector.
X	The random variable.
X_i	i^{th} random variable/ i^{th} trajectory.

$X(t)$ Functional variable at time t.

$\mathbf{X}$ The $n \times p$ data matrix.

z_k Rowwise k^{th} principal component.

w_k Columnwise k^{th} principal component.

W_j j^{th} weight function.

Y_{ij} Functional representation of j^{th} object on the i^{th} trajectory.

Y_{il}^{*} l^{th} stretched interpolated moving average value on the i^{th} trajectory.

λ_k k^{th} eigenvalue of eigenvector z_k / k^{th} eigenvalue of the covariance matrix.

ϕ_k Eigenfunction of $G(t, t^{+})$.

ξ_k k^{th} principal component scores.

$\mathbf{\Phi}$ Matrix formed by basis vectors in fitting of ordinary least squares.

$\boldsymbol{\mu}$ The mean vector.

$\mathbf{\Sigma}$ Theoretical covariance matrix.

Chapter 1

INTRODUCTION

Studies in statistical data analysis gained momentum at the beginning of the 20^{th} century. Substantial foundation work laid during the first half, and with the advent of computers in the second half of the same century, wide applications into all disciplines became common ground. Today the development of statistical theory and application of developed ideas using the continuously advancing computer technology has enabled the testing of abstract statistics theory, previously not possible. This resulted in rapid development of nonparametric statistical data analysis.

For the analysis of multivariate data that is considered in this study, one main issue is to smooth the data before processing in order to eliminate the effect of extreme values. Alternately smoothing the mean and covariance functions are commonly used in many data analysis methods. Functional data analysis (FDA) forms the theoretical foundation for multivariate data analysis (MDA). Theory related with MDA and FDA is summarized in Chapter 2 and Chapter 3 respectively. One of the main concepts used in MDA is the Principle Component Analysis (PCA) that enables dimension reduction of multivariate data. Pioneering work on this topic was initially carried out by Hotelling (Hotelling, 1933), who built his theory on the foundations laid out by Karl Pearson. One main issue in FDA and MDA is the smoothing of raw data, mean and covariance functions. There are many smoothing methods developed over the years. Amongst some widely used ones are Kernel group of smoothers, Spline smoothers, Regression smoothers and Moving Average smoothers. In this work, Epanechnikov kernel, and local linear regression smoothers are used and a specific method for the moving average smoothers is proposed.

The Epanechnikov kernel and spline smoothers are used in the PCA through conditional expectation method (PACE) for smoothing the mean and covariance functions, Müller (2005). Further details on kernel smoothing is given by Härdle (1992). Structure of a smoothing spline is explained in detail by Ramsay and Silverman (2006). Principles of smoothing PCA is given in (Ramsay and Silverman, 2002). Local linear regression smoothing (LLR) is explained by Loader (1999).

The main idea of estimating a trajectory from available data, through MDA is widely studied by many different researchers. One note worthy method in this respect is PACE, which is summarized in Section 2.3.

Smoothing a trajectory, the mean function or the covariance matrix (surface) is possible using any of the above mentioned methods. The proposed moving average smoothing method is named as Stretched Interpolated Moving Average (SIMA), mainly because of the steps involved in its computation. Details are given in Chapter 4.

The proposed SIMA smoothing method is applied to two distinct data sets together with kernel, and LLR smoothers in Chapter 5. Obtained results from SIMA are compared with those from other methods. SIMA performed better than LLR under the condition of weak correlation between the variables involved. Since kernel and spline smoothers are used in the PACE method, these smoothers are sometimes referred to as PACE smoothers. In the application to the data sets for the smoothing of mean and covariance, SIMA performed equally well with PACE smoothers. The measure used in comparing different smoothers is Mean Square Error (MSE) between an observed trajectory and its smooth estimate.

Chapter 2

MULTIVARIATE DATA ANALYSIS

2.1 Introduction to the Multivariate Data Analysis

Analyzing data where more than one variable is involved requires the use of MDA techniques. Representing the data in matrix format is essential in the process. Let $x_{ij}:\ i=1,\ldots,n;\ j=1,\ldots,p$ be the set of $n\times p$ observations or data set. Then each column of the data belongs to the j^{th} random variable X_j and denoted by the column vector $\mathbf{x}_{(j)}=\begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix};\ j=1,\ldots,p$. Similarly the rows represent the data values belonging to each trajectory, denoted by the row vector $\mathbf{x}_i=\begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix};\ i=1,\ldots,n$. In this setup the row vectors $\mathbf{x}_1^T,\ldots,\mathbf{x}_n^T$ represents a random sample of the trajectories while the column vectors $\mathbf{x}_{(1)},\ldots,\mathbf{x}_{(p)}$ represents the values of the random variables X_j.

The vector representation of data given above forms the $n\times p$ data matrix $\mathbf{X}$ as given below.

$$\mathbf{X} = \text{objects}\left\{ \overbrace{\begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix}}^{\text{variables}} \right. .$$

When the number of variables and the trajectories are large, numerical processing needed in analyzing such data becomes prohibitive. On the other hand it is a fact that not every variable will have the same impact on the process under consideration. Therefore, one main concern of MDA is to identify the variables having major influence on the process under study. This in turn will enable the exclusion of the variables with minor or marginal effect, hence alleviating load of data processing while maintaining high level of accuracy (Mardia, et. al., 1979).

2.1.1 The Sample Statistics

The mean vector and covariance matrix of the multivariate data can be written by extending the univariate case to the multivariate form. The sample mean and sample variance of the j^{th} variable are given as in equations (2.1.1) and (2.1.2).

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \tag{2.1.1}$$

The vector of means for p variables is $\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i = \frac{1}{n}\mathbf{X}^T \mathbf{1}_n$. Here $\mathbf{1}_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

$$s_j^2 = \frac{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}{n-1}, \ (j = 1, \dots, p). \tag{2.1.2}$$

The sample covariance between the j^{th} and the j^{*th} variables is

$$s_{jj^*} = \frac{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(x_{ij^*} - \overline{x}_{j^*})}{n-1}, \ j, j^* = 1, \dots, p \tag{2.1.3}$$

It is evident that, $s_{jj^*} = s_j^2$, when $j = j^*$.

The $p \times p$ covariance matrix can be written as

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T \ .$$

Using the centring matrix $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$, the covariance can also be denoted as

$$\mathbf{S} = \frac{1}{n}\mathbf{X}^T\mathbf{H}\mathbf{X} \ . \tag{2.1.4}$$

Since $\mathbf{H}$ is symmetric and idempotent, using a p-vector **v,** $\mathbf{v}^T\mathbf{S}\mathbf{v} = \frac{1}{n}\mathbf{v}^T\mathbf{X}^T\mathbf{H}^T\mathbf{H}\mathbf{X}\mathbf{v} \geq 0$ can be written, meaning the covariance matrix $\mathbf{S}$ in equation (2.1.4) is positive semi-definite.

2.1.2 Linear Transformation

Often in MDA linear transformation of data becomes necessary before analysis, due to linearly transformed data results in dimension reduction, simplifying computations. Hence, computation of statistics for linearly transformed data has to be formulated. Letting $\mathbf{a}^T = (a_1, \ldots, a_p)$ be the vector of coefficients to be used in the transformation, transformed data will be $y_i = \mathbf{a}^T \mathbf{x}_{ij}$; $\mathrm{i} = 1, \ldots, n,\ \mathrm{j} = 1, \ldots, p$. Transformed data will have a mean

$$\bar{y} = \frac{1}{n} \mathbf{a}^T \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{a}^T \bar{\mathbf{x}}$$

and variance

$$s_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{a}^T (\mathbf{x}_i - \bar{\mathrm{x}})(\mathbf{x}_i - \bar{\mathrm{x}})^T \mathbf{a} = \mathbf{a}^T \mathbf{S} \mathbf{a} \ .$$

In a q dimensional linear transformation, $\mathbf{A}_{q \times p}$ being the matrix of coefficients and $\mathbf{b}_q$ vector of constants, then,

$$\mathbf{y}_i = \mathbf{A}\mathbf{x} + \mathbf{b},\ i = 1, \ldots, n \rightarrow \mathbf{Y} = \mathbf{X}\mathbf{A}^T + \mathbf{1}\mathbf{b}^T$$

can be written. Then, the mean vector and covariance matrix of transformation will be

$$\bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{b} \text{ and } \mathbf{S}_y = n^{-1} \sum_{i=1}^{n} (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T = \mathbf{A}\mathbf{S}\mathbf{A}^T \ .$$

It can be shown that linear combinations of a multinormal vector are univariate normal (Mardia, et.al., 1979).

2.2 Multivariate Normal Theory

The univariate normal distribution is the most widely used distribution in many statistical application problems. Its multivariate version distribution similarly enables the solution of many multivariate estimation problems. Therefore, it plays a major role in MDA. It is wholly defined by its first and second moments and the *p*-variate normal distribution is given by

$$f(\mathbf{x}) = \left|2\pi\mathbf{\Sigma}\right|^{-1/2} \exp(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

where $\mathbf{\Sigma} > 0$ is the positive definite covariance matrix, $\mathbf{x}_p$ and $\boldsymbol{\mu}_p$ are the vectors of random variables and their means, respectively. Then, $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ denotes a multivariate normal distribution with parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ (Park, 2008).

In multivariate normal distribution, for pairs (X_i, X_j), correlation $\rho_{X_i X_j} = 0,\ i \neq j$ implies independence and pairwise independence implies total independence.

Corollary 2.2: If $\mathbf{x}$ has a *p*-variate normal distribution, and if $\mathbf{y} = \mathbf{Ax} + \mathbf{c}$ is the linear combination of the variables, with $q \times p$ dimensional matrix $\mathbf{A}$ and *q*-vector $\mathbf{c}$, then $\mathbf{y}$ has *q*-variate normal distribution. That is, if $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, then $\mathbf{y} \sim \mathcal{N}_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A\Sigma A}^T)$, where $\boldsymbol{\mu}$ is the mean vector and $\mathbf{\Sigma}$ is the covariance matrix (Mardia et al., 1979).

Certain applications may require the use of partitioned matrices, due to the nature of the process. For example in a p variate process, if it is required to show the independence of k variables in a process from remaining $t = p - k$ elements of the process, then the covariance vector of variables $\mathbf{x}$ can be partitioned into two sub-vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ with k and t elements, respectively as given in the following Theorem 2.2.1.

Theorem 2.2.1 (Mardia et al., 1979): Assume $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\mathbf{x}_1 \in \mathbb{R}^k$, $\mathbf{x}_2 \in \mathbb{R}^t$, and $\mathbf{x}_{2.1} = \mathbf{x}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{x}_1$ defined from the partitioned covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, then, $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{x}_{2.1} \sim \mathcal{N}_t(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1})$ are statistically independent, i.e. $Cov(\mathbf{x}_1, \mathbf{x}_{2.1}) = 0$, with $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Proof: Let $\mathbf{x}_1 = \begin{pmatrix} \mathbf{I}_k & 0 \end{pmatrix}\mathbf{x}$ and $\mathbf{x}_{2.1} = \begin{pmatrix} -\Sigma_{21}\Sigma_{11}^{-1} & \mathbf{I}_t \end{pmatrix}\mathbf{x}$. By Corollary 2.2, the multivariate normalities, $\mathbf{x}_1 \sim \mathcal{N}_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{x}_{2.1} \sim \mathcal{N}_t(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$ are clear.

Consider the covariance for the independency.

$$\operatorname{cov}(\mathbf{x}_1, \mathbf{x}_{2.1}) = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_t \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{11}} & \boldsymbol{\Sigma}_{\mathbf{12}} \\ \boldsymbol{\Sigma}_{\mathbf{21}} & \boldsymbol{\Sigma}_{\mathbf{22}} \end{pmatrix} \begin{pmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_t \end{pmatrix}$$

$$= \left(-\boldsymbol{\Sigma}_{11} \left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11} \right)^T + \boldsymbol{\Sigma}_{12} \right)$$

$$=\left(-\mathbf{\Sigma}_{11}\mathbf{\Sigma}_{11}^{T}\mathbf{\Sigma}_{12}+\mathbf{\Sigma}_{12}\right)$$

$$=0,$$

i.e. $\mathbf{x}_1$ and $\mathbf{x}_{2.1}$ are statistically independent. ■

This theorem paves the way to the idea of conditional relationship between two sub-vectors of a vector of random variables. This is highlighted in the following theorem.

Theorem 2.2.2 (Mardia et al., 1979): From Corollary 2.2 and Theorem 2.2.1, conditional distribution of $\mathbf{x}_2$ for a certain value of $\mathbf{x}_1$ is approximately normally distributed, that is

$$\mathbf{x}_2 \mid \mathbf{x}_1 \sim \mathcal{N}_t(\boldsymbol{\mu}_2+\mathbf{\Sigma_{21}\Sigma_{11}^{-1}(x_1-\boldsymbol{\mu}_1),\Sigma_{22.1}}). \tag{2.2.1}$$

with conditional mean

$$E[\mathbf{x}_2 \mid \mathbf{x}_1]=\boldsymbol{\mu}_{2.1}+\mathbf{\Sigma_{21}\Sigma_{11}^{-1}x_1}=\boldsymbol{\mu}_2+\mathbf{\Sigma_{21}\Sigma_{11}^{-1}(x_1-\boldsymbol{\mu}_1)}. \tag{2.2.2}$$

Proof: Since $\mathbf{x}_{2.1}$ is independent of $\mathbf{x}_1$, its conditional distribution for a given value of $\mathbf{x}_1$ is same as its marginal distribution. Now $\mathbf{x}_2=\mathbf{x}_{2.1}+\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{x}_1$ and this term is constant when $\mathbf{x}_1$ is given. Therefore the conditional distribution of $\mathbf{x}_2 \mid \mathbf{x}_1$ is normal with conditional mean $E[\mathbf{x}_2 \mid \mathbf{x}_1]=\boldsymbol{\mu}_{2.1}+\mathbf{\Sigma_{21}\Sigma_{11}^{-1}x_1}$ (Mardia et al., 1979).

On the other hand, let $x_1 \sim \mathcal{N}(\mu_1,\sigma_1^2)$ and $x_2 \sim \mathcal{N}(\mu_2,\sigma_2^2)$. Since, by Normal Distribution Theory the marginal distribution of x_1 is,

$$f_1(x_1) = \int_{-\infty}^{\infty} \frac{\exp\left(\frac{-1}{2\text{-}2\rho^2} \left(\frac{\left(x_1\text{-}\mu_1\right)^2}{\sigma_1^2} \text{-}2\rho \frac{x_1\text{-}\mu_1}{\sigma_1} \frac{x_2\text{-}\mu_2}{\sigma_2} + \frac{\left(x_2\text{-}\mu_2\right)^2}{\sigma_2^2} \right) \right)}{2\pi\sigma_1\sigma_2\sqrt{1\text{-}\rho^2}} dx_2 = \frac{e^{-\frac{1}{2}(\frac{x_1-\mu_1}{\sigma_1})^2}}{\sigma_1\sqrt{2\pi}} .$$

Moreover,

$$E\left[x_2 \mid x_1\right] = \int_{-\infty}^{\infty} x_2 f(x_2 \mid x_1) dx_2 ,$$

where the conditional distribution is as,

$$f(x_2 \mid x_1) = \frac{e^{-\frac{1}{2}\left(\frac{x_2 - (\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1))}{\sigma_2\sqrt{1-\rho^2}} \right)^2}}{\sigma_2\sqrt{2\pi(1\text{-}\rho^2)}} .$$

Clearly by Normal theory, the conditional expectation can be written as,

$$E\left[x_2 \mid x_1\right] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}\left(x_1 - \mu_1\right) \tag{2.2.3}$$

and the general variance/covariance is as,

$$\mathrm{var}\left[x_2 \mid x_1\right] = \sigma_2^2\left(1 - \rho^2\right). \tag{2.2.4}$$

Now, in terms of matrices; consider $\mathbf{x}_1 \sim N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{x}_{2.1} \sim N_t(\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$, then, substitute $\boldsymbol{\Sigma}_{ij} = \rho\sigma_i\sigma_j$ and $\boldsymbol{\Sigma}_{ii} = \sigma_i^2$ in (2.2.3) and (2.2.4) we have, (2.2.2) and $\mathrm{cov}\left[x_2 \mid x_1\right] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{22.1}$.■

This theorem will help in understanding the concept of principal component analysis under some conditional forms.

2.3 Eigenvalues and Eigenvectors

The symmetric covariance matrix is extensively used in MDA by means of eigenvalues and eigenvectors. Therefore two of the theorems pertaining to the eigenvalues and eigenvectors are presented for information.

The theorem that establishes links between the structure of a symmetric matrix and its eigenvalues and eigenvectors which can be found in most serious linear algebra books, is as follows.

Theorem 2.3.1: (Spectral or Jordan Decomposition) Any symmetric $p \times p$ matrix $\mathbf{A}$ can be written in terms of $\mathbf{\Lambda}$, the diagonal matrix of its eigenvalues and $\mathbf{\Gamma}$, an orthogonal matrix whose columns are standardized eigenvectors of $\mathbf{A}$,

$$\mathbf{A} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T = \sum_{j=1}^{p} \lambda_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T \qquad (2.3.1)$$

where, $\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_p)$ and $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_p)$.

Therefore, (2.3.1) shows that a symmetric matrix is uniquely determined by its eigenvalues and eigenvectors. If λ_i are distinct and written in decreasing order, then $\mathbf{\Gamma}$ is also uniquely determined.

Another important concept used in MDA under certain circumstances, i.e. when a matrix is not square, is the singular value decomposition theorem given below.

Theorem 2.3.2: (Singular Value Decomposition) Each $n \times p$ matrix **A** with rank r can be decomposed by column orthonormal matrices, $\mathbf{\Gamma}\,(n \times r)$ and $\mathbf{\Delta}\,(p \times r)$, satisfying $\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{\Delta}^T\mathbf{\Delta} = \mathbf{I}_r$ and diagonal $\mathbf{\Theta}$ matrix of positive elements, so that $\mathbf{A} = \mathbf{\Gamma}\mathbf{\Theta}\mathbf{\Delta}^T$.

This is the generalization of the Jordan decomposition theorem.

2.4 Factoring the Data Matrices

In this section factoring data matrices is reviewed, since principal component analysis depends on the concepts developed here. The aim is to reduce the dimension of the data matrix by means of geometric approach with respect to a least-squares criterion. As a result, low dimensional graphical pictures of the data matrix area obtained. This is the process of decomposing the data matrix into factors, a concept used in many multivariate techniques. Dimension reduction facilitates the easy interpretation of the process estimated by the data.

2.4.1 Projecting Data from Higher to Lower Dimensional Space

Representing a multivariate data set in matrix format $\mathbf{X}_{n \times p}$ was introduced in Section 2.1. Projection of data values can be performed row or columnwise. In other words the column space **C(X)** or the row space $\mathbf{C}(\mathbf{X}^T)$ can be approximated by smaller subspaces. An important point in dimension reduction is not to lose much from the variation and structure of the data.

2.4.1.1 Projecting Data onto $\mathbb{R}^q$ from $\mathbb{R}^p$

Procedure for projecting the *p*-dimensional *n* data points onto a subspace $\mathbb{R}^q$, $(q \le p)$ is explained. For simplicity details of projection from $\mathbb{R}^p$ onto $\mathbb{R}$ will be given. Projection onto $\mathbb{R}^q$ becomes an extension of the procedure undertaken for the one dimensional subspace.

Let F_1 be the line that passes through the origin, onto which data is to be projected. Direction of the line F_1 is determined by the unit vector $\mathbf{u}_1$. Projection is achieved by projecting the i^{th} individual $x_i \in \mathbb{R}^p$ onto $\mathbf{u}_1$. Hence, the projection point p_{x_i} will have the coordinate $p_{x_i} = \mathbf{x}_i^T \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \mathbf{x}_i^T \mathbf{u}_1$ on F_1. On the other hand F_1 has to be located such that $\mathbf{u}_1 \in \mathbb{R}^p$ and $\sum_{i=1}^{n} \| x_i - p_{x_i} \|^2$ is minimum. This is equivalent to maximizing $\sum_{i=1}^{n} \| p_{x_i} \|^2$, reducing the problem to finding $\mathbf{u}_1 \in \mathbb{R}^p$ such that $\sum_{i=1}^{n} \| p_{x_i} \|^2$ is maximum subject to the constraint $\|\mathbf{u}_1\| = 1$. Then projection is written as

$$\begin{pmatrix} p_{\mathbf{x}_1} \\ p_{\mathbf{x}_2} \\ \vdots \\ p_{\mathbf{x}_n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \mathbf{u}_1 \\ \mathbf{x}_2^T \mathbf{u}_1 \\ \vdots \\ \mathbf{x}_n^T \mathbf{u}_1 \end{pmatrix} = \mathbf{X}\mathbf{u}_1 \ . \qquad (2.4.1)$$

and can also be expressed in quadratic form as

$$\max_{\mathbf{u}_1^T\mathbf{u}_1=1} \mathbf{u}_1^T (\mathbf{X}^T\mathbf{X})\mathbf{u}_1 \tag{2.4.2}$$

(Härdle and Simar, 2003). For details see Appendix A.

The vector $\mathbf{u}_1$ is the eigenvector of $\mathbf{X}^T\mathbf{X}$ corresponding to the largest eigenvalue λ_1 of $\mathbf{X}^T\mathbf{X}$. $\mathbf{u}_1$ minimizes, $\sum_{i=1}^{n} \| x_i - p_{x_i} \|^2$. When the data is centered ($\bar{x} = 0$) and $\mathbf{S} = n^{-1}\mathbf{X}^T\mathbf{X}$ is the covariance matrix, then the quadratic form (2.4.2) is maximized with respect to **S.**

Representation of n trajectories on F_1 are given by $\mathbf{Xu}_1$ which is the first factorial variable $\mathbf{z}_1 = \mathbf{Xu}_1$ and $\mathbf{u}_1$ is the first factorial axis. Then, $\mathbf{z}_1 = u_{11}\mathbf{x}_{(1)} + \cdots + u_{p1}\mathbf{x}_{(p)}$ represents a linear combination of trajectories with coefficients being the elements of $\mathbf{u}_1$ (Härdle and Simar, 2003).

Projection of data on $\mathbb{R}^q$ from $\mathbb{R}^p$ where $q \le p$, is the extension of the above explained process from $\mathbb{R}^p$ to $\mathbb{R}$, except minimization of $\sum_{i=1}^{n} \| x_i - p_{x_i} \|^2$ will produce the best subspace $u_1, \ldots, u_q$ which are the orthonormal eigenvectors of $\mathbf{X}^T\mathbf{X}$. Corresponding eigenvalues of $\mathbf{X}^T\mathbf{X}$ are $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_q$. Then the coordinates of n trajectories on the k^{th} factorial axis u_k are given by $\mathbf{z}_k = \mathbf{Xu}_k$; $k = 1, \ldots, q$. The linear combination of the original variables $x_{(1)}, \ldots x_{(p)}$ whose coefficients are given by the k^{th} vector

$\mathbf{u}_k: z_{ik} = \sum_{m=1}^{p} x_{im} u_{mk}$ forms the factorial variables $\mathbf{z}_k = (z_{1k}, \ldots, z_{nk})^T$ (Härdle and Simar, 2003).

2.4.1.2 Projecting Data onto $\mathbb{R}^q$ from $\mathbb{R}^n$

The columns of $n \times p$ data matrix **X** represents the data as *p* points in $\mathbb{R}^n$. That is each column or variable is a vector $\mathbf{x}_{(j)} = (x_{1j}, \ldots, x_{nj})^T \in \mathbb{R}^n$. Projecting the *n*-dimensional *p* data points onto a subspace $\mathbb{R}^q$ $(q \le n)$ is sought. A similar approach used in Section 2.4.1.1 will be followed. Starting with projection onto a one dimensional space or a straight line G_1 defined by the unit vector $\mathbf{v}_1 \in \mathbb{R}^n$ that will best fit for the *n* dimensional *p* points. It means finding $\mathbf{v}_1$ such that $\sum_{j=1}^{p} \| p_{x_{(j)}} \|^2$ is maximized. In other words the unit vector $\mathbf{v}_1$ should maximize, $\left(\mathbf{X}^T \mathbf{v}_1\right)^T \left(\mathbf{X} \mathbf{v}_1\right) = \mathbf{v}_1^T \left(\mathbf{X}\mathbf{X}^T\right) \mathbf{v}_1$. Then, the coordinates of *p* variables on G_1 are given by $w_1 = \mathbf{X}^T \mathbf{v}_1$. Hence, $w_{1j} = v_{11} x_{1j} + \cdots + v_{1n} x_{nj}$; $j = 1, \ldots, p$, can be written (Härdle and Simar, 2003).

Extending to projection onto $\mathbb{R}^q$; $q \le n$ means generating a subspace through the orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_q$ corresponding to the eigenvalues $\ell_1 \ge \cdots \ge \ell_q$ obtained from $\mathbf{X}\mathbf{X}^T$. Coordinates of the *p* variables on the k^{th} factorial axis are $\mathbf{w}_k = \mathbf{X}^T \mathbf{v}_k$; $k = 1, \ldots, p$ and $\mathbf{w}_k = (w_{k1}, \ldots, w_{kp})$. Then, $w_{kj} = \sum_{m=1}^{n} v_{km} x_{mj}$.

2.4.1.3 Relationship Between the Projection of Data from $\mathbb{R}^p$ onto $\mathbb{R}^q$ and $\mathbb{R}^n$ onto $\mathbb{R}^q$

The $q\ (q \le p)$ dimensional (column) subspace onto which data points are projected, is generated by the orthonormal eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_q$ of $\mathbf{X}^T\mathbf{X}$. Respective eigenvalues are $\lambda_1,\ldots,\lambda_q$.

Projection of data from $\mathbb{R}^n$ onto $q\ (q \le n)$, the row subspace is generated by the orthonormal eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_q$ of $\mathbf{X}\mathbf{X}^T$, $\ell_1,\ldots,\ell_q$ being the respective eigenvalues.

Taking into account the similar logic used in both projections, the eigenvector equations $(\mathbf{X}^T\mathbf{X})\mathbf{u}_k = \lambda_k\mathbf{u}_k$ in $\mathbb{R}^p$ and $(\mathbf{X}\mathbf{X}^T)\mathbf{v}_k = \ell_k\mathbf{v}_k$ in $\mathbb{R}^n$ can be written. They satisfy the condition that, for $k \le r$, r being the rank of **X**, the eigenvalues of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$ are the same, and their eigenvectors are related by

$$\mathbf{u}_k = \frac{\mathbf{X}^T\mathbf{v}_k}{\sqrt{\lambda_k}}, \quad \mathbf{v}_k = \frac{\mathbf{X}\mathbf{u}_k}{\sqrt{\lambda_k}}, \qquad (2.4.3)$$

(Härdle and Simar, 2003).

Let, $\mathbf{U} = \left[\mathbf{u}_1,\ldots,\mathbf{u}_r\right], \mathbf{V} = \left[\mathbf{v}_1,\ldots,\mathbf{v}_r\right]$, and $\mathbf{\Lambda} = diag\left[\lambda_1,\ldots,\lambda_r\right]$, then singular value decomposition of the data matrix is $\mathbf{X} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$. From here, $x_{ij} = \sum_{k=1}^{r} \lambda_k^{1/2} v_{ik} u_{jk}$.

Chapter 3

FUNCTIONAL DATA ANALYSIS

3.1 Functional Modelling

In general many processes are continuous in nature, while available data is discrete. The question is how to express discrete observations in functional form for the assessment of the process in question. Processing of discrete data is dealt with in multivariate data analysis (MDA), which is explained in Section 2. A random variable X is a functional variable if it takes values in infinite dimensional space also called functional space.

Observations over X are denoted as x. If T is a subset of $\mathbb{R}^k$; $k = 1, 2, \ldots$ representing the range of time or space within which the process is taking place, then the random function $X = \{X(t);\ t \in T\}$ and its realizations are $x = \{x(t);\ t \in T\}$. Then the functional data set $x_1, \ldots, x_n$ is a particular realization of the n functional variables $X_1, \ldots, X_n$ having the identical distribution as X.

In MDA a linear combination of variable values (the k^{th} principal components) is taken by $z_{ik} = \sum_{m=1}^{p} x_{im} u_{mk}$, u_{mk} being the weights applied to the x_{im} observed value.

Obtaining the functional data from available discrete observations for each trajectory, would mean linear interpolation between successive observations and smoothing of trajectories. Hence each trajectory can be denoted by $X_1(t),\ldots,X_n(t)$, t representing the time or space coordinate of a trajectory. The discrete index used in MDA is replaced by the continuous index t in functional PCA. Then, the principal component score in functional data becomes,

$$\xi_i = \int u x_i = \int u(t) x_i(t) dt\,. \tag{3.1.1}$$

Estimating the mean function from n trajectories is possible. The obtained mean function $\bar{X}(t)$ should be smoothed to avoid undesirable fluctuations in $\bar{X}(t)$ stemming from noisy data.

The functional principal components *weight function* $u(t)$ is defined for each component over the range of t such that $\int u(t)^2 dt = 1$. Then, the principal component scores ξ_i in (3.1.1) for the sample data is given by $\xi_i = \int u(t) X_i(t) dt$.

In the first functional principal component the aim is to determine $u_1(t)$ such that it maximizes the variance $Var(\xi_i) = n^{-1}\sum_i \xi_{i1}^2 = n^{-1}\sum_i \left(\int_t u_1(t) x_i dt\right)^2$ under the constraint $\int u(t)^2 dt = 1$. Second and higher order principal components can be defined in the same way except, they have to satisfy the additional mutual orthogonality property. That is,

$$\int u_j(t)u_1(t)dt = \int u_j(t)u_2(t)dt = \cdots = \int u_j(t)u_{j-1}(t)dt = 0$$

(Ramsay and Silverman, 2002).

An optimal empirical orthonormal basis is needed for application purpose. That is finding K orthonormal functions u_m, enabling the expansion of each curve or trajectory in terms of these basis functions, such that the trajectory is approximated as accurate as possible. As a result seeking an expansion of the form,

$$\hat{x}_i(t) = \sum_{k=1}^{K} \xi_{ik} u_k(t) \text{ where } \xi_{ik} = \int x_i(t) u_k(t) dt \qquad (3.1.2)$$

is necessary.

The fitting criterion

$$\| x_i - \hat{x}_i \|^2 = \int [x(t) - \hat{x}(t)]^2 dt \qquad (3.1.3)$$

is used.

Then, from equation (3.1.3) the sum of the squares of errors for PCA, $SSE_{PCA} = \sum_{i=1}^{n} \| x_i - \hat{x}_i \|^2$, can be used as a measure of approximation. The basis that minimizes SSE_{PCA} corresponds to the same set of PC weight functions that maximize variance components.

3.1.1. Brief Comparison of MDA and FDA

Given a data matrix $\mathbf{X}_{n\times p}$, its covariance matrix $\mathbf{S}_{p\times p} = n^{-1}\mathbf{X}^T\mathbf{X}$ or as in equation (2.1.4) with eigenvalues, $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_q$ and corresponding orthogonal eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_q$, are computed. Projection of data onto column subspace is subject to the constraint, $\max_{\|\mathbf{u}\|=1}\left(\mathbf{u}^T\mathbf{S}\mathbf{u}\right)$. Solution of the eigen-equation $\mathbf{S}\mathbf{u} = \boldsymbol{\lambda}\mathbf{u}$ gives the eigenvalues in descending order, $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_q$. Due to centering of **X**, its rank is at most, $n-1$, meaning that the $\mathbf{S}_{p\times p}$ covariance matrix will have $\min\{p, n-1\}$ nonzero eigenvalues.

In functional PCA, the covariance function $s(t,t^+)$ can be written for n available data as, $s(t,t^+) = n^{-1}\sum_{i=1}^{n} x_i(t)x_i(t^+)$.

Finding the principal components weights $u_j(t)$ requires the following concepts (Ramsay and Silverman, 2006, Appendix A.5).

1. In a general inner product space, the symmetric matrix is replaced by a self-adjoint linear operator, A, which satisfies $\langle x, Ay\rangle = \langle Ax, y\rangle$, $\forall x, y$.

2. A is compact (completely continuous) symmetric transformation on Hilbert space.

Maximizing the inner product space $\langle x, Ay\rangle$ subject to the constraint $\|x\| = 1$ is similar to maximizing $\mathbf{x}^T\mathbf{A}\mathbf{x}$ subject to $\mathbf{x}^T\mathbf{x} = 1$ in the finite dimensional space.

Here the sequence u_j is defined as the solutions to the set of optimization problems

$$\max\langle x, Ax\rangle \text{, subject to } \lVert x\rVert = 1 \text{ and } \langle x, u_i\rangle = 0 \text{ for } i < j\,. \qquad (3.1.4)$$

The solution is obtained under the given conditions by assessing the eigenfunction problem $Au = \lambda u$ and normalizing the eigenfunctions u to satisfy, $\lVert u\rVert = 1$. Then, the first eigenfunction u_1 solves the optimization problem given in (3.1.4) resulting in a maximum value equal to λ_1. Subsequent eigenfunctions u_j solve the constrained problem given by (3.1.4). Then, maximum at the j^{th} stage is, $\langle u_j, Au_j\rangle = \lambda_j \lVert u_j\rVert^2 = \lambda_j$.

Each of the principal components weight functions $u_j(t)$ should satisfy

$$\int s(t,t^+)u(t^+)dt^+ = \lambda u(t) \qquad (3.1.5)$$

for a certain eigenvalue λ. Left hand side of equation (3.1.5) is the integral transform of the weight function u by the covariance function S given by $Su = \int s(.,t^+)u(t^+)dt^+$. This is called the covariance operator, S. Hence, $Su = \lambda u$, can be written.

In MDA there are p eigenvalue-eigenvector pairs, whereas in FDA it becomes the number of function values which are infinitely many.

3.1.2. Smoothing in Functional PCA

Data matrix described in MDA, rows are the subjects or trajectories, representing one realization of the random process governed by p random variables. In FDA a trajectory represents a random function in continuum. The continuous environment T is a bounded time or space interval over which the domain of the random process $X(.)$ lies.

The weight functions $u(t)$ used in the computation of principal components needs to be smoothed by controlling their roughness. This results in the smoothing of the principal components. For the first PC the function $u_1(t)$ maximizes the variance of the principal component scores subject to

$$\int \{u(t)\}^2 dt + \alpha \int \{u''(t)\}^2 dt = 1;\ \alpha \geq 0\,. \tag{3.1.6}$$

α in (3.1.6) is a control factor over the amount of smoothing required. For smoothing the second and higher order principal component scores in addition to constraint (3.1.6), the constraint

$$\int u_i(t)u_j(t)dt + \alpha \int u_i''(t)u_j''(t)dt = 0;\ i \neq j$$

is required.

Smoothing the sample mean function has a major importance in FDA. A functional data set $X_1(t),\ldots,X_n(t)$ can be expanded in terms of basis functions, $f_1(t),\ldots,f_m(t)$. A coefficient matrix $\mathbf{A}_{n\times m}$ can be defined such that, $X_i(t)=\sum_{j=1}^{m} a_{ij} f_j(t)$. Then the smoothed sample mean becomes,

$$\bar{X}(t)=\sum_{j=1}^{m}\bar{a}_j f_j(t),\ \text{if}\ \bar{a}_j = n^{-1}\sum_i a_{ij}\ .$$

3.1.3 Storing Functional Data/Observations

Taking a basis to mean a standard set of functions $\left(f_1(t),\ldots,f_m(t)\right)$ such that, any function of interest can be expanded in terms of $f_j(t)$. Then, a functional datum $x(t)$ can be expressed in terms of m basis functions and the principal component weight (u) as $x(t)=\sum_{j=1}^{m} u_j f_j(t)$.

In FDA basis functions are designed to represent the nature of the process under study. This is achieved by fitting basis coefficients to the observed data. Given values $x_1,\ldots,x_n$ observed at locations $t_1,\ldots,t_n$, basis functions $f_j(t)$ can be represented by the matrix $\mathbf{B}=f_j(t_i),\ i=1,\ldots,n,\ j=1,\ldots,m$.

If $\mathbf{u}$ is the vector of coefficients, then the vector of values at the observed locations will be $\mathbf{Bu}$. When the number of basis functions are at most the same as the observation locations ($m\le n$), then the basis functions can be fit by minimizing the sum of squares of deviations, $\left[x_i-\sum_j u_j f_j(t)\right]^2$.

For $m = n$ the expansion $x(t) = \sum_j u_j f_j(t)$ gives an exact interpolation of the x_i values.

For $m < n$ the expansion is the smooth version of the initial data.

If $m > n$, more basis functions than observed locations, then, a choice of u values gives exact interpolation of the x_i values. That is, $x_k(t) = \sum_{j=1}^{m} u_j f_j(t_k),\ k = 1, \ldots, n$.

Then, the interpolation that includes the parameters that minimizes the roughness of the curve is selected.

3.2 Expressing the Random Process in Terms of Global Mean and Covariance

Let $\{X_t\}_{t\in[a,b]}$ be a random process in the interval $[a,b]$. Then the random trajectories X in $L^2(I)$ assumed to have mean function $\mu(t) = E(X(t))$ and covariance function $G(t,t^+) = \mathrm{cov}(X(t), X(t^+))$, where $t, t^+ \in I$. It is assumed that the operators on covariance have a sequence of orthonormal eigenfunctions ϕ_k with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$.

The Hilbert- Schmidt kernel, G is very useful in the expression of a random process in terms of global mean and covariance and is given as

$$G(t,t^{+}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(t^{+}),\ \text{t,t}^{+} \in T\ . \qquad (3.1.7)$$

This is the orthogonal expansion of the covariance in L^2 in terms of eigenfunctions ϕ_k and corresponding eigenvalues $\lambda_k,\ k = 1,2,\ldots$ and $\lambda_1 \geq \lambda_2 \geq \ldots.$ Karhunen-Loéve Expansion and, Mercer's theorems guarantee the expansion given in equation (3.1.7) and its spectral decomposition. Details of the Karhunen-Loéve theorems are in Apendix C.

In classical functional PCA, for the process, $\{X_t\}_{t \in [a,b]}$, the random curve (trajectory) $X(t)$ can be expressed as,

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \phi_k(t) \xi_k,\ \text{t} \in \text{T}\ . \qquad (3.1.8)$$

Here, ξ_k are uncorrelated random variables with, $E(\xi_k) = 0$ and $E(\xi_k^2) = \lambda_k,\ \sum_k \lambda_k < \infty$. For any finite K, (3.1.8) can be written as,

$$X(t) = \mu(t) + \sum_{k=1}^{K} \phi_k(t) \xi_k\ . \qquad (3.1.9)$$

K, determines the fraction of variance, $F(K) = \sum_{i=1}^{K} \lambda_i / \sum_{k=1}^{\infty} \lambda_k$, in the process under consideration and is required to be as high as possible (preferably above 0.8).

3.3 The Principal Components Analysis Through Conditional Expectation

Functional principal component scores, ξ_{ik}, play a major role in the estimation of a trajectory. They are uncorrelated random variables with mean zero and variances being the eigenvalues of covariance matrix $\mathbf{G}$. The functional principal components scores given directly by equation (3.1.1), and it works well when the density of the grid of measurements for each subject is sufficiently large.

The functional representation of a trajectory, $X(\cdot)$, for the j^{th} observation of the i^{th} subject made at time t_{ij}, $i=1,...,n$, $j=1,...,p_i$ is represented by Y_{ij}. Number of observations, p_i, made on each of the i^{th} subjects are assumed to be i.i.d. random variables. If no error is involved, then, $Y_{ij}=X(t_{ij})$. However, observations will inherently include some measurement errors ε_{ij} that are also assumed to be i.i.d. with $E(\varepsilon_{ij})=0$ and constant variance σ^2. Then, the model representing the i^{th} subject based on observations can be written as

$$Y_{ij}=\mu(t_{ij})+\sum_{k=1}^{\infty}\xi_{ik}\phi_k(t_{ij})+\varepsilon_{ij}\ ,\ t_{ij}\in T\ . \qquad (3.1.10)$$

Thus, instead of the integral, the estimated principal components scores are given by $\xi_{ik}=\sum_{j=1}^{p_i}(Y_{ij}-\mu(t_{ij}))\phi_k(t_{ij})(t_{ij}-t_{ij-1})$. However, this does not work well when the data are sparse, since substituting Y_{ij} for $X_i(t_{ij})$ causes biased FPC scores. Therefore, an alternative method, the Principal Components Analysis through Conditional Expectation (PACE) is proposed by Yao, et. al. (2005). The PACE method, assumes that the principal component scores ξ_{ik} and error ε_{ik} are jointly Gaussian. Let $X_i = (X_i(t_{i1}),..., X_i(t_{ip_i}))^T$, $Y_i = (Y_i(t_{i1}),..., Y_i(t_{ip_i}))^T$, $\mu_i = (\mu(t_{i1}),..., \mu(t_{ip_i}))^T$, and $\phi_{ik}^T = (\phi_k(t_{i1}),...\phi_k(t_{ip_i}))$. Thus, under Gaussian assumptions, the best prediction of the PC scores can be found by conditional expectation.

$$\xi_{ik} = E[\xi_{ik} \mid Y_i] = \lambda_k \phi_{ik}^T \Sigma_{Y_i}^{-1} (Y_i - \mu_i), \tag{3.1.11}$$

where $\Sigma_{Y_i} = \mathrm{cov}(Y_i, Y_i) = \mathrm{cov}(X_i, X_i) + \sigma^2 \mathbf{I}_{p_i} = G(T_{ij}, T_{il}) + \sigma^2 \delta_{jl}$.

Substituting the estimates of λ_k, ϕ_{ik}, Σ_{Y_i} and μ_i in equation (3.1.11), leading to

$$\xi_{ik} = E[\xi_{ik} \mid Y_i] = \lambda_k \phi_{ik}{}^T \Sigma_{Y_i}^{-1} (Y_i - \mu_i), \tag{3.1.12}$$

where, Σ_{Y_i} is obtained from the whole data. Then, the infinite-dimensional processes are approximated by the projection on the functional space spanned by the first K eigenfunctions of estimated Σ_{Y_i} covariance matrix. Thus, in practice the estimated i^{th} trajectory is

$$X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik} \phi_k(t). \tag{3.1.13}$$

The conditional method (3.1.12) under the Gaussian assumptions works well in both case of sparse and dense data and yields the best estimates. It is also worth mentioning that (3.1.11) is the best linear prediction of principal components scores ξ_{ik} and works well weather the Gaussian assumption holds or not (Yao, et. al. , 2005). See Yao, et.al., (2003) for another estimation for functional principal component scores.

Chapter 4

SMOOTHING

4.1 Smoothing The Mean

In a random process $X(\cdot)$ the underlying function f is generally unknown. Available data collected or observations made at p different time or space locations, $Y_i(t_1), Y_i(t_2), \cdots, Y_i(t_p)$ give some idea about the likely behaviour of the function. Using available observations to predict the underlying random function representing the random process is not an easy task, since data is mostly contaminated by errors due to various agents. If there is sufficient evidence to indicate that data is error free, then some simple linear interpolation may be adequate to represent $X(\cdot)$ with available observations. However, in the presence of error ε smoothing will be required to represent $X(\cdot)$. Here, ε_i is i.i.d. with $E(\varepsilon_i) = 0$ and $\text{var}(\varepsilon_i) = \sigma^2$.

Let $Y_i = X(t_i) + \varepsilon_i,\ i = 1, \ldots, n$ be the noisy representation of $X(\cdot)$. If $\mathbf{Y}$ is the matrix of observations, then $\text{var}(\mathbf{Y}) = \mathbf{\Sigma}_e = \sigma^2 \mathbf{I}$.

There are many different smoothing techniques used in various application fields. Some of the important ones are introduced in the following section.

4.1.1. Commonly Used Smoothing Methods

A simple smoothing will be the fitting of ordinary least squares function to the data defined by the basis function expansion,

$$X(t_j) = \sum_{k}^{K} b_k \phi_k(t_j) = \mathbf{b}^T \mathbf{\Phi} .$$

$\left\{\phi(t_j)\right\}_k$ are the basis functions and the coefficients vector $\mathbf{b}_{1\times K}$ is determined by minimizing the sum of square errors (SSE),

$$SSE(Y \mid b) = \sum_{j=1}^{n} \left[Y_j - \sum_{k}^{K} b_k \phi_k(t_j) \right]^2 = (\mathbf{Y} - \mathbf{\Phi b})^T (\mathbf{Y} - \mathbf{\Phi b}) = \left\| \mathbf{Y} - \mathbf{\Phi b} \right\|^2 .$$

Then, the estimated (fitted) vector $\hat{\mathbf{y}}$ is found by

$$\hat{\mathbf{y}} = \mathbf{\Phi}(\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{Y} .$$

This assumes equal weight assignment to all observations, regardless of observation time/space t. In the case of regular grid this may be acceptable, but for irregularly observed data a weighing method is necessary. Then, the weighted least squares (WLS) fit offers a solution given by

$$SSE(Y \mid b) = (\mathbf{Y} - \mathbf{\Phi b})^T \mathbf{W} (\mathbf{Y} - \mathbf{\Phi b}) .$$

Weights W_j to be assigned can be computed using different methods. Kernel functions, splines, moving averages are just a few that can be used. In the case of kernel functions, some of widely used ones are

$$\text{Uniform: } W_j(t) = K\left(\frac{t_s - t_p}{h}\right) = \begin{cases} 0.5 & \text{for } \left|\frac{t_s - t_p}{h}\right| \le 1 \\ 0, & \text{otherwise} \end{cases}.$$

$$\text{Quadratic: } W_j(t) = K\left(\frac{t_s - t_p}{h}\right) = \begin{cases} 0.75\left(1 - \left(\frac{t_s - t_p}{h}\right)^2\right) & \text{for } \left|\frac{t_s - t_p}{h}\right| \le 1 \\ 0\ , & \text{otherwise} \end{cases}.$$

$$\text{Gaussian: } W_j(t) = K\left(\frac{t_s - t_p}{h}\right) = \frac{e^{-\frac{1}{2}\left(\frac{t_s - t_p}{h}\right)^2}}{\sqrt{2\pi}}.$$

$$\text{Nadaraya-Watson: } W_j(t) == \frac{K\left(\frac{t_s - t_p}{h}\right)}{\sum_s K\left(\frac{t_s - t_p}{h}\right)}.$$

Widely used *Epanechnikov* kernel function takes the form $K(x) = 0.75(1 - x^2)\mathbf{1}_{[-1,1]}(x)$ is the univariate case and $K(x, y) = 0.5625(1 - x^2)(1 - y^2)\mathbf{1}_{[-1,1]}(x)\mathbf{1}_{[-1,1]}(y)$ is the bivariate case with $\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise for any set A.

Smoothing the mean $\mu(t)$ and covariance $Cov(X(t), X(t^+))$ of the set of observed curves is necessary in many applications (Rice and Silverman, 1991).

Using the spline smoothing to smooth the mean is to use the penalized least squares. Given the i^{th} data vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^T$, the estimated mean curve $\hat{\mu}$ should minimize

$$n^{-1}\sum_i \lVert \mathbf{x}_i - \mu \rVert^2 + \alpha \int \mu''(t)^2\, dt\,. \qquad (4.1.1)$$

Here, α is a positive smoothing parameter and the integral represents the roughness of μ. (4.1.1) is equivalent to $\sum_j [\bar{X}_j - \mu(t_j)]^2 + \alpha \int \mu''(t)^2\, dt$, which is the *spline smoothing* applied to pointwise averages. Choice of the smoothing parameter α is subjective, but methods such as cross validation (CV) are available to automate the choice of optimum α value (Heckman, 1986).

Similarly various moving averages techniques can also be used as the smoother of the mean and covariance functions. The stretches interpolated moving average technique developed during the course of this study and given in Section 4.3 is successfully applied to the smoothing of the mean. It is shown to be more robust than the local linear smoothing when the correlation between trajectories are weak.

Local Linear Regression is also used as a smoother and some details of this method are given in Section 4.3.3.

4.1.2. Kernel Smoothing

One other method used for smoothing the mean is the scatter plot smoothing utilizing kernel smoothers. Let $Y_{ij}(t),\ i = 1,\ldots,n$ and $j = 1,\ldots, p_i$, be the j^{th} observation on the i^{th} subject made at time t. If the model has no additional error then shortly $Y_{ij} = X(t_{ij})$ can be written.

The local linear scatter plot smoother at time t for $\mu(t)$ is obtained from the scatter plot (t_{ij}, Y_{ij}) by minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{p_i} \kappa\left(\tfrac{t_{ij}-t}{h_\mu}\right)\left\{Y_{ij} - \beta_0 - \beta_1(t - t_{ij})\right\}^2, \tag{4.1.2}$$

with respect to β_0 and β_1, where h_μ is the bandwidth and univariate density κ is the kernel function. Then the estimate of $\mu(t)$ is $\hat{\mu}(t) = \hat{\beta}_0(t)$. The minimization can be done by taking the derivative of (4.1.2) with respect to β_0 and β_1. Solution of the obtained equations yields the local linear estimator (smoothed) $\hat{\mu}(t)$

$$\hat{\mu}(t) = \hat{\beta}_0(t) = \frac{\sum_i \frac{1}{Ep} \sum_j w_{ij} Y_{ij}}{\sum_i \frac{1}{Ep} \sum_j w_{ij}} - \frac{\sum_i \frac{1}{Ep} \sum_j w_{ij}(t_{ij} - t)}{\sum_i \frac{1}{Ep} \sum_j w_{ij}} \hat{\beta}_1(t), \tag{4.1.3}$$

where,

$$\hat{\beta}_1(t) = \frac{\sum_i \frac{1}{Ep} \sum_j w_{ij}(t_{ij} - t) Y_{ij} - \dfrac{\sum_i \frac{1}{Ep} \sum_j w_{ij}(t_{ij} - t) \sum_i \frac{1}{Ep} \sum_j w_{ij} Y_{ij}}{\sum_i \frac{1}{Ep} \sum_j w_{ij}}}{\sum_i \frac{1}{Ep} \sum_j w_{ij}(t_{ij} - t)^2 - \dfrac{\left(\sum_i \frac{1}{Ep} \sum_j w_{ij}(t_{ij} - t)\right)^2}{\sum_i \frac{1}{Ep} \sum_j w_{ij}}}.$$

Here, $Ep = \frac{\sum p_i}{n}$ (p_i: Number of observations on the i^{th} trajectory) and

$w_{ij} = \kappa\left(\frac{t_{ij} - t}{h_\mu}\right) \Big/ nh_\mu$ is the weight function, $\kappa(x) = 0.75(1 - x^2)\mathbf{1}_{[-1,1]}(x)$ is the uni-variate *Epanechnikov* kernel function with $\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise for any set A where t is the starting data value of bandwidth h_μ (Yao et. al. 2005). Alternative formulas for (4.1.3) can be found in Hall et.al. (2006), and in Müller (2005).

Choi and Hall (1998) give interesting ideas about bias reduction in local linear smoothing depending on the kernel function.

4.2 Moving Average Approach to Smoothing

Moving average is a well-known technique in a broad range of disciplines and initially used for trend generation in mainly time dependent variables. Various versions of moving averages are in implementation, depending on the nature of the process under consideration. Simple, cumulative, weighted, exponential moving average techniques are some of the commonly used ones for trend generation.

Cumulative moving average is used when the average starting from a given time point up to the present is required for decision making. Weighted moving average is mostly used in cases where higher weights are to be assigned for the latest or new data. Different weight determination techniques are employed. Widely used weighing techniques are linearly or exponentially decreasing, by assigning the high weights to most recent data and lower weights as data becomes old (Durbin, 1959). In finance, economy or in some medical applications weighted moving average is used.

In principle moving average is taken to be a simple smoothing method, but its examination in detail indicates that sophisticated moving average methods can be developed. In part of the research devoted to this thesis, some new ideas are developed, and due to the process involved, it is named as "Stretched Interpolated Moving Average" (SIMA). Obtained smoothing results are compared with those of Local Linear Smoothing (LLR) and PACE results, to verify the validity of SIMA. First step is to locate the computed moving averages at equal distance over the range of the data. Hence, named as, *stretched* moving average, details of which are given in the following section.

4.3. Stretched Interpolated Moving Average (SIMA)

4.3.1. Stretched Moving Average

The method developed handles the moving average process when data are available on a regular grid basis in terms of time or space coordinates. That means, each datum is located at equal time or space interval from its immediate neighboring data values. Considering the one dimensional case, when p data values are regularly spaced on a trajectory, m data points within a fixed lag interval h are used for averaging. The number of averaged values will be, $M = p - m + 1$. Process starts with the first or oldest data averaging the first m data values (Kenny and Durbin, 1982). Let obtained average values be denoted by the random variable Y. Given p_i observations located at equal distance on the i^{th} trajectory, the moving average for each lag with m data values will be

$$Y_{il} = \frac{1}{m}\sum_{j=l}^{m+l-1} X_{ij}, \ 2 \le m \le p_i, \ i = 1,\ldots,n, \ l = 1,\ldots,M. \quad (4.3.1)$$

Every averaged value has to be assigned to a coordinate within the lag interval it belongs to. Assuming every averaged value Y is assigned to the first point's coordinate of each lag, the final average obtained from the last or M^{th} lag will fall h units distance behind the present or newest datum. Alternately, averaged values can be assigned to the mid-point or last point of each lag. Assignment of the upper lag boundary would mean ignoring a time or space interval equivalent to h from the beginning point of the data values. The idea proposed to overcome this handicap, is to assign the coordinate of the first datum to the first average value, and the coordinate of the last datum to the last (M^{th}) average value. Remaining $M-2$ moving average values will be equally spaced over the span of data.

Mean and variance of the moving average values for the i^{th} trajectory is

$$E(Y_i) = \frac{1}{M}\sum_{l=1}^{M} Y_{il} = \frac{1}{mM}\sum_{l=1}^{M}\sum_{j=l}^{m+l-1} X_{ij}$$

$$\sigma^2_{Y_i} = \frac{1}{Mm^2}\sum_{l=1}^{M}\left(\sum_{j=l}^{m+l-1} X_{ij}\right)^2 - \left(\frac{1}{Mm}\sum_{l=1}^{M}\sum_{j=l}^{m+l-1} X_{ij}\right)^2.$$

Relationship between the moving average values Y_{il}, with data values X_{il} can be found as follows (Mentz, 1975).

Consider the case, $m = 2$, then

$$Y_{il} = \frac{1}{m}\left(X_{il} + X_{il+1}\right).$$

Via mathematical induction, $X_{il+1} = mY_{il} - X_{il}$, follows with,

$$\begin{aligned} X_{il+1} &= mY_{il} - mY_{il-1} + X_{il-1} \\ &= mY_{il} - mY_{il-1} + mY_{il-2} - X_{il-2} \\ &= mY_{il} - mY_{il-1} + mY_{il-2} - mY_{il-3} + X_{il-3} \\ &= \vdots \\ &= m\sum_{j=1}^{l}(-1)^{j+l}Y_{ij} + (-1)^{l}X_{i1}. \end{aligned}$$

For the formulation of the Stretched Moving Average, let random variable S be the time or space interval covered by p_i data values on the i^{th} trajectory. It is required to locate the averaged values equally spaced on the same interval S. In other words, p_i data values X_{ij}, $j = 1,..., p_i$ on the i^{th} trajectory covers an interval S, the same interval will also be covered by M_i moving average values. Without loss of generality the data values can be assumed to be uniformly distributed on a trajectory with equal interval between points.

Then, distance between data values is $d = S/(p-1)$ and the distance between the moving average points will be $s = S/(M-1)$. Note that, $d \leq s$. Alternately if p data values are uniformly located on S with unit interval in between ($d = 1$), then $s = (p-1)/(M-1)$.

Moving average values will be located at s distance apart such that the first one Y_{i1} will correspond to the location of X_{i1}, and last one Y_{iM} corresponding to the location of X_{ip}. Other moving average values ranging from Y_{i2} to Y_{iM-1} will occupy locations on S accordingly at equal distances. Steps followed in the assignment of coordinates to compute moving average values, is named as the *Stretched Moving Average.*

Clearly as the lag interval becomes larger, in simple moving average the computed final Y_{il} value will lag farther behind the latest data values. Magnitude of this distance is $(m-1)d$. Stretched moving average eliminates the handicap by assigning the computed Y_{il} values uniformly over the full interval S covered by the trajectory.

Let $t \in S$, $s = \dfrac{s_M - s_1}{M-1} = \dfrac{S}{M-1}$, where $s_1 = \min(t)$ and $s_M = \max(t)$, then the stretched moving average can be written as

$$Y_{il}(s_{il}) = \frac{1}{m} \sum_{j=l}^{m+l-1} X_{ij}(t), \; l = 1, \dots, M$$

(Tandoğdu and İyikal, 2013).

To highlight the stretched moving average concept, a hypothetical data set consisting of p=10 data values X_{ij}, $j = 1, \dots, 10$ given at unit interval apart over the i^{th} trajectory is used. Cumulative distances from the first datum towards the last one will be, $t_1 = 0,\ t_2 = 1, \dots, t_{10} = 9$. This means, data is spread over $S = 9$ units of time or space interval. Hence, data can be represented by $X_{ij}(t_j)$. If a lag interval of $h = 3$ units is used, it will include $m=4$ data values for the averaging process. Then, $s=(p-1)/(M-1)=9/6=1.5$ units. First moving average value $Y_{i1}(s_1)$ will be assigned to the same location as, t_1, i.e. $s_1=t_1$. Subsequent moving average values will be $Y_{il}(s_l)$. The graph showing the raw data, simple moving average and stretched moving average is as shown in Figure 4.1.

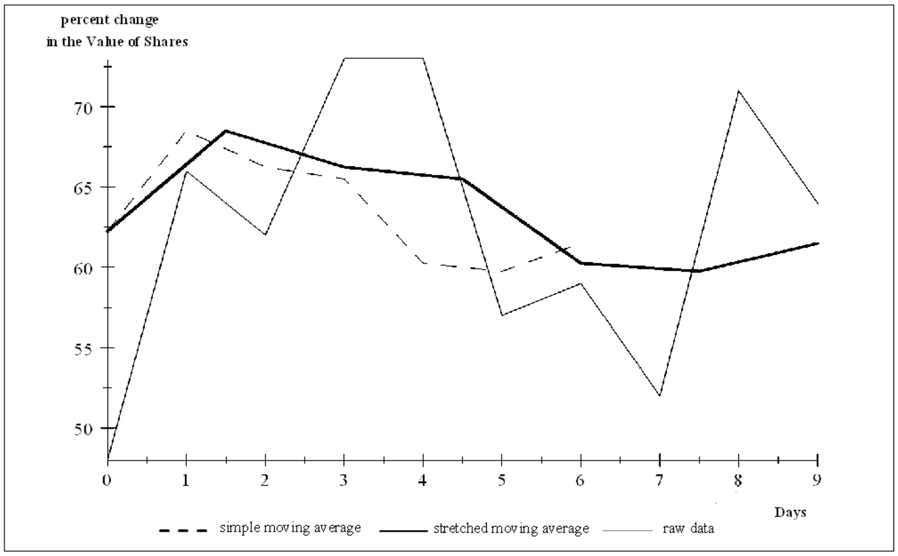

Figure 4.1: Raw, Simple Moving and Stretched Moving Averages.

The simple moving average graph lags behind the latest raw data points X_{ij} by $m-1$ unit distance. Stretching the locations of moving average values to cover the whole space S, is the first step towards the computation of the error involved in smoothing committed by the moving average process. Accurate computation of this error would require a one to one correspondence between the data points and the moving average values. However, the number of moving average values is less than the number of data points on a trajectory by $m-1$. A linear interpolation is introduced to equalize the number of moving average values with the number of data values on a trajectory.

Further, the normality of a trajectory, leads to the conclusion that the stretched moving average trajectory is also normal as explained below.

It is known that the random process X_i has a p-variate normal distribution if and only if $\mathbf{a}^T X_i$ is univariate normal for all fixed p vectors $\mathbf{a}$. Then a linear combination $\mathbf{y} = \mathbf{A}X_i + \mathbf{c}$ of X_i has a q-variate normal distribution, where $\mathbf{A}$ is a $q \times p$ matrix of coefficients and $\mathbf{c}$ is a $q \times 1$ vector of constants (Mardia, et.al., 1979). In place of $\mathbf{y}$ the stretched moving average $Y_i(t)$ can be written, leading to the fact expressed in the following theorem.

Theorem 4.3.1: If the i^{th} trajectory $X_i(t)$ of the random process has p-variate normal distribution, then the corresponding stretched moving average trajectory $Y_i(t)$ has an (p-m+1)-variate normal distribution.

Proof of Theorem 4.3.1 is given in the Appendix B.

4.3.2. Linear Interpolation of the Stretched Moving Averages

Since the idea is to use the moving average as a smoother, the error committed in the smoothing process should be measurable. This measure is usually expressed in some form of the difference between the observations and the corresponding moving average values. However, as number of observations on a trajectory is always greater than the number of computed moving average values ($p_i > M_i$), it is deemed necessary to compute a moving average value corresponding to each observation. This is possible by linear interpolation of the stretched moving average values. Method followed to obtain these averages is named as "*Stretched Interpolated Moving Average"* (SIMA). Obtained new averages for the i^{th} trajectory are denoted by Y_i^* and given by

$$Y_{ij}^*(t) = [Y_{ij}(s_j) - Y_{ij-1}(s_{j-1})]w_j + Y_{ij-1}(s_{j-1}) \qquad (4.3.2)$$

where, $w_j = \dfrac{t - s_{j-1}}{s_j - s_{j-1}}$. In (4.3.2), the first stretched interpolated moving average value Y_{i1}^* is equal to the first moving average value Y_{i1}, the last stretched interpolated moving average value Y_{ip}^* is equal to the last moving average value Y_{ip} (Tandoğdu and İyikal, 2013).

Sample covariance function's limit distribution when the variance is finite in the moving average process, is derived by Davis and Resnick (1986). Obviously the difference between the observed and smoothed values or the error is required to be a minimum.

In SIMA, random variable p_i represents the number of observations on the i^{th} trajectory and p_i's are i.i.d. Assuming t_j and $X_{ij} : j \in J_i$ are independent of p_i, $E(p_i) < \infty$, $P(p_i > 1) > 0$, and $E(X_{ij}) = \mu$, then it is a well known fact that as $p_i \to \infty$ in a fixed interval T, and $m \to 1$,

$$\sup_{t \in T} E\,|\,Y_{il}^*(t) - X_{il}(t)\,| \to 0\,.$$

For more details see (Davis and Resnick, 1986) and (Furrer et.al, 2006). The following theorem shows the variance of the error approaches to σ^2 under similar assumptions.

Theorem 4.3.2: Let p_i be the number of observations on the i^{th} trajectory, and also under the assumption that $X_{ij}(t)$'s are i.i.d., with $E(X_{ij})=0$ and $E(X_{ij})^2=\sigma^2$, when $t \to s_j$, i.e. $w_j \to 1$, then,

$$\sup_{t\in T} E[Y_{il}^*(t)-X_{il}(t)]^2 \to \frac{m-1}{m}\sigma^2.$$

Further, when $p_i \to \infty$ and $t \to s_j$ with very large m,

$$\sup_{t\in T} E[Y_{il}^*(t)-X_{il}(t)]^2 \to \sigma^2.$$

Proof is given in Appendix B.

Application of SIMA concept to data trajectories has yielded a smooth trend. Details of the application are explained in Chapter 5. Given a data set consisting of several variables and multiple observations on each variable, smoothing enables the construction of a functional relationship among the variables. Degree of smoothing obtained via SIMA is compared with the smoothing results obtained from LLR. Hence, a brief summary of LLR is given in Section 4.3.3.

4.3.3. Local Linear Regression

The Local Linear Regression (LLR) is a smoothing method primarily based on the idea that a smooth function can be approximated by a low degree polynomial (Fan and Gijbels, 1996). This translates into minimizing

$$\sum_{i=1}^{n} w\left(\frac{x_i - x}{h}\right)\left(Y_i - (a_0 + a_1(x_i - x))\right)^2 ,$$

to obtain estimates for the coefficients a_0 and a_1. Here $w((x_i - x)/h)$ is the kernel weight function with bandwidth h. The LLR estimate at point x is given by $\hat{\mu}(x) = \sum_{i=1}^{n} b_i(x)Y_i$. The coefficients b_i are found by

$$\mathbf{b}(x)^T = (b_1(x),...,b_n(x)) = q^T (\mathbf{X}^T \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}$$

where, $q^T = (1 \quad 0)$, and $\mathbf{X} = \begin{bmatrix} 1 & x_1 - x \\ \vdots & \vdots \\ 1 & x_n - x \end{bmatrix}$ is the design matrix. Further details are given in Loader (1999).

LLR smoothing follows the observed trend closely when the correlation (r_{ij}) between the predictor and response variables is high (preferably, $|r_{ij}| > 0.7$), and deviations from actual trend increase as this correlation becomes lower (Breiman and Friedman, 1985) (Hoover et. al.,1998, Härdle, 1992).

Let r_{ij} be the correlation coefficient between different trajectories, where $i = 1,...,n$ is the trajectory counter and $j = 1,...,p$ is the column counter. Then obtained $p \times p$

correlation coefficient matrix $\mathbf{R}$ is symmetric with, $r_{ij} = r_{ji}$. Close inspection of $\mathbf{R}$ gives an idea in between which trajectories one can expect good or poor LLR smoothing results.

Figures 5.5 and Figure 5.7 are the graphical representation of matrix $\mathbf{R}$ for the data sets used in this study. In examples where natural phenomenon is involved, such as those in Earth Sciences, trajectories spatially closer will tend to have high r values.

4.4. Smoothing the Covariance

Observations made over a trajectory are likely to include errors or noise. Smoothing by expanding into orthogonal eigenfunctions will help to reduce the noise involved during sampling. Therefore, estimation of eigenfunctions becomes a crucial step in smoothing. Covariance matrix from observed data is defined as, $G_{ijl}(Y_{ij} - \hat{\mu}(t_{ij}))(Y_{il} - \hat{\mu}(t_{il}))$. The obtained covariance matrix can be plotted as a scatter plot surface $((t_{ij}, t_{il}), G_{ijl}),\ l, j = 1, \ldots, p_i$, and $i = 1, \ldots, n$ in $\mathbb{R}^3$. Smoothing is applied to this surface under certain regularity assumptions as given in Fang et.al., (2005). Then the local linear surface smoother becomes

$$\sum_{i=1}^{n} \sum_{1 \le j \ne l \le p_i} \kappa\left(\frac{t_{ij} - t}{h_G}, \frac{t_{il} - t}{h_G}\right)\left\{G_i(t_{ij}, t_{il}) - f(\beta, (t, t^+), (t_{ij}, t_{il})\right\}^2 \quad (4.4.1)$$

where, $f(\beta, (t, t^+), (t_{ij}, t_{il})) = \beta_0 + \beta_{11}(t - t_{ij}) + \beta_{12}(t - t_{il})$. Minimization of (4.4.1) with respect to β_0, β_{11}, and β_{12} yields the smooth covariance matrix $\hat{G}(t, t^+) = \hat{\beta}_0(t, t^+)$.

Based on SIMA concept in Sections 4.3.1 and 4.3.2, smoothing the covariance matrix can be performed with a similar thought expressed above.

The covariance surface can be smoothed by applying the SIMA concept in various directions. Smooth surfaces obtained from each direction are then averaged to obtain the final smooth surface. That is equivalent to averaging smooth values obtained for a given point, smoothed by different directions.

Assume k is the number of different directions used in directional smoothing and $\mathbf{G}_1^*,\ldots,\mathbf{G}_k^*$ be directional smoothed covariance matrices. Smoothing in each direction is carried out according to SIMA methodology. Final smoothed covariance matrix $\mathbf{G}^*$ is obtained by averaging the corresponding elements of directional smoothed matrices.

$$\mathbf{G}^* = k^{-1}\sum_k \mathbf{G}_i^*,\ i = 1,\ldots,k \tag{4.4.2}$$

The *directional moving average smoothing* concept in (4.4.2) is still under study. Thus, only one direction is used to smooth the covariance in Chapter 5.

Chapter 5

APPLICATIONS

Multivariate and functional data analysis concepts introduced in Chapter 2 and Chapter 3 together with the new method of smoothing (SIMA) developed during the course of this study given in Chapter 4, sets up the foundation of this thesis. Estimation of the unknown values or parameters in a certain process is of prime importance. Equation (3.1.13) is the key to the estimation of trajectories. Terms involved in this equation can all be estimated as explained in appropriate sections. The global mean function $\hat{\mu}(t)$ is estimated from all available data. $\hat{\phi}(t)$, the eigenfunction of the covariance function is also estimated using all available data. $\hat{\xi}$, represents the principal component scores computed from the data. During the estimation process, some types of smoothing are used. Hence, smoothing becomes an important aspect in the estimation process. There exists a range of different smoothing techniques developed over the years. Kernel, splines, LLR are some of the widely used smoothers. However, there are still certain situations that require special attention. One such point was the use of moving averages as a smoother. When multivariate data are observed on a regular grid and especially when the correlation between variables are low, use of SIMA as explained in detail under Chapter 4, is at least as efficient as the kernel or LLR smoothing methods. Therefore, in the application of the aforementioned methodologies were tried on data sets from different fields to test their validity, and sometimes compare different methodologies.

5.1 Smoothing the Global Mean Function Using Different Smoothing Techniques

Part of this thesis work involved evaluation of the new smoothing technique developed. Smoothing the global mean was undertaken using SIMA, PACE, and LLR methods, and results compared. SIMA is also applied to smoothing the covariance surface and graphically compared with the kernel smoothed covariance surface. Details of the applications are presented in the following sections of this chapter.

5.1.1. Using SIMA and PACE for Smoothing the Mean

There are several smoothing techniques used in various stages of data processing, i.e. direct smoothing of data values, smoothing the mean and covariance functions. These are the cases considered and used in this thesis. As an example, a data set consisting of the daily percent change in the value of shares over 30 consecutive working-days, belonging to 20 companies (trajectories) from İstanbul Stock Exchange is used to estimate the smooth mean function. Raw data and the estimated/smoothed data are shown in Figure 5.1. Negative values refer to a loss and positive values refer to a gain in the value of a share. Actual data set obtained from the online records of İstanbul Stock Exchange web site is given in Appendix D, Table 1.

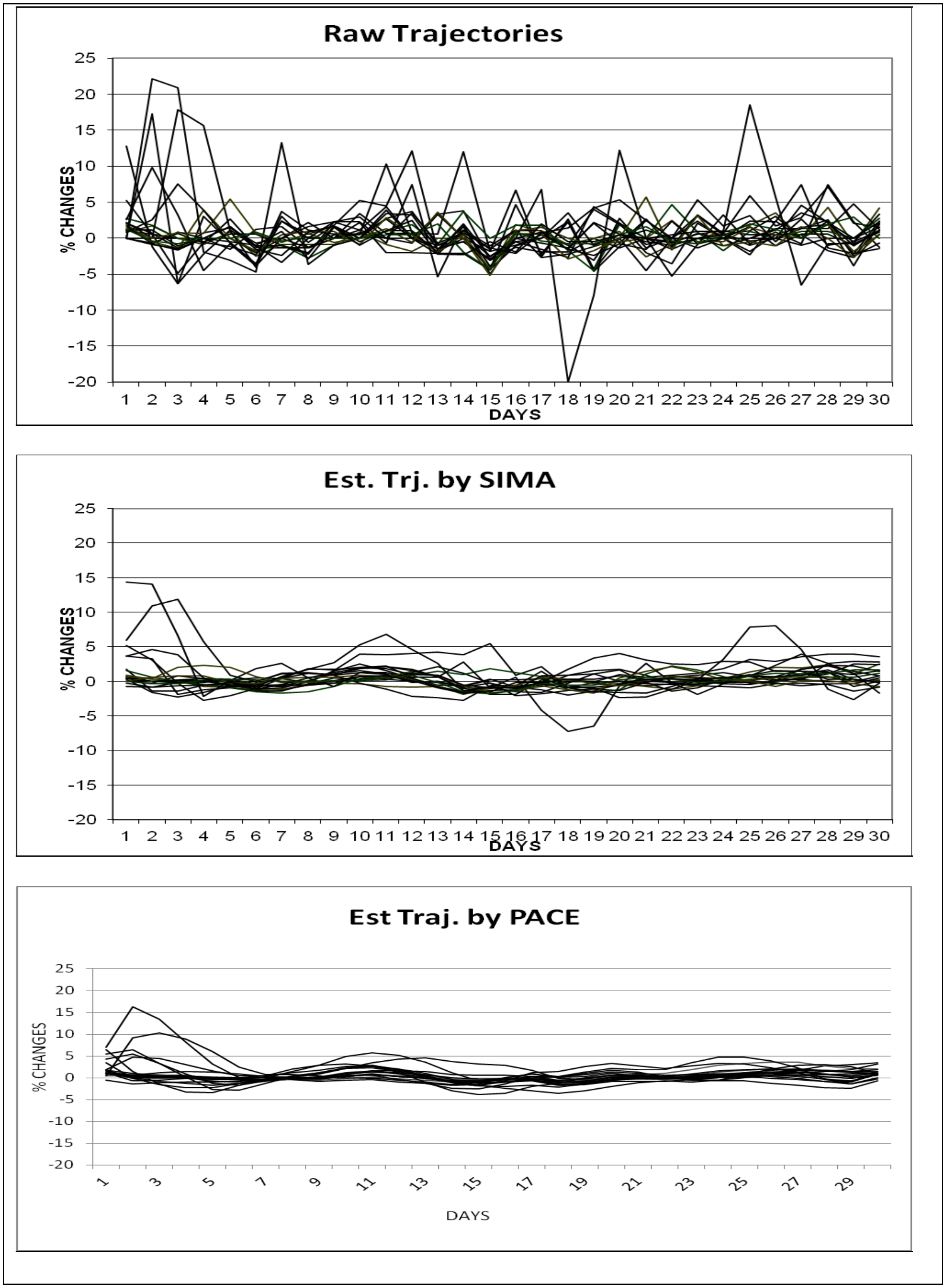

Figure 5.1: Raw Trajectories of the Daily Percent Change in the Value of Shares Over 30 Consecutive Working-Days, Trajectories Smoothed by SIMA, and PACE Methods.

Extreme values are visible in the first few days as well as days 18 and 24. It is a fact that the observation of an extreme value in a process would mean that there is/are unexpected situation(s) temporarily affecting the process. They pose a danger if they are included in a study by shifting the trend (mean and variance) out of its expected path. Therefore, smoothing of data prior to the generation of the estimated mean function is beneficial.

The Epanechnikov kernel smoother used in the PACE method, and the SIMA method are implemented to smooth the trajectories of the data. For each case the average of trajectories are computed and compared with the average of the raw trajectories. MSE functions between average raw and smooth averages obtained through PACE and SIMA are given in Figure 5.2. In general the MSE function obtained from SIMA appears to be lower than that of PACE MSE function, indicating better performance in smoothing by the SIMA method.($MSE_{SIMA} = 0.80$ and $MSE_{PACE} = 1.09$)

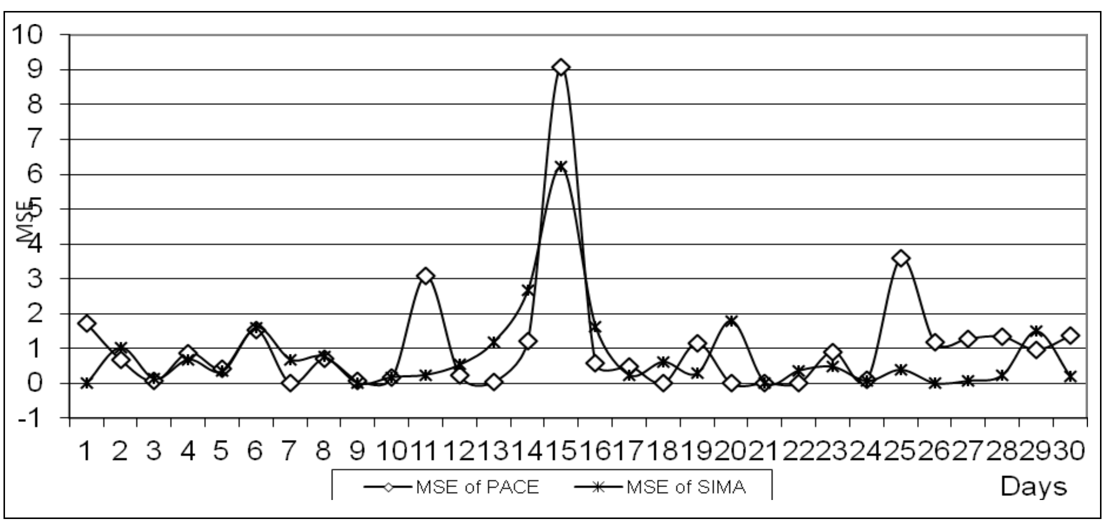

Figure 5.2: MSE Functions between Average Raw and Smooth Averages obtained through PACE and SIMA.

A second data set from a different field of study is chosen for smoothing the mean using SIMA and PACE methodologies. This will give an idea about the performance of the two methods for comparison.

The data set selected represents the coordinates of points in three dimensional space (x, y, z). x is the coordinate of a point in East –West, y is the coordinate in the North – South direction, while z is the elevation from sea level in meters.

Data is obtained from 1:2500 scale topographic maps. Selected area containing gentle slopes as well as locations where rapid changes in elevation are observed, and are selected for the purpose of this study.

The data set consists of 30 parallel trajectories at 50 meters apart and each trajectory containing 30 measurements taken on a regular grid bases at every 50 meters in the East-West direction.

Figure 5.3 shows the raw trajectories of this data in the East-West direction, but there are no significant deviations in the values forming each trajectory. Smoothed trajectories by SIMA and PACE are also given in Figure 5.3 for comparison.

Data used in this study is given in Appendix D, Table 2. Hence, no extreme values.

Smoothing of the mean is undertaken using SIMA and PACE methods and associated MSE functions are generated. Figure 5.4 shows the MSE functions between raw trajectories and smoothed by SIMA and PACE methods. From Figure 5.4 it can be seen that the error functions are following a similar trend, indicating no significant difference between the two methodologies. In fact the overall MSE of SIMA method and PACE method are $MSE_{SIMA} = 8.47$ and $MSE_{PACE} = 7.97$ suporting the idea of close performance.

However, for the data values at locations 3, 4, 28, and 29 the increase in MSE for SIMA is due to rapid change in elevation in the Nort – South direction.

A similar increase in MSE at data locations 14, 15, 17, and 18 in PACE method is considered attributable to over smoothing of the mean, foloowing the detailed study of the raw and smooth mean values.

The shares data set with extreme values and the elevation data set without extreme values are both smoothed using SIMA and PACE methods. For each data set, associated MSE values indicates no significant difference between the two methodologies.

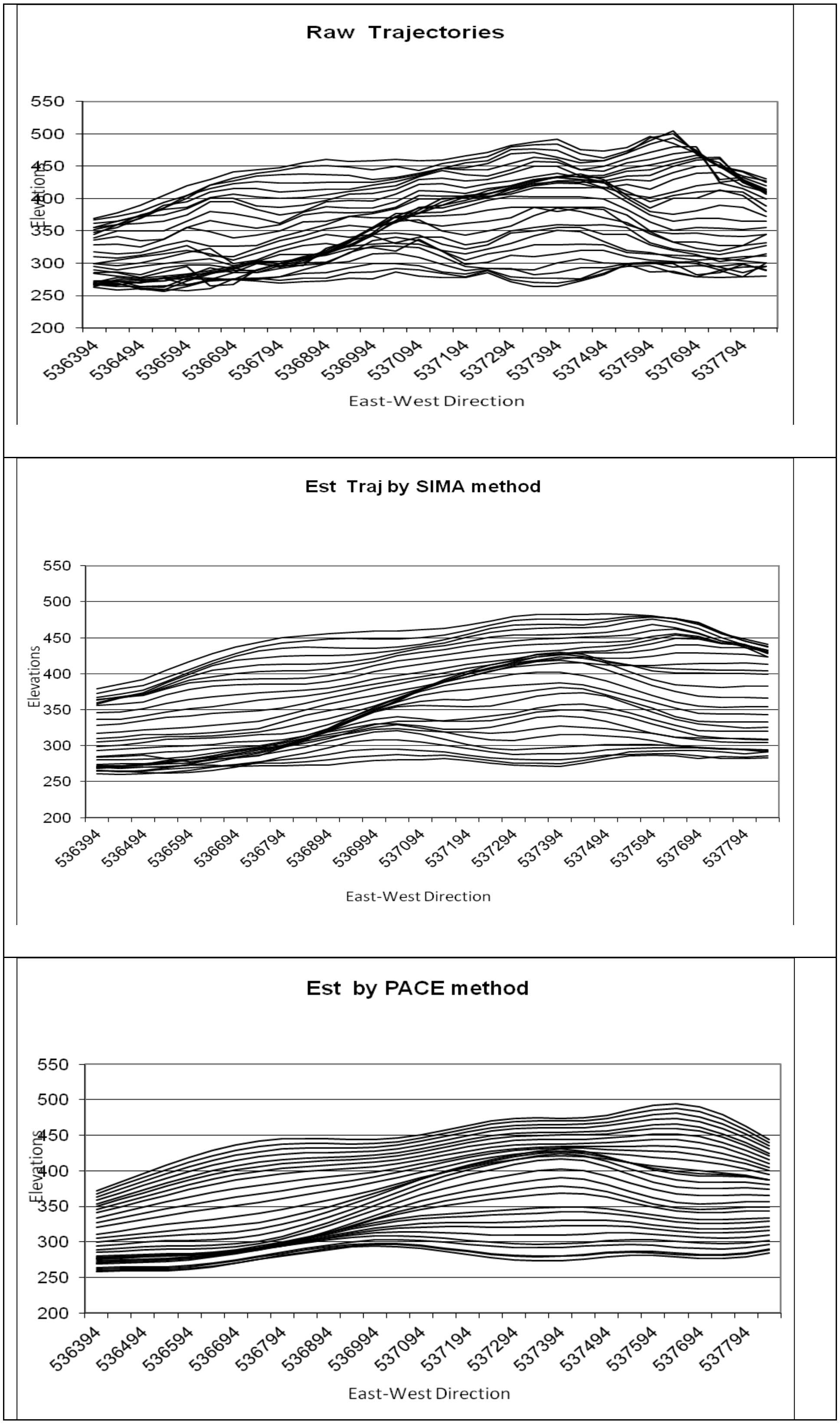

Figure 5.3: Raw and Smoothed Trajectories of Elevation Data.

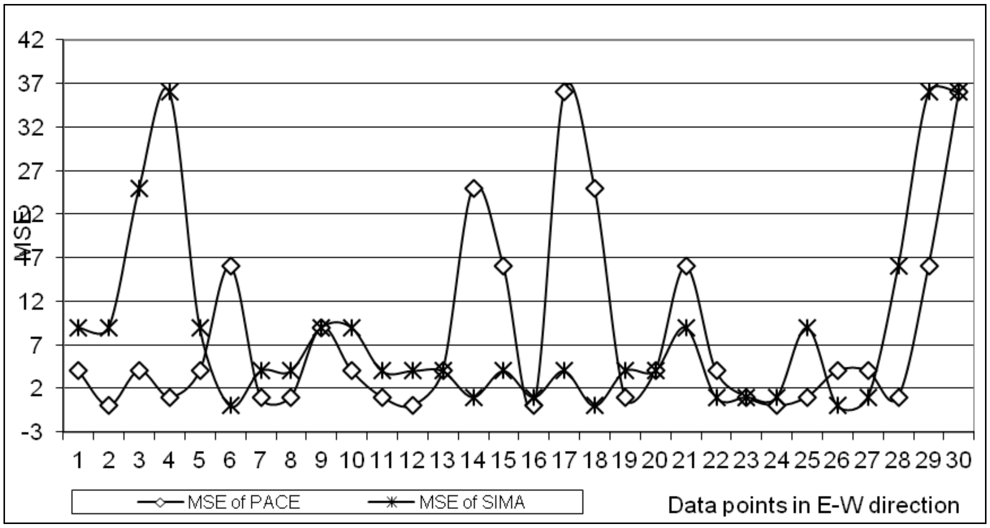

Figure 5.4: MSE Functions for SIMA and PACE Exhibits a Similar Behaviour with a Few Exceptions.

5.1.2 Using SIMA and LLR for Smoothing the Mean.

In this section same data sets used in Section 5.1.1 are used to compare the performance of SIMA with the LLR smoothing method. It is known that the accuracy of smoothing with LLR gets better as the correlation coefficient between the variables increase. For the elevation data set the high correlation between trajectories is expected, since the nature forces shaping the topography are the same within close proximity.

In the shares data set, high correlation between trajectories should not be expected due to market conditions affecting shares differently.

5.1.2.1 High Correlation Between Variables Case

Area from where the elevation data is taken contains gentle slopes as well as locations where rapid changes in elevation are observed. Correlation between trajectories is mostly high to very high $|r_{ij}| > 0.8$. Correlation matrix surface given in Figure 5.5 clearly shows this feature. On the diagonal correlation values are 1 and decline towards the ends as distances between trajectories increase.

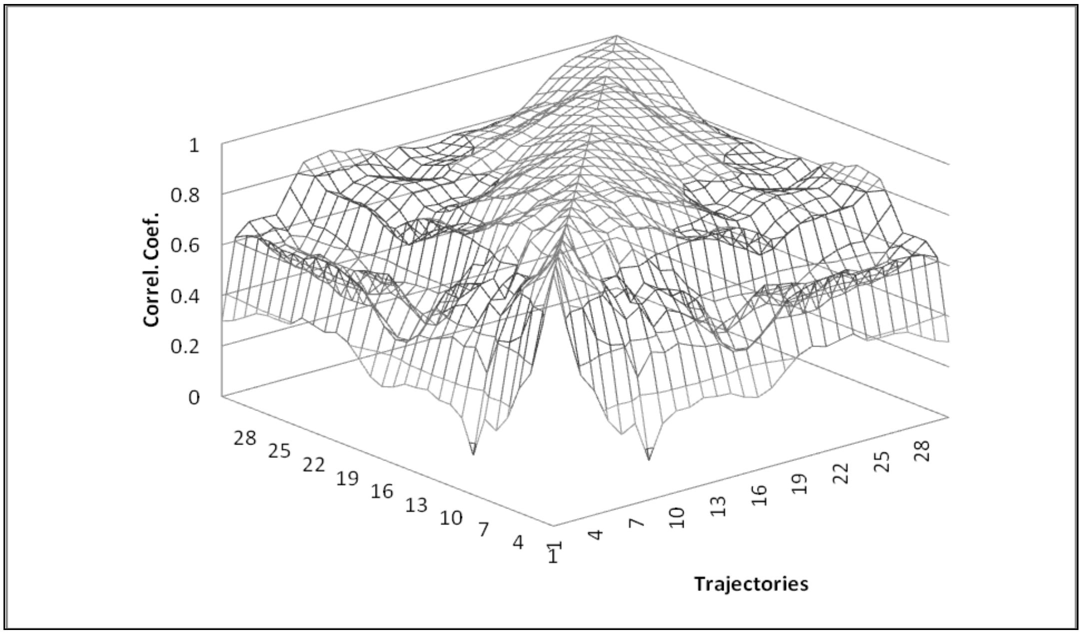

Figure 5.5: Correlation Surface between Different Trajectories of Elevation Data.

Smoothing of trajectories through SIMA is compared with the smoothing results of LLR. For comparison the Root Mean Square Deviation (RMSD) or Root Mean Square Error (RMSE) between the observed and the smooth trajectories for SIMA and LLR are considered. Figure 5.6 shows the RMSD functions for the two smoothing methods.

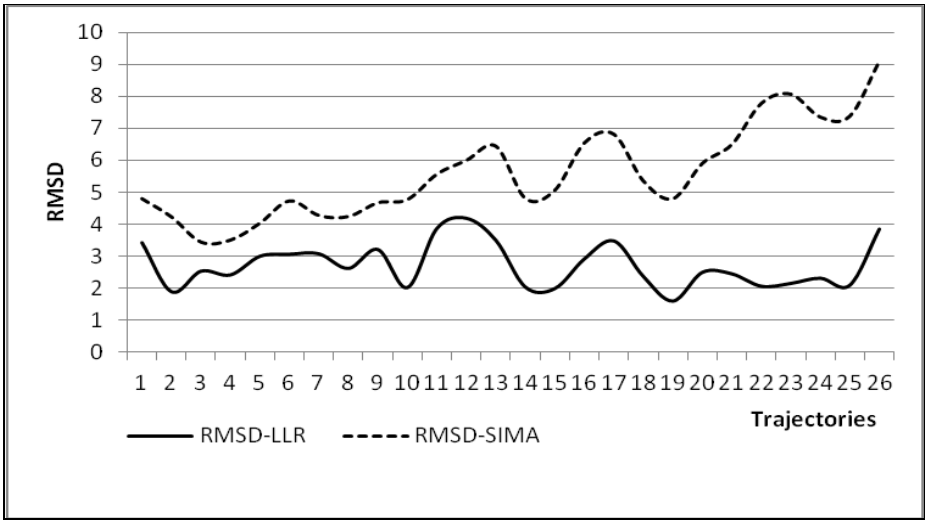

Figure 5.6: RMSD Functions for SIMA and LLR Smoothing Methods.

As expected in the presence of high correlation between trajectories, smoothing carried out by LLR has lower trend, indicating smooth mean following closely the observed mean function. Higher RMSD values obtained in SIMA, meaning when the correlation is high between the variables LLR performs better than SIMA.

5.1.2.2 When Correlation is Low Between Variables

Based on the theory of linear regression it is known that as the correlation between two variables decrease, estimation using the obtained regression equation deteriorates, i.e. becomes less reliable. LLR is used as a smoother. Smoothing in the case when pairwise correlation between variables under study is low will result in high errors. Through trials using data sets where correlation between variables are low, agreed with this concept. The stock exchange shares data set introduced in Section 5.1.1 is used to compute the correlation between the daily performances of different shares. Results are shown in Figure 5.7. Apart from the diagonal elements where the correlation values are

$r_{ij} = 1$, when $i = j$. Of the remaining 870 correlation values only 13% are above 50%, leading to the conclusion of low correlation between 30 variables.

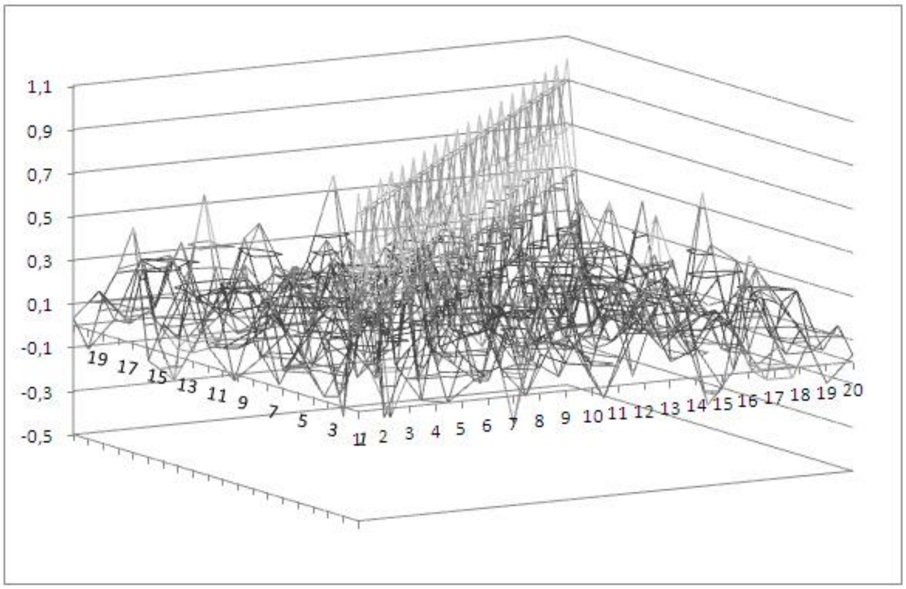

Figure 5.7: Correlation Values Between the Daily Performance of Different Shares.

Smooth mean functions are computed via SIMA and LLR. The RMSD between the average obtained from raw data and the smooth means are shown in Figure 5.8.

Clearly, SIMA resulted in lower errors compared with LLR, indicating better smoothing results can be obtained from a data set where the correlation between variables are low.

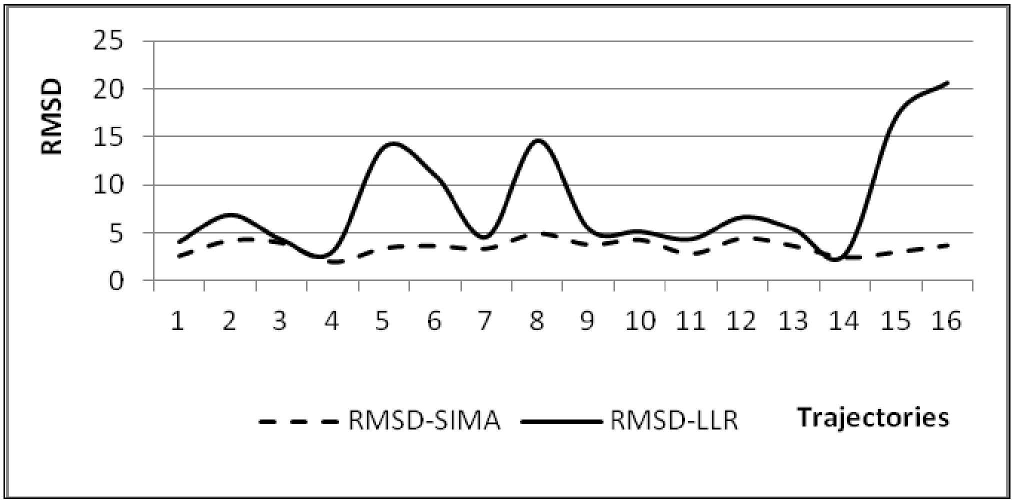

Figure 5.8: RMSD Functions Obtained for SIMA and LLR Smoothing.

In Figure 5.9, correlation coefficient (r_{ij}) values between predictor and response trajectories, and RMSD values between observed and LLR smoothed trajectories are given for comparison. It can be seen that for correlation r_{ij} values under 0.45, RMSD values tend to be not very sensitive to changes in r_{ij}, while RMSD starts decreasing for values of r_{ij} above 45%. This is an expected result due to the nature of the shares of different companies.

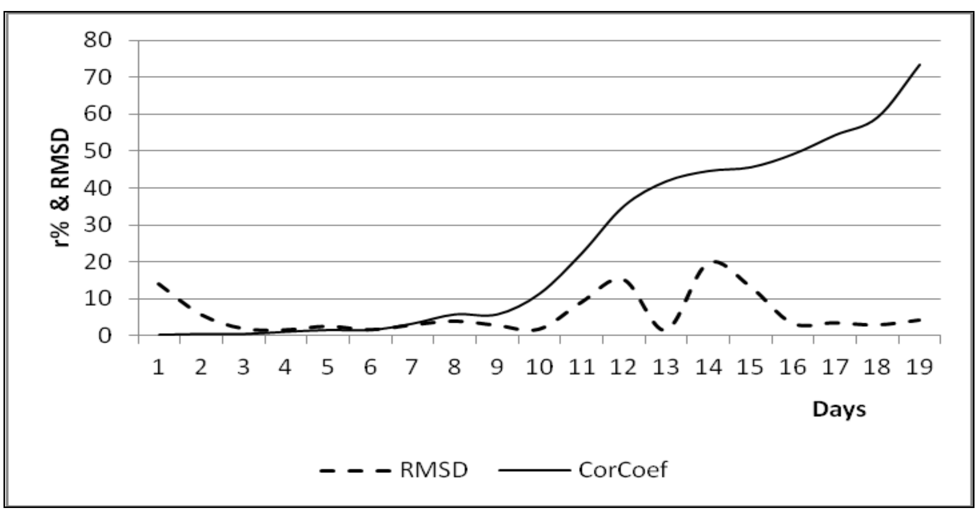

Figure 5.9: Relationship Between Correleation Coefficient and RMSD for the Shares Data Set.

5.2. Smoothing the Covariance Surface Using SIMA and Kernel Smoothers

The covariance matrix plays a major role in MDA and FDA. Smoothing is essential especially when high local variability is observed. Amount of smoothing in SIMA depends on the number of data points (*m*) falling into each lag interval (*h*), while in kernel smoothing used in PACE depends on the band width (*h*).

In the studied elevation and shares data sets, for SIMA $m=3$ and for PACE *h* is automatically selected by generalized cross validation (GCV) method (Müller and Prewitt, 1993). As a result both covariance surfaces are highly smoothed by the kernel smoother, while smoothing by SIMA remained relatively mild compared with kernel smoothing. Increasing the lag interval will increase the degree of smoothing, and similarly reducing the size of the bandwidth in kernel will reduce the degree of smoothing.

Figure 5.10 and Figure 5.11 show the raw, smoothed by SIMA and smoothed by kernel covariance surfaces for the elevation and shares data, respectively.

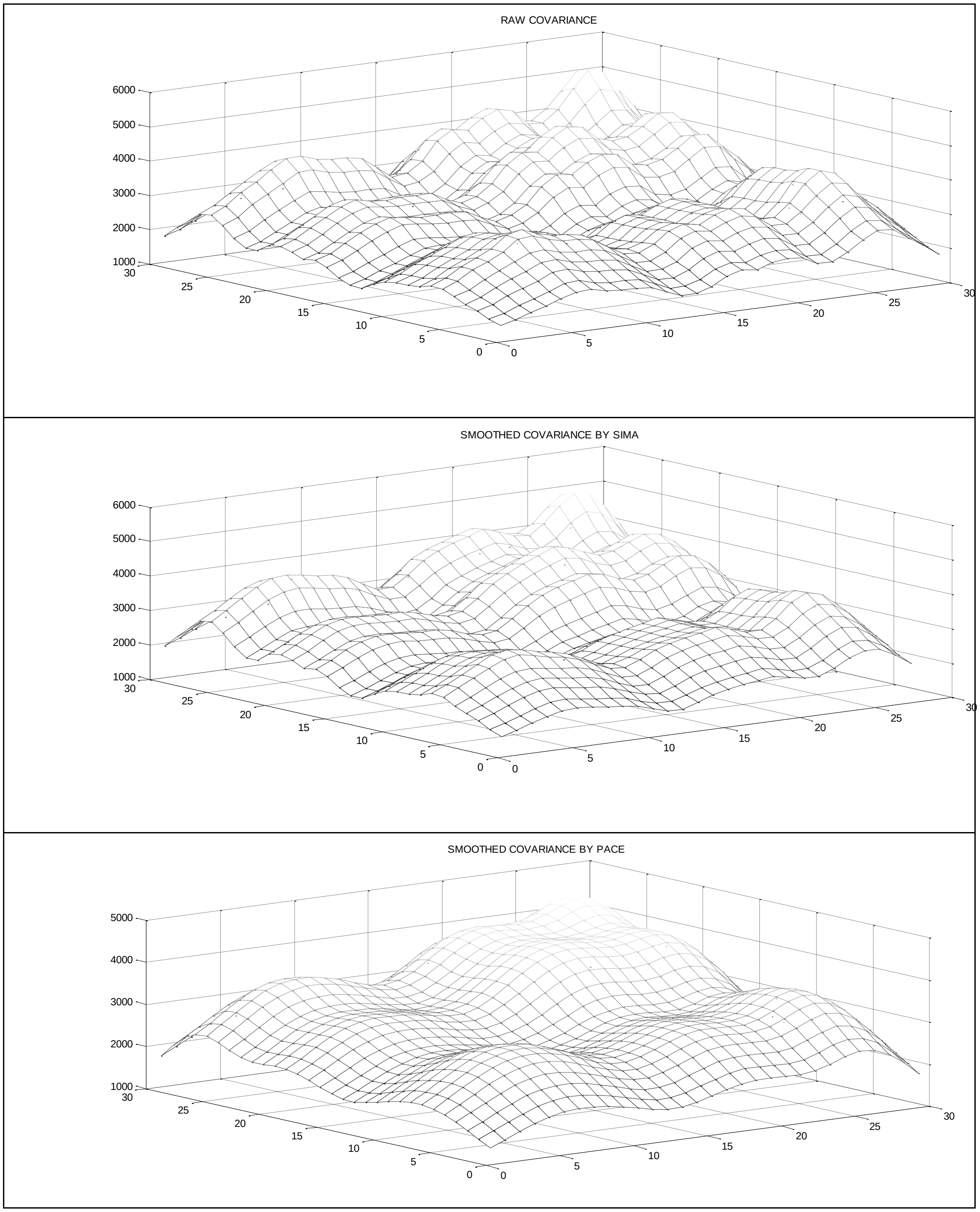

Figure 5.10: Raw Covariance Surface and Covariance Surfaces Smoothed by SIMA, and by Epanechnikov Kernel Methods for the Elevation Data.

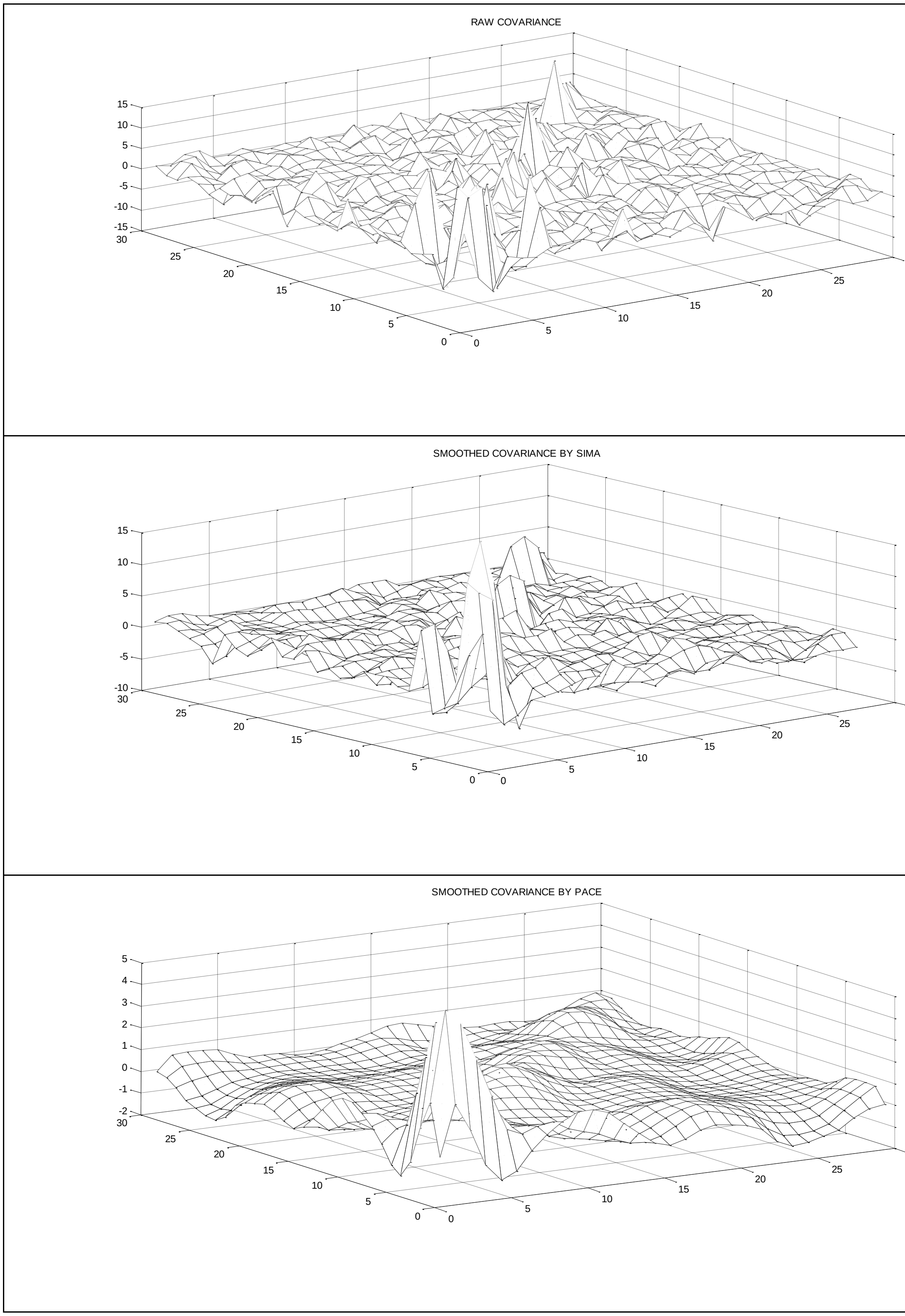

Figure 5.11: Raw Covariance Surface and Covariance Surfaces Smoothed by SIMA, and by Epanechnikov Kernel Methods for the Shares Data.

5.3 Robustness of SIMA Method

An estimator is said to be robust if it is insensitive to changes in the underlying distribution and resistant against the presence of outliers.

A good robust estimator as given in Abu-Shawiesh (2008) should have,

- high efficiency, meaning minimum variance.
- high breakdown point, which is a measure of the maximum fraction of outliers.
- redescending influence function measuring the reaction of an estimator to a small fraction of outliers.
- low gross error sensitivity which measures the worst influence a small amount of contamination of fixed size can have on the value of the estimator.

The median absolute deviation from the sample median (MAD) is considered as one of the good robust estimators since it satisfies these requirements to notable level of significance. Other noteworthy references on robustnes are Lax (1985), and Parr and Schucany (1980).

MAD values are computed using the formula given in Abu-Shawiesh, (2008)

$$MAD = 1.4826 MD\left\{\left|X_i - MD\right|\right\}, \text{ i=1,2,...,n..}$$

Here, MD is the sample median. The factor 1.4826 is only used when samples come from Gaussian distribution, and should not be used when sample comes from non-

Gaussian distributions. The lower control limit (L_{cl}) and upper control limit (U_{cl}) of the S-control chart are given as,

$$
\begin{aligned}
L_{cl} &= c_4 S + 3S\sqrt{1-c_4^{\,2}} \\
U_{cl} &= c_4 S - 3S\sqrt{1-c_4^{\,2}}
\end{aligned}
\tag{5.3.1}
$$

where,

$$c_4 = \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma(n-1/2)} \;,$$

(He and Grigoryan, 2002).

The robustness of the proposed SIMA method of smoothing is checked by using the Shewhart S-control chart. Column-wise standard deviations S_j of SIMA values of the share and elevation data sets are calculated and plotted together with respective control limits from (5.3.1).

Figure 5.12 shows the standard deviation function together with the control limits for the elevation data, while Figure 5.13 shows the same for the shares data. In both cases the standard deviation function falls completely within the control limits indicating the robustness of the SIMA smoothing method.

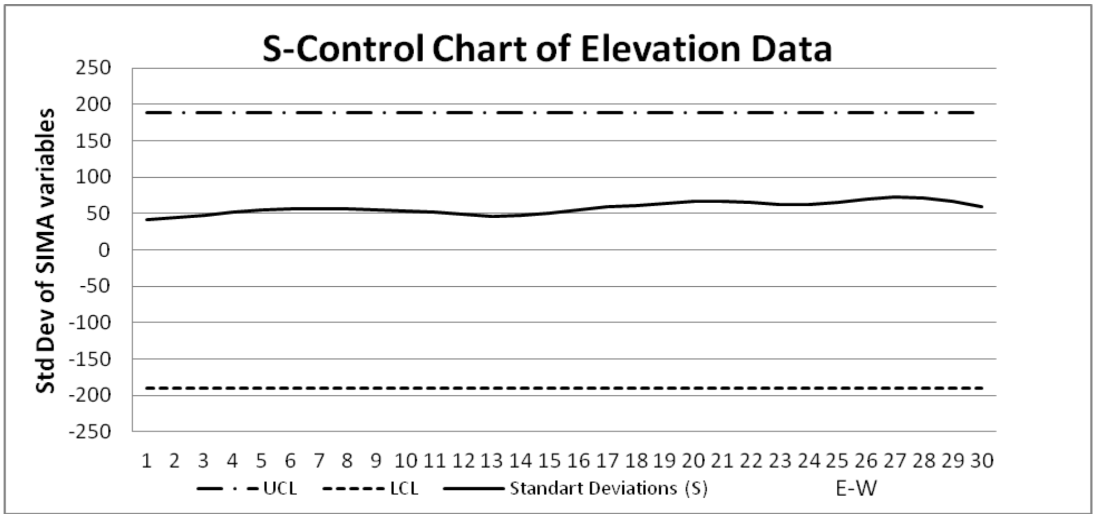

Figure 5.12: Standard Deviation Function for the Elevation Data Totally within the Control Limits Indicating the Robustness of the SIMA Smoothing Method.

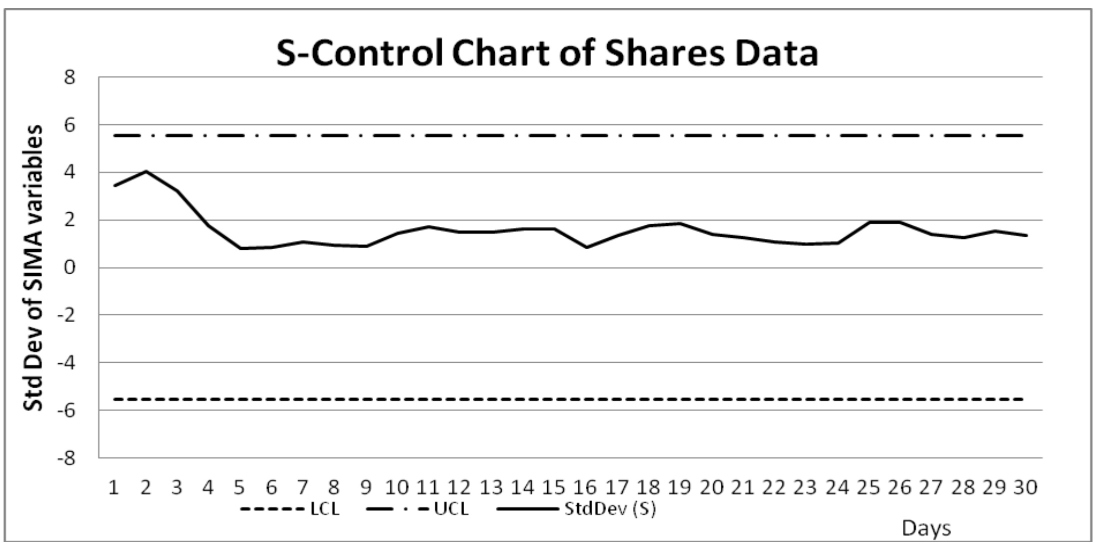

Figure 5.13: Standard Deviation Function for the Shares Data Set Completely within the Control Limits Indicating the Robustness of the SIMA Smoothing Method.

Chapter 6

CONCLUSION AND FURTHER STUDY

6.1 Conclusion

Application of the proposed SIMA smoothing method is applied to two distinct data sets together with kernel (PACE) and LLR smoothers. Obtained results from SIMA are compared with those from PACE and LLR.

SIMA performed better than LLR under the condition of weak correlation between variables involved. In the application to the data sets for the smoothing of mean and covariance, SIMA performed equally well with PACE method which uses Epanechnikov kernel smoother.

The measures used in comparing different smoothers are Mean Square Error (MSE) and Root Mean Square Deviation (RMSD) between an observed trajectory and its smoothed estimate.

SIMA smoothing method is checked for robustness using the standard deviation function control limits method (Shewhart S-control chart) and is found to be a robust smoother.

6.2 Further Research

Based on the research that led to the preparation of this thesis, the following topics are identified as possible areas of further research.

1. Minimizing the error for the proposed SIMA smoothing method by using areas between a trajectory and its estimated SIMA trajectory.
2. The *directional moving average smoothing* with SIMA concept is another area for further investigation. One immediate application area can be the smoothing of the covariance surface given in equation (4.4.2).

REFERENCES

Abu- Shawiesh, A. M. O. (2008). A Simple Robust Control Chart Based on MAD. *Journal of Mathematics and Statistics*, Vol. 4, No 2, 102-107.

Breiman, L. and Friedman. J. H., (1985). Estimating Optimal Transformations for Multiple Regression and Correlation, *Journal of American Statistical Association*, Vol. 80, No 391, 580-598.

Choi, E., & Hall, P. (1998). On Bias Reduction in Local Linear Smoothing. *Biometrika*, Vol. 85, 333-345.

Davis, R., & Resnick, S., (1986). Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages. *The Annals of Statistics*, Vol. 14, No 2, 533-558.

Durbin, J. (1959). Efficient Estimation of Parameters in Moving Average Models. *Biometrica*, Vol. 46, No ¾, 306-316.

Fan, J., & Gijbels, I. (1996). Local Polynomial Modelling and its Applications, Chapman & Hall, London.

Fang, Y., Müller, H. G., & Wang, J. L. (2005). Functional Linear Regression Analysis for Longitudinal Data. *The Annals of Statistics*, Vol.33, No 6, 2873-2903.

Furrer, R., Genton M. G., & Nychka, D. (2006). Covariance Tapering for Interpolation of Large Spatial Datasets. *Journal of Computational and Graphical Statistics*, Vol. 15, No 3, 502-523.

Hall, P., Müller, H. G., & Wang, J. L. (2006). Properties of Principal Component Methods for Functional and Longitudinal Data Analysis, *The Annals of Statistics*, Vol. 34, 1493-1517.

Härdle, W. (1992). Applied Nonparametric Regression. Cambridge University Press.

Härdle, W.,& Simar, L. (2003). Applied Multivariate Statistical Analysis, Berlin and Louvain-la-Neuve.

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components, *J. Edu. Psychol.* Vol. 24, 417-441, 498- 520.

Heckman, N. E. (1986). Spline Smoothing in a Partly Linear Model, *J. R. Statist. Soc. B.* Vol. 48, No 2, 244-248.

He, D., & Grigoryan, A. (2002). Construction of Double Sampling S-Control Charts for Agile Manufacturing, Quality and Reliability Engineering International, Vol. 18, No 4, 343-355.

Hoover, D. R, Rice, J. A., Wu, C. O. & Yang L.P. (1998). Nonparametric Smoothing Estimates of Time Varying Coefficient Models with Longitudinal Data. *Biometrica*, Vol. 85, No 4, 809 – 822.

Karhunen, K. (1946). Zur Spektraltheorie Stochastischer Prozess. *Ann. Acad. Sci. Fennicae*, A I, 37.

Kenny, P. B. & Durbin, J. (1982). Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series. *J. R. Statist. Soc. A.* Vol. 145, 1 – 45.

Lax, D. A., (1985). Robust Estimators of Scale: Sample Performance in Long – Tailed Symmetric Distributions. *Journal of American Statistical Association*, Vol. 80, No. 391, 736 – 741.

Loader, C. (1999). Local Regression and Likelihood. Springer, New York.

Mardia, K. V., Kent. T., & Bibby. J. M. (1979). Multivariate Analysis, London Academic Press.

Mentz, R. P. (1975). Asymptotic Properties of Some Estimators in Moving Average Models. *Technical Report* No. 21, Department of Statistics, Stanford University, California, 6 – 12.

Müller, H. G. (2005). Functional Modeling and Classification of Longitudinal Data. *Scandinavian Journal of Statistics*. Vol. 32, 223 – 240.

Müller, H. G., & Prewitt, K. (1993). Multiparameter Bandwidth Processes and Adaptive Surface Smoothing. *Journal of Multivariate Analysis*. Vol. 47, 1 – 21.

Park, H. M. (2008). Univariate Analysis and Normality Test Using SAS, Stata, and SPSS, *working paper*. The UITS centre for Statistical and Mathematical Computing, Indiana University.

Parr, W. C., & Schucany, W. R. (1980), Minimum Distance and Robust Estimation, *Journal of the American Statistical Association*, Vol. 75, No. 371, 616 – 624.

Rice, J., & Silverman B. (1991). Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves, *J. R. Statist. Soc. Ser. B* Vol. 53, 233 – 243.

Ramsay, J. O., & Silverman B. W. (2002). Applied Functional Data Analysis; Methods and Case Studies. Springer Series in Statistics, Springer - Verlag, New York.

Ramsay, J. O., & Silverman, B. W. (2006). Functional Data Analysis. Springer, New York, United States.

Tandoğdu, Y., & İyikal, Ö. Ç. (2013). Stretched Interpolated Moving Average Technique as a Smoother. *Pakistan Journal of Statistics*. (Accepted for publishing in 2013).

Yao, F. , Müller, H. G., Andrew, J. C., Steven, R. D., Jennifer, F.., Yumei, L., Bruce, A. B., & John S. V. (2003). Shrinkage Estimation for Functional Principal Component Scores with Application to Population Kynetics of Plasma Folate. *Biometrics*,Vol. 59, 676 – 685.

Yao, F. , Müller, H. G. & Wang, J. L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of American Statistical Associati*on. Vol. 100, 577 – 590.

APPENDICES

APPENDIX A: Projecting the Data

It is known that the angle θ between vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ is defined by

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}. \tag{A.1}$$

If $\mathbf{x}^T\mathbf{y} = 0$, then $\theta = \pi/2$, from (A.1). In a right angled triangle, the cosine of θ is equal to the length of the base $\|\mathbf{p}_x\|$, over, the length of the hypotenuse $\|\mathbf{x}\|$. Thus,

$$|\cos\theta| = \frac{\|\mathbf{p}_x\|}{\|\mathbf{x}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|},$$

can be written. Hence

$$\|\mathbf{p}_x\| = \|\mathbf{x}\| |\cos\theta| = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}. \tag{A.2}$$

Here, $\mathbf{p}_x$ is the projection (coordinate) of $\mathbf{x}$ on $\mathbf{y}$. Therefore, the coordinate of $\mathbf{x}_i$ on L_1 is the projection point p_{x_i},

$$p_{x_i} = \mathbf{x}_i^T \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \mathbf{x}_i^T\mathbf{u}_1. \tag{A.3}$$

In "least-square" meaning; the best line L_1 is defined by finding a unit vector $\mathbf{u}_1 \in \mathbb{R}^p$ which minimizes

$$\sum_{i=1}^{n} \left\| \mathbf{x}_i - p_{x_i} \right\|^2 \tag{A.4}$$

or, maximizes $\sum_{i=1}^{n} \left\| p_{x_i} \right\|^2$ under the constraint $\|\mathbf{u}_1\| = 1$. From (A.3) the following can be written

$$\begin{pmatrix} p_{x_1} \\ p_{x_2} \\ \vdots \\ p_{x_n} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \mathbf{u}_1 \\ \mathbf{x}_2^T \mathbf{u}_1 \\ \vdots \\ \mathbf{x}_n^T \mathbf{u}_1 \end{pmatrix} = \mathbf{X}\mathbf{u}_1 \in \mathbb{R}^n .$$

The following theorem will be used to reformulate the maximization problem (Härdle and Simar, 2003).

Theorem A.1 : If $\mathbf{A}$ and $\mathbf{B}$ are symmetric and $\mathbf{B} > 0$, then maximum of $\dfrac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}}$ is given by the largest eigenvalue of $\mathbf{B}^{-1}\mathbf{A}$. In general terms this can be expressed as,

$$\max_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \min_{\mathbf{x}} \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}},$$

where, λ_i's are the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$. If $\mathbf{x}^T \mathbf{B}\mathbf{x} = 1$ we obtain

$$\max_{\{\mathbf{x}:\mathbf{x}^T\mathbf{B}\mathbf{x}=1\}} \mathbf{x}^T \mathbf{A}\mathbf{x} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \min_{\{\mathbf{x}:\mathbf{x}^T\mathbf{B}\mathbf{x}=1\}} \mathbf{x}^T \mathbf{A}\mathbf{x},$$

Proof: By definition $\mathbf{B}^{1/2} = \mathbf{\Gamma_B}\mathbf{\Lambda_B}^{1/2}\mathbf{\Gamma_B}^T$ is symmetric. Then $\mathbf{x}^T\mathbf{B}\mathbf{x} = \|\mathbf{x}^T\mathbf{B}^{1/2}\|^2 = \|\mathbf{B}^{1/2}\mathbf{x}\|^2$.

Let $\mathbf{y} = \frac{\mathbf{B}^{1/2}\mathbf{x}}{\|\mathbf{B}^{1/2}\mathbf{x}\|^2}$, then

$$\max_{\mathbf{x}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = \max_{\{\mathbf{y}:\mathbf{y}^T\mathbf{y}=1\}} \mathbf{y}^T\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}\mathbf{y} \tag{A.5}$$

From spectral decomposition theorem for symmetric square matrices

$\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$. Letting $\mathbf{z} = \mathbf{\Gamma}^T\mathbf{y}$ we get $\mathbf{z}^T\mathbf{z} = \mathbf{y}^T\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{y} = \mathbf{y}^T\mathbf{y}$.

Thus (A.5) can be written as

$$\max_{\{\mathbf{z}:\mathbf{z}^T\mathbf{z}=1\}} \mathbf{z}^T\mathbf{\Lambda}\mathbf{z} = \max_{\{\mathbf{z}:\mathbf{z}^T\mathbf{z}=1\}} \sum_{i=1}^{p} \lambda_i \mathbf{z}_i^2 \text{ and } \max_{\mathbf{z}} \sum \lambda_i \mathbf{z}_i^2 \le \lambda_1 \underbrace{\max_{\mathbf{z}} \sum_{i=1}^{p} \mathbf{z}_i^2}_{=1} = \lambda_1$$

When $\mathbf{z} = (1, 0, \ldots, 0)^T$ gives the maximum. That is $\mathbf{y} = \boldsymbol{\gamma}_1$ the first column of $\mathbf{\Gamma}$, hence $\mathbf{x} = \mathbf{B}^{-1/2}\boldsymbol{\lambda}_1$.

As $\mathbf{B}^{-1}\mathbf{A}$ and $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$ have the same eigenvalues, proof is complete.■

In Theorem A.1, substituting $\mathbf{A} = \mathbf{X}^T\mathbf{X}$ and $\mathbf{B} = \mathbf{I}$, we have:

If $\mathbf{X}^T\mathbf{X}$ and $\mathbf{I} > 0$ are symmetric, then the maximum of $\mathbf{u}_1^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1$ under the constraint $\mathbf{u}_1^T\mathbf{u}_1 = 1$ is given by the largest eigenvalue of $\mathbf{X}^T\mathbf{X}$.

$$\max_{\{\mathbf{u}_1:\mathbf{u}_1^T\mathbf{u}_1=1\}} \mathbf{u}_1^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \min_{\{\mathbf{u}_1:\mathbf{u}_1^T\mathbf{u}_1=1\}} \mathbf{u}_1^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1 \qquad \text{(A.6)}$$

Thus, by (A.6) the problem reformulated as: Find $\mathbf{u}_1 \in \mathbb{R}^p$, with $\|\mathbf{u}_1\| = 1$ that maximizes the quadratic form $(\mathbf{X}\mathbf{u}_1)^T(\mathbf{X}\mathbf{u}_1)$ or $\max_{\{\mathbf{u}_1:\mathbf{u}_1^T\mathbf{u}_1=1\}} \mathbf{u}_1^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1$. The unit vector $\mathbf{u}_1$ can be found by the following theorem.

Theorem A.2: The vector $\mathbf{u}_1$ is the eigenvector of $\mathbf{X}^T\mathbf{X}$ associated with the largest eigenvalue λ_1 of $\mathbf{X}^T\mathbf{X}$.

Proof: If eigenvector equation is used for $\mathbf{X}^T\mathbf{X}$, we have $(\mathbf{X}^T\mathbf{X})\mathbf{k} = (\mathbf{u}_1{}^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1)\mathbf{k}$, where $\mathbf{k}$ is the eigenvector of $\mathbf{X}^T\mathbf{X}$. The eigenvector equation can be rewritten as follows

$$(\mathbf{X}^T\mathbf{X} - \mathbf{u}_1^T\mathbf{X}^T\mathbf{X}\mathbf{u}_1)\mathbf{k} = \mathbf{X}^T\mathbf{X}(k - \mathbf{u}_1^T\mathbf{k}\mathbf{u}_1) = 0. \qquad \text{(A.7)}$$

From (A.7) it is clear that $\mathbf{k} = \mathbf{u}_1$. ■

APPENDIX B: Some Proofs

Proof of Theorem 4.3.1: Let $\mathbf{J}$ be any ($p-m+1$)-vector and $Y_i(t)=\frac{1}{m}\mathbf{A}_m X_i(t)$. Then, $\mathbf{J}^T Y_i(t)=\boldsymbol{\theta}^T X_i(t)$, where $\boldsymbol{\theta}=\frac{1}{m}\mathbf{A}_m{}^T\mathbf{J}$ and the $(p-m+1)\times p$ matrix $\mathbf{A}_m$ has the general form.

$$\mathbf{A}_m=\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & & \ddots & & \vdots \\ \vdots & & & \ddots & & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Let a_{ij} be the elements of $\mathbf{A}_m$. Then

$$a_{ij}=\begin{cases} 1, & for, i\le j\le i+m-1 \\ 0, & for, elsewhere \end{cases}.$$

Since $X_i(t)$ is p-variate normal, then $\boldsymbol{\theta}^T X_i(t)$ is univariate normal. Thus $\mathbf{J}^T Y_i(t)$ is also univariate normal for all fixed vectors $\mathbf{J}$, and $Y_i(t)$ is multivariate normal.■

Proof of Theorem 4.3.2: Under the assumptions on p_i and t,

$$E[Y_{il}^*(t)-X_{il}(t)]^2=E\left[\frac{1}{m}\left(\sum_{j=l}^{m+l-1}X_{ij}(t)-\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right)w_j+\frac{1}{m}\sum_{j=l-1}^{m+l-2}X_{ij}(t)-X_{il}(t)\right]^2$$

can be rewritten as,

$$E[Y_{il}^*(t)-X_{il}(t)]^2=\frac{w_j^2}{m^2}E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)-\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]^2+\frac{2w_j}{m^2}E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)\sum_{j=l-1}^{m+l-2}X_{ij}(t)-\left(\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right)^2\right]$$
$$-\frac{2w_j}{m}E\left[X_{il}(t)\sum_{j=l}^{m+l-1}X_{ij}(t)-X_{il}\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]-\frac{2}{m}E\left[X_{il}\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]+\frac{1}{m^2}E\left[\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]^2+E[X_{il}]^2$$

$$=\frac{w_j^2}{m^2}E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)\right]^2+\frac{2w_j(1-w_j)}{m^2}E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]+\frac{w_j(w_j-2)}{m^2}E\left[\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]^2$$
$$-\frac{2}{m}E\left[X_{il}\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]+\frac{1}{m^2}E\left[\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]^2+E[X_{il}]^2. \qquad \text{(B.1)}$$

Under the i.i.d assumption on X_{ij}, and $E(X_{ij})=0$ and $E(X_{ij})^2=\sigma^2$, then

$$E\left[\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]^2=E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)\right]^2=E\left[\sum_{j=l}^{m+l-1}X_{ij}(t)\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right]=m\sigma^2.$$

Therefore, (B.1) can be written as,

$$\frac{1}{m^2}E\left[\sum_{j=l-1}^{m+l-1}X_{ij}(t)\right]^2+E\left[X_{il}(t)\right]^2-\frac{2}{m}E\left[X_{il}(t)\sum_{j=l-1}^{m+l-2}X_{ij}(t)\right].$$

As $t\to s_j$,

$$E[Y_{il}^*(t)-X_{il}(t)]^2\to\frac{m-1}{m}\sigma^2.$$

Moreover, for $p_i\to\infty$ and $t\to s_j$, with very large m,

$$E[Y_{il}^*(t)-X_{il}(t)]^2\to\sigma^2. \blacksquare$$

APPENDIX C: Karhunen–Loève Theorem

The Karhunen–Loève theorem is an infinite linear combination of orthogonal functions representation of a stochastic process, as a Fourier series expansion of a function on a bounded interval. The theorem yields the best basis that minimizes the total mean squared error.

The coefficients in the Karhunen–Loève theorem are random variables and the expansion basis depends on the process. Hence, the orthogonal basis functions are determined by the covariance function of the process.

In the case of a centered stochastic process $\{X_t\}_{t\in I}$, i.e. $E(X_t)=0$ for all $t \in I$, where I is a closed and bounded time or space interval, can be decomposed as

$$X_t = \sum_{k=1}^{\infty} \xi_k \phi_k(t) \tag{C.1}$$

where, ξ_k are pairwise uncorrelated random variables and the functions ϕ_k are continuous real-valued functions on I that are pairwise orthogonal in $L^2(I)$. (C.1) is called Karhunen–Loève expansion or Karhunen–Loève decomposition. If the process is not centered, then in place of X_t, $X_t - E(X_t)$ is used.

The empirical version of the Karhunen–Loève theorem is called as the Karhunen–Loève transform, principal component analysis, proper orthogonal decomposition, Empirical orthogonal functions or the Hotelling transform.

A linear operator $T_{G_X} : L^2(I) \to L^2(I)$ defined for covariance function G_k, as follows,

$$f(t) \to \int_I G_X(s,t) f(s) ds .$$

Since T_{G_X} is a linear operator, its eigenvalues λ_k and eigenfunctions ϕ_k can be found by solving the following equation.

$$\int_I G_X(s,t)\phi(s) ds = \lambda_k \phi_k(t) \tag{C.2}$$

Mercer's Theorem: Let X_t be a zero-mean square integrable stochastic process over $t \in I$, with continuous covariance function $G_X(s,t)$. Then, $G_X(s,t)$ is a *Mercer kernel*. Let ϕ_k be an orthonormal basis of $L^2(I)$ formed by the eigenfunctions of T_{G_X} with respect to eigenvalues λ_k, X_t has the representation $X_t = \sum_{k=1}^{\infty} \xi_k \phi_k(t)$ where the convergence is in L^2, uniform in t, and $\xi_k = \int_I X_t \phi_k(t) dt$. Further, ξ_k are uncorrelated with $E(\xi_k) = 0$ and $\operatorname{var}(\xi_k) = \lambda_k$.

Hence, Mercer's theorem says that, there exists a set of eigenvalues and eigenfunctions of T_{G_X} from (C.2), forming an orthonormal basis of $L^2(I)$, such as,

$$G_X(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t) .$$

More specifically, given any orthonormal basis $\{\varphi_k\}$ of $L^2(I)$, the process X_t may decomposed as $X_t(s) = \sum_{k=1}^{\infty} \zeta_k(s) \varphi_k(t)$ where $\zeta_k(s) = \int_I X_t(s) \varphi_k(t) dt$ and may approximate by the finite sum

$$\hat{X}_t(s) = \sum_{k=1}^{K} \zeta_k(s) \varphi_k(t) . \tag{C.3}$$

The integer K can be found by using the function $F(K) = \sum_{k=1}^{K} \lambda_k / \sum_{k=1}^{\infty} \lambda_k$. Claim of all such approximations (C.3) is that; the Karhunen–Loève approximation is the one that minimizes the total mean square error provided that eigenvalues are arranged in decreasing order.

APPENDIX D: Data Sets.

Table 1. Data values of the shares in Istanbul Stock Exchange.

	04/08/09	05/08/09	06/08/09	07/08/09	11/08/09	12/08/09	13/08/09	18/08/09	19/08/09	20/08/09	21/08/09	25/08/09	01/09/09	02/09/09	03/09/09
ACIBADE	0.00	-0.67	0.67	-0.67	2.67	-1.30	-1.32	-1.32	0.00	0.67	0.00	-0.66	3.27	3.80	-3.05
ARCELIK	12.74	-1.13	-6.29	3.05	-0.58	-3.53	3.05	-0.59	1.78	5.23	4.42	12.12	-1.83	0.00	-1.86
AVIVA SI	1.05	0.52	0.00	-0.52	0.52	0.51	-1.02	-2.02	1.55	-0.51	1.02	1.00	-2.01	1.03	-1.52
BIM	0.86	1.74	0.00	-0.85	0.85	-2.52	-0.86	-0.85	-0.86	0.87	-0.89	-1.80	0.00	0.00	-4.46
CREDITWI	0.00	-0.82	0.83	0.00	-1.59	1.21	1.60	-1.23	0.41	0.41	10.29	0.77	0.53	12.04	-1.87
DYO	2.13	0.00	0.00	0.00	0.00	-2.08	0.00	2.17	0.00	2.13	-2.08	-2.08	-2.13	-2.17	-4.44
ECZACIB	2.10	0.68	-1.36	0.69	1.37	-0.68	0.00	-1.36	0.00	3.45	0.67	0.64	-1.94	1.97	-2.58
EGE SER	0.00	17.28	-6.32	-2.25	1.11	-1.10	-3.33	0.00	0.00	1.11	0.00	0.00	3.33	0.00	-4.30
GENTAS	2.78	1.80	0.00	0.00	0.00	-1.77	-0.90	-2.80	-0.96	0.00	0.97	1.90	-0.93	3.74	0.00
HURRIYE	2.52	9.84	3.22	-4.51	-0.76	-3.82	0.00	1.55	2.29	2.24	-0.73	7.41	-5.42	1.91	-5.00
IZOCAM	0.00	0.00	17.86	15.66	1.00	-4.08	3.72	1.09	0.54	0.53	2.66	0.00	-1.02	-0.52	-3.11
KELEBEK	1.15	1.14	-1.12	0.00	1.55	-1.11	-1.12	-2.22	2.27	1.11	4.40	0.00	-2.20	1.12	-1.11
KOC	5.20	0.55	-4.92	-0.57	-1.13	-2.29	2.34	0.57	1.69	0.00	2.78	3.70	-1.00	0.50	-2.50
LUKS	1.92	-0.47	-1.42	3.85	0.00	-1.83	0.47	0.95	0.94	-0.93	2.82	0.46	-1.90	0.49	-4.34
MIGROS	0.00	22.12	20.93	-1.92	-3.06	-4.74	13.26	-3.70	-0.96	0.00	3.88	0.00	-1.44	-0.49	-2.93
MUTLU	1.40	0.00	0.69	0.68	5.41	0.64	-0.64	0.64	-0.63	0.00	1.27	-0.63	3.61	0.00	-5.23
OTOKAR	1.68	-0.83	-1.67	0.00	1.63	-1.60	-2.44	0.87	1.72	0.00	3.39	3.20	-1.57	1.60	-3.94
PINAR SU	1.22	-0.40	-0.81	0.00	0.00	0.82	-0.40	0.21	0.00	0.00	0.41	0.41	1.96	-1.92	-4.31
TUBORG	0.00	-0.93	-1.42	0.00	-0.48	-0.96	0.97	-2.87	1.48	-0.97	0.98	2.44	-2.27	-2.33	-0.48
THY	0.86	2.56	7.50	3.88	-0.74	-3.73	2.33	0.00	2.26	2.94	0.00	3.52	-0.69	0.69	-1.38
Days	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RawAver	1.88	2.65	1.32	0.83	0.39	-1.70	0.79	-0.55	0.68	0.91	1.81	1.62	-0.68	1.07	-2.92

	04/09/09	08/09/09	09/09/09	10/09/09	11/09/09	15/09/09	16/09/09	17/09/09	18/09/09	24/09/09	25/09/09	29/09/09	30/09/09	01/10/09	06/01/09
ACIBADE	-1.26	-1.90	-1.94	0.00	-1.32	0.00	-0.66	0.66	0.00	0.00	1.32	1.30	2.56	-2.50	-0.65
ARCELIK	-0.95	0.49	-0.98	3.94	1.90	-0.90	0.00	-1.36	0.00	-1.82	0.46	0.45	-1.78	-2.71	1.90
AVIVA SI	1.03	1.02	-1.01	-3.06	2.11	0.00	-1.53	0.00	0.52	1.03	3.06	0.00	1.94	-3.81	2.03
BIM	1.87	0.91	-0.90	0.00	0.91	-2.56	-0.88	0.00	0.88	1.80	3.54	0.00	4.27	-0.82	4.13
CREDITWI	6.67	-2.60	2.67	-4.33	2.71	-1.33	0.00	3.15	0.44	-0.44	0.00	1.34	1.32	4.78	1.25
DYO	4.65	-2.22	2.27	-4.44	0.00	-4.55	0.00	2.38	0.00	-2.33	2.38	0.00	2.33	-2.27	2.38
ECZACIB	0.00	-0.66	-1.33	2.03	-0.66	2.61	-1.27	3.23	0.63	5.88	1.67	2.81	7.10	2.04	1.52
EGE SER	-1.12	0.00	2.27	-4.65	-1.22	2.41	0.00	0.00	0.00	0.00	1.16	3.53	2.27	-1.11	1.15
GENTAS	1.80	1.82	-1.79	-4.55	1.90	0.00	4.63	0.88	-1.75	0.91	0.00	0.90	0.89	0.00	2.75
HURRIYE	0.00	6.76	-20.89	-8.00	12.17	0.00	-5.30	-0.88	3.23	-0.79	0.80	-0.81	7.38	2.29	-1.31
IZOCAM	0.53	0.53	-1.58	2.14	0.52	0.00	-0.52	0.52	1.04	0.48	0.48	1.46	1.92	-0.47	3.27
KELEBEK	1.12	-2.68	-2.25	-1.15	0.00	1.15	0.00	0.00	-1.04	0.00	0.00	4.60	2.20	0.00	2.22
KOC	-0.51	-1.57	-2.66	4.36	2.09	-2.00	-3.57	2.12	0.00	1.56	-1.03	0.51	0.51	-1.02	1.04
LUKS	1.01	1.50	0.00	-0.99	-0.50	0.00	2.48	-0.48	-0.49	1.47	0.48	0.00	1.91	-1.88	0.45
MIGROS	-0.51	0.00	3.54	-0.98	0.49	-0.49	-0.98	0.00	0.49	-0.98	0.00	-0.98	0.00	0.00	1.49
MUTLU	0.61	0.58	-2.89	-1.76	0.00	5.75	-1.63	3.13	-1.07	-0.54	-1.08	1.64	0.00	-2.69	0.58
OTOKAR	0.00	0.80	-1.59	-2.42	-0.83	-1.57	2.40	-0.78	0.00	2.36	2.31	7.41	-1.38	-2.10	-1.47
PINAR SU	0.82	-0.40	-0.81	-0.81	0.41	1.62	-0.40	-0.80	0.40	2.02	1.19	0.00	2.00	2.94	0.00
TUBORG	-1.91	1.96	-0.48	-0.48	0.49	0.00	0.48	0.95	0.94	18.49	5.67	-6.57	-0.78	-0.79	0.81
THY	-2.10	0.71	1.42	4.20	5.37	2.42	0.00	5.33	1.69	3.16	-0.51	2.62	-1.02	-0.52	1.06
Days	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
RawAver	0.59	0.25	-1.45	-1.05	1.33	0.13	-0.34	0.90	0.30	1.61	1.10	1.01	1.68	-0.53	1.23

Table 2. Elevation coordinates data.

	536394	536444	536494	536544	536594	536644	536694	536744	536794	536844	536894	536944	536994	537044
3903787	266	266	272	277	267	273	275	273	269	272	273	277	276	287
3903837	263	259	261	257	264	274	276	277	274	277	277	286	287	292
3903887	267	264	264	265	266	275	275	275	279	284	286	293	300	300
3903937	268	268	262	260	258	261	276	277	279	290	300	303	314	316
3903987	273	273	264	263	277	275	285	285	292	297	300	317	326	326
3904037	268	269	260	259	273	283	293	301	294	301	302	316	335	340
3904087	268	271	269	278	279	277	275	289	300	308	312	322	324	319
3904137	264	273	270	277	282	284	289	289	300	308	318	328	342	331
3904187	270	273	270	273	279	287	291	302	300	306	315	334	345	347
3904237	284	280	277	280	283	289	295	304	300	301	317	326	330	332
3904287	267	275	274	273	280	293	286	300	296	304	315	332	355	365
3904337	270	278	275	277	275	286	293	285	296	296	323	334	355	371
3904387	290	285	278	287	295	265	267	290	300	310	317	335	353	372
3904437	285	287	282	293	297	297	292	303	303	307	320	336	355	377
3904487	295	290	295	300	304	307	296	303	313	313	315	327	347	372
3904537	300	296	301	308	317	323	300	305	319	327	331	325	344	366
3904587	300	305	310	314	320	315	305	316	325	335	345	340	345	364
3904637	310	308	312	319	327	311	310	323	335	342	353	359	355	363
3904687	318	315	317	327	335	326	326	327	340	350	360	372	370	370
3904737	328	330	325	337	355	349	339	343	350	364	365	373	377	384
3904787	336	341	335	337	355	366	360	354	360	375	379	378	379	387
3904837	339	350	350	350	365	380	377	370	362	379	388	385	385	395
3904887	345	360	362	365	375	395	395	383	375	380	395	397	396	407
3904937	355	365	373	375	384	400	401	389	387	390	397	407	408	414
3904987	355	360	375	383	387	402	407	403	400	402	406	411	415	420
3905037	348	355	370	390	395	410	415	415	410	412	414	414	421	426
3905087	352	360	369	385	405	415	423	426	424	424	425	425	425	431
3905137	362	365	374	385	401	421	427	435	437	438	437	436	430	433
3905187	367	373	380	394	405	421	432	439	443	450	451	449	445	450
3905237	369	378	390	405	420	430	441	445	448	455	461	457	458	461
raw avrg	306	309	311	316	324	330	331	334	337	343	350	356	363	371

	537094	537144	537194	537244	537294	537344	537394	537444	537494	537544	537594	537644	537694	537744	537794	537844
3903787	280	278	278	286	272	264	264	273	282	298	287	288	279	278	279	301
3903837	290	284	279	289	275	270	269	275	284	298	299	299	282	285	298	290
3903887	296	287	283	290	279	277	277	276	290	295	296	285	279	289	279	280
3903937	309	298	289	292	292	282	285	293	293	299	302	294	296	295	299	289
3903987	324	310	295	292	291	290	301	307	301	300	302	302	303	294	285	295
3904037	335	319	298	302	312	310	315	320	320	310	305	303	299	290	294	295
3904087	337	319	305	315	320	328	329	330	328	317	315	312	311	307	310	311
3904137	330	320	314	322	331	342	351	349	334	320	317	309	305	307	304	301
3904187	344	330	322	328	347	350	355	355	345	345	328	320	313	303	307	314
3904237	340	340	328	334	350	355	360	359	360	355	332	322	318	313	320	327
3904287	367	350	343	350	358	374	385	380	370	363	346	333	323	319	327	332
3904337	363	357	362	365	370	386	380	385	383	360	349	333	333	327	327	345
3904387	373	374	376	382	386	387	387	387	387	375	350	346	347	342	341	345
3904437	378	379	388	397	404	403	403	403	399	383	364	351	355	354	352	355
3904487	385	385	394	402	414	420	424	423	415	395	375	365	369	369	365	365
3904537	387	390	397	407	416	425	434	434	425	400	380	376	381	390	386	373
3904587	382	401	400	412	420	429	434	438	430	405	385	400	400	413	405	380
3904637	378	395	405	408	414	423	426	425	425	411	395	405	410	412	410	389
3904687	377	390	403	404	407	417	417	417	418	407	402	410	420	424	415	399
3904737	399	399	403	402	410	421	427	428	417	422	414	430	439	440	420	410
3904787	405	404	404	407	418	427	433	430	424	429	427	440	450	452	430	410
3904837	412	411	408	415	420	430	433	430	427	435	436	450	460	462	430	407
3904887	422	423	417	423	427	434	439	433	433	442	445	454	463	464	425	409
3904937	430	432	427	429	439	450	450	435	436	447	452	460	467	455	425	410
3904987	434	443	437	435	445	457	456	445	442	455	461	469	474	450	427	412
3905037	435	447	445	444	454	464	461	445	450	462	471	480	480	425	430	414
3905087	437	447	448	454	469	471	464	455	459	470	484	494	475	430	433	420
3905137	439	447	455	459	476	477	475	460	459	470	494	500	471	449	435	427
3905187	443	454	460	465	480	483	484	469	463	474	490	505	469	449	441	425
3905237	459	460	466	471	482	488	492	476	473	479	496	485	470	450	442	431
raw avrg	376	376	374	379	386	391	394	391	389	387	383	384	381	375	368	362