TESTING SEPARABILITY OF FUNCTIONAL DATA

A Dissertation in
Statistics
by
Panayiotis Constantinou

© 2017 Panayiotis Constantinou

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2017
The dissertation of Panayiotis Constantinou was reviewed and approved* by the following:

Matthew Reimherr  
Assistant Professor of Statistics  
Dissertation Adviser  
Chair of Committee

Michael Akritas  
Professor of Statistics

Bing Li  
Professor of Statistics

John Liechty  
Professor of Marketing and Statistics

Aleksandra Slavković  
Professor of Statistics  
Head of Graduate Program

*Signatures are on file in the Graduate School.
Abstract

The assumption of separability is used heavily in spatiotemporal statistics. Separability means that the spatiotemporal covariance structure factors into the product of two functions, one depending only on space and the other only on time. Separability is a property which can dramatically improve computational efficiency by substantially reducing model complexity. It is especially useful for functional data as it implies that the functional principal components are the same for each spatial location.

In Chapter 1, we give a brief introduction to functional data, separability and introduce the data sets that motivate this dissertation.

In Chapter 2, we present a new methodology to test for separability of spatiotemporal functional data. We present three tests, one being a functional extension of the Monte Carlo likelihood method of Mitchell et al. (2005), while the other two are based on quadratic forms. Our tests are based on asymptotic distributions of maximum likelihood estimators, and do not require Monte Carlo or bootstrap replications. The specification of the joint asymptotic distribution of these estimators is the main theoretical contribution in this chapter. It can be used to derive many other tests. The main methodological finding is that one of the quadratic form methods, which we call a norm approach, emerges as a clear winner in terms of finite sample performance in nearly every setting we considered. The norm approach focuses directly on the Frobenius distance between the spatiotemporal covariance function and its separable approximation. We demonstrate the efficacy of our methods via simulations, and applications to Irish wind data and Nitrogen Dioxide levels on the east coast of the United States.
In Chapter 3, we derive and study a significance test for determining if a panel of functional time series is separable. In this context, separability means that the covariance structure factors into the product of two functions, one depending only on time and the other depending only on the coordinates of the panel. In this case, under the assumption of separability, the functional principal components are the same for each member of the panel. However such an assumption must be verified before proceeding with further inference. Our approach is based on functional norm differences and provides a test with well controlled size and high power. In addition to an asymptotic justification, our methodology is validated by a simulation study. It is applied to functional panels of particulate pollution and stock market data.
## Table of Contents

List of Figures vii

List of Tables viii

Chapter 1  
Introduction 1  
1.1 Functional data 1  
1.2 Separability 4

Chapter 2  
Testing separability of space-time functional processes 9  
2.1 Introduction 9  
2.2 Multivariate theory 12  
2.3 Tests for functional data 15  
2.3.1 The test procedure 15  
2.3.2 Large–sample justification 19  
2.4 Simulation study and applications 20  
2.4.1 A simulation study 20  
2.4.2 Irish wind data 25  
2.5 Appendix-A 27  
2.5.1 Tests for a small number of spatial locations 27  
2.5.1.1 Procedure 2: fixed spatial locations, fixed temporal basis 28  
2.5.1.2 Procedure 3: fixed spatial locations, data driven temporal basis 30  
2.5.2 Derivation of the Q matrices 32  
2.5.3 Proof of Theorem 2.3 42
2.5.4 Proof of Theorem 2.4 ........................................ 48
2.5.5 Additional empirical rejection rates, more discussion in Irish wind data and application to pollution data ............... 58
  2.5.5.1 Additional simulations .................................. 58
  2.5.5.2 Irish wind data .......................................... 60
  2.5.5.3 Air quality data ......................................... 62

Chapter 3
Testing separability of functional time series 64
  3.1 Introduction ................................................. 64
  3.2 Assumptions and problem formulation ........................ 65
  3.3 Derivation of the test and its asymptotic justification ......... 68
  3.4 Details of implementation ................................... 75
  3.5 Simulation study and applications ............................ 79
    3.5.1 A simulation study ....................................... 79
      3.5.1.1 Case 1: dimension reduction in time only .......... 81
      3.5.1.2 Case 2: dimension reduction in both time and panel coordinates ........................................... 82
    3.5.2 Applications to pollution and stock market data .......... 83
  3.6 Appendix-B .................................................. 86
    3.6.1 Proof of Theorem 3.2 .................................... 87
    3.6.2 Proof of Theorem 3.3 .................................... 93
    3.6.3 Details on how $W^*$ and $Q^*$ are actually computed in practice .................................................. 94
    3.6.4 Proof of Theorem 3.6 .................................... 96

Bibliography 102

vi
List of Figures

2.1 QQ-plots of the p-values of the test statistics $T_L$, $T_F$, $T_W$, from left to right. .................................................. 61
2.2 Irish wind speed curves for January 1961. Each curve corresponds to a different location. The curves are offset since they overlap. .................. 61
2.3 Maximum one hour nitrogen dioxide curves for December 2012. The curves are offset since they overlap. ................................. 63
3.1 Maximum one–hour nitrogen dioxide curves for December 2012 at the nine locations. ................................................................. 84
3.2 Cumulative intraday return curves for the ten companies for April 2nd, 2007. ................................................................. 86
List of Tables

2.1 Rejection rates (%), based on 1000 replications, for \( N = 100, L = J = 2, \beta = 0, 0.5, 1 \) and \( c_0 = 1, 5, 10 \). The parameters \( \beta = 0 \) and \( c_0 = 1 \) correspond to the null hypothesis of separability. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses. 22

2.2 Empirical power (%), based on 1000 replications, for \( N = 100, 150 \) and 200, \( L = J = 2, \beta = 1 \) or \( c_0 = 10 \). The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses. 23

2.3 Rejection rates (%), based on 1000 replications, for \( \beta = 0, 1 \) and \( c_0 = 1, 10, N = 100 \), and the varying degrees of spatial and temporal dimension reduction. \( L \) is the number of spatial and \( J \) of temporal principal components. The proportion of variance explained (CPV) from the temporal principal components is given in the last column. The standard errors are given in the parentheses. 24

2.4 Rejection rates (%), based on 1000 replications, for data whose covariance is the sum of a large separable component and a small nonseparable component; \( N = 100 \). The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses. 25

2.5 P-values for the separability tests applied to deseasonalized wind speed data. 26

2.6 Rejection rates (%), based on 1000 replications, for \( N = 100, L, J \in \{2, 3, 4, 5, 6, 7\}, \beta = 0 (H_0) \) and \( \beta = 1 (H_A) \). \( L \) is the number of spatial principal components used in the dimension reduction and \( J \) the number of temporal principal components. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses. 59
2.7 Rejection rates (%), based on 1000 replications, for the method of Section A, temporal dimension reduction only; \( N = 100, J \in \{2, 3, 4\}, \beta = 0 \) and \( \beta = 1 \). \( J \) is the number of temporal principal components. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

3.1 Rejection rates under \( H_0 (c = 0) \) at the nominal 5 percent level; \( J \) the number of temporal PC’s. The explained variance of the temporal FPCA is given in parentheses.

3.2 Empirical power \((c = 1)\); \( J \) the number of temporal PC’s. The explained variance of the temporal FPCA is given in parentheses.

3.3 Rejection rates under \( H_0 (c = 0) \); \( K \) is the reduced panel dimension and \( J \) the number of temporal PC’s. The explained variance of the dimension reduction is given in parentheses.

3.4 Empirical power \((c = 1)\); \( K \) and \( J \) are as in Table 3.3. The explained variance of the dimension reduction is given in parentheses.
Introduction

1.1 Functional data

Functional data analysis, FDA, has expanded rapidly the last years due to the demand of analyzing high dimensional data sets and is introduced in many textbooks such as Ramsay and Silverman (2002), Ramsay and Silverman (2005), Ferraty (2011), Ferraty and Romain (2011), Horváth and Kokoszka (2012), Hsing and Eubank (2015), Srivastava and Klassen (2016), Kokoszka and Reimherr (2017).

Functional data are data that can be viewed as smooth curves or functions and they have the following form:

$$X_n(t), \quad t \in [T_1, T_2], \quad n = 1, 2, \ldots, N,$$

for which the values $X_n(t)$ exist at any point $t$, but may be observed only at selected points $t_{n,j}$. For example, financial records can be naturally treated as functional outcomes since they are observed every minute or more. In that case, $X_n(t)$ can be the stock price at minute $t$ of the $n$th trading day. Another example, is the average
temperature on day $t$ of the $n$th month or the maximum nitrogen dioxide levels of day $t$ of the $n$th year. Functional data are objects of infinite dimension and are the next step of multivariate data analysis. Existing methods used in traditional statistics such as regression, analysis of variance, classification and clustering can be extended to functional inputs, making the field of functional data analysis very broad for research. There are three categories of functional regression models, depending on whether the responses or the predictors or both are curves. Most of the books mentioned above introduce all three functional regression models. In addition, the reader can refer to the following articles; Faraway (1997), Carey et al. (2002), Cuevas et al. (2002), James (2002), Chiou et al. (2004), Ferraty and Vieu (2004), M"uller and Stadtma"uller (2005) as well as Chiou and M"uller (2007) for diagnostics for functional regression. Analysis of variance for functional data is introduced in the textbooks; Zoglat (1994), Zhang (2013) as well as in the articles; Cuevas et al. (2004) and Zhang and Liang (2014). Finally, for more information in classification and clustering of functional data, the reader is referring to the following articles; James and Hastie (2001), Biau et al. (2005), M"uller (2005), Alonso et al. (2012), Delaigle and Hall (2012), Delaigle et al. (2012) among many others.

Typically the first step when working with functional data is to express them by means of a basis expansion:

$$X_n(t) \approx \sum_{m=1}^{M} c_{nm} B_m(t) \quad n = 1, 2, ..., N,$$

(1.1)

where $B_m$ are standard basis functions, like splines or sine/cosine functions and $c_{nm}$ are coefficients. This is done mainly for computational convenience. This step can be easily implemented in R. For more details, on how to implement this, the reader is referring to Ramsay et al. (2009).
For functional data the sample mean, the sample standard deviation and the sample covariance function are defined as follows:

\[
\hat{\mu}(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t)
\]

\[
\hat{\sigma}(t) = \left[ \frac{1}{N} \sum_{n=1}^{N} (X_n(t) - \hat{\mu}(t))^2 \right]^{1/2}
\]

\[
\hat{c}(t, t') = \frac{1}{N} \sum_{n=1}^{N} (X_n(t) - \hat{\mu}(t))(X_n(t') - \hat{\mu}(t'))
\]

From the above estimators, the sample covariance function \( \hat{c}(t, t') \) plays an important role since it is used to estimate functional principal components (EFPC's), one of the most useful tools of functional data analysis. Functional principal components allow us to reduce the infinite dimension of functional data to a small finite dimension in an optimal way. The idea of the functional principal component expansion is to find functions \( \hat{v}_j(t) \) such that the centered functions \( X_n(t) - \hat{\mu}(t) \) are represented as:

\[
X_n(t) - \hat{\mu}(t) \approx \sum_{j=1}^{J} \xi_{nj} \hat{v}_j(t) \quad n = 1, 2, ..., N,
\]

with \( J \) much smaller than \( M \) in (1.1). The above decomposition is called the Karhunen-Loève expansion. The coefficient \( \xi_{nj} \) is called the score of \( X_n(t) \) with respect to \( \hat{v}_j(t) \) and is defined by \( \xi_{nj} = \langle X_n(t), \hat{v}_j(t) \rangle \). We define the estimated functional principal components (EFPC’s) \( \hat{v}_j(t) \) by:

\[
\int \hat{c}(t, t') \hat{v}_j(t') dt' = \hat{\lambda}_j \hat{v}_j(t),
\]

where \( \hat{\lambda}_j \) is the corresponding eigenvalue of the eigenfunction \( \hat{v}_j(t) \). The estimated
functional principal components (EFPC’s) $\hat{v}_j(t)$ are orthonormal, that is:

$$\int \hat{v}_j(t)\hat{v}_i(t)dt = \begin{cases} 0 & \text{if } j \neq i \\ 1 & \text{if } j = 1 \end{cases}.$$ 

The R function `pca.fd` computes the estimated functional principal components (EFPC’s) $\hat{v}_j(t)$, the corresponding eigenfunctions $\hat{\lambda}_j$ and the scores $\xi_{nj}$. For more details, on the R function `pca.fd`, the reader is referring to Ramsay et al. (2009).

Theory of operators in a Hilbert space and properties of random samples in the space $L^2$ of square integrable functions are heavily used in functional data analysis. The reader may refer to Rudin (1987), Riesz and Sz.-Nagy (1990), Akhiezier and Glazman (1993), Debnath and Mikusinski (2005), Gohberg et al. (1990) for more details on Hilbert spaces. Recall that the space $L^2 = L^2([0,1])$ is the set of measurable real-valued functions $x$ defined on $[0,1]$ satisfying $\int_0^1 x^2(t)dt < \infty$ and is a separable Hilbert space with the inner product:

$$\langle x, y \rangle = \int x(t)y(t)dt.$$ 

**1.2 Separability**

The assumption of separability is used heavily in spatiotemporal statistics and is introduced in many textbooks such as Schabenberger and Gotway (2004), Sherman (2011) and Cressie and Wikle (2015). Separability means that the spatiotemporal covariance structure is the product of two functions, one depending only on space, the other only on time. This provides a number of benefits. From a modeling perspective, it allows one to draw on the large literature on covariance structures for spatial or temporal data. The simpler structure induced by separability is
then much easier to estimate than a nonseparable structure. Separability has
been extensively used in spatiotemporal statistics because it leads to theoretically
tractable models and computationally feasible procedures, some recent references
are Hoff (2011), Paul and Peng (2011), Sun et al. (2012). Before separability
is assumed, it must be tested. If separability is incorrectly assumed, it leads to
serious biases and misleading conclusions. Tests of separability for spatiotemporal
covariances of scalar fields are reviewed in Mitchell et al. (2005, 2006) and Fuentes
(2006). If the spatiotemporal covariance has a specific parametric form, a likelihood
test can be derived. Tests based on composite likelihood can also be used
(Bevilacqua et al. 2010). A similar parametric approach, in conjunction with
bootstrap, is taken by Liu et al. (2016) in the context of functional data. Lu and
Zimmerman (2005) and Mitchell et al. (2006) introduce a nonparametric likelihood
test, which requires that the number of repeated measurements be greater
than the product of the number of spatial and temporal locations. Mitchell et al.
(2005) explain how to deal with this restriction by dividing the temporal domain
into blocks. The test of Fuentes (2006) is based on the spectral representation,
which assumes that the data are available on a spatial grid.

By using statistical notation, assume we observe \( N \) independent and identically
distributed scalar fields at temporal points \( t_i \) and at spatial locations \( s_k \), i.e, observations of the form:

\[
X_n(s_k; t_i), \quad 1 \leq k \leq K, \quad 1 \leq i \leq I, \quad 1 \leq n \leq N.
\]

The covariances (assuming zero mean) of these observations are given by:

\[
\text{Cov}(s_k, t_i, s_\ell, t_j) = E \{X(s_k; t_i) \{X(s_\ell; t_j)\} \}.
\]
Separability means that the covariance structure factors into the product of two functions:

\[
\text{Cov}(s_k, t_i, s_\ell, t_j) = C_1(s_k, s_\ell)C_2(t_i, t_j),
\]

one depending only on space and the other only on time. To illustrate the tremendous benefits of separability assume that we have \( I = 10 \) temporal points and \( K = 10 \) spatial locations. A non-separable covariance will have \( \frac{KI(KI+1)}{2} = 5050 \) unknown parameters. On the other hand, a separable covariance will have only \( \frac{K(K+1)}{2} + \frac{I(I+1)}{2} - 1 = 109 \) unknown parameters (the \(-1\) is used because we have a constraint). It is clearly how useful is the assumption of separability.

In this dissertation we will test separability for two different forms of data sets.

**Chapter 2 - Testing separability of space-time functional processes:**

The work presented in Chapter 2 is motivated by geostatistical functional data; functions are observed at a number of spatial locations. Such data are quite common in environmental and climate studies. Perhaps the best-known example is annual temperature and log-precipitation curves, averaged over several decades, at locations distributed over a region of interest. Such data are used in Ramsay and Silverman (2005), and Delicado et al. (2010) provide further references. Gromenko et al. (2012) and Gromenko and Kokoszka (2012, 2013) study curves describing the evolution of certain ionospheric parameters measured at globally distributed locations at which radar-type instruments called ionosondes operate. Gromenko et al. (2016) study precipitation measurements extending over several decades at about sixty locations in the midwestern United States. Particulate pollution data (Krall et al. 2015) are also of this form; hourly measurements over many years of concentrations of various pollutants at a number of measurement stations within urban
areas and larger regions form functions defined at fixed geographical locations.

The data that motivate the first part of this dissertation are $N$ independent replications of the field

$$X_n(s_k, t), \quad 1 \leq k \leq K, \quad t \in T.$$  

At each spatial location $s_k$, a function with argument $t$ is observed. For example, $X_n(s_k, t)$ can be the maximum daily temperature on day $t$ of year $n$ at location $s_k$.

**Chapter 3 - Testing separability of functional time series:** Time series of weather or pollution related measurements obtained at spatial locations typically exhibit strong periodic patterns. An approach to accommodate this periodicity is to divide the time series of such type into segments, each segment corresponding to a natural period. For example, a periodic time series of maximum daily temperatures at some location can be viewed as a stationary time series of functions, one function per year. If the measurements are available at many locations $s_k$, this gives rise to the data structure of the form

$$X_n(s_k; t_i), \quad k = 1, \ldots, K, \quad i = 1, \ldots, I(= 365), \quad n = 1, \ldots, N,$$

where $n$ indexes year, and $t_i$ day within a year. Time series of functions are now discussed in several books, e.g. Bosq (2000), Horváth and Kokoszka (2012), Kokoszka and Reimherr (2017), but research on spatial fields of time series of functions is relatively new, e.g. Tupper et al. (2015), Liu et al. (2016), Kokoszka et al. (2016), Gromenko et al. (2016, 2016), French et al. (2016). Testing separability of spatiotemporal functional data of the above form is investigated in Constantinou et al. (2015) and Aston et al. (2015), under the assumption that the fields
$X_n(\cdot, \cdot), 1 \leq n \leq N$, are independent. No tests are currently available for testing separability in the presence of temporal dependence across $n$. In a broader setting, the data that motivate the second part of this dissertation have the form of functional panels:

$$X_n(t) = [X_{n1}(t), X_{n2}(t), ..., X_{ns}(t)]^T, \quad 1 \leq n \leq N.$$  

Each $X_{ns}(\cdot)$ is a smooth curve, and all curves are defined on the same time interval. The index $n$ typically stands for day, week, month or year. For instance, $X_{ns}(t)$ can be the exchange rate (against the Euro or the US Dollar) of currency $s$ at minute $t$ of the $n$th trading day, or $X_{ns}(t)$ can be the stock price of company $s$ at minute $t$ of the $n$th trading day. Another extensively studied example is daily or monthly yield curves for a panel of countries, e.g. Ang and Bekaert (2002), Bowsher and Meeks (2008), Hays et al. (2012), Kowal et al. (2016), among others. A significance test, which accounts for the temporal dependence present in all examples listed above, is therefore called for. The derivation of such a test, and the examination of its properties, is the purpose of Chapter 3. We will demonstrate that it works well in situations where the tests of Constantinou et al. (2015) and Aston et al. (2015) fail.

As for scalar data, the assumption of separability has numerous benefits for functional data as well. In addition to the benefits described, in the context of geostatistical functional data (functional time series), separability implies that the optimal functions used for temporal dimension reduction are the same for each spatial location (member (coordinate) of the panel); information can then be pooled across spatial locations (coordinates) to get better estimates of these functions.
Testing separability of space-time functional processes

2.1 Introduction

We begin this chapter by summarizing the procedure of Mitchell et al. (2006). Suppose we observe \( N \) independent and identically distributed scalar fields at temporal points \( t_i \) and spatial locations \( s_k \), so that the data are of the form

\[
X_n(s_k; t_i), \quad k = 1, \ldots, K, \quad i = 1, \ldots, I, \quad n = 1, \ldots, N.
\]

The independent and identically distributed assumption implies that the mean function \( E\{X_n(s; t)\} = \mu(s; t) \) and the covariances

\[
\sigma_{k\ell; ij} = E \left[ \{X_n(s_k; t_i) - \mu(s_k; t_i)\} \{X_n(s_\ell; t_j) - \mu(s_\ell; t_j)\} \right]
\]

do not depend on \( n \). The assumption of separability implies that \( \sigma_{k\ell; ij} = u_{k\ell}v_{ij} \), where \( u_{k\ell} \) does not depend on time, \( t \), and \( v_{ij} \) does not depend on space, \( s \). This
relation is stated in matrix form as

\[ \Sigma = V \otimes U = \begin{bmatrix} v_{11}U & v_{12}U & \ldots & v_{1I}U \\ v_{21}U & v_{22}U & \ldots & v_{2I}U \\ \vdots & \vdots & \ddots & \vdots \\ v_{I1}U & v_{I2}U & \ldots & v_{II}U \end{bmatrix}, \tag{2.1} \]

where \( U \) is the \( K \times K \) matrix with entries \( u_{k\ell} \) and \( V \) is the \( I \times I \) matrix with entries \( v_{ij} \). The matrix \( \Sigma \) is \( KI \times KI \) and can be viewed as the covariance matrix of the vectorized matrix \( \{ X_n(s_k, t_i) \} \), with \( k \) indexing rows and \( i \) columns. Lu and Zimmerman (2005) and Mitchell et al. (2006) use the test statistic

\[ \hat{T} = N \left\{ K \log \det(\hat{V}) + I \log \det(\hat{U}) - \log \det(\hat{\Sigma}) \right\}, \tag{2.2} \]

where \( \hat{V}, \hat{U}, \hat{\Sigma} \) are Gaussian likelihood estimates defined in Theorem 2.2. Their approach is based on Theorem 2.1, which justifies a Monte Carlo approximation for the null distribution of \( \hat{T} \). One can, e.g., use \( \mu = 0, U = I_K, V = I_I \) to obtain a large number of replicates of \( \hat{T} \), and so approximate its null distribution. The following theorem was introduced, apparently independently, by Lu and Zimmerman (2005) and Mitchell et al. (2006).

**Theorem 2.1.** If the observations are normal, (2.1) holds, and \( N > KI \), then the distribution of \( \hat{T} \) defined in (2.2) does not depend on \( \mu, U, V \).

The choice of statistic (2.2) is thus justified by the invariance property stated in Theorem 2.1. Perhaps more natural test statistics should be based on some distance between the matrices \( \hat{\Sigma} \) and \( \hat{V} \otimes \hat{U} \). It might be hoped that a more direct comparison would lead to tests with better power. However, such statistics are not
invariant in the sense of Theorem 2.1 and their asymptotic distribution has not been found. The first contribution of this work is to derive the joint asymptotic null distribution of \( \hat{\Sigma}, \hat{V}, \hat{U} \) and show how it enables one to derive the limit distribution of several natural test statistics in the multivariate context. This is addressed in Section 2.2. The proofs are presented in the Appendix-A.

The second contribution, which motivated our research, is related to functional data that are \( N \) replicates of the field \( X_n(s_k, t) \), \( k = 1, \ldots, K, \ t \in T \). At each spatial location \( s_k \), a function with argument \( t \) is observed. For example, \( X_n(s_k, t) \) can be the average daily wind speed on day \( t \) of month \( n \) at location \( s_k \), like the Irish wind data we consider in Section 2.4.2. For this data set, \( N = 216 \). The number of spatial locations, \( K \), can be anything from a dozen to a few hundred. For the Irish wind data \( K = 11 \), and the number of measurements per month is \( I = 28, 29, 30, 31 \). The approach of Mitchell et al. (2006) thus cannot be applied because the condition \( N > KI \) is violated. For pollution data, \( I \) can be even larger, as hourly and half hourly measurements are often available. Our approach exploits the functional structure of the data and uses a dimension reduction. Since for climate and environmental data sets \( N \) is fairly large, we derive asymptotic tests as \( N \to \infty \) using the results of Section 2.2. These developments are described in Section 2.3. A related, independently and concurrently developed, paper concerned with testing the separability of functional data is Aston et al. (2015).

The remainder of this chapter is organized as follows. In Section 2.2, we derive the test statistics and their asymptotic distributions. Section 2.3 focuses on test procedure of functional data and its large-sample justification. Finally, in Section 2.4 we present results of a simulation study and apply our procedure to Irish
2.2 Multivariate theory

This section clarifies the behavior of several test statistics based on normal maximum likelihood estimators. The theory is valid under the following assumption.

Assumption 1. Assume $X_1, \ldots, X_N$ are independent and identically normally distributed $K \times I$ matrices with $E(X_n) = M$ and Cov$\{\text{vec}(X_n)\} = \Sigma$, where $\Sigma$ is a $KI \times KI$ positive definite matrix. Assume that $N > KI$.

We begin with Theorem 2.2 (Dutilleul 1999). Recall that if $A$ is a $K \times I$ matrix, then vec$(A)$ is a column vector of length $KI$ obtained by stacking the columns of $A$ on top of each other.

Theorem 2.2. Under Assumption 1, the maximum likelihood estimators of $M$ and $\Sigma$ are

$$\hat{M} = \frac{1}{N} \sum_{n=1}^{N} X_n, \quad \hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \text{vec}(X_n - \hat{M})\{\text{vec}(X_n - \hat{M})\}^\top.$$

If $\Sigma$ admits the decomposition

$$\Sigma = V \otimes U, \quad \dim(U) = K \times K, \quad \dim(V) = I \times I, \tag{2.3}$$

where $U$ and $V$ are positive definite, then the maximum likelihood estimators of $U$ and $V$ satisfy

$$\hat{U} = \frac{1}{NI} \sum_{n=1}^{N} (X_n - \hat{M})\hat{V}^{-1}(X_n - \hat{M})^\top, \quad \hat{V} = \frac{1}{NK} \sum_{n=1}^{N} (X_n - \hat{M})^\top\hat{U}^{-1}(X_n - \hat{M}).$$
Roś et al. (2016) show that if $N > KI$, then the maximum likelihood estimates of $U$ and $V$ exist and are unique. The estimators $\hat{U}$ and $\hat{V}$ are defined indirectly and must be computed using an iterative procedure with some normalization to ensure identifiability. The following algorithm (Dutilleul 1999) produces $O_P(N^{-1/2})$ consistent estimators. It uses the normalization $\text{tr}(U_i) = K$.

Initialize with $U_0 = I_K$, the $K \times K$ identity matrix. For $i = 0, 1, 2, \ldots$, calculate

\[
\hat{V}_i = \frac{1}{NK} \sum_{n=1}^{N} (X_n - \hat{M})^\top \hat{U}_i^{-1} (X_n - \hat{M}),
\]

\[
\tilde{U}_{i+1} = \frac{1}{NI} \sum_{n=1}^{N} (X_n - \hat{M}) \hat{V}_i^{-1} (X_n - \hat{M})^\top,
\]

\[
\hat{U}_{i+1} = \frac{K}{\text{tr}(\tilde{U}_{i+1})} \tilde{U}_{i+1},
\]

until convergence is reached.

This algorithm is commonly referred to as the flip-flop algorithm for estimating $U$ and $V$. Alternatives can be found in Van Loan and Pitsianis (1993), Strobach (1995), Yu et al. (2004) and Werner et al. (2008). The work of Werner et al. (2008) is particularly relevant as they base their estimation on the minimization of the Frobenius norm $\|\hat{V} \otimes \hat{U} - \hat{\Sigma}\|_F$, which appears below in the definition of the statistic $\hat{T}_F$. As we will see, this statistic outperformed the other tests for the non-separable alternatives we considered in Section 2.4.

The most natural statistic to test separability, i.e., $\Sigma = V \otimes U$, should be based on a difference between $\hat{V} \otimes \hat{U}$ and $\hat{\Sigma}$. We will show that the statistic

\[
\hat{T}_F = N \|\hat{V} \otimes \hat{U} - \hat{\Sigma}\|_F^2,
\]

where $\| \cdot \|_F^2$ is the squared Frobenius matrix norm, i.e., the sum of squares of all entries, converges, and will find the asymptotic distribution. This distribution
involves the asymptotic covariance matrix of vec($\hat{V} \otimes \hat{U} - \hat{\Sigma}$), which we denote by $W$. The form of $W$ is complex; see the Appendix-A. To obtain a chi–square limit distribution, a suitable quadratic form must be used. This suggests using the statistic

$$\hat{T}_W = N \text{vec}((\hat{V} \otimes \hat{U} - \hat{\Sigma})^\top \hat{W}^+ \text{vec}(\hat{V} \otimes \hat{U} - \hat{\Sigma})),$$

where $\hat{W}$ is an estimator of $W$ and $\hat{W}^+$ is its Moore–Penrose generalized inverse and therefore is uniquely determined. A Moore–Penrose generalized inverse is used because $U$, $V$, $\Sigma$, and the corresponding estimates, are all symmetric, and this implies many linear constraints on the entries of vec($\hat{V} \otimes \hat{U} - \hat{\Sigma}$). Using a Moore–Penrose generalized inverse is equivalent to dropping redundant entries. Alternatively one can use vech to eliminate redundant entries. Recall that vech of a matrix is a column vector obtained by vectorizing only the lower triangular part.

We will also show that the likelihood ratio statistic

$$\hat{T}_L = N \left\{ T \log \det(\hat{U}) + K \log \det(\hat{V}) - \log \det(\hat{\Sigma}) \right\}$$

discussed in Section 2.1 has the same limit as $\hat{T}_W$, i.e., is asymptotically chi–square with known degrees of freedom. Lu and Zimmerman (2005) derived the asymptotic distribution of $\hat{T}_L$. However they recommended use of a Monte Carlo approximation to the distribution of $\hat{T}_L$. Also Mitchell et al. (2006) used a Monte Carlo finite–sample approximation based on Theorem 2.1. We collect our results in Theorem 2.3.

**Theorem 2.3.** Suppose Assumption 1 and decomposition (2.3) hold. Let $W$ be the $(KI)^2 \times (KI)^2$ matrix defined in (2.16), with eigenvalues $\gamma_1, \ldots, \gamma_R$ and $R = (KI)^2$. Then,

$$\hat{T}_L \to \chi^2_d, \quad \hat{T}_W \to \chi^2_d, \quad \hat{T}_F \to \sum_{r=1}^R \gamma_r \chi^2_1(r), \quad N \to \infty,$$
where
\[
d = \frac{KI(KI + 1)}{2} - \frac{K(K + 1)}{2} - \frac{I(I + 1)}{2} + 1,
\]
and \(\chi^2_1(r)\) are independent chi-square random variables with one degree of freedom.

Theorem 2.3 is proven in the Appendix-A. The three statistics listed in Theorem 2.3 are not the only ones that our theory covers. Theorem 2.9, in the Appendix-A, which specifies the joint asymptotic distribution of \(\hat{\mathbf{V}}, \hat{\mathbf{U}}\) and \(\hat{\Sigma}\), can be used to derive the asymptotic distribution of many other reasonable test statistics. Alternatively, instead of normalizing by the trace of \(U\), it is also possible to take the first main diagonal element of either \(V\) or \(U\) to one, and not include that element in the vector; this is the normalization that Srivastava et al. (2008) use. Then one can obtain the limiting distribution and covariance matrix of \(\{\text{vech}^*(\hat{\mathbf{V}}^*), \text{vech}(\hat{\mathbf{U}})\}\) where \(\hat{\mathbf{V}}^*\) is the normalized version of \(\hat{\mathbf{V}}\), using the new normalization, and \(\text{vech}^*(a)\) is the subvector of \(\text{vech}(a)\) obtained by deleting its uppermost element.

### 2.3 Tests for functional data

#### 2.3.1 The test procedure

We now show how the results of Section 2.2 are applied to testing separability of geostatistical functional data. We consider independent spatiotemporal random fields \(X_n(\cdot, \cdot)(n = 1, \ldots, N)\), which have the same distribution as the field \(\{X(s, t), s \in \mathcal{S}, t \in \mathcal{T}\}\), which satisfies \(E\{\int_{\mathcal{S}} \int_{\mathcal{T}} X^2(s, t)dsdt\} < \infty\). Then \(X_n(s, t) = \mu(s, t) + \varepsilon_n(s, t)\), where \(\mu \in L^2(\mathcal{S} \times \mathcal{T})\), and the \(\varepsilon_n\) are independent and identically distributed random elements of \(L^2(\mathcal{S} \times \mathcal{T})\) which satisfy

\[
E \left\{ \int_{\mathcal{S}} \int_{\mathcal{T}} \varepsilon^2_n(s, t)dsdt \right\} < \infty, \quad E\{\varepsilon_n(s, t)\} = 0.
\]
We consider the covariances
\[
\sigma(s, s'; t, t') = \text{Cov}\{X(s, t), X(s', t')\} = E\{\varepsilon_n(s; t)\varepsilon_n(s', t')\}.
\]

Our objective is to test
\[
H_0 : \quad \sigma(s, s'; t, t') = U(s, s') V(t, t').
\] (2.4)

We assume that complete functions of time are available. This means that they are either densely observed or reconstructed from sparse data. In the latter case, the tests makes a statement about the reconstructed functions. In contrast, there can be only a handful of spatial locations. If the nonseparable structure is not reflected in the functions available at these locations, it will not be detected. As in the multivariate case, the functions \(U\) and \(V\) are uniquely determined only up to multiplicative constants, and any testing algorithm must include some arbitrary normalization. However, the p-values of our testing procedures do not depend on this choice. In the Appendix-A, we describe procedures suitable for curves defined at a small number of spatial locations. We describe the most general procedure that involves both spatial and temporal dimension reduction. Its asymptotic justification is provided in Section 2.3.2. We assume that for each \(n\) the field \(X_n\) is observed at the same spatial locations \(s_k(k = 1, \ldots, K)\). We estimate \(\mu(s_k, t)\) by the sample average \(\hat{\mu}(s_k, t) = N^{-1} \sum_{n=1}^{N} X_n(s_k, t)\), and focus in the following on the covariance structure. Under \(H_0\), the covariances of the observations are
\[
\text{Cov}\{X_n(s_k, t), X_n(s_\ell, t')\} = U(k, \ell) V(t, t'),
\] (2.5)

with entries \(U(k, \ell) = U(s_k, s_\ell)\) forming a \(K \times K\) matrix \(U\), and \(V\) being the tempo-
general covariance function over $\mathcal{T} \times \mathcal{T}$. We thus assume that the data are observed at fixed spatial locations. Since the null hypothesis (2.4) is a statement about separability over the entire domain $\mathcal{S}$, our procedures, like any other procedures based on such data, will have no power unless, under the alternative, the covariance structure is not separable at the observed locations. Practically implementable tests must also involve dimension reduction in time, so that we can work with finite-dimensional objects on which computations can be done. If the nonseparability is present only in the orthogonal complement of the temporal dimension reduction subspace, it will not be detected either.

The testing procedure with dimension reduction in both space and time is as follows:

**Procedure 1.**

1. Under the assumption of separability, the optimal functions used for temporal dimension reduction are the same for each spatial location. To find the estimated temporal functional principal components, $\hat{v}_1(t), \ldots, \hat{v}_J(t)$, we pool across space to obtain the following estimated covariance function:

$$\tilde{C}(t, t') = \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} X_n(s_k, t)X_n(s_k, t').$$

Then $\hat{v}_1, \ldots, \hat{v}_J$ are the first $J$ eigenfunctions of $\tilde{C}$.

2. Approximate each curve $X_n(s_k, t)$ by

$$X_n^{(j)}(s_k, t) = \hat{\mu}(s_k, t) + \sum_{j=1}^{J} \xi_{jn}(s_k) \hat{v}_j(t),$$

where $J$ is finite and represents the number of principal components used in the
temporal dimension reduction. Construct $K \times J$ matrices

$$\Xi_n = [\xi_{jn}(s_k), \ k = 1, \ldots, K, \ j = 1, \ldots, J],$$

where $J$ is chosen so large that the first $J$ sample eigenvalues explain at least 85\% of the temporal variance, i.e., of $\hat{C}$. This is functional principal components analysis carried out on the sample pooled across space.

3. Approximate each vector $\{\xi_{jn}(s_1), \ldots, \xi_{jn}(s_K)\}$ using

$$\xi_{jn}(s_k) = \sum_{l=1}^{L} \zeta_{lj} \hat{u}_l(s_k).$$

The vectors $\{\hat{u}_l(s_1), \ldots, \hat{u}_l(s_K)\}$ are the estimated eigenvectors of the matrix

$$\hat{U}_i(k, \ell) = (NJ_i)^{-1} \sum_{j=1}^{J_i} \sum_{n=1}^{N} \xi_{jn}(s_k) \xi_{jn}(s_\ell) \lambda_{ij}.$$

Construct the $L \times J$ matrices $Z_n = [\zeta_{lj}, \ l = 1, \ldots, L, \ j = 1, \ldots, J]$ where $L$ is chosen large enough so that the first $L$ eigenvalues explain at least 85\% of the spatial variance, i.e., of $\hat{U}$. This is multivariate principal components analysis on the sample pooled across time.

4. Compute the matrix

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \text{vec}(Z_n - \hat{M}) \{\text{vec}(Z_n - \hat{M})\}^\top, \quad \hat{M} = \frac{1}{N} \sum_{n=1}^{N} Z_n.$$

Using Dutilleul's Algorithm with $Z_n$ in place of $X_n$, $J$ in place of $I$ and $L$ in place of $K$ compute the matrices $\hat{U}$ and $\hat{V}$.

5. Estimate the matrix $W$ defined by (2.16) in the Appendix-A by replacing $\Sigma, U, V$ by their estimates.
6. Calculate the p-value using the limit distribution specified in Theorem 2.3, with
I replaced by J and K replaced by L.

Step 2 can be easily implemented using R function pca.fd and step 3 by using
R function prcomp. To find the p-value of \( \hat{T}_F \), i.e. the quantile of a mixture of
chi–squares, we use the R package CompQuadForm and the function imhof.

2.3.2 Large–sample justification

In this section we provide asymptotic theory justifying the testing procedure of
Section 2.3.1. All proofs can be found in the Appendix-A. Our theory is developed
in the setting in which in step 4 of Procedure 1 only the first iteration step is
applied, cf. Dutilleul’s Algorithm. In other words, the spatial and temporal co-
variances are estimated separately by pooling across time and space respectively,
and without iterating:

\[
V(t, t') = \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} X_n(s_k, t)X_n(s_k, t')
\]

\[
\hat{U}(k, k') = \frac{K \sum_{n} \int X_n(s_k, t)X_n(s_{k'}, t) \, dt}{N \text{tr}(\hat{\sigma})},
\]

where \( \hat{\sigma} \) is the empirical covariance. For ease of reference, we state the following
assumption on the data generating process.

Assumption 2. Assume that \( X_1(s, t), \ldots, X_N(s, t) \) are independent and identi-
cally distributed Gaussian processes in \( L^2(S \times T) \) with mean zero and a separable
covariance function \( \sigma(s, t, s', t') = U(s, s')V(t, t') \). Assume that the temporal func-
tions \( X_i(s, \cdot) \) are observed at spatial locations \( \{s_k : k = 1, \ldots, K\} \). Let \( U \) be the
matrix resulting in evaluating \( U \) at the \( \{s_k\} \). Assume, without loss of generality,
that \( \text{tr}(U) = K \).
The key to establishing the equivalence between procedures based on the population and estimated principal components is the derivation of $O_P(N^{-1})$ bounds for differences between the population covariance operator $\mathbf{V}$ and the operator $\widehat{\mathbf{V}}$ and an analogous bound for the difference between $\mathbf{U}$ and $\widehat{\mathbf{U}}$. We can conclude that replacing the population principal components to which the multivariate theory applies by the estimated functional principal components does not affect the limiting distribution.

**Theorem 2.4.** Denote by $\widehat{T}_F$ the statistic computed using the estimators $\widehat{u}_k$ and $\widehat{v}_j$ and by $T_F^\dagger$ the random variable computed using the population functions $u_k$ and $v_j$. If Assumption 2 holds, then $\widehat{T}_F - T_F^\dagger = O_P(N^{-1/2})$.

As we show in the Appendix-A, the eigenvalues $\gamma$ can be consistently estimated, so we can use them to approximate the null distribution in Theorem 2.3, and so in practice we can compute rejection regions that are asymptotically justified.

### 2.4 Simulation study and applications

#### 2.4.1 A simulation study

We compare the performance of the tests based on the statistics introduced in Section 2.2 and the procedure described in Section 2.3. We include in the comparison the modified approach of Mitchell et al. (2006), which is based on Theorem 2.1 and Monte Carlo replications, and the spatial principal components introduced in Section 2.3. It can typically be applied only in conjunction with dimension reduction so that the condition $N > LJ$ holds. We thus consider four test procedures applicable to space–time functional data, which we denote $T_L, T_F, T_W$ and $T_{L-MC}$. The first three procedures use asymptotic critical values of limit distributions spec-
ified in Theorem 2.3; $T_{L-MC}$ uses the Monte Carlo critical values computed using $\mu = 0, U = I_L, V = I_J$.

The generated data follow multivariate normal distributions with mean $M = 0$ and one of the following two covariance functions. The first is the spatiotemporal covariance function introduced by Gneiting (2002):

$$
\sigma(s, s', t, t') = \frac{\sigma^2}{(a|t - t'|^{2\alpha} + 1)^{\tau}} \exp \left\{ - \frac{c\|s - s'|^{2\gamma}}{(a|t - t'|^{2\alpha} + 1)^{\beta\gamma}} \right\}.
$$

(2.7)

In this covariance function, $a$ and $c$ are nonnegative scaling parameters of time and space respectively, $\alpha$ and $\gamma$ are smoothness parameters which take values in $(0, 1]$, $\beta$ is the separability parameter which takes values in $[0, 1]$, $\sigma^2 > 0$ is the point-wise variance and finally $\tau \geq \beta d/2$, where $d$ is the spatial dimension. We focus on the effect of the space–time interaction parameter, $\beta \in [0, 1]$. If $\beta = 0$, the covariance function is separable. As $\beta$ increases, the space-time interaction becomes stronger. We set $\gamma = 1, \alpha = 1/2, \sigma^2 = 1, a = 1, c = 1$ and $\tau = 1$.

The second is the spatiotemporal covariance function introduced by Cressie and Huang (1999):

$$
\sigma(s, s', t, t') = \frac{\sigma^2 c_0^{d/2}}{\left\{a_0^2(t - t')^2 + 1\right\}^{1/2} \left\{a_0^2(t - t')^2 + c_0\right\}^{d/2}} \times \exp \left[ -b_0 \left\{ \frac{a_0^2(t - t')^2 + 1}{a_0^2(t - t')^2 + c_0} \right\}^{1/2} \|s - s'| \right].
$$

(2.8)

In this covariance function, $a_0$ and $b_0$ are nonnegative scaling parameters of time and space respectively, $c_0 > 0$ is the separability parameter, $\sigma^2 > 0$ is the point-wise variance and $d$ is the spatial dimension. We focus on the effect of the space–time interaction parameter, $c_0 > 0$. If $c_0 = 1$, the covariance function is separable. We set $a_0 = 2, b_0 = 1, \sigma^2 = 1$ and $d = 2$. 
We use $I = 100$ time points equally spaced on $[0,1]$ and $K = 11$ space points in $[0,1] \times [0,1]$, the same number of points as in the Irish wind data considered by Gneiting (2002), which we study in Section 2.4.2. We consider different numbers of spatial principal components, $L$, and temporal functional principal components, $J$. In all our simulation settings, we achieve 85% of variance with $J = 2$ temporal functional principal components. We will also consider different values for the sample size $N$. All empirical rejection rates are based on 1000 replications, so their standard deviation is about 0.7% percent for size, we use a nominal significance level of 5%, and about 2% for power.

We study three different scenarios. Table 2.1 considers different values of the space–time interaction parameters, $\beta$ and $c_0$. Table 2.2 examines the effect of the sample size $N$, while Table 2.3 explores the effect of the number of principal components. Each table reports the rejection rates, the standard errors and the proportion of variance explained (CPV) by temporal principal components.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_F$</th>
<th>$T_W$</th>
<th>CPV(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.1 (0.70)</td>
<td>5.9 (0.75)</td>
<td>4.5 (0.65)</td>
<td>3.7 (0.60)</td>
<td>89</td>
</tr>
<tr>
<td>0.5</td>
<td>12.3 (1.04)</td>
<td>13.4 (1.08)</td>
<td>15.2 (1.14)</td>
<td>5.4 (0.71)</td>
<td>89</td>
</tr>
<tr>
<td>1</td>
<td>54.1 (1.58)</td>
<td>55.8 (1.57)</td>
<td>63.4 (1.52)</td>
<td>32.3 (1.48)</td>
<td>88</td>
</tr>
<tr>
<td>$c_0$</td>
<td>$T_{L-MC}$</td>
<td>$T_L$</td>
<td>$T_F$</td>
<td>$T_W$</td>
<td>CPV(%)</td>
</tr>
<tr>
<td>1</td>
<td>6.0 (0.75)</td>
<td>6.6 (0.79)</td>
<td>3.8 (0.60)</td>
<td>3.8 (0.60)</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>17.1 (1.19)</td>
<td>19.1 (1.24)</td>
<td>36.7 (1.52)</td>
<td>22.1 (1.31)</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>47.9 (1.58)</td>
<td>50.4 (1.58)</td>
<td>77.7 (1.32)</td>
<td>60.2 (1.55)</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 2.1: Rejection rates (%), based on 1000 replications, for $N = 100$, $L = J = 2$, $\beta = 0, 0.5, 1$ and $c_0 = 1, 5, 10$. The parameters $\beta = 0$ and $c_0 = 1$ correspond to the null hypothesis of separability. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

Table 2.1 shows that the test $T_F$ wins in terms of balance between correct size and power in our settings. The two likelihood methods do not exhibit significantly
Table 2.2: Empirical power (%), based on 1000 replications, for $N = 100, 150$ and $200$, $L = J = 2$, $\beta = 1$ or $c_0 = 10$. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

different rejection rates. Table 2.2 shows that, as expected, the empirical power increases as the sample size increases, with $T_F$ preserving its lead in terms of power, despite its conservative empirical size. In Table 2.3, we see that only the tests $T_{L-MC}$ and $T_F$ have empirical size robust to the number of the principal components used. This is desirable, as there is always uncertainty regarding the number of functional principal components that should be used. The test $T_F$ is more powerful than $T_{L-MC}$.

As a final demonstration, which is particularly relevant to the analysis in Section 2.4.2, we consider a covariance structure with two components, one which is separable and one which is not. The separable component gives the majority of the variation, so it is more likely to dominate the lower principal components, since these are the directions of highest variability. For this scenario, we use a covariance which is the sum of (2.7) with $\beta = 0$, separable, and (2.8) with $c_0 = 0.3$, nonseparable. Table 2.4 shows that the rejection fraction is 5% for $L = J = 2$, and as the number of spatial and temporal principal components increases the rejection fraction increases too.

Our overall conclusion is that the norm–based test, $T_F$, emerges as a clear win-
Table 2.3: Rejection rates (%), based on 1000 replications, for $\beta = 0, 1$ and $c_0 = 1, 10$, $N = 100$, and the varying degrees of spatial and temporal dimension reduction. $L$ is the number of spatial and $J$ of temporal principal components. The proportion of variance explained (CPV) from the temporal principal components is given in the last column. The standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$L$</th>
<th>$J$</th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_F$</th>
<th>$T_W$</th>
<th>CPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2</td>
<td>5.1 (0.70)</td>
<td>5.9 (0.75)</td>
<td>4.5 (0.65)</td>
<td>3.7 (0.60)</td>
<td>89</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>3</td>
<td>4.4 (0.65)</td>
<td>6.9 (0.80)</td>
<td>4.7 (0.67)</td>
<td>39.4 (1.55)</td>
<td>93</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>4</td>
<td>5.2 (0.70)</td>
<td>13.6 (1.08)</td>
<td>5.2 (0.70)</td>
<td>98.1 (0.43)</td>
<td>95</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>5</td>
<td>4.1 (0.63)</td>
<td>32.2 (1.48)</td>
<td>5.7 (0.73)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>6</td>
<td>2.5 (0.49)</td>
<td>81.3 (1.23)</td>
<td>5.3 (0.71)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>7</td>
<td>5.3 (0.71)</td>
<td>100 (0)</td>
<td>4.0 (0.62)</td>
<td>100 (0)</td>
<td>98</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>54.1 (1.58)</td>
<td>55.8 (1.57)</td>
<td>63.4 (1.52)</td>
<td>32.3 (1.48)</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>54.7 (1.57)</td>
<td>61.9 (1.54)</td>
<td>89.1 (0.99)</td>
<td>95.8 (0.63)</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>79.0 (1.29)</td>
<td>90.5 (0.93)</td>
<td>99.7 (0.17)</td>
<td>100 (0)</td>
<td>94</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>81.5 (1.23)</td>
<td>97.9 (0.45)</td>
<td>99.9 (0.10)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>91.0 (0.90)</td>
<td>100 (0)</td>
<td>99.9 (0.10)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>93.2 (0.80)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>$c_0$</td>
<td>$L$</td>
<td>$J$</td>
<td>$T_{L-MC}$</td>
<td>$T_L$</td>
<td>$T_F$</td>
<td>$T_W$</td>
<td>CPV (%)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6.0 (0.75)</td>
<td>6.6 (0.79)</td>
<td>3.8 (0.60)</td>
<td>3.8 (0.60)</td>
<td>87</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5.2 (0.70)</td>
<td>8.1 (0.86)</td>
<td>4.1 (0.63)</td>
<td>64.5 (1.51)</td>
<td>95</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4.8 (0.68)</td>
<td>13.2 (1.07)</td>
<td>5.4 (0.71)</td>
<td>100 (0)</td>
<td>99</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5.4 (0.71)</td>
<td>34.6 (1.50)</td>
<td>3.6 (0.59)</td>
<td>100 (0)</td>
<td>99.7</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>6</td>
<td>2.7 (0.51)</td>
<td>85.5 (1.11)</td>
<td>2.3 (0.47)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>5.0 (0.69)</td>
<td>100 (0)</td>
<td>2.2 (0.46)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>47.9 (1.58)</td>
<td>50.4 (1.58)</td>
<td>77.7 (1.32)</td>
<td>60.2 (1.55)</td>
<td>96</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>4</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>5</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>99.6</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>6</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>7</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>99.9</td>
</tr>
</tbody>
</table>

The empirical size of $T_L$ is not robust to the number of principal components used, consistent with the findings of Lu and Zimmerman (2005) for their non-functional data version of that test. The empirical size of $T_{L-MC}$ is competitive, but the test is less powerful. The fact that $T_W$ test cannot control the type I error is consistent with
Table 2.4: Rejection rates (%), based on 1000 replications, for data whose covariance is the sum of a large separable component and a small nonseparable component; $N = 100$. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

<table>
<thead>
<tr>
<th>L</th>
<th>J</th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_F$</th>
<th>$T_W$</th>
<th>CPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>6.6 (0.79)</td>
<td>7.9 (0.85)</td>
<td>5.2 (0.70)</td>
<td>1.5 (0.38)</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7.9 (0.85)</td>
<td>10.4 (0.97)</td>
<td>10.5 (0.97)</td>
<td>68.0 (1.48)</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>16.3 (1.17)</td>
<td>33.4 (1.49)</td>
<td>34.6 (1.50)</td>
<td>99.2 (0.28)</td>
<td>96</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>24.2 (1.35)</td>
<td>73.3 (1.40)</td>
<td>69.8 (1.45)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>19.6 (1.26)</td>
<td>97.8 (0.46)</td>
<td>87.7 (1.04)</td>
<td>100 (0)</td>
<td>98</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>24.7 (1.36)</td>
<td>100 (0)</td>
<td>89.5 (0.97)</td>
<td>100 (0)</td>
<td>98</td>
</tr>
</tbody>
</table>

the results of Srivastava et al. (2008).

Appropriate QQ-plots can be found in the Appendix-A to verify the asymptotic distributions of the test statistics introduced in this chapter.

### 2.4.2 Irish wind data

Consider the data of Haslett and Raftery (1989), which consist of daily averages of wind speeds, measured in knots, at 11 meteorological stations in Ireland during the period 1961–1978. The data are available at [http://lib.stat.cmu.edu/datasets/wind.data](http://lib.stat.cmu.edu/datasets/wind.data). The stations are fairly uniformly distributed over Ireland. Each functional observation $X_n(s_k,t)$ consists of the average wind speed for day $t$, month $n$, $N = 216$, and at location $s_k$. A plot of these data can be found in the Appendix-A. Gneiting (2002) estimated model (2.7) on deseasonalized and transformed data and obtained $\hat{\beta} = 0.61$, which indicates a nonseparable covariance structure. His conjecture was verified by Bevilacqua et al. (2010) using a parametric significance test.

We first applied our tests to deseasonalized curves obtained after removing the monthly mean from each curve; we center all Januaries, Februaries, etc., separately.
Table 2.5: P-values for the separability tests applied to deseasonalized wind speed data.

<table>
<thead>
<tr>
<th></th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_F$</th>
<th>$T_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = J = 2$</td>
<td>0.06</td>
<td>0.055</td>
<td>0.039</td>
<td>0.034</td>
</tr>
<tr>
<td>$L = J = 3$</td>
<td>0.467</td>
<td>0.436</td>
<td>0.149</td>
<td>0.203</td>
</tr>
<tr>
<td>$L = J = 4$</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

This simple transformation removes the seasonality to a large extent. The p-values for selected combinations of $L$ and $J$ are shown in Table 2.5. Only when $L$ and $J$ are larger than 3, do we see clear evidence for nonseparability, perhaps because the covariance structure has two components, one separable and one not. The separable component gives the majority of the variation, so it is more likely to dominate the lower principal components, since these are the directions of highest variability. This would then explain why separability is not rejected for smaller values of $L$ and $J$. This effect is illustrated on simulated data; see Table 2.4 and its discussion.

Gneiting (2002) applied a square root transformation to stabilize the variance and to make the marginal distributions approximately normal, and deseasonalized the data in slightly different way. He estimated the seasonal effect by calculating the average of the square roots of the daily means over all years and stations for each day of the year, and then regressing the results on a set of annual harmonics. Subtraction of the estimated seasonal effect and the estimated spatially varying mean, as given in Gneiting’s Table 3, from the square roots of the daily means results in the deseasonalized data, referred to as velocity measures. Using Gneiting’s procedure, our tests produce p-values smaller than 0.001 even for $L = J = 2$, which indicates a nonseparable covariance structure.

Since the number of meteorological stations is relatively small it is natural to apply our tests without dimension reduction in space. For this approach, all tests
produce p-values smaller than 0.001, for $J = 2, 3, 4$.

2.5 Appendix-A

Appendix-A contains mostly proofs of the large sample results used in this chapter. However, we begin with the detailed description of testing procedures which do not involve spatial dimension reduction, Section 2.5.1. Then, in Section 2.5.2, we proceed with the derivation of the form of several matrices which are needed to formulate the joint asymptotic distribution of the maximum likelihood estimators. The form of these matrices is complex, and in some cases can be defined only via an algorithm, which can however be coded. With these matrices defined, we present in Section 2.5.3 the proof of our main result, Theorem 2.3. Section 2.5.4 contains the proof of Theorem 2.4, and finally Section 2.5.5 has additional empirical rejection rates, more discussion in the Irish wind data and application to pollution data.

2.5.1 Tests for a small number of spatial locations

In Section 2.3.1, we described the testing procedure which involves dimension reduction in both space and time, using estimated principal components. In this section, we describe two alternative versions of the tests. Both are suitable for curves defined at a small number of spatial locations. In Section 2.5.1.1, we describe the simplest approach which uses a fixed basis of time domain functions. Section 2.5.1.2 considers temporal dimension reduction via functional principal components.
2.5.1.1 Procedure 2: fixed spatial locations, fixed temporal basis

We assume that for each \( n \) the field \( X_n \) is observed at the same spatial locations \( s_k, k = 1, \ldots, K \). We estimate \( \mu(s_k, t) \) by the sample average \( \hat{\mu}(s_k, t) = N^{-1} \sum_{n=1}^{N} X_n(s_k, t) \), and focus on the covariance structure. Under \( H_0 \), the covariances of the observations are

\[
\text{Cov}\{X_n(s_k, t)X_n(s_\ell, t')\} = U(k, \ell)V(t, t'),
\]

with entries \( U(k, \ell) = U(s_k, s_\ell) \) forming a \( K \times K \) matrix \( U \), and \( V \) being the temporal covariance function over \( \mathcal{T} \times \mathcal{T} \). As in the multivariate setting, the estimation of the matrix \( U \) and the covariance function \( V \) must involve an iterative procedure. In the functional setting, a dimension reduction is also needed. Suppose \( \{v_j, j \geq 1\} \) is a basis system in \( L^2(\mathcal{T}) \) such that for sufficiently large \( J \), the functions

\[
X_n^{(J)}(s_k, t) = \mu(s_k, t) + \sum_{j=1}^{J} \xi_{jn}(s_k)v_j(t)
\]

are good approximations to the functions \( X_n(s_k) \). We thus replace a large number of time points by a moderate number \( J \), and seek to reduce the testing of \( H_0 \) to testing the separability of the covariances of the transformed observations given as \( K \times J \) matrices

\[
\Xi_n = [\xi_{jn}(s_k), \ k = 1, \ldots, K, \ j = 1, \ldots, J]. \tag{2.9}
\]

The index \( j \) should be viewed as a transformed time index. The number \( I \) of the actual time points \( t_i \) can be very large, \( J \) is usually much smaller. The following proposition is easy to prove. It establishes the connection between the original testing problem and testing the separability of the transformed data (2.9).
assumption that the $v_j$ are orthonormal cannot be removed.

**Proposition 1.** For some orthonormal $v_j$, set

\[
X^{(J)}(s_k, t) = \mu(s_k, t) + \sum_{j=1}^{J} \xi_j(s_k)v_j(t).
\]

If

\[
E\{\xi_j(s_k)\xi_i(s_\ell)\} = U(k, \ell)V(j, i), \tag{2.10}
\]

then

\[
\text{Cov}\{X^{(J)}(s_k, t), X^{(J)}(s_\ell, s)\} = U(k, \ell)V(t, s). \tag{2.11}
\]

Conversely, (2.11) implies (2.10). The entries $V(j, i)$ and $V(t, s)$ are related via

\[
V(j, i) = \int \int V(t, s)v_j(t)v_i(s)dt\,ds, \quad V(t, s) = \sum_{j,i=1}^{J} V(j, i)v_j(t)v_i(s).
\]

We assume that \{v_j, j = 1, \ldots, J\} is a fixed orthonormal system, for example the first $J$ trigonometric basis functions. Slightly abusing notation, consider the matrices

\[
\hat{\Sigma} (KJ \times KJ), \quad \hat{U} (K \times K), \quad \hat{V} (J \times J),
\]

defined as in Theorem 2.2, but with matrices $X_n$ replaced by the matrices $\Xi_n$. The index $j \leq J$ now plays the role of the index $i \leq I$ of Section 2.2. To apply tests based on Theorem 2.3, we must recursively calculate $\hat{U}$ and $\hat{V}$ using the relations stated in Theorem 2.2. This can be done using Dutilleul’s Algorithm with $\Xi_n$ in place of $X_n$. This approach leads to the following test procedure. The test statistic can be one of the three statistics introduced in Section 2.2.

**Procedure 2.**
1. Choose a deterministic orthonormal basis $\hat{v}_j, j \geq 1$.

2. Approximate each curve $X_n(s_k, t)$ by

$$X^{(j)}_n(s_k, t) = \hat{\mu}(s_k, t) + \sum_{j=1}^{J} \xi_{jn}(s_k)\hat{v}_j(t).$$

Construct $K \times J$ matrices $\Xi_n$ defined in (4).

3. Compute the matrix

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \text{vec}(\Xi_n - \hat{M})\{\text{vec}(\Xi_n - \hat{M})\}^\top, \quad \hat{M} = \frac{1}{N} \sum_{n=1}^{N} \Xi_n.$$

Using Dutilleul’s Algorithm with $\Xi_n$ in place of $X_n$, compute the matrices $\hat{U}$ and $\hat{V}$.

4. Estimate the matrix $W$ defined by (2.16) by replacing $\Sigma, U, V$ by their estimates.

5. Calculate the $p$-value using the limit distribution specified in Theorem 2.3, with $I$ replaced by $J$.

Step 2 can be easily implemented using R function `pca.fd`, see Ramsay et al. (2009) Chapter 7. Several methods of choosing $J$ are available. We used the cumulative variance rule requiring that $J$ be so large that at least 85% of variance is explained for each location $s_k$.

### 2.5.1.2 Procedure 3: fixed spatial locations, data driven temporal basis

In Section 2.5.1.1, we used a deterministic orthonormal system. To achieve the most efficient dimension reduction, it is usual to project on a data driven system, with the functional principal components being used most often. Since the sequences of functions are defined at a number of spatial locations, it is not a priori clear how a suitable orthonormal system should be constructed, as each sequence
\{X_n(s_k), n = 1, \ldots, N\} has different functional principal components \(v_j(s_k), j \geq 1\), and Proposition 1 requires that a single system be used. Our next algorithm proposes an approach which leads to suitable estimates \(\hat{U}, \hat{V}\) and \(\hat{\Sigma}\). It is not difficult to show that these estimators are \(O_P(N^{-1/2})\) consistent.

Initialize with \(U_0 = I_K\). For \(i = 1, 2, \ldots\), perform the following two steps until convergence is reached.

1. Calculate
   \[
   \mathcal{V}_i(t, t') = (NK)^{-1} \sum_{n=1}^N \{X_n(\cdot, t) - \hat{\mu}(\cdot, t)\}^\top U_{i-1}^{-1} \{X_n(\cdot, t') - \hat{\mu}(\cdot, t')\}.
   \]

Denote the eigenfunctions and eigenvalues of \(\mathcal{V}_i\) by \(\{v_{ij}\}\) and \(\{\lambda_{ij}\}\). Determine \(J_i\) such that the first \(J_i\) eigenfunctions of \(\mathcal{V}_i\) explain at least 85% of the variance.

2. Project each function \(X_n(s_k, \cdot)\) on the first \(J_i\) eigenfunctions of \(\mathcal{V}_i\). Denote the scores of these projections by
   \[
   Z_{in}(s_k, j) = \langle X_n(s_k, \cdot) - \hat{\mu}(s_k, \cdot), v_{ij} \rangle
   \]
   and calculate
   \[
   U_i(k, \ell) = (NJ_i)^{-1} \sum_{j=1}^{J_i} \sum_{n=1}^N \frac{Z_{in}(s_k, j)Z_{in}(s_\ell, j)}{\lambda_{ij}}.
   \]

Normalize \(U_i\) so that \(\text{tr}(U_i) = K\).

Let \(\{\hat{v}_j, j = 1, \ldots, J\}\) denote the final eigenfunctions. Carry out the final projection \(\hat{Z}_n(s_k, j) = \langle X_n(s_k, \cdot) - \hat{\mu}(s_k, \cdot), \hat{v}_j \rangle\). For each \(n\), denote by \(\hat{Z}_n\) the \(K \times J\) matrix with these entries. Set
   \[
   \hat{\Sigma} = \frac{1}{N} \sum_{n=1}^N \text{vec}(\hat{Z}_n) \text{vec}(\hat{Z}_n)^\top.
   \]

Using the above algorithm, the testing procedure is as follows:

**Procedure 3.**

1. Calculate matrices \(\hat{\Sigma}, \hat{U}, \hat{V}\) according to the above Algorithm.
2. Perform steps 4 and 5 of Procedure 2.
2.5.2 Derivation of the Q matrices

This section introduces four matrices, that describe the covariance structure of products of various vectorized matrices consisting of standard normal variables. We refer to them collectively as Q matrices, as we use the symbol $Q$ with suitable subscripts and superscripts to denote them. These matrices appear in the asymptotic distribution of the vectorized matrices $\hat{U}, \hat{V}, \hat{\Sigma}$, which, in turn, is used to prove Theorem 2.3. In particular, the asymptotic distribution of statistic $\hat{T}_F$, which we recommended in Section 2.2, is expressed in terms of these Q matrices. Some of them are defined though an algorithm.

**Theorem 2.5.** If $E$ is an $K \times I$ matrix of independent standard normals, then

$$\text{Cov}\{\text{vec}(EE^\top)\} = 2I Q_K.\quad (2.12)$$

where

$$Q_K(i, j) = \begin{cases} 
1, & i = j = k + (k - 1)K, \quad k = 1, \ldots, K \\
\frac{1}{2}, & i = j \neq k + (k - 1)K, \quad k = 1, \ldots, K \\
\frac{1}{2}, & i \neq j = 1 + \left[\frac{(i-1) - ((i-1) \mod K)}{K}\right] + \{(i - 1) \mod K\}K, \\
0, & \text{otherwise},
\end{cases}$$

Proof. Denote by $e_{kl}$ the independent standard normals, and set

$$e_k = (e_{k1}, e_{k2}, \ldots, e_{kI})^\top, \quad k = 1, \ldots, K,$$
so that

\[
E = \begin{pmatrix}
e_1^T \\
\vdots \\
e_K^T
\end{pmatrix}.
\]

Then for any \(i, j\) in \(\{1, \ldots, K\}\) we have that the \((i, j)\) entry of \(\text{Cov}\{\text{vec}(EE^\top)\}\) can be written as

\[
\text{Cov}(e_{k_1}^\top e_{l_1}, e_{k_2}^\top e_{l_2}),
\]

where we have the relationships

\[
\begin{align*}
i &= k_1 + (l_1 - 1)K, k_1 = \{ (i - 1) \mod K \} + 1, l_1 = 1 + \frac{(i - 1) - (i - 1) \mod K}{K} \\
j &= k_2 + (l_2 - 1)K, k_2 = \{ (j - 1) \mod K \} + 1, l_2 = 1 + \frac{(j - 1) - (j - 1) \mod K}{K}.
\end{align*}
\]

For the diagonal terms, i.e. \(i = j\), we have two settings \(k_1 = k_2 = l_1 = l_2\), in which case the covariance is \(2I\), or alternatively \(k_1 = k_2 \neq l_1 = l_2\) in which case the covariance is \(I\). Since the former occurs in every \(K\)th term, we have established the proper pattern for the diagonal.

We now need only establish the pattern for the off diagonal. Every term in the off diagonal can be expressed as \(\text{Cov}(e_i^\top e_j, e_k^\top e_l)\), for some \(i, j, k, l = 1, \ldots, K\). Clearly, if any one index is different from the other 3, then the covariance is 0. We can’t have all four indices being equal as that would be a diagonal element, and we can’t have \(i = k \neq j = k\) as that would also be a diagonal element. If \(i = j\) and \(k = l\) then two inner products are independent, and thus the covariance is zero. Therefore, the only nonzero off diagonal entries occur when \(i = l \neq j = k\), and the covariance would be \(I\). To determine where in the \(K^2 \times K^2\) matrix these occur, we use the change of base formulas. 

\(\square\)
We illustrate the form of the matrices $Q_K$:

$$Q_2 = \begin{pmatrix} 1.0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1.0 \end{pmatrix},$$

$$Q_3 = \begin{pmatrix} 1.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.0 \end{pmatrix}.$$  

**Theorem 2.6.** If $E$ is an $K \times I$ matrix of independent standard normals, then

$$\text{Cov}\{\text{vec}(EE^\top), \text{vec}(E^\top E)\} = 2(IK)^{1/2}Q_{K,I},$$

(2.13)

where $Q_{K,I}$ is an $K^2 \times I^2$ matrix given by

$$Q_{K,I}(i,j) = \begin{cases} (KI)^{-1/2} & i = k_1 + K(k_1 - 1), j = k_2 + I(k_2 - 1) \\ 0 & \text{for } k_1 = 1, \ldots, K \text{ } k_2 = 1, \ldots, I \\ 0 & \text{otherwise}, \end{cases}$$

**Proof.** Here, each entry of the above covariance matrix is obtained by taking two
rows of $E$, possibly the same row, forming the inner product, taking two columns, taking their inner product, and then computing the covariance between the two. Due to the symmetry of this calculation, there are only three possible resulting values: when the two rows are different, then the two columns are different, or when both the rows and columns are different. When both rows and columns are different, we can, without loss of generality, take the first two rows and columns. In that case, the covariance becomes

$$\text{Cov} \left( \sum_{k=1}^{K} e_{1,k}e_{2,k}, \sum_{i=1}^{I} e_{i,1}e_{i,2} \right) = \sum_{k=1}^{K} \sum_{i=1}^{I} \text{Cov} \left( e_{1,k}e_{2,k}, e_{i,1}e_{i,2} \right).$$

However, every summand above is zero when $i > 2$ or $k > 2$ since they will then involve independent variables. Therefore, we can express the above as

$$\text{Cov} (e_{1,1}e_{2,1} + e_{1,2}e_{2,2}, e_{1,1}e_{1,2} + e_{2,1}e_{2,2}) = 0.$$ 

Hence, any term with two different rows and columns is zero. A similar result will hold when there are either two different rows or two different columns. The only nonzero term will stem from taking the same row and same column, in which case the value becomes

$$\text{Cov}(e_{1,1}e_{1,1} + e_{1,2}e_{1,2}, e_{1,1}e_{1,1} + e_{2,1}e_{2,1}) = \text{Var}(e_{1,1}^2) = 2.$$ 

Therefore, every nonzero entry will be 2. We now only need to determine which entries of the covariance matrix correspond to taking the same row and same column. Considering the structure induced by vectorizing, the first row of the covariance matrix and every subsequent $K$ rows will correspond to matching the same row of $E$. Similarly, the first and every subsequent $I$ column will correspond to matching
the same column of $E$. This corresponds to our definition and the result follows.

Some examples $Q_{K,I}$ are

$Q_{2,2} = \begin{pmatrix}
0.5 & 0 & 0 & 0.5 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0.5 & 0 & 0 & 0.5 \\
\end{pmatrix}$

and

$Q_{2,3} = \begin{pmatrix}
6^{-1/2} & 0 & 0 & 6^{-1/2} & 0 & 0 & 0 & 6^{-1/2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6^{-1/2} & 0 & 0 & 6^{-1/2} & 0 & 0 & 0 & 6^{-1/2} \\
\end{pmatrix}$

Before the next theorem, we define the matrix $\tilde{Q}_{R,K}$ via pseudo code.

End Code

Set $\tilde{Q}_{R,K}$ to be an $R = K^2 I^2$ by $K^2$ matrix of zeros

For $i = 1, \ldots, R$

\begin{align*}
l &= 1 + \lceil (i - 1)/(K^2 I) \rceil \\
k &= 1 + \lceil \{i - 1 - (l - 1)K^2 T\}/(KI) \rceil \\
p &= 1 + \lceil \{i - 1 - (l - 1)K^2 I - (k - 1)KI\}/(K) \rceil \\
m &= i - (l - 1)K^2 I - (k - 1)KI - (p - 1)K \\
\end{align*}

If $p = l$ and $m \neq k$ then $\tilde{Q}_{R,K}\{i, m + (k - 1)K\} = 1/(2I^{1/2})$

and $\tilde{Q}_{R,K}\{i, k + (m - 1)K\} = 1/(2I^{1/2})$

If $p = l$ and $m = k$ then $\tilde{Q}_{R,K}\{i, m + (m - 1)K\} = 1/I^{1/2}$

End For Loop
Theorem 2.7. If $E$ is an $K \times I$ matrix of independent standard normals and $E_\circ = \text{vec}(E)$, then

$$\text{Cov}\{\text{vec}(E_\circ E_\circ^\top), \text{vec}(EE^\top)\} = 2I^{1/2}\tilde{Q}_{R,K},$$

(2.14)

where $\tilde{Q}_{R,K}$ is the $R \times K^2$ matrix defined by the pseudo code above.

Proof. Begin by considering the $(i,j)$ of the desired covariance matrix. There exists indices such that the $(i,j)$ entry is equal to

$$\text{Cov}(e_m,e_{k,l},e_r,e_s),$$

where $m,k,r,s$ take values $1,\ldots,K$ and $p,l$ take values $1,\ldots,I$. Moving from $(r,s)$ to $j$ we have that

$$j = 1 + (r - 1) + (s - 1)K,$$

and the reverse is obtained using

$$s = 1 + \lfloor (j - 1)/K \rfloor$$

$$r = j - (s - 1)K.$$  

Moving from $(m,p,k,l)$ to $i$ we have that

$$i = 1 + (m - 1) + (p - 1)K + (k - 1)K I + (l - 1)K^2 I.$$
We can move back to \((m, p, k, l)\) from \(i\) using

\[
\begin{align*}
l &= 1 + \lfloor (i - 1)/(K^2 I) \rfloor \\
k &= 1 + \lfloor \{i - 1 - (l - 1)K^2 I\}/(K I) \rfloor \\
p &= 1 + \lfloor \{i - 1 - (l - 1)K^2 I - (k - 1)KI\}/(K) \rfloor \\
m &= i - (l - 1)K^2 I - (k - 1)KI - (p - 1)K.
\end{align*}
\]

We can see that the covariance will be zero if any one of \(m, k, r\) or \(s\) is distinct. Thus, the only nonzero entries will correspond to either \(m = r = k = s\), \(m = r \neq k = s\), or \(m = s \neq k = r\) when \(m = k \neq r = s\) we get zero. When all four are equal we get that

\[
\text{Cov}(e_{m,p} e_{m,l}, e_m^\top e_m) = \text{Cov}(e_{m,p} e_{m,l}, e_{m,p}^2 + e_{m,l}^2 1_{p \neq l}),
\]

which will be zero unless \(p = l\), in which case it equals

\[
\text{Cov}(e_{m,p}^2, e_{m,p}^2) = \{E(e_{m,p}^4) - E(e_{m,p}^2)^2\} = 2.
\]

We have therefore established the first if-statement in the pseudo code.

Turning to the next case, when \(m = r \neq k = s\), we have that

\[
\text{Cov}(e_{m,p} e_{k,l}, e_m^\top e_k) = \text{Cov}(e_{m,p} e_{k,l}, e_{m,p} e_{k,p} + e_{m,l} e_{k,l} 1_{l \neq p}),
\]

which is again only nonzero when \(p = l\), in which case it will be

\[
\text{Cov}(e_{m,p} e_{k,p}, e_{m,p} e_{k,p}) = E(e_{m,p}^2) E(e_{k,p}^2) = 1.
\]

An identical result will hold for when \(m = s \neq k = r\), which gives both the second
and third if-statements in the pseudo code, and the proof is established.

One example of $\tilde{Q}_{R,K}$ is

$$
\tilde{Q}_{16,2} = 
\begin{pmatrix}
2^{-1/2} & 0 & 0 & 0 \\
0 & 8^{-1/2} & 8^{-1/2} & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 8^{-1/2} & 8^{-1/2} & 0 \\
0 & 0 & 0 & 2^{-1/2} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
2^{-1/2} & 0 & 0 & 0 \\
0 & 8^{-1/2} & 8^{-1/2} & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 8^{-1/2} & 8^{-1/2} & 0 \\
0 & 0 & 0 & 2^{-1/2}
\end{pmatrix}.
$$

For the last Q matrix, we also use pseudo code which is only slightly different from the code defining $\tilde{Q}_{R,K}$.

Begin Code

Set $\tilde{Q}_{R,I}$ to be an $R = K^2 I^2$ by $I^2$ matrix of zeros

For $i = 1, \ldots, R$

$$
l = 1 + \lfloor (i - 1)/(K^2 I) \rfloor
$$

$$
k = 1 + \lfloor \{i - 1 - (l - 1)K^2 I\}/(KI) \rfloor
$$
\[ p = 1 + \lfloor \{ i - 1 - (l - 1)K^2I - (k - 1)KI \}/(K) \rfloor \]
\[ m = i - (l - 1)K^2I - (k - 1)KI - (p - 1)K \]

If \( m = k \) and \( p \neq l \) then \( \tilde{Q}_{R,I}\{i, p + (l - 1)I\} = 1/(2K^{1/2}) \)
and \( \tilde{Q}_{R,I}\{i, l + (p - 1)I\} = 1/(2K^{1/2}) \)

If \( m = k \) and \( p = l \) then \( \tilde{Q}_{R,I}\{i, l + (l - 1)I\} = 1/K^{1/2} \)

End For Loop

End Code

**Theorem 2.8.** If \( E \) is an \( K \times I \) matrix of independent standard normals and 
\( E_\circ = \text{vec}(E) \), then

\[
\text{Cov}\{\text{vec}(E_\circ E_\circ^\top), \text{vec}(E^\top E)\} = 2K^{1/2}\tilde{Q}_{R,I},
\]

(2.15)

where \( \tilde{Q}_{R,I} \) is the \( R \times I^2 \) matrix defined by the pseudo code above.

**Proof.** Begin by considering the \((i, j)\) of the desired covariance matrix. There exists indices such that the \((i, j)\) entry is equal to

\[
\text{Cov}\{e_{m,p}e_{k,l}, e_{(r)}^\top e_{(s)}\},
\]

where \( m, k \) take values \( 1, \ldots, K \) and \( p, l, r, s \) take values \( 1, \ldots, I \). Moving from \((r, s)\) to \( j \) we have that

\[ j = 1 + (r - 1) + (s - 1)I, \]

and the reverse is obtained using

\[ s = 1 + \lfloor (j - 1)/I \rfloor, \]
\[ r = j - (s - 1)I. \]
Moving from \((m, p, k, l)\) to \(i\) we have that

\[
i = 1 + (m - 1) + (p - 1)K + (k - 1)KI + (l - 1)K^2I.
\]

We can move back to \((m, p, k, l)\) from \(i\) using

\[
l = 1 + \lfloor (i - 1)/(K^2I) \rfloor,
\]

\[
k = 1 + \lfloor \{i - 1 - (l - 1)K^2I\}/(KI) \rfloor,
\]

\[
p = 1 + \lfloor \{i - 1 - (l - 1)K^2I - (k - 1)KI\}/(K) \rfloor,
\]

\[
m = i - (l - 1)I^2K - (k - 1)KI - (p - 1)K.
\]

We can see that the covariance will be zero if any one of \(p, l, r\) or \(s\) is distinct. Thus, the only nonzero entries will correspond to either \(p = r = l = s\), \(p = r \neq l = s\), or \(p = s \neq l = r\) when \(p = l \neq r = s\) we get zero. When all four are equal we get that

\[
\text{Cov}\{e_{m,p}e_{k,p}, e^{\top}_{(p)}e_{(p)}\} = \text{Cov}(e_{m,p}e_{k,p}, e^2_{m,p} + e^2_{k,p}1_{m \neq k}),
\]

which will be zero unless \(m = k\), in which case it equals \(2\).

We have therefore established the first if-statement in the pseudo code.

Turning to the next case, when \(p = r \neq l = s\), we have that

\[
\text{Cov}\{e_{m,p}e_{k,l}, e^{\top}_{(p)}e_{(l)}\} = \text{Cov}(e_{m,p}e_{k,l}, e_{m,l} + e_{k,p}e_{k,l}1_{m \neq k}),
\]

which is again only nonzero when \(m = k\), in which case it will be

\[
\text{Cov}(e_{m,p}e_{m,l}, e_{m,p}e_{m,l}) = E(e^2_{m,p}) E(e^2_{m,l}) = 1.
\]

An identical result will hold for when \(p = s \neq l = r\), which gives both the second
and third if-statements in the pseudo code, and the proof is established.

2.5.3 Proof of Theorem 2.3

We begin by establishing in Theorem 2.9 the joint null limit distribution of the vectors vec(\(\hat{U} - U\)), vec(\(\hat{V} - V\)), and vec(\(\hat{\Sigma} - \Sigma\)). We first define several matrices that appear in this distribution. Recall the Q matrices derived in Section 2.5.2: the matrix \(Q_K\) is defined in (2.12), \(Q_{K,I}\) in (2.13), \(\tilde{Q}_{R,K}\) in (2.14), and \(\tilde{Q}_{R,I}\) in (2.15).

Denote by \((\cdot)^{\dagger}\) the Moore-Penrose generalized inverse. We define the following generalized information matrices:

\[
\mathcal{I}_{U,V} = \frac{1}{2} \begin{pmatrix} U^{-1/2} \otimes U^{-1/2} & 0 \\ 0 & V^{-1/2} \otimes V^{-1/2} \end{pmatrix}
\begin{pmatrix} IQ_K & (IK)^{1/2}Q_{K,I} \\ (IK)^{1/2}Q_{I,K} & KQ_I \end{pmatrix}
\begin{pmatrix} U^{-1/2} \otimes U^{-1/2} & 0 \\ 0 & V^{-1/2} \otimes V^{-1/2} \end{pmatrix},
\]

\[
\mathcal{I}_{U,V}^c = \{D(D^\top \mathcal{I}_{U,V}D)^{\dagger}D^\top\}^+,\]

where \(D\) is an \((K^2 + I^2) \times (K^2 + I^2 - 1)\) matrix whose columns are orthonormal and are perpendicular to vec(\(I_{K^2+I^2}\)), and

\[
\mathcal{I}_{\Sigma} = \frac{1}{2} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) Q_{R^{1/2}}(\Sigma^{-1/2} \otimes \Sigma^{-1/2}),
\]

where \(R^{1/2} = KI\).

**Theorem 2.9.** Suppose Assumption 1 and decomposition 2.3 hold. Assume further
that \( \text{tr}(U) = K \). Then

\[
(N)^{1/2} \begin{bmatrix}
\text{vec}(\hat{U} - U) \\
\text{vec}(\hat{V} - V) \\
\text{vec}(\hat{\Sigma} - \Sigma)
\end{bmatrix} \to N(0, \Gamma), \quad N \to \infty
\]

The asymptotic covariance matrix \( \Gamma \) is defined as follows. The asymptotic covariance of \( \{\text{vec}(\hat{U} - U), \text{vec}(\hat{V} - V)\} \) is given by \((\mathcal{I}_{U,V}^c)^+\), of \( \text{vec}(\hat{\Sigma} - \Sigma) \) is given by \( \mathcal{I}_{\Sigma}^+ \), and the cross covariance matrix between the two is

\[
\frac{1}{2}(\mathcal{I}_{U,V}^c)^+ \begin{pmatrix}
U^{-1/2} \otimes U^{-1/2} & 0 \\
0 & V^{-1/2} \otimes V^{-1/2}
\end{pmatrix}
\begin{pmatrix}
I^{1/2} \tilde{Q}_{R,K}^\top \\
K^{1/2} \tilde{Q}_{R,I}^\top
\end{pmatrix}(\Sigma^{-1/2} \otimes \Sigma^{-1/2})\mathcal{I}_{\Sigma}^+.
\]

**Proof.** From standard theory for maximum likelihood estimators, Ferguson (1996) Chapter 18, we can use the partial derivatives of the log likelihood function, score equations, to find the Fisher information as well as asymptotic expressions for the maximum likelihood estimators. One can show that the cross terms of the Fisher information involving \( M \) and \( V,U \) and \( \Sigma \) are all zero, meaning that the estimate of the \( M \) is asymptotically independent of \( \hat{U}, \hat{V} \) and \( \hat{\Sigma} \). We therefore treat in the following \( M \) as known.

We start by working with \( U \) and \( V \). Applying the constrained likelihood methods described in Moore et al. (2008), asymptotically, \( \hat{U} \) and \( \hat{V} \) are jointly normally distributed with means \( U \) and \( V \) and covariance given by the generalized inverse of the constrained Fisher information matrix. Starting with \( U \) we have that unconstrained score equation is given by

\[
\frac{1}{N^{1/2}} \frac{\partial l(M,U,V)}{\partial U} = \frac{1}{2N^{1/2}} \sum_{n=1}^{N} \{U^{-1}(X_n - M)V^{-1}(X_n - M)^\top U^{-1} - IU^{-1}\}
\]
\[
\begin{align*}
&= \frac{1}{2N^{1/2}} \sum_{n=1}^{N} (U^{-1/2} E_n E_n^T U^{-1/2} - IU^{-1}) \\
&= \frac{1}{2N^{1/2}} \sum_{n=1}^{N} \{U^{-1/2}(E_n E_n^T - II_{K \times K})U^{-1/2}\}.
\end{align*}
\]

To get a handle on the unconstrained Fisher information matrix, and therefore the covariance matrix, it will be easier to work with the vectorized version

\[
\frac{1}{2N^{1/2}} (U^{-1/2} \otimes U^{-1/2}) \sum_{n=1}^{N} \text{vec}(E_n E_n^T - II_{K \times K}).
\]

Notice that we will have a complete handle on the above if we can understand the form for the covariance of \(\text{vec}(E_n E_n^T - II_{K \times K})\). However, this is a term that in no way depends on the underlying parameters as it is composed entirely of iid standard normals. We label \(Q_K = (2I)^{-1} \text{cov}\{\text{vec}(E_n E_n^T - II_{K \times K})\}\) and its explicit form is given in (2.12).

The part of the Fisher information matrix for \(\text{vec}(U)\) is given by

\[
\frac{I}{2} (U^{-1/2} \otimes U^{-1/2}) Q_K (U^{-1/2} \otimes U^{-1/2}).
\]

Identical arguments give that the part of the Fisher information matrix for \(\text{vec}(V)\) is given by

\[
\frac{K}{2} (V^{-1/2} \otimes V^{-1/2}) Q_I (V^{-1/2} \otimes V^{-1/2}).
\]

The joint unconstrained Fisher information matrix for \(\hat{U}\) and \(\hat{V}\) is given by

\[
\mathcal{I}_{U,V} = \frac{1}{2} \begin{pmatrix}
U^{-1/2} \otimes U^{-1/2} & 0 \\
0 & V^{-1/2} \otimes V^{-1/2}
\end{pmatrix}
\times \begin{cases}
IQ_K & (IK)^{1/2} Q_{I,K} \\
(IP)^{1/2} Q_{I,K} & IQ_I
\end{cases}
\begin{pmatrix}
U^{-1/2} \otimes U^{-1/2} & 0 \\
0 & V^{-1/2} \otimes V^{-1/2}
\end{pmatrix}.
\]
where $Q_{K,I}$ is defined in (2.13). The constrained version is then given by

$$\mathcal{I}_{U,V}^c = \{D(D^\top \mathcal{I}_{U,V} D)D^+\}^+.$$  

Recall that $D$ is an $(K^2+I^2) \times (K^2+I^2-1)$ matrix whose columns are orthonormal and are perpendicular to $\text{vec}(I_{K^2+I^2})$. The form for $D$ come from the gradient of the constraint $\text{tr}(U) = K$.

The last piece we need is the joint behavior of $\hat{U}$ or $\hat{V}$ and the estimator $\hat{\Sigma}$. The score equation for $\Sigma$ can be expressed as

$$\frac{\partial l(\mu, \Sigma)}{\partial \Sigma} = -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left\{ \sum_{n=1}^{N} (Y_n - \mu)(Y_n - \mu) \right\} \Sigma^{-1}$$

$$= -\frac{N}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1/2} \left( \sum_{n=1}^{N} E_n E_n^\top \right) \Sigma^{-1/2}$$

$$= \frac{1}{2} \Sigma^{-1/2} \left( \sum_{n=1}^{N} E_n E_n^\top - I_{KI \times KI} \right) \Sigma^{-1/2}.$$  

Using the same arguments as before, we get that Fisher information matrix for $\text{vec}(\Sigma)$ is

$$\mathcal{I}_\Sigma = \frac{1}{2} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) Q_R (\Sigma^{-1/2} \otimes \Sigma^{-1/2}).$$  

For the joint behavior, we use the following asymptotic expression for the maximum likelihood estimators, Ferguson (1996) Chapter 18,

$$N^{1/2} \begin{bmatrix} \text{vec}(\hat{U} - U) \\ \text{vec}(\hat{V} - V) \end{bmatrix} = N^{-1/2} (\mathcal{I}_{U,V}^c)^+ \begin{bmatrix} \text{vec} \left\{ \frac{\partial l(M,U,V)}{\partial U} \right\} \\ \text{vec} \left\{ \frac{\partial l(M,U,V)}{\partial V} \right\} \end{bmatrix} + o_P(1),$$  

$$N^{1/2} \begin{bmatrix} \text{vec}(\hat{U} - U) \\ \text{vec}(\hat{V} - V) \end{bmatrix} = N^{-1/2} (\mathcal{I}_{U,V}^c)^+ \begin{bmatrix} \text{vec} \left\{ \frac{\partial l(M,U,V)}{\partial U} \right\} \\ \text{vec} \left\{ \frac{\partial l(M,U,V)}{\partial V} \right\} \end{bmatrix} + o_P(1).$$
and

\[ N^{1/2} \text{vec}(\hat{\Sigma} - \Sigma) = N^{-1/2} I^+_\Sigma \frac{\partial l(\mu, \Sigma)}{\partial \Sigma} + o_P(1). \]

For the covariance between \( \hat{\Sigma} \) and \( \hat{U} \) or \( \hat{V} \) we obtain two more matrices, called \( \tilde{Q}_{R,K} \) and \( \hat{Q}_{R,I} \) which satisfy

\[
\text{Cov} \left\{ \text{vec}(E_n E_n^\top - I_{R \times R}), \text{vec}(E_n E_n^\top - I_{K \times K}) \right\} = 2I^{1/2} \tilde{Q}_{R,K},
\]

\[
\text{Cov} \left\{ \text{vec}(E_n E_n^\top - I_{R \times R}), \text{vec}(E_n^\top E_n - K I_{I \times I}) \right\} = 2K^{1/2} \hat{Q}_{R,I}.
\]

Recall that the diamond subscript indicates vectorization and the definitions of \( \tilde{Q}_{R,K} \) and \( \hat{Q}_{R,I} \) can be found in (2.14) and (2.15), respectively. The cross-covariance matrix for \( N^{1/2} \text{vec}(\hat{\Sigma} - \Sigma) \) and \( N^{1/2} \text{vec}(\hat{U} - U) \) is then given by

\[
(I^c_{U,V})^+ \text{cov} \left[ \begin{bmatrix} \text{vec} \left( N^{-1/2} \frac{\partial l(M,U,V)}{\partial U} \right) \\ \text{vec} \left( N^{-1/2} \frac{\partial l(M,U,V)}{\partial V} \right) \end{bmatrix}, \text{vec} \left( N^{-1/2} \frac{\partial l(M,\Sigma)}{\partial \Sigma} \right) \right] \Sigma^+_\Sigma
\]

\[
= \frac{1}{2} (I^c_{U,V})^+ \begin{pmatrix} U^{-1/2} \otimes U^{-1/2} & 0 \\ 0 & V^{-1/2} \otimes V^{-1/2} \end{pmatrix} \begin{pmatrix} I^{1/2} \tilde{Q}_{R,K}^\top \\ K^{1/2} \hat{Q}_{R,I}^\top \end{pmatrix} (\Sigma^{-1/2} \otimes \Sigma^{-1/2}) \Sigma^+_\Sigma
\]

\( \square \)

**Proof of Theorem 2.3:** Since we have the joint asymptotic distribution for \( \hat{U}, \hat{V}, \) and \( \hat{\Sigma} \), we can use the delta method to find the asymptotic distributions of desired test statistics, and in particular, we can find the form of \( W \), the asymptotic covariance matrix of \( \text{vec}(\hat{V} \otimes \hat{U}) - \text{vec}(\hat{\Sigma}) \). To apply the delta method, we need the partial derivatives. Taking the derivative with respect to \( V_{i,j} \) yields

\[
\text{vec}(1_{i,j} \otimes U)
\]
and with respect to $U_{k,l}$

$$\text{vec}(V \otimes 1_{k,l}).$$

So the matrix of partials with respect to vec($V$) is

$$G_V = \begin{cases} 
\text{vec}(1_{1,1} \otimes U)^\top \\
\text{vec}(1_{2,1} \otimes U)^\top \\
\vdots \\
\text{vec}(1_{I,I} \otimes U)^\top 
\end{cases},$$

with respect to vec($U$) is

$$G_U = \begin{cases} 
\text{vec}(V \otimes 1_{1,1})^\top \\
\text{vec}(V \otimes 1_{2,1})^\top \\
\vdots \\
\text{vec}(V \otimes 1_{K,K})^\top 
\end{cases},$$

and with respect to vec($\Sigma$) is just $(-1)$ times the $KI \times KI$ identity matrix. We therefore have that

$$\text{vec}(\hat{V} \otimes \hat{U}) - \text{vec}(\hat{\Sigma}) \approx \begin{pmatrix} 
G_{\hat{U}} \\
G_{\hat{V}} \\
-I_{KI \times KI}
\end{pmatrix}^\top \begin{pmatrix} 
\text{vec}(\hat{U}) \\
\text{vec}(\hat{V}) \\
\text{vec}(\hat{\Sigma})
\end{pmatrix}.$$ 

This implies that

$$W = \begin{pmatrix} 
G_U \\
G_V \\
-I_{KI \times KI}
\end{pmatrix}^\top \Gamma \begin{pmatrix} 
G_U \\
G_V \\
-I_{KI \times KI}
\end{pmatrix}. \quad (2.16)$$

The degrees of freedom are obtained by noticing that under the alternative $\Sigma$
has \( KI(KI + 1)/2 \) free parameters, while under the null there \( K(K+1)/2 + I(I + 1)/2 - 1 \), where the last \(-1\) is included because we have one constraint, \( \text{tr}(U) = K \).

### 2.5.4 Proof of Theorem 2.4

We begin by restating Assumption 2 in Section 2.3.2. We assume that the processes have mean zero for simplicity.

**Assumption 2.** Assume that \( X_1(s, t), \ldots, X_N(s, t) \) are independent and identically distributed Gaussian processes in \( L^2(S \times T) \) with mean zero and a separable covariance function \( \sigma(s, t, s', t') = U(s, s')V(t, t') \). Assume that the temporal functions \( X_i(s, \cdot) \) are observed at spatial locations \( \{s_k : k = 1, \ldots, K\} \). Let \( U \) be the matrix resulting in evaluating \( U \) at the \( \{s_k\} \). Assume, without loss of generality, that \( \text{tr}(U) = K \).

The following corollary essentially follows from Isserlis’ theorem applied to our stochastic processes.

**Corollary 2.10.** Under Assumption 2, we have

\[
\text{Cov}\{X_n(s_k, t)X_n(s_k, t'), X_n(s_i, t)X_n(s_i, t')\} = \\
E\{X_n(s_k, t)X_n(s_i, t)\}E\{X_n(s_k, t')X_n(s_i, t')\} \\
+ E\{X_n(s_k, t)X_n(s_i, t')\}E\{X_n(s_k, t')X_n(s_i, t)\},
\]

and

\[
\text{Cov}\{\xi_{jn}(s_k)\xi_{jn}(s_t), \xi_{in}(s_k)\xi_{in}(s_t)\} = E\{\xi_{jn}(s_k)\xi_{in}(s_k)\}E\{\xi_{jn}(s_t)\xi_{in}(s_t)\} \\
+ E\{\xi_{jn}(s_k)\xi_{in}(s_t)\}E\{\xi_{jn}(s_t)\xi_{in}(s_k)\}.
\]

The key to establishing the equivalence of procedures based on population and
estimated principal components is the derivation of $O_P(N^{-1})$ bounds for differences between the population covariance operator $\mathcal{V}$ and the operator $\hat{\mathcal{V}}$ defined in the procedure of Section 2.3, and an analogous bound for the difference between $U$ and $\hat{U}$. To emphasize the different dimensionalities, we refer to the Hilbert-Schmidt norm of a matrix as a Frobenius norm, and use the subscript $\text{F}$ to indicate it. The Hilbert-Schmidt norm of operators is indicated with $\text{S}$.

**Theorem 2.11.** Under Assumption 2,

$$E \left( \| \hat{\mathcal{V}} - \mathcal{V} \|_S^2 \right) \leq \frac{2}{NK^2} \| U \|_\text{F}^2 \| \mathcal{V}^{1/2} \|_\text{S}^4$$

where $\| \cdot \|_\text{F}$ is the Frobenius norm and $\| \cdot \|_\text{S}$ is the Hilbert-Schmidt norm.

**Proof.** The expected value of the estimator $\hat{\mathcal{V}}(t, t')$ is given by

$$E\{ \hat{\mathcal{V}}(t, t') \} = E \left\{ \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} X_n(s_k, t)X_n(s_k, t') \right\}$$

$$= \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} E\{ X_n(s_k, t)X_n(s_k, t') \}$$

$$= \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} U(k, k)\mathcal{V}(t, t')$$

$$= \mathcal{V}(t, t'),$$

since $\sum_{k=1}^{K} U(k, k) = \text{tr}(U) = K$, which means the estimator $\hat{\mathcal{V}}(t, t')$ is unbiased. Next observe that

$$E \left( \| \hat{\mathcal{V}} - \mathcal{V} \|_S^2 \right) = E \left[ \int_t \int_{t'} \{ \hat{\mathcal{V}}(t, t') - \mathcal{V}(t, t') \}^2 dt dt' \right]$$

$$= \int_t \int_{t'} E\{ \hat{\mathcal{V}}(t, t') - \mathcal{V}(t, t') \}^2 dt dt'.$$

Since the estimator $\hat{\mathcal{V}}(t, t')$ is unbiased, $E\{ \hat{\mathcal{V}}(t, t') - \mathcal{V}(t, t') \}^2 = \text{Var}\{ \hat{\mathcal{V}}(t, t') \}$. Ap-
plying Corollary 2.10, we obtain

\[
\text{Var}\{\hat{V}(t, t')\} = \text{Cov}\{\hat{V}(t, t'), \hat{V}(t, t')\} \\
= \text{Cov}\left\{ \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} X_n(s_k, t)X_n(s_k, t'), \frac{1}{NK} \sum_{m=1}^{N} \sum_{i=1}^{K} X_m(s_i, t)X_m(s_i, t') \right\} \\
= \frac{1}{NK^2} \sum_{k=1}^{K} \sum_{i=1}^{K} \text{Cov}\{X_n(s_k, t)X_n(s_k, t'), X_n(s_i, t)X_n(s_i, t')\} \\
\overset{\text{Corollary 2.10}}{=} \frac{1}{NK^2} \sum_{k=1}^{K} \sum_{i=1}^{K} \left[ E\{X_n(s_k, t)X_n(s_i, t)\} E\{X_n(s_k, t')X_n(s_i, t')\} \right. \\
\left. + E\{X_n(s_k, t)X_n(s_i, t')\} E\{X_n(s_k, t')X_n(s_i, t)\}\right] \\
= \frac{1}{NK^2} \sum_{k=1}^{K} \sum_{i=1}^{K} U(k, i)\mathcal{V}(t, t)U(k, i)\mathcal{V}(t', t') + U(k, i)\mathcal{V}(t, t')U(k, i)\mathcal{V}(t', t)\right] \\
= \frac{1}{NK^2} \sum_{k=1}^{K} \sum_{i=1}^{K} [\mathcal{V}(t, t)\mathcal{V}(t', t') + \mathcal{V}^2(t, t')] \\
= \frac{1}{NK^2} \|U\|_F^2 \{\mathcal{V}(t, t)\mathcal{V}(t', t') + \mathcal{V}^2(t, t')\}.
\]

Applying the Cauchy-Schwarz inequality we conclude that

\[
E\left(\|\hat{V} - \mathcal{V}\|_S^2\right) = \frac{1}{NK^2} \|U\|_F^2 \int_t \int_{t'} \{\mathcal{V}(t, t)\mathcal{V}(t', t') + \mathcal{V}^2(t, t')\} dt dt' \\
= \frac{1}{NK^2} \|U\|_F^2 \left\{ \int_t \mathcal{V}(t, t) dt \int_{t'}\mathcal{V}(t', t') dt' + \int_t \int_{t'} \mathcal{V}^2(t, t') dt dt' \right\} \leq c.s. \frac{2}{NK^2} \|U\|_F^2 \left\{ \int_t \mathcal{V}(t, t) dt \int_{t'}\mathcal{V}(t', t') dt' \right\} \\
= \frac{2}{NK^2} \|U\|_F^2 \text{tr}^2(\mathcal{V}) = \frac{2}{NK^2} \|U\|_F^2 \|\mathcal{V}^{1/2}\|_S^4
\]

as desired. \(\square\)
As corollaries, we have that the estimated eigenvalues and eigenfunctions can also be consistently estimated. The proofs are directly based on well-known inequalities from operator theory, e.g., Horváth and Kokoszka (2012) Chapter 2.

**Corollary 2.12.** Let \( \hat{v}_j, v_j \) be the eigenfunctions of \( \hat{V}, V \) respectively. Then

\[
E \left( \| \hat{v}_j - v_j \|_2 \right) \leq \frac{2(2)^{1/2}}{\alpha_j} E \left( \| \hat{V} - V \|_S^2 \right) \leq \frac{4(2)^{1/2}}{NK^2 \alpha_j} \| U \|_F^2 \| V \|^{1/2}_S^4
\]

where \( \alpha_1 = (\lambda_1 - \lambda_2) \), \( \alpha_j = \min\{ (\lambda_j - \lambda_{j+1}), (\lambda_{j-1} - \lambda_j) \} \) for \( j \geq 2 \).

**Corollary 2.13.** Let \( \hat{\lambda}_j, \lambda_j \) be the eigenvalues of \( \hat{V}, V \) respectively. Then

\[
E \left( \| \hat{\lambda}_j - \lambda_j \|_2 \right) \leq E \left( \| \hat{V} - V \|_S^2 \right) \leq \frac{2}{NK^2} \| U \|_F^2 \| V \|^{1/2}_S^4.
\]

**Theorem 2.14.** Under Assumption 2, \( \| \hat{U} - U \|_F^2 = O_P(N^{-1}) \).

**Proof.** Recall that

\[
\hat{U}(k, k') = \frac{K \sum_n \int X_n(s_k, t) X_n(s_{k'}, t) \ dt}{N \text{tr}(\hat{\sigma})}.
\]

Notice that

\[
\text{tr}(\hat{U}) = \frac{K \sum_k \sum_k \int X_n(s_k, t) X_n(s_{k'}, t) \ dt}{N \text{tr}(\hat{\sigma})} = \frac{K \text{tr}(\hat{\sigma})}{\text{tr}(\hat{\sigma})} = K,
\]

which explains the chosen normalization. Define the intermediate term

\[
\tilde{U}(k, k') = \frac{K \sum_n \int X_n(s_k, t) X_n(s_{k'}, t) \ dt}{N \text{tr}(\sigma)},
\]

based on the true trace of \( \sigma \).
To establish the desired result, we will first separate it into two terms

\[ \|\hat{U} - U\|_F \leq \|\hat{U} - \tilde{U}\|_F + \|\tilde{U} - U\|_F. \]

The square of the first term is given by

\[
|\hat{U}(k, k') - \tilde{U}(k, k')|^2 \\
= \left| \frac{K \sum_n \int X_n(s_k, t)X_n(s_{k'}, t) \, dt}{N \text{tr}(\hat{\sigma})} - \frac{K \sum_n \int X_n(s_k, t)X_n(s_{k'}, t) \, dt}{N \text{tr}(\sigma)} \right|^2 \\
= \frac{K}{N} \sum_n \int \left| X_n(s_k, t)X_n(s_{k'}, t) \right|^2 \, dt \left| \left\{ \frac{1}{\text{tr}(\hat{\sigma})} - \frac{1}{\text{tr}(\sigma)} \right\} \right|^2 \\
\leq \frac{K^2}{N^2} \sum_n \left( \sum_n \left\| X_n(s_k) \right\|_S^2 \right) \left( \sum_n \left\| X_n(s_{k'}) \right\|_S^2 \right) \left| \left\{ \frac{1}{\text{tr}(\hat{\sigma})} - \frac{1}{\text{tr}(\sigma)} \right\} \right|^2 \\
= O_p(K^2 N^{-1/2}),
\]

since \( \hat{\sigma} \) is \( N^{1/2} \) consistent and therefore by using delta method with the function \( f(\sigma) = 1/\text{tr}(\sigma) \) we have that \( 1/\text{tr}(\hat{\sigma}) \) is also \( N^{1/2} \) consistent.

Turning to the second term, the expected value of the \( \tilde{U}(k, k') \) is given by

\[
E\{\tilde{U}(k, k')\} = E \left\{ \frac{K \sum_n \int X_n(s_k, t)X_n(s_{k'}, t) \, dt}{N \text{tr}(\sigma)} \right\} \\
= \frac{K}{N \text{tr}(\sigma)} \sum_n \int E \{X_n(s_k, t)X_n(s_{k'}, t)\} \, dt \\
= \frac{K}{N \text{tr}(\sigma)} \sum_n \int \mathcal{V}(t, t) \, dt \ U(k, k') \\
= \frac{K}{\text{tr}(\sigma)} U(k, k') \int \mathcal{V}(t, t) \, dt
\]
\[
\begin{align*}
&= \frac{K}{\text{tr}(\sigma)} U(k, k') \text{tr}(\mathcal{V}) \\
&= U(k, k'),
\end{align*}
\]

since \(\text{tr}(\sigma) = \text{tr}(\mathcal{V}) \text{tr}(U)\) and \(\text{tr}(U) = K\), which means the estimator \(\tilde{U}(k, k')\) is unbiased. Comparing \(\tilde{U}\) to \(U\) we have

\[
E\left(\|\tilde{U} - U\|^2\right) = E\left[\sum_{k=1}^{K} \sum_{k'=1}^{K} (\tilde{U}(k, k') - U(k, k'))^2\right] = \sum_{k=1}^{K} \sum_{k'=1}^{K} E\{\tilde{U}(k, k') - U(k, k')\}^2.
\]

Since the estimator \(\tilde{U}(k, k')\) is unbiased we have

\[
E\{\tilde{U}(k, k') - U(k, k')\}^2 = \text{Var}\{\tilde{U}(k, k')\},
\]

and so the difference is given by

\[
\text{Var}\{\tilde{U}(k, k')\} = \text{Cov}\{\tilde{U}(k, k'), \tilde{U}(k, k')\}
\]

\[
= \text{Cov}\left\{\frac{K}{N \text{tr}(\sigma)} \int_t^t X_n(s, t) X_n(s', t') \ dt, \frac{K}{N \text{tr}(\sigma)} \int_t^t X_m(s, t'') X_m(s', t'') \ dt''\right\}
\]

\[
= \frac{K^2}{N \text{tr}^2(\sigma)} \int_t^t \int_t^t \text{Cov}\left\{X_n(s, t) X_n(s', t), X_n(s, t') X_n(s', t')\right\} dt \ dt'
\]

Corollary 2.10

\[
= \frac{K^2}{N \text{tr}^2(\sigma)} \int_t^t \int_t^t \left[ E\{X_n(s, t) X_n(s, t')\} E\{X_n(s', t) X_n(s', t')\} dt \ dt' \\
+ E\{X_n(s, t) X_n(s', t')\} E\{X_n(s', t) X_n(s, t')\} dt \ dt' \\
= \frac{K^2}{N \text{tr}^2(\sigma)} \int_t^t \int_t^t \left\{U(k, k') \mathcal{V}(t, t') U(k', k') \mathcal{V}(t, t') \\
+ U(k, k') \mathcal{V}(t, t') U(k, k') \mathcal{V}(t, t')\right\} dt \ dt'
\]
\[
\begin{align*}
E\left(\|\tilde{U} - U\|^2\right)
&= \frac{K^2}{N \text{tr}^2(\sigma)} \sum_{k=1}^{K} \sum_{k'=1}^{K} \{U(k, k)U(k', k') + U^2(k, k')\} \int_{t}^{t'} \mathcal{V}^2(t, t') dt \, dt' \\& \leq \frac{2K^2}{N \text{tr}^2(\sigma)} \sum_{k=1}^{K} \sum_{k'=1}^{K} U(k, k)U(k', k') \int_{t}^{t'} \mathcal{V}(t, t') dt \int_{t'}^{t''} \mathcal{V}(t', t'') dt'' \\
&= \frac{2K^2}{N \text{tr}^2(\sigma)} \text{tr}^2(U) \text{tr}^2(\mathcal{V}) \\
&= O_p(K^2 N^{-1})
\end{align*}
\]

since \(\text{tr}(\sigma) = \text{tr}(U) \text{tr}(\mathcal{V}) = K \text{tr}(\mathcal{V})\).

Using Theorems 2.11 and 2.14, one can show that replacing the \(u_k\) and the \(v_j\) by their estimates \(\hat{u}_k\) and \(\hat{v}_j\) has an asymptotically negligible effect. We provide the argument in the case of statistic \(\hat{T}_F\), which generally leads to a test that performs better than the other tests. We also assume that \(T_F\) is computed without iterating using the following estimates

\[
\begin{align*}
\hat{\mathcal{V}}(t, t') &= \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} X_n(s_k, t) X_n(s_k, t') , \\
\hat{U}(k, k') &= \frac{1}{N \text{tr}(\hat{\sigma})} \sum_{n=1}^{N} \int X_n(s_k, t) X_n(s_{k'}, t) \, dt.
\end{align*}
\]

This simplifies the asymptotic arguments greatly as the test statistics will depend
linearly on eigenfunctions. In this case, the test statistic is given by

$$T_F = \sum_{k_1} \sum_{j_1} \sum_{k_2} \sum_{j_2} \langle \hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\Sigma}, u_{k_1} \otimes v_{j_1} \otimes u_{k_1} \otimes v_{j_1} \rangle^2.$$ 

**Theorem 2.4.** Denote by $\hat{T}_F$ the statistic computed using the estimators $\hat{u}_k$ and $\hat{v}_j$ and by $T_F^\dagger$ the random variable computed using the population functions $u_k$ and $v_j$. If Assumption 2 holds, then $\hat{T}_F - T_F^\dagger = O_P(N^{-1/2})$.

**Proof.** To lighten the notation, we drop the superscript $F$. First notice that

$$\hat{T} = \sum_{k_1} \sum_{j_1} \sum_{k_2} \sum_{j_2} N \langle \hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\Sigma}, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle^2,$$

with an analogous formula for $T_F^\dagger$. Let $T^* = \hat{\mathbf{V}} \otimes \hat{\mathbf{U}} - \hat{\Sigma}$. By using the triangle inequality we have that

$$\left| \sum_{k_1} \sum_{j_1} \sum_{k_2} \sum_{j_2} N \langle T^*, u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} \rangle^2 - \sum_{k_1} \sum_{j_1} \sum_{k_2} \sum_{j_2} N \langle T^*, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle^2 \right| \leq N \sum_{k_1} \sum_{j_1} \sum_{k_2} \sum_{j_2} \left| \langle T^*, u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} \rangle^2 - \langle T^*, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle^2 \right|.$$

By using the formula $a^2 - b^2 = (a - b)(a + b)$ and the linearity of the inner product we obtain

$$\left| \langle T^*, u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} \rangle^2 - \langle T^*, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle^2 \right| = \left| \langle T^*, u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle \right| \times \left| \langle T^*, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle \right|.$$
Applying the Cauchy-Schwarz inequality, the above is bounded by

\[
\left\| T^* \right\|^2 \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\times \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} + \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\leq 2 \left\| T^* \right\|^2 \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \|.
\]

We have that

\[
\| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
= \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} \\
+ \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\leq \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} \| \\
+ \| \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
= \| u_{k_1} - \hat{u}_{k_1} \| + \| \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
= \| u_{k_1} - \hat{u}_{k_1} \| + \| \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} \\
+ \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\leq \| u_{k_1} - \hat{u}_{k_1} \| + \| \hat{u}_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} \| \\
+ \| \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
= \| u_{k_1} - \hat{u}_{k_1} \| + \| v_{j_1} - \hat{v}_{j_1} \| + \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
= \| u_{k_1} - \hat{u}_{k_1} \| + \| v_{j_1} - \hat{v}_{j_1} \| \\
+ \| \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} \| + \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\leq \| u_{k_1} - \hat{u}_{k_1} \| + \| v_{j_1} - \hat{v}_{j_1} \| + \| \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes u_{k_2} \otimes v_{j_2} \| + \| u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} - \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \| \\
\leq \| u_{k_1} - \hat{u}_{k_1} \| + \| v_{j_1} - \hat{v}_{j_1} \| + \| u_{k_2} - \hat{u}_{k_2} \| + \| v_{j_2} - \hat{v}_{j_2} \|
\[ \leq \frac{2(2)^{1/2}}{\beta_{k_1}} \| \hat{U} - U \| + \frac{2(2)^{1/2}}{\alpha_{j_1}} \| \hat{V} - V \| + \frac{2(2)^{1/2}}{\beta_{k_2}} \| \hat{U} - U \| + \frac{2(2)^{1/2}}{\alpha_{j_2}} \| \hat{V} - V \| , \]

where \( \alpha_1 = (\lambda_1 - \lambda_2) \), \( \alpha_{j_1} = \min\{ (\lambda_{j_1} - \lambda_{j_1+1}), (\lambda_{j_1-1} - \lambda_{j_1}) \} \) for \( j_1 \geq 2 \), \( \beta_1 = (\mu_1 - \mu_2) \) and \( \beta_{k_1} = \min\{ (\mu_{k_1} - \mu_{k_1+1}), (\mu_{k_1-1} - \mu_{k_1}) \} \) for \( k_1 \geq 2 \). Here \( \lambda_1, \ldots, \lambda_J \) are the eigenvalues of \( V \) and \( \mu_1, \ldots, \mu_K \) are the eigenvalues of \( U \).

To sum up we have

\[
\left| \sum_{k_1}^{K} \sum_{j_1}^{J} \sum_{k_2}^{K} \sum_{j_2}^{J} N \langle T^*, u_{k_1} \otimes v_{j_1} \otimes u_{k_2} \otimes v_{j_2} \rangle^2 - \sum_{k_1}^{K} \sum_{j_1}^{J} \sum_{k_2}^{K} \sum_{j_2}^{J} N \langle \hat{T}^*, \hat{u}_{k_1} \otimes \hat{v}_{j_1} \otimes \hat{u}_{k_2} \otimes \hat{v}_{j_2} \rangle^2 \right|
\leq N \sum_{k_1}^{K} \sum_{j_1}^{J} \sum_{k_2}^{K} \sum_{j_2}^{J} 2 \| T^* \|^2 \times \left( \frac{2(2)^{1/2}}{\beta_{k_1}} \| \hat{U} - U \| + \frac{2(2)^{1/2}}{\alpha_{j_1}} \| \hat{V} - V \| + \frac{2(2)^{1/2}}{\beta_{k_2}} \| \hat{U} - U \| + \frac{2(2)^{1/2}}{\alpha_{j_2}} \| \hat{V} - V \| \right) \]
\[= \|(N)^{1/2}T^*\|^28(2)^{1/2}JK \left( J \sum_{k_1}^{K} \frac{1}{\beta_{k_1}} \| \hat{U} - U \| + K \sum_{j_1}^{J} \frac{1}{\alpha_{j_1}} \| \hat{V} - V \| \right) \]
\[= O_P(N^{-1/2}), \]

and the claim holds. \( \square \)
2.5.5 Additional empirical rejection rates, more discussion in Irish wind data and application to pollution data

2.5.5.1 Additional simulations

We now study the effect of increasing the number of spatial locations and the number of spatial and temporal principal components. We will use the first spatiotemporal covariance function introduced by Gneiting (2002). We use $I = 100$ time points equally spaced on $[0, 1]$ and $K = 16, 25$ equally spaced points in $[0, 1] \times [0, 1]$. We display the results, based on 1000 replications, for two extreme values of the space-time interaction parameter, $\beta = 0$ and $\beta = 1$. Table 2.6 shows that increasing the number of spatial points and the number of principal components does not change our previous conclusions, i.e., only the tests $T_{L-MC}$ and $T_F$ are robust to the number of the principal components used with the $T_F$ having more power.

We now explore the behavior of the method which involves dimension reduction in time only. Details are presented in Section 2.5.1. We continue to work with the covariance function of Gneiting (2002). Table 2.7 shows that only the tests $T_F$ and $T_{L-MC}$ are robust to the selection of $J$, but in some settings, the empirical size of this method is not as well calibrated as that of the general method of Section 2.3. Test $T_{L-MC}$ tends to be too conservative, with a corresponding loss of power. If $K = 9$, and $J \geq 3$, test $T_F$ overrejects. It appears that the cut off point for spatial dimension reductions is about $K = 10$.

Finally in Figure 2.1 we can see the QQ-plots of the p-values of the three test statistics introduced in this chapter, $T_L, T_F, T_W$ respectively. Based on the QQ-plots we don’t have a clear evidence that the approximate null distributions of the test statistics are the one claimed. Therefore, we apply Kolmogorov-Smirnov tests and obtain p-values 0.5361, 0.8593 and 0.0971 for $T_L, T_F, T_W$ respectively. Based on the p-values we can conclude that the approximate null distributions are the
<table>
<thead>
<tr>
<th>$\beta$</th>
<th>K</th>
<th>L</th>
<th>J</th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_R$</th>
<th>$T_W$</th>
<th>CPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>5.0 (0.69)</td>
<td>5.8 (0.74)</td>
<td>5.8 (0.74)</td>
<td>4.6 (0.66)</td>
<td>90</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>3</td>
<td>3</td>
<td>4.4 (0.65)</td>
<td>6.7 (0.79)</td>
<td>6.0 (0.75)</td>
<td>38.7 (1.54)</td>
<td>93</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>3.9 (0.61)</td>
<td>13.7 (1.08)</td>
<td>6.2 (0.76)</td>
<td>98.2 (0.42)</td>
<td>95</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>6.2 (0.76)</td>
<td>33.7 (1.49)</td>
<td>7.1 (0.81)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>3.3 (0.56)</td>
<td>84.2 (1.15)</td>
<td>4.3 (0.64)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
<td>7</td>
<td>7</td>
<td>5.7 (0.73)</td>
<td>100 (0)</td>
<td>5.0 (0.69)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>2</td>
<td>2</td>
<td>54.0 (1.58)</td>
<td>57.4 (1.56)</td>
<td>67.1 (1.49)</td>
<td>32.0 (1.48)</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>3</td>
<td>3</td>
<td>50.0 (1.58)</td>
<td>56.7 (1.57)</td>
<td>89.7 (0.96)</td>
<td>94.6 (0.71)</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>71.5 (1.43)</td>
<td>87.4 (1.05)</td>
<td>99.6 (0.20)</td>
<td>100 (0)</td>
<td>95</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>5</td>
<td>5</td>
<td>80.5 (1.25)</td>
<td>98.1 (0.43)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>92.5 (0.83)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>4.2 (0.63)</td>
<td>4.9 (0.68)</td>
<td>4.2 (0.63)</td>
<td>3.0 (0.54)</td>
<td>89</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>3</td>
<td>3</td>
<td>4.5 (0.66)</td>
<td>6.9 (0.80)</td>
<td>4.9 (0.68)</td>
<td>38.0 (1.53)</td>
<td>93</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>3.8 (0.60)</td>
<td>12.4 (1.04)</td>
<td>5.4 (0.71)</td>
<td>98.4 (0.40)</td>
<td>95</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>5</td>
<td>5</td>
<td>5.0 (0.69)</td>
<td>31.4 (1.47)</td>
<td>5.7 (0.73)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>6</td>
<td>6</td>
<td>2.5 (0.49)</td>
<td>82.8 (1.19)</td>
<td>4.2 (0.63)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>7</td>
<td>7</td>
<td>5.5 (0.72)</td>
<td>99.9 (0.10)</td>
<td>4.8 (0.68)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>52.1 (1.58)</td>
<td>56.0 (1.57)</td>
<td>65.5 (1.50)</td>
<td>25.2 (1.37)</td>
<td>87</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>3</td>
<td>3</td>
<td>48.0 (1.58)</td>
<td>56.1 (1.57)</td>
<td>89.6 (0.97)</td>
<td>94.8 (0.70)</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>4</td>
<td>4</td>
<td>65.9 (1.50)</td>
<td>84.1 (1.16)</td>
<td>99.8 (0.14)</td>
<td>99.9 (0.10)</td>
<td>94</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>5</td>
<td>5</td>
<td>76.9 (1.33)</td>
<td>97.0 (0.54)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>6</td>
<td>6</td>
<td>90.8 (0.91)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>96</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>7</td>
<td>7</td>
<td>94.4 (0.73)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>97</td>
</tr>
</tbody>
</table>

Table 2.6: Rejection rates (%), based on 1000 replications, for $N = 100$, $L, J \in \{2, 3, 4, 5, 6, 7\}$, $\beta = 0$ ($H_0$) and $\beta = 1$ ($H_A$). $L$ is the number of spatial principal components used in the dimension reduction and $J$ the number of temporal principal components. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

one claimed. To construct the QQ-plots we use the first spatiotemporal covariance function introduced by Gneiting (2002), $I = 100$ time points equally spaced on $[0, 1]$, $K = 11$ space points in $[0, 1] \times [0, 1]$, $\beta = 0$, $N = 100$ and $L = J = 2$. The results are based on 1000 replications.
<table>
<thead>
<tr>
<th>$\beta$</th>
<th>K</th>
<th>J</th>
<th>$T_{L-MC}$</th>
<th>$T_L$</th>
<th>$T_F$</th>
<th>$T_W$</th>
<th>CPV(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>2</td>
<td>4.2 (0.63)</td>
<td>7.1 (0.81)</td>
<td>5.6 (0.73)</td>
<td>5.4 (0.71)</td>
<td>89</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>3</td>
<td>5.3 (0.71)</td>
<td>9.3 (0.92)</td>
<td>5.5 (0.72)</td>
<td>85.4 (1.12)</td>
<td>93</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3.2 (0.56)</td>
<td>12.2 (1.03)</td>
<td>5.8 (0.74)</td>
<td>99.9 (0.10)</td>
<td>95</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
<td>71.5 (1.43)</td>
<td>81.7 (1.22)</td>
<td>91.5 (0.88)</td>
<td>75.7 (1.36)</td>
<td>89</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>86.6 (1.08)</td>
<td>92.1 (0.85)</td>
<td>98.5 (0.38)</td>
<td>80.7 (1.25)</td>
<td>93</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>79.8 (1.27)</td>
<td>92.4 (0.84)</td>
<td>99.6 (0.20)</td>
<td>100 (0)</td>
<td>95</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>2</td>
<td>1.5 (0.38)</td>
<td>8.8 (0.90)</td>
<td>4.4 (0.65)</td>
<td>65.0 (1.51)</td>
<td>88</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>3</td>
<td>4.3 (0.64)</td>
<td>17.1 (1.19)</td>
<td>4.8 (0.68)</td>
<td>100 (0)</td>
<td>93</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>4</td>
<td>3.8 (0.60)</td>
<td>29.5 (1.44)</td>
<td>5.1 (0.70)</td>
<td>100 (0)</td>
<td>94</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>2</td>
<td>79.5 (1.28)</td>
<td>95.8 (0.63)</td>
<td>98.4 (0.40)</td>
<td>100 (0)</td>
<td>88</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3</td>
<td>96.5 (0.58)</td>
<td>99.7 (0.17)</td>
<td>99.9 (0.10)</td>
<td>100 (0)</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
<td>97.0 (0.54)</td>
<td>99.9 (0.10)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>94</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>2</td>
<td>4.5 (0.66)</td>
<td>17.6 (1.20)</td>
<td>5.2 (0.70)</td>
<td>100 (0)</td>
<td>88</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>3</td>
<td>2.1 (0.45)</td>
<td>43.1 (1.57)</td>
<td>6.6 (0.79)</td>
<td>100 (0)</td>
<td>92</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>4</td>
<td>3.7 (0.60)</td>
<td>84.6 (1.14)</td>
<td>7.2 (0.82)</td>
<td>100 (0)</td>
<td>94</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
<td>96.4 (0.59)</td>
<td>99.1 (0.30)</td>
<td>99.4 (0.24)</td>
<td>100 (0)</td>
<td>90</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>3</td>
<td>99.6 (0.20)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>92</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
<td>99.3 (0.26)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>100 (0)</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 2.7: Rejection rates (%), based on 1000 replications, for the method of Section A, temporal dimension reduction only; $N = 100$, $J \in \{2, 3, 4\}$, $\beta = 0$ and $\beta = 1$. $J$ is the number of temporal principal components. The proportion of variance explained (CPV) by the temporal principal components is given in the last column. The standard errors are given in the parentheses.

### 2.5.5.2 Irish wind data

In addition to the p-values obtained by the procedures describe in Section 2.4.2 we applied our tests to the deseasonalized square root transformed wind speed data and to the deseasonalized way of Haslett and Raftery (1989). Haslett and Raftery (1989) estimated the seasonal effect by calculating the average of the square roots of the daily means over all years and stations for each day of the year, and then regressing the results on a set of annual harmonics. Subtraction of the estimated seasonal effect from the square roots of the daily means yields the deseasonalized data, referred to as velocity measures. Both these procedures follow the same
Figure 2.1: QQ-plots of the p-values of the test statistics $T_L, T_F, T_W$, from left to right.

pattern as in Table 5.

Figure 2.2: Irish wind speed curves for January 1961. Each curve corresponds to a different location. The curves are offset since they overlap.
### 2.5.5.3 Air quality data

We conclude this section by considering air quality data from the east coast of the United States. The Environmental Protection Agency, EPA, collects massive amounts of air quality data which are available through its website [http://www3.epa.gov/airdata/ad_data_daily.html](http://www3.epa.gov/airdata/ad_data_daily.html). The records consist of data for 6 common pollutants, collected by outdoor monitors in hundreds of locations across the United States. The number and frequency of the observations varies greatly by location, but some locations have as many as 3 decades worth of daily measurements. We focus on nitrogen dioxide, a common pollutant emitted by combustion engines and power stations, which has been linked to adult and infant deaths, and to incidence of lung cancer. In addition, nitrogen dioxide contributes to ozone formation with all its health impacts, as well as impacts on both terrestrial and aquatic ecosystems.

We consider nine locations that have relatively complete records since 2000: Allentown, Baltimore, Boston, Harrisburg, Lancaster, New York City, Philadelphia, Pittsburgh, and Washington D.C. We use the data for the years 2000-2012. Each functional observation $X_n(s_k, t)$ consists of the daily maximum one-hour nitrogen dioxide concentration measured in ppb, parts per billion, for day $t$, month $n$, $N = 156$, and at location $s_k$. Figure 2.3 shows the data for the nine locations for December 2012.

Applying the tests to the deseasonalized curves obtained after removing the monthly mean from each curve, as we did for the Irish wind data, we obtained p-values smaller than 0.001 for all $L, J \in \{2, 3, 4, 5\}$. In addition since the number of spatial locations is relatively small, only 9 stations, we applied our tests without taking dimension reduction in space. Again, we obtained p-values smaller than 0.001. We see that nonseparability is a feature of pollution data which is difficult
Figure 2.3: Maximum one hour nitrogen dioxide curves for December 2012. The curves are offset since they overlap.

to ignore. This means that simplifying the spatiotemporal dependence structure of many pollution data sets by assuming separability is questionable. Conclusions obtained under the assumption of separability may be incorrect, and may lead to incorrect public health policies.
Testing separability of functional time series

3.1 Introduction

Recall from Chapter 1 that the data that motivate this chapter have the form of functional panels:

\[ X_n(t) = [X_{n1}(t), X_{n2}(t), ..., X_{nS}(t)]^T, \quad 1 \leq n \leq N. \]  (3.1)

For this kind of data the presence of temporal dependence across \( n \) is a feature that cannot be ignored. Therefore, separability tests for independent data, which are currently available, cannot be used and a significance test, which accounts for temporal dependence, is called for. In this chapter we present such a test. As we mentioned in Chapter 1, we will demonstrate that our test works well in situations where the tests of Constantinou et al. (2015) and Aston et al. (2015) fail.

The remainder of this chapter is organized as follows. In Section 3.2, we formulate the assumptions, the definitions, and the problem. In Section 3.3, we derive
the test and provide the required asymptotic theory. Section 3.4 focuses on details of the implementation. In Section 3.5.1, we present results of a simulation study, and, finally, in Section 3.5.2 we apply our procedure to functional panels of Nitrogen Dioxide levels on the east coast of the United States and to US stock market data.

### 3.2 Assumptions and problem formulation

We assume that the $X_n$ in (3.1) form a strictly stationary functional time series of dimension $S$. To simplify notation, we assume that all functions are defined on the unit interval $[0,1]$. We assume that they are square integrable in the sense that

$$E\|X_{ns}\|^2 = E\int X^2_{ns}(t) dt < \infty \quad 1 \leq s \leq S.$$  

Stationarity implies that the covariances

$$\text{Cov}(X_{ns}(t), X_{ns'}(t')) = c(s, t, s', t')$$

are well-defined (do not depend on $n$). We want to test the null hypothesis

$$H_0 : \quad c(s, t, s', t') = c_1(s, s')c_2(t, t'), \quad s, s' \in \{1, 2, \ldots, S\}; \quad t, t' \in [0, 1]. \quad (3.2)$$

To derive the asymptotic distribution of our test statistic we impose a weak dependence condition on the $X_n$. We use the concept of $L^p$–$m$–approximability introduced in Hörmann and Kokoszka (2010), see also Chapter 16 of Horváth and Kokoszka (2012). Suppose $\mathcal{H}$ is a separable Hilbert space. Let $p \geq 1$ and let $L^p_{\mathcal{H}}$
be the space of $\mathbb{H}$–valued random elements $X$ such that

$$\nu_{p}(X) = \left( E \|X\|^p \right)^{1/p} < \infty.$$ 

**Definition 3.1.** The sequence, $\{Z_n\}$, $Z_n \in L^p_{\mathbb{H}}$, is $L^p$–approximable if the following conditions hold:

1. There exists a sequence $\{u_n\}$ of iid elements in an abstract measurable space $\mathcal{U}$ such that

   $$Z_n = f(u_n, u_{n-1}, \ldots),$$

   for a measurable function $f : \mathcal{U}^\infty \to \mathbb{H}$.

2. For each integer $M > 0$, consider an approximating sequence $Z_{n,M}$ defined by

   $$Z_{n,M} = f(u_n, u_{n-1}, \ldots, u_{n-M}, u^*_n, u^*_{n-M-1}, u^*_{n-M-2}, \ldots),$$

   where the sequences $\{u^*_n\} = \{u^*_n(n, m)\}$ are copies of $\{u_n\}$ independent across $m$ and $n$ and independent of the original sequence $\{u_n\}$. We assume that $Z_{n,M}$ well approximates $Z_n$ in the sense that

   $$\sum_{M=1}^{\infty} \nu_{p}(Z_n - Z_{n,M}) < \infty. \quad (3.3)$$

Condition 1 of Definition 3.1 implies that the sequence is strictly stationarity and ergodic. The essence of Condition 2 is that the dependence of $f$ on the innovations far in the past decays so fast that these innovations can be replaced by their independent copies. Such a replacement is asymptotically negligible in the sense quantified by (3.3). Similar conditions, which replace the more restrictive assumption of a linear moving average with summability conditions on its coefficients, have been used for at least a decade, see e.g. Shao and Wu (2007) and
references therein. We work with Definition 3.1 as it is satisfied by most time series models, including functional time series, and provides a number of desirable asymptotic properties including the central limit theorem. We therefore make the following assumption.

**Assumption 3.** The $X_n$ form an $L^4$–$m$–approximable sequence in

\[ \mathbb{H} = (L^2([0, 1]))^S. \]

We now introduce definitions and notation used in this chapter. We use tensor notation analogous to Aston et al. (2015). Let $\mathbb{H}_1$ and $\mathbb{H}_2$ denote two real separable Hilbert spaces with bases \{u\}_i and \{v\}_j respectively. We define $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$ to be the tensor product Hilbert space. The tensors \{u \otimes v\} form a basis for $\mathbb{H}$. In other words, the tensor product Hilbert space can be obtained by completing of the set span\{u \otimes v : i = 1, \ldots \ j = 1, \ldots \}, under the following inner product:

\[
\langle u \otimes v_j, u_k \otimes v_\ell \rangle = \langle u_i, u_k \rangle \langle v_j, v_\ell \rangle, \quad u_i, u_k \in \mathbb{H}_1, v_j, v_\ell \in \mathbb{H}_2.
\]

In the context of our study $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0, 1])$. Therefore the tensor product Hilbert space in our context is $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2 = \mathbb{R}^S \otimes L^2([0, 1]) = (L^2([0, 1]))^S = L^S_2$, where we omit [0, 1] for simplicity. Each $X_n$ is thus an element of a tensor space, formed by the tensor product between two real separable Hilbert spaces, $X_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$. We denote by $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$ the space of Hilbert-Schmidt operators acting on $\mathbb{H}_1 \otimes \mathbb{H}_2$. Note that \{u_i \otimes v_j \otimes u_k \otimes v_\ell\} is a basis for $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$. The covariance operator of $X_n \in \mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$, $C = \mathbb{E}[X_n \otimes X_n] \in \mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2)$, is called separable if

\[
C = C_1 \tilde{\otimes} C_2, \quad (3.4)
\]

where $C_1$ is a covariance operator over $\mathbb{H}_1$ and $C_2$ is a covariance operator over $\mathbb{H}_2$. 
We define $C_1 \otimes C_2$ as a linear operator on $H_1 \otimes H_2$ satisfying
\[(C_1 \otimes C_2)(u \otimes v) = (C_1 u) \otimes (C_2 v), \quad \forall u \in H_1, \forall v \in H_2.\]

In the context of our study, the covariance operator of $X_n \in L_2^S$ is $C = E[X_n \otimes X_n] \in S(L_2^S)$. In terms of covariance functions (i.e. the kernels of the covariance operators), (3.4) is equivalent to $H_0$ stated as (3.2) above.

### 3.3 Derivation of the test and its asymptotic justification

To test hypothesis (3.4), we propose a statistic which quantifies the difference between $\hat{C}_1 \otimes \hat{C}_2$ and $\hat{C}$:
\[
\hat{T} = N \| \hat{C}_1 \otimes \hat{C}_2 - \hat{C} \|_S^2,
\]
where $\hat{C}_1, \hat{C}_2, \hat{C}$ are estimates defined below, and $\| \cdot \|_S$ is the Hilbert-Schmidt norm. The asymptotic null distribution involves the covariance operator of $\hat{C}_1 \otimes \hat{C}_2 - \hat{C}$, which we denote by $Q$. Note that $Q \in S(S(H_1 \otimes H_2))$, i.e. it is an operator acting on $S(H_1 \otimes H_2)$. Therefore, it can be expanded using the basis functions of the form $\{u_i \otimes v_j \otimes u_k \otimes v_\ell \otimes u_m \otimes v_n \otimes u_p \otimes v_q\}$. In the context of (3.1), $Q \in S(S(L_2^S))$.

We now define the estimators appearing in (3.5) and obtain their limiting behavior even in the case where $C$ is not separable. A natural estimator for the general covariance, $C$, is given by
\[
\hat{C} = \frac{1}{N} \sum_{n=1}^{N} (X_n - \hat{\mu}) \otimes (X_n - \hat{\mu}) \in S(L_2^S),
\]
where $\mathbf{X}_n(t) = [X_{n1}(t), \ldots, X_{nS}(t)]^T$, $1 \leq n \leq N$, and $\mathbf{\mu}(t) = [\mu_1(t), \ldots, \mu_S(t)]^T$ with $\hat{\mu}_s(t) = \frac{1}{N} \sum_{n=1}^{N} X_{ns}(t)$, $1 \leq s \leq S$. Since centering by the sample mean is asymptotically negligible, we assume, without loss of generality and to ease the notation, that our data are centered, so the estimator takes the form

$$\hat{C} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{X}_n \otimes \mathbf{X}_n,$$

equivalently, the kernel of $\hat{C}$ is

$$\hat{c}(s, t, s', t') = \frac{1}{N} \sum_{n=1}^{N} X_{ns}(t) X_{ns'}(t').$$

Under $H_0$, $C = C_1 \tilde{\otimes} C_2$ with $C_1 \in S(\mathbb{H}_1) = S(\mathbb{R}^S)$, $C_2 \in S(\mathbb{H}_2) = S(L^2([0, 1]))$ and $C \in S(\mathbb{H}) = S(\mathbb{H}_1 \otimes \mathbb{H}_2) = S(L^S_2)$. To obtain estimators for $C_1$ and $C_2$, we utilize the trace and the partial trace operators. For any trace-class operator $T$, see e.g. Section 13.5 of Horváth and Kokoszka (2012) or Section 4.5 of Hsing and Eubank (2015), its trace is defined by

$$\text{Tr}(T) := \sum_{i=1}^{\infty} \langle Te_i, e_i \rangle,$$

where $(e_i)_{i \geq 1}$ is an orthonormal basis. It is invariant with respect to the basis. The partial-trace operators are defined as

$$\text{Tr}_1(A \tilde{\otimes} B) = \text{Tr}(A)B, \quad A \in \mathbb{H}_1, \ B \in \mathbb{H}_2,$$

and

$$\text{Tr}_2(A \tilde{\otimes} B) = \text{Tr}(B)A, \quad A \in \mathbb{H}_1, \ B \in \mathbb{H}_2.$$

This means that $\text{Tr}_1$ and $\text{Tr}_2$ are bilinear forms that satisfy $\text{Tr}_1 : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_2$ and $\text{Tr}_2 : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_1$. 
and $\text{Tr}_2 : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_1$. In general, the trace of any element of $T \in \mathbb{H}_1 \otimes \mathbb{H}_2$ can be defined using proper basis expansions. More specifically, let $u_1, u_2, \ldots$ be an orthonormal basis for $\mathbb{H}_1$ and $v_1, v_2, \ldots$ an orthonormal basis for $\mathbb{H}_2$. Then a basis for $\mathbb{H}_1 \otimes \mathbb{H}_2$ is given by \{ $u_i \otimes v_j : i = 1, 2, \ldots, j = 1, 2, \ldots$ \}. Let $T : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{H}_1 \otimes \mathbb{H}_2$. Then the trace of $T$ is defined by:

$$\text{Tr}(T) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \langle T(u_i \otimes v_j), u_i \otimes v_j \rangle,$$

$\text{Tr} : \mathbb{H}_1 \otimes \mathbb{H}_2 \to \mathbb{R}$.

If $T = A \tilde{\otimes} B$, the partial-trace operators in terms of a basis are defined as

$$\text{Tr}_1(T) = \text{Tr}_1(A \tilde{\otimes} B) = \text{Tr}(A) B = \sum_{i=1}^{\infty} \langle Au_i, u_i \rangle B$$

$$= \sum_{i=1}^{\infty} \langle Au_i, u_i \rangle \sum_{j=1}^{\infty} B_j v_j, \quad \forall A \in \mathbb{H}_1, \quad \forall B \in \mathbb{H}_2,$$

and

$$\text{Tr}_2(T) = \text{Tr}_2(A \tilde{\otimes} B) = \text{Tr}(B) A = \sum_{j=1}^{\infty} \langle Bv_j, v_j \rangle A$$

$$= \sum_{j=1}^{\infty} \langle Bv_j, v_j \rangle \sum_{i=1}^{\infty} A_i u_i, \quad \forall A \in \mathbb{H}_1, \quad \forall B \in \mathbb{H}_2.$$

In the context of functional panels, let $u_1, u_2, \ldots, u_S$ be an orthonormal basis for $\mathbb{R}^S$ and $v_1, v_2, \ldots$ an orthonormal basis for $L^2([0, 1])$. Then a basis for $L^S_2$ is given by \{ $u_i \otimes v_j : i = 1, 2, \ldots, S, j = 1, 2, \ldots$ \}. Recall that \{ $u_i \otimes u_k$ \}, when viewed as operators, is a basis for $\mathcal{S}(\mathbb{R}^S)$, that is a basis for the space of Hilbert-Schmidt operators acting on $\mathbb{R}^S$. Similarly \{ $v_j \otimes v_\ell$ \} is a basis for $\mathcal{S}(L^2([0, 1]))$. Finally \{ $u_i \otimes v_j \otimes u_k \otimes v_\ell$ \} is a basis for $\mathcal{S}(L^S_2)$. The basis expansion of $C$ is given by:

$$\sum_{i} \sum_{j} \sum_{k} \sum_{\ell} C_{ijkl} u_i \otimes v_j \otimes u_k \otimes v_\ell.$$
Therefore its trace is given by

$$\text{Tr}(C) = \sum_i \sum_j C_{ijij}.$$ 

Under the assumption of separability, i.e. $C = C_1 \tilde{\otimes} C_2$, the partial trace with respect to $\mathbb{H}_1$ in terms of a basis is given by:

$$\text{Tr}_1(C) = \text{Tr}_1(C_1 \tilde{\otimes} C_2) = \text{Tr}(C_1)C_2 = \sum_j \sum_\ell (\sum_i C_{ij\ell})v_j \otimes v_\ell$$

with $C_{2, j\ell} = \sum_i C_{ij\ell}$, and with respect to $\mathbb{H}_2$ is given by

$$\text{Tr}_2(C) = \text{Tr}_1(C_1 \tilde{\otimes} C_2) = \text{Tr}(C_2)C_1 = \sum_i \sum_k (\sum_j C_{ijk})u_i \otimes u_k$$

with $C_{1, ik} = \sum_j C_{ijk}$. In addition, under the assumption of separability, i.e. $C = C_1 \tilde{\otimes} C_2$, we define estimators of $C_1$ and $C_2$ as

$$\hat{C}_1 = \frac{1}{\text{Tr}(\hat{C})} \text{Tr}_2(\hat{C}) \quad \text{and} \quad \hat{C}_2 = \text{Tr}_1(\hat{C}),$$

where $\hat{C}_1$ is an $S \times S$ matrix and $\hat{C}_2$ is a temporal covariance operator. The intuition behind the above estimators is that $\text{Tr}(C)C = \text{Tr}_2(C) \tilde{\otimes} \text{Tr}_1(C)$. Note that the decomposition $C = C_1 \tilde{\otimes} C_2$ is not unique since $C_1 \tilde{\otimes} C_2 = (\alpha C_1) \tilde{\otimes} (\alpha^{-1} C_2)$ for any $\alpha \neq 0$, however the product $C_1 \tilde{\otimes} C_2$ is.

To derive the asymptotic distribution of the test statistic $\hat{T}$ defined in (3.5), we must first derive the joint asymptotic distribution of $\hat{C}, \hat{C}_1, \hat{C}_2$. A similar strategy was used in Constantinou et al. (2015). However, there the observations were assumed to be independent and more traditional likelihood methods were used to derive the asymptotic distributions. Here, we take a different approach, instead
using the CLT for $\hat{C}$, and then leveraging a Taylor expansion over Hilbert spaces to obtain the joint asymptotic distribution of $\hat{C}, \hat{C}_1, \hat{C}_2$. In this way, we are able to relax both the independence and Gaussian assumptions from Constantinou et al. (2015). The result is provided in Theorem 3.2. Due to the temporal dependence, the covariance operator of the limit normal distribution is a suitably defined long–run covariance operator. It has a very complex, but explicit and computable, form, which is displayed in the Appendix-B. All theorems that follow are proven in the Appendix-B.

Recall that we are interested in testing

$$H_0 : C = C_1 \tilde{\otimes} C_2 \quad vs \quad H_A : C \neq C_1 \tilde{\otimes} C_2.$$ 

In the following theorems notice that Theorem 3.2 and Theorem 3.3 hold without the assumption of separability, i.e. they hold under $H_0$ and under $H_A$. These two theorems are used to establish the behavior of our test statistic under both the null, Theorems 3.5 and 3.7, and the alternative, Theorems 3.6. Under the alternative both $C_1$ and $C_2$ are still defined as partial traces of $C$, it is just that their tensor product no longer recovers the original $C$.

**Theorem 3.2.** Under Assumption 3, one can explicitly define a long–run covariance operator $W$ such that

$$\sqrt{N} \begin{pmatrix} \hat{C}_1 - C_1 \\ \hat{C}_2 - C_2 \\ \hat{C} - C \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, W).$$

The definition of $W$ is given in (3.13). The operator $W$ has a block structure, it is built up of the operators $W_{ij}, i, j = 1, 2, 3$, which belong to Hilbert–Schmidt.
Armed with Theorem 3.2, we can derive the asymptotic distribution of \( \hat{C}_1 \tilde{\otimes} \hat{C}_2 - \hat{C} \), which forms the basis of the test statistic \( \hat{T} \).

**Theorem 3.3.** Under Assumption 3,

\[
\sqrt{N}((\hat{C}_1 \tilde{\otimes} \hat{C}_2 - \hat{C}) - (C_1 \tilde{\otimes} C_2 - C)) \xrightarrow{L} N(0, Q)
\]

The covariance operator \( Q \in S(S(H_1 \otimes H_2)) \) is defined in (3.14).

As corollary, we have the asymptotic distribution of \( \hat{C}_1 \tilde{\otimes} \hat{C}_2 - \hat{C} \) under \( H_0 \).

**Corollary 3.4.** Suppose Assumption 3 holds. Then, under \( H_0 \),

\[
\sqrt{N}(\hat{C}_1 \tilde{\otimes} \hat{C}_2 - \hat{C}) \xrightarrow{L} N(0, Q)
\]

The covariance operator \( Q \in S(S(H_1 \otimes H_2)) \) is defined in (3.14).

As noted above, in the context of (3.1), \( Q \in S(S(L_2^S)) \), i.e. it is a Hilbert-Schmidt operator acting on a space of Hilbert-Schmidt operators over \( L_2^S \). The following result is a direct consequence of Theorem 3.3. While the weighted chi-square expansion is standard, to compute the weights, the operator \( Q \) must be estimated, so \( W \) must be estimated. Formula (3.13) defining \( W \) is new and nontrivial.

**Theorem 3.5.** Suppose Assumption 3 holds. Let \( Q \) be the covariance operator appearing in Theorem 3.3, with eigenvalues \( \gamma_1, \gamma_2, \ldots \). Then, under \( H_0 \), as \( N \to \infty \),

\[
\hat{T} \xrightarrow{L} \sum_{r=1}^{\infty} \gamma_r \chi_1^2(r),
\]

where the \( \chi_1^2(r) \) are iid chi-square random variables with one degree of freedom.
To describe the behavior of the test statistic under the alternative, some specific form of the alternative must be assumed, as the violation of (3.4) can take many forms. A natural approach corresponding to a fixed alternative to $C_1 \tilde{\otimes} C_2 - C = 0$, is to assume that

$$C_1 \tilde{\otimes} C_2 - C =: \Delta \neq 0. \tag{3.6}$$

**Theorem 3.6.** Suppose Assumption 3 holds. If (3.6) holds, then

$$\hat{T} = N\|\Delta\|^2 + O_P(N^{1/2}) \overset{P}{\rightarrow} \infty.$$

In our applications, each $X_n$ is an element of a tensor space, formed by the tensor product between two real separable Hilbert spaces, $X_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$, where $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0,1])$. Therefore, in practice, we must first project these random elements onto a truncated basis by using a dimension reduction procedure. Note that $\mathbb{H}_1 = \mathbb{R}^S$ is already finite. However, if the number of coordinates in the panel is large, then a dimension reduction in $\mathbb{H}_1 = \mathbb{R}^S$ is also recommended. Here we present the general case where we use dimension reduction in both $\mathbb{H}_1 = \mathbb{R}^S$ and $\mathbb{H}_2 = L^2([0,1])$. Assume that this truncated basis is of the form $\hat{u}_k \otimes \hat{v}_j$ with $1 \leq k \leq K$, $1 \leq j \leq J$ where $K < S$ and $J < \infty$. We can approximate each $X_n \in \mathbb{H}_1 \otimes \mathbb{H}_2$ by a $K \times J$ random matrix $Z_n \in \mathbb{R}^{K \times J}$, where $Z_n(k,j) = \langle X_n, \hat{u}_k \otimes \hat{v}_j \rangle$, $1 \leq k \leq K$, $1 \leq j \leq J$. Therefore, from now on, we have observations in the form of random $K \times J$ matrices defined as

$$Z_n = [\zeta_{kj,n}, \ 1 \leq k \leq K, \ 1 \leq j \leq J],$$

where $\zeta_{kj,n} = \langle X_n, \hat{u}_k \otimes \hat{v}_j \rangle$. Let $\hat{T}_F$ be the truncated test statistic $\hat{T}$, i.e.

$$\hat{T}_F = N\|\hat{C}_1^* \tilde{\otimes} \hat{C}_2^* - \hat{C}^*\|^2_F,$$
where $\hat{C}^*_{1}$ is a $K \times K$ matrix, $\hat{C}^*_{2}$ is a $J \times J$ matrix, $\hat{C}^*$ is a fourth order array of dimension $K \times J \times K \times J$, and $\| \cdot \|_F$ is the Frobenius norm, which is the Hilbert–Schmidt norm in finite dimensions. Finally, let $Q^*$ be the truncated covariance operator $Q$, i.e. $Q^*$ is the asymptotic covariance operator in the convergence

$$\sqrt{N}((\hat{C}^*_{1} \otimes \hat{C}^*_{2} - \hat{C}^*) - (C^*_{1} \otimes C^*_{2} - C^*)) \xrightarrow{\mathcal{D}} N(0, Q^*).$$

Note that $Q^*$ is an array of order eight with finite dimensions, i.e., $Q^* \in \mathbb{R}^{K \times J \times K \times J \times K \times J}$. More details are given in Remark 2 in Appendix-B. As a finite array, it has only a finite number of eigenvalues, which with denote $\gamma_1, \gamma_2, \ldots, \gamma_R$. The arguments leading to Theorem 3.5, yield the following counterpart.

**Theorem 3.7.** Suppose Assumption 3 holds. Under $H_0$, as $N \to \infty$,

$$\hat{T}_F \xrightarrow{\mathcal{D}} \sum_{r=1}^{R} \gamma_r \chi_1^2(r),$$

where the $\chi_1^2(r)$ are as in Theorem 3.5.

In Theorem 3.7 we stress that the asymptotic holds when using the estimated eigenfunctions. This requires showing that the projecting onto the estimated or population level eigenfunctions is asymptotically equivalent.

### 3.4 Details of implementation

Recall that we assume that all functions have been rescaled so that their domain is the unit interval $[0, 1]$, and that they have mean zero. The testing procedure consists of dimension reduction in time and, for large panels, a further dimension reduction in coordinates. After reducing the dimension our "observations" are of
the form of $K \times J$ matrices which are used to compute estimators of we need to perform our test. The remainder of this section explains the details in an algorithmic form. The reader will notice that most steps have obvious variants, for example, different weights and bandwidths can be used in Step 6. Procedure 4 describes the exact implementation used in Sections 3.5.1 and 3.5.2.

Procedure 4.

1. [Pool across $s$ to get estimated temporal FPC’s.] Under the assumption of separability, i.e., under the $H_0$ stated in Section 3.2, the optimal functions used for temporal dimension reduction are the same for each member (coordinate) of the panel; information can then be pooled across the coordinates to get better estimates of these functions. In other words under separability we can use simultaneously all the $N \times S$ functions to estimate temporal FPC’s $\hat{v}_1(t), \ldots, \hat{v}_J(t)$, which are the eigenfunctions of the following covariance function:

$$\hat{C}_2(t,t') = \frac{1}{NS} \sum_{n=1}^{N} \sum_{s=1}^{S} X_{ns}(t)X_{ns}(t').$$

2. Approximate each curve $X_{ns}(t)$ by

$$X_{ns}^{(J)}(t) = \hat{\mu}_s(t) + \sum_{j=1}^{J} \xi_{nsj} \hat{v}_j(t),$$

where $\xi_{nsj} = \langle X_{ns}(t), \hat{v}_j(t) \rangle$. Construct $S \times J$ matrices $\Xi_n$ defined as

$$\Xi_n = [\xi_{nsj}, \ 1 \leq s \leq S, \ 1 \leq j \leq J],$$

where $J$ is chosen large enough so that the first $J$ FPC’s explain at least 85% of the variance. This is Functional Principal Components Analysis carried out on the
pooled (across coordinates) sample.

3. [Pool across time to get panel PC’s.] Under the assumption of separability the panel principal components are the same for each time. In other words the panel PC’s $\hat{u}_1, \ldots, \hat{u}_K$ are the principal components of the following covariance matrix:

$$\hat{C}_1(s,s') = \sum_n \int X_{ns}(t) X_{ns'}(t) \, dt \quad \frac{N \text{tr}(\hat{C})}{N}.$$  

However since we have already reduced the dimension of the observed functions the panel PC’s $\hat{u}_1, \ldots, \hat{u}_K$ are the principal components of the following covariance matrix:

$$\tilde{C}_1(s,s') = \frac{1}{N J} \sum_{n=1}^{N} \sum_{j=1}^{J} \frac{\xi_{nsj} \xi_{ns'j}}{\lambda_j}.$$  

4. Approximate each column $\xi_{n,j} = (\xi_{n1j}, \xi_{n2j}, \ldots, \xi_{nSj})$ of the $\Xi_n$ matrices by

$$\xi_{n,j} = \sum_{k=1}^{K} \zeta_{kj,n} \hat{u}_k,$$

where $\zeta_{kj,n} = \langle \xi_{n,j}, \hat{u}_k \rangle$. Construct the $K \times J$ matrices $Z_n = [\zeta_{kj,n}, 1 \leq k \leq K, 1 \leq j \leq J]$, where $K$ is chosen large enough so that the first $K$ eigenvalues explain at least 85% of the variance. This is a multivariate PCA on the pooled (across time) variance adjusted sample.

If the number of panel coordinates is small, then a multivariate dimension reduction is not necessary, so one can skip steps 3 and 4 and use the $\Xi_n$ matrices instead of the $Z_n$ matrices, and replace $K$ with $S$ in the following steps.
5. Compute the fourth order array of dimensions $K \times J \times K \times J$

\[
\hat{C}^* = \frac{1}{N} \sum_{n=1}^{N} Z_n \otimes Z_n.
\]

Compute the estimators of $C_1$ and $C_2$ by using the partial trace and trace operators

\[
\hat{C}_1^*(k, k') = \frac{\sum_{j=1}^{J} \hat{C}_1^*(k, j, k', j)}{\sum_{k=1}^{K} \sum_{j=1}^{J} C_1^*(k, j, k, j)} \quad \text{and} \quad \hat{C}_2^*(j, j') = \sum_{k=1}^{K} \hat{C}_2^*(k, j, k', j'),
\]

where $\hat{C}_1^*$ is a $K \times K$ matrix and $\hat{C}_2^*$ is a $J \times J$ matrix.

6. Estimate $\Gamma^* \in \mathbb{R}^{K \times J \times K \times J \times K \times J}$. A Bartlett-type estimator of $\Gamma^*$ is given by

\[
\hat{\Gamma}^* = \hat{R}_0^* + \sum_{h=1}^{N-1} \omega_h (\hat{R}_h^* + \hat{R}_h^{**})
\]

where

\[
\hat{R}_0^* = \frac{1}{N} \sum_{n=1}^{N} [(Z_n \otimes Z_n - \hat{C}^*) \otimes (Z_n \otimes Z_n - \hat{C}^*)],
\]

\[
\hat{R}_h^* = \frac{1}{N} \sum_{n=1}^{N-h} [(Z_n \otimes Z_n - \hat{C}^*) \otimes (Z_{n+h} \otimes Z_{n+h} - \hat{C}^*)],
\]

\[
\hat{R}_h^{**} = \frac{1}{N} \sum_{n=1}^{N-h} [(Z_{n+h} \otimes Z_{n+h} - \hat{C}^*) \otimes (Z_n \otimes Z_n - \hat{C}^*)].
\]

The $\omega_h$ are the Bartlett’s weights, i.e.,

\[
\omega_h = \begin{cases} 
1 - \frac{h}{1+q}, & \text{if } h \leq q \\
0, & \text{otherwise}
\end{cases}
\]

where $h$ is the number of lags and $q$ is the bandwidth which is assumed to be a func-
tion of the sample size, i.e., \( q = q(N) \). In our simulations, in Section 3.5.1, we use the formula \( q \approx 1.1447\left(\frac{N}{4}\right)^{1/3} \) (Chapter 16 of Horváth and Kokoszka (2012)).

7. **Estimate the arrays** \( \mathbf{W}^* \) (the truncated analog of \( \mathbf{W} \)) and \( \mathbf{Q}^* \) defined in Section 3.3. More details are given in Remark 2 in the Appendix-B.

8. **Calculate the p–value using the limit distribution specified in Theorem 3.7.**

Step 2 can be easily implemented using \( \text{R} \) function \( \text{pca.fd} \), and step 3 by using \( \text{R} \) function \( \text{prcomp} \). The matrix \( \mathbf{Q}^* \) can be computed using the \( \text{R} \) package ”tensorA” by van den Boogaart (2007).

### 3.5 Simulation study and applications

#### 3.5.1 A simulation study

The purpose of this section is to provide information on the performance of our test procedure in finite samples. Before we proceed with the details, we comment on the performance of existing tests. Constantinou et al. (2015) derived several separability tests based on the assumption of independent \( \mathbf{X}_n \). For the functional panels which exhibit temporal dependence (we define them below), the empirical sizes are close to zero; the tests of Constantinou et al. (2015) are too conservative to be usable, unless we have independent replications of the spatio–temporal structure. Aston et al. (2015) proposed three tests, also for independent \( \mathbf{X}_n \). In the presence of temporal dependence, their tests are not useable either; they can severely overreject, the empirical size can approach 50% at the nominal level of 5%. We give some specific numbers at the end of this section.
For our empirical study, we simulate functional panels as the moving average process

\[ X_{ns}(t) = \sum_{s' = 1}^{S} \Psi_{ss'}[e_{ns'}(t) + e_{n-1s'}(t)], \]

which is a 1-dependent functional time series. Direct verification, shows that it is separable as long the \( e_{ns}(t) \) are separable. We generate \( e_{ns}(t) \) as Gaussian processes with the following covariance function, which is a modified version of Example 2 of Cressie and Huang (1999):

\[
\sigma_{ss'}(t, t') = \frac{\sigma^2}{(a|t - t'| + 1)^{1/2}} \exp \left( -\frac{b^2|s - s'|/(S - 1)^2}{(a|t - t'| + 1)^c} \right).
\]

In this covariance function, \( a \) and \( b \) are nonnegative scaling parameters of time and space, respectively, and \( \sigma^2 > 0 \) is an overall scale parameter. The most important parameter is the separability parameter \( c \) which takes values in \([0, 1]\). If \( c = 0 \), the covariance function is separable, otherwise it is not. We set \( a = 3, b = 2, \sigma^2 = 1 \).

To simulate the functions, we use \( T = 50 \) time points equally spaced on \([0, 1]\), and \( S \in \{4, 6, 8, 10, 12, 14\} \) coordinates in the panel. The MA coefficients are taken as:

\[
\Psi_{ss'} = \exp \left( -\frac{25(s - s')^2}{(S - 1)^2} \right).
\]

Notice that in the covariance above, the differences in the coordinates of the panel, i.e. \( |s - s'| \), are rescaled to be within the interval \([0, 1]\), i.e. we use \( |s - s'|/(S - 1) \).

We set

\[
c = 0 \text{ under } H_0; \quad c = 1 \text{ under } H_A.
\]

We consider two different cases. The first one with dimension reduction only in time and the second one with dimension reduction in both time and coordinates. For each case, we study two different scenarios. The first scenario is under the null
hypothesis (separability) and the second scenario under the alternative hypothesis. We consider different numbers of temporal FPC's, \( J \), in the first case and different numbers of coordinate PC's, \( K \), and temporal FPC's, \( J \), in the second case. We will also consider different values for the series length \( N \). All empirical rejection rates are based on one thousand replications, so their SD is about 0.7 percent for size (we use nominal significance level of 5%), and about two percent for power.

### 3.5.1.1 Case 1: dimension reduction in time only

We examine the effect of the series length \( N \) and the number of principal components \( J \) on the empirical size (Table 3.1) and power (Table 3.2) for \( S \in \{4, 6, 8\} \). Each table reports the rejection rates in percent. In parentheses, the proportion of variance explained by the \( J \) PC’s is given.

In Table 3.1, we can see that the size of our test is robust to the number of the principal components used. This is a very desirable property, as in all procedures of FDA there is some uncertainty about the optimal number of FPC's that should be used. While still within two standard errors of the nominal size, the empirical size becomes inflated for \( S = 8 \). We recommend dimension reduction in panel coordinates if \( S \geq 10 \). In Table 3.2, we see that the empirical power increases as \( N \) and \( J \) increase. The power increase with \( N \) is expected; its increase with \( J \) reflects the fact that projections on larger subspaces better capture a departure from \( H_0 \). However, \( J \) cannot be chosen too large so as not to increase the dimensionality of the problem, which negatively affects the empirical size.
3.5.1.2 Case 2: dimension reduction in both time and panel coordinates

The general setting is the same as in Section 3.5.1.1, but we consider larger panels, $S \in \{10, 12, 14\}$, and reduce their dimension to $K \in \{2, 3, 4\}$ coordinates. The proportion of variance explained is now computed as

$$ CPV(J, K) = \frac{\sum_{j=1}^{J} \lambda_j}{\sum_{j=1}^{\infty} \lambda_j} \times \frac{\sum_{k=1}^{K} \mu_k}{\sum_{k=1}^{S} \mu_k}, $$

(3.7)

where the $\lambda_1, \lambda_2, \ldots$, and $\mu_1, \mu_2, \ldots, \mu_S$ are, respectively, the estimated eigenvalues of the time and panel PCA’s.

Tables 3.3 and 3.4 show that the reduction of the panel dimension does not negatively affect the properties of the tests. The conclusions are the same as in Section 3.5.1.1. Either approach leads to a test with well controlled size, which is does not depend on $J$ ($J, K$) as long the the proportion of explained variance remains within the generally recommended range of 85%–95%. If $J = 2$ or $K = 2$ are used, this requirement is generally not met, resulting in a size distortion, which is however acceptable and decreases with $N$.

As noted at the beginning of this section, the tests of Constantinou et al. (2015) are too conservative, they almost never reject under the null for all scenarios considered in this section. The tests of Aston et al. (2015) reject too often under the null. For example, in the settings considered in Table 3.3, the rejection rates for their asymptotic test, Gaussian parametric bootstrap test, and Gaussian parametric bootstrap test using Hilbert–Schmidt distance, range between 19.0% – 49.4%, 14.6%–32.2% and 38.1%–44.9%, respectively. By contrast, the test derived in this work, in its both versions and under all reasonable choices of tuning parameters, has precise empirical size at the standard 5% nominal level and useful power.
3.5.2 Applications to pollution and stock market data

We begin by applying our method to air quality data studied by Constantinou et al. (2015) under the assumption that the monthly curves are independent and identically distributed. These curves however form a time series, so it is important to check if a test that accounts for the temporal dependence leads to the same or a different conclusion.

The Environmental Protection Agency (EPA) collects massive amounts of air quality data which are available through its website http://www3.epa.gov/airdata/ad_data_daily.html. The records consist of data for 6 common pollutants, collected by outdoor monitors in hundreds of locations across the United States. The number and frequency of the observations varies greatly by location, but some locations have as many as 3 decades worth of daily measurements. We focus on nitrogen dioxide, a common pollutant emitted by combustion engines and power stations.

We consider nine locations along the east coast that have relatively complete records since 2000: Allentown, Baltimore, Boston, Harrisburg, Lancaster, New York City, Philadelphia, Pittsburgh, and Washington D.C. We use the data for the years 2000-2012. Each functional observation $X_{ns}(t)$ consists of the daily maximum one-hour nitrogen dioxide concentration measured in ppb (parts per billion) for day $t$, month $n$ ($N = 156$), and at location $s$. We thus have a panel of $S = 9$ functional time series (one at every location), $X_{ns}(t)$, $s = 1, 2, \ldots, 9$, $n = 1, 2, \ldots, 156$. Figure 3.1 shows the data for the nine locations for December 2012. Before the application of the test, the curves were deseasonalized by removing the monthly mean from each curve.

We applied both versions of Procedure 4 (dimension in time only and double dimension reduction). Requiring 85% to 95% of explained variance yielded the
Figure 3.1: Maximum one-hour nitrogen dioxide curves for December 2012 at the nine locations.

values $J, K = 2, 3, 4$, similarly as in our simulated data example. For all possible combinations, we obtained p-values smaller than 10E-4. This indicates a nonseparable covariance function and confirms the conclusion obtained by Constantinou et al. (2015); nonseparability is an intrinsic feature of pollution data, simplifying the covariance structure by assuming separability may lead to incorrect conclusions.

We now turn to an application to a stock portfolio. Cumulative intradaily returns have recently been studied in several papers, including Kokoszka and Reimherr (2013b), Kokoszka et al. (2015) and Lucca and Moench (2015). If $P_n(t)$ is the price of a stock at minute $t$ of the trading day $n$, then the cumulative intraday return curve on day $n$ is defined by

$$R_n(t) = \log(P_n(t)) - \log(P_n(0)),$$
where time 0 corresponds to the opening of the market (9:30 EST for the NYSE). Horváth et al. (2014) showed that such time series of functions are stationary. Figure 3.2 shows the curves $R_n$ for ten companies on April 2nd, 2007. This portfolio of $S = 10$ stocks produces a panel of functional time series studied in this work. We selected ten US blue chip companies, and want to determine if the resulting panel can be assumed to have a separable covariance function. The answer is yes, as we now explain.

We consider stock values, recorded every minute, from October 10, 2001 to April 2, 2007 (1,378 trading days) for the following 10 companies: Bank of America (BOA), Citi Bank, Coca Cola, Chevron Corporation (CVX), Walt Disney Company (DIS), International Business Machines (IBM), McDonald’s Corporation (MCD), Microsoft Corporation (MSFT), Walmart Stores (WMT) and Exxon Mobil Corporation Common (XOM). On each trading day, there are 390 discrete observations. There is an outlier on August 26, 2004 for Bank of America, which is due to a stock split. That day is discarded from further analysis, so the sample size is $N = 1377$.

We now discuss the results of applying Procedure 4. Using dimension reduction in time only, we obtained p-values 0.234 for $J = 2$ (CPV = 92%) and 0.220 (CPV = 95%). Using the double dimension reduction, we obtained the following values:

<table>
<thead>
<tr>
<th>$K$ = $J$</th>
<th>P-value</th>
<th>CPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.272</td>
<td>45%</td>
</tr>
<tr>
<td>3</td>
<td>0.217</td>
<td>62%</td>
</tr>
<tr>
<td>4</td>
<td>0.224</td>
<td>67%</td>
</tr>
<tr>
<td>6</td>
<td>0.223</td>
<td>80%</td>
</tr>
<tr>
<td>7</td>
<td>0.221</td>
<td>85%</td>
</tr>
</tbody>
</table>

These remarkably similar p-values indicate that panels of cumulative intraday re-
Figure 3.2: Cumulative intraday return curves for the ten companies for April 2nd, 2007.

turn curves can in some cases be assumed to have a separable covariance function. This could be useful for portfolio managers as it indicates that they can exploit separability of the data for more efficient modeling.

3.6 Appendix-B

In Constantinou et al. (2015), the asymptotic distributions of their test statistics were derived under the assumption of independent and identically distributed Gaussian data so that maximum likelihood estimators could be used to estimate the covariance and its separable analog. However, here we make no normality assumptions and we allow the sequence to be weakly dependent across $n$, thus entirely different proof techniques are employed. In particular, we utilize multiple stochastic Taylor expansions to leverage the asymptotic normality of $\hat{C}$ to derive
the joint asymptotic distribution of \((\hat{C}_1, \hat{C}_2, \hat{C})\) as well as the asymptotic behavior of our test statistics under both the null and alternative hypotheses. These arguments become quite technical due to the fact that we are deriving asymptotic distributions of random operators.

### 3.6.1 Proof of Theorem 3.2

The starting point is the asymptotic distribution of \(\hat{C}\). It follows from Theorem 3 of Kokoszka and Reimherr (2013a) that under Assumption 3,

\[
\sqrt{N}(\hat{C} - C) \xrightarrow{L} N(0, \Gamma),
\]

where \(\Gamma\) is given by

\[
\Gamma = R_0 + \sum_{h=1}^{\infty} [R_h + R_h^*]
\text{ with } R_h = \mathbb{E}[(X_1 \otimes X_1 - C) \otimes (X_1+h \otimes X_1+h - C)].
\]

Here \(R_h^*\) denotes the adjoint of \(R_h\). The operator \(\Gamma\) is a Hilbert-Schmidt operator acting on the space of Hilbert-Schmidt operators over \(H_1 \otimes H_2\). Since we have the asymptotic distribution of \(\hat{C}\), in the following steps we use a one term Taylor expansion of the partial trace operators to find the joint asymptotic distribution of \(\hat{C}_1, \hat{C}_2, \hat{C}\). Consider the operator:

\[
f(\hat{C}) = \begin{pmatrix} f_1(\hat{C}) \\ f_2(\hat{C}) \\ f_3(\hat{C}) \end{pmatrix} = \begin{pmatrix} \frac{\text{Tr}_2(\hat{C})}{\text{Tr}(C)} \\ \text{Tr}_1(\hat{C}) \\ \hat{C} \end{pmatrix} = \begin{pmatrix} \hat{C}_1 \\ \hat{C}_2 \\ \hat{C} \end{pmatrix}.
\]

So \(f(\hat{C})\) is an element of the Cartesian product space \(H_1 \times H_2 \times H\) with \(f_1 : H_1 \otimes H_2 \to H_1, f_2 : H_1 \otimes H_2 \to H_2, f_3 : H_1 \otimes H_2 \to H_1 \otimes H_2\). We will apply
at Taylor expansion to $f(\hat{C})$ about the true parameter value $C$. To do this, we require the Fréchet derivative of $f(\hat{C})$ which can be computed coordinate-wise as

$$
\frac{\partial}{\partial C} f = \begin{pmatrix}
\frac{\partial f_1}{\partial C} \\
\frac{\partial f_2}{\partial C} \\
\frac{\partial f_3}{\partial C}
\end{pmatrix}.
$$

Here $\partial f_i / \partial C$ denotes the Fréchet derivative of $f_i$ with respect to $C$. Since $f_i$ is an operator, this means its derivative is a linear operator acting on the space of operators. Our goal is to use a Taylor expansion for Hilbert spaces to obtain the joint asymptotic distribution of $\hat{C}_1, \hat{C}_2, \hat{C}$. We approximate $f_1(\hat{C}), f_2(\hat{C}), f_3(\hat{C})$ by:

$$
f_1(\hat{C}) = f_1(C) + \frac{\partial f_1}{\partial C} (\hat{C} - C) + O_P(N^{-1}),
$$

$$
f_2(\hat{C}) = f_2(C) + \frac{\partial f_2}{\partial C} (\hat{C} - C) + O_P(N^{-1}),
$$

$$
f_3(\hat{C}) = f_3(C) + \frac{\partial f_3}{\partial C} (\hat{C} - C) + O_P(N^{-1}),
$$

where the last term is $O_P(N^{-1})$ because $\hat{C} - C = O_P(N^{1/2})$. In terms of the cartesian product form, this is equivalent to:

$$
f(\hat{C}) = f(C) + \nabla f(C)(\hat{C} - C) + O_P(N^{-1}).
$$

We therefore have that the variance operator of $f(\hat{C})$ is asymptotically given by

$$
\text{Var}(f(\hat{C})) \approx \text{Var} \left( f(C) + \nabla f(C)(\hat{C} - C) \right) \tag{3.9}
$$

$$
= \text{Var} \left( f(C) + \nabla f(C)(\hat{C}) - \nabla f(C)(C) \right)
$$

$$
= \text{Var} \left( \nabla f(C)(\hat{C}) \right)
$$

$$
= \nabla f(C) \left( \text{Var}(\hat{C})(\nabla f(C)^*) \right)
$$
\[
\n\n\n= \nabla f(C) \circ \Gamma \circ \nabla f(C)^* \\
\n:= W,
\n\]

where \((\cdot)^*\) denotes the adjoint operator. We stress again that each term written above is a linear operator, and thus \(W\) is actually a composition \((\circ)\) of three linear operators. This implies that the joint asymptotic distribution of \(\hat{C}_1, \hat{C}_2, \hat{C}\) is given by:

\[
\sqrt{N} (f(\hat{C}) - f(C)) = \sqrt{N} \begin{pmatrix} \hat{C}_1 - C_1 \\ \hat{C}_2 - C_2 \\ \hat{C} - C \end{pmatrix} \xrightarrow{L} \mathcal{N}(0, W).
\]

To complete the proof we need to find the Fréchet derivatives. This turns out to be easier if we work with a basis for the Hilbert spaces. For example, the actions of a continuous linear operator are completely determined by its actions on individual basis elements. Let \(u_1, u_2, \ldots\) be a basis for \(\mathbb{H}_1\) and \(v_1, v_2, \ldots\) a basis for \(\mathbb{H}_2\). Then a basis for \(\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2\) is given by \(\{u_i \otimes v_j : i = 1, 2, \ldots, j = 1, 2, \ldots\}\). Since \(C_1 : \mathbb{H}_1 \to \mathbb{H}_1\) is a compact operator, we can express it as

\[
C_1 = \sum_{ik} C_{1;ik} u_i \otimes u_k,
\]

where \(C_{1;ik} \in \mathbb{R}\) and \(\sum_{ik} C_{1;ik}^2 < \infty\). Similarly we have that

\[
C_2 = \sum_{jl} C_{2;lj} v_j \otimes v_l,
\]

and

\[
C = \sum_{ijkl} C_{ijkl} u_i \otimes v_j \otimes u_k \otimes v_l.
\]
These forms will be useful as we will be able to determine derivatives by taking derivatives with respect to the basis coordinate system. In the following, \( \delta_{ik} \) is the usual Kronecker delta. We begin with \( f_2(C) \) as it is simpler than \( f_1 \). Note that, by definition we have

\[
C_2 = f_2(C) = \sum_{jl} \left( \sum_i C_{ijl} \right) v_j \otimes v_l,
\]

that is, we take the trace over the \( u \) coordinates. So we have hat \( f_2 \) is a linear mapping from \( \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \mathbb{H}_1 \otimes \mathbb{H}_2 \rightarrow \mathbb{H}_2 \otimes \mathbb{H}_2 \). If we take the derivative of this expression with respect to \( C_{ijkl} \), then we get that

\[
\frac{\partial f_2(C)}{\partial C_{ijkl}} = (v_j \otimes v_l) \delta_{ik}.
\]

If \( i \neq k \), then \( C_{ijkl} \) does not appear in the expression for \( f_2(C) \) and thus the derivative would be zero. So we have that

\[
\frac{\partial f_2(C)}{\partial C} = \sum_{ijkl} \frac{\partial f_2(C)}{\partial C_{ijkl}} u_i \otimes v_j \otimes u_k \otimes v_l = \sum_{ijkl} \delta_{ik}(v_j \otimes v_l) \otimes (u_i \otimes v_j \otimes u_k \otimes v_l),
\]

(3.10)

where again, this is interpreted as a linear operator from \( \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \mathbb{H}_1 \otimes \mathbb{H}_2 \rightarrow \mathbb{H}_2 \otimes \mathbb{H}_2 \). Note that the above operator is nearly the identity, e.g. \( \frac{\partial f_2(C)}{\partial C}(x) = x \), but returns 0 for the off-diagonal \( u \) coordinates. We denote this operator as

\[
M_2 = \sum_{ijkl} \delta_{ik}(v_j \otimes v_l) \otimes (u_i \otimes v_j \otimes u_k \otimes v_l).
\]

The partial derivative of \( f_1 \) with respect to \( C \) is a bit more complicated as it
is a nonlinear function of \( C \). We can express \( f_1 \) as

\[
C_1 = f_1(C) = \frac{\text{Tr}_2(C)}{\text{Tr}(C)} = \sum_{ik} \sum_{j} C_{ijkj} u_i \otimes u_k.
\]

Again, taking the derivative with respect to the \( C_{ijkl} \) coordinate, we get that

\[
\frac{\partial f_1(C)}{\partial C_{ijkl}} = \delta_{jl} \frac{\text{Tr}(C) - \delta_{ik} \sum_{j'} C_{ij'k'} u_i \otimes u_k}{\text{Tr}(C)^2} := M_1.
\]

Therefore we have that

\[
\frac{\partial f_1(C)}{\partial C} = \sum_{ijkl} \delta_{jl} \frac{\text{Tr}(C) - \delta_{ik} \sum_{j'} C_{ij'k'} u_i \otimes u_k \otimes (u_i \otimes v_j \otimes u_k \otimes v_l)}{\text{Tr}(C)^2} := M_1.
\]

(3.11)

Finally, the partial derivative of \( f_3 \) with respect to \( C \), i.e., the partial derivative of \( C \) with respect to \( C \) is simply the identity, therefore

\[
\frac{\partial f_3}{\partial C} = \sum_{ijkl} (u_i \otimes v_j \otimes u_k \otimes v_l) \otimes (u_i \otimes v_j \otimes u_k \otimes v_l) := M_3.
\]

(3.12)

By (3.8), (3.9), (3.11), (3.10), and (3.12), we obtain

\[
\sqrt{N} \begin{pmatrix} \hat{C}_1 - C_1 \\ \hat{C}_2 - C_2 \\ \hat{C} - C \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, W),
\]
where $W$ is given by

$$W = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} \circ \Gamma \circ \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix}^*.$$  \hspace{1cm} (3.13)

**Remark 1.** The operator $W$ has the following block structure form:

$$W = \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix},$$

where

$$W_{11} = M_1 \Gamma M_1^*, \quad W_{12} = M_1 \Gamma M_2^*, \quad W_{13} = M_1 \Gamma M_3^*$$

$$W_{21} = M_2 \Gamma M_1^*, \quad W_{22} = M_2 \Gamma M_2^*, \quad W_{23} = M_2 \Gamma M_3^*$$

$$W_{31} = M_3 \Gamma M_1^*, \quad W_{32} = M_3 \Gamma M_2^*, \quad W_{33} = M_3 \Gamma M_3^* = \Gamma.$$

The operator $W_{11}$ is the covariance operator of $\hat{\mathcal{C}}$ and $W_{11} \in \mathcal{S}(\mathcal{H}_1)$, $W_{12}$ is the covariance between $\hat{\mathcal{C}}_1$ and $\hat{\mathcal{C}}_2$ and $W_{12} \in \mathcal{S}(\mathcal{H}_1) \otimes \mathcal{S}(\mathcal{H}_2)$, $W_{13}$ is the covariance between $\hat{\mathcal{C}}_1$ and $\hat{\mathcal{C}}$ and $W_{13} \in \mathcal{S}(\mathcal{H}_1) \otimes \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, $W_{21}$ is the covariance between $\hat{\mathcal{C}}_2$ and $\hat{\mathcal{C}}_1$ and $W_{21} \in \mathcal{S}(\mathcal{H}_2) \otimes \mathcal{S}(\mathcal{H}_1)$, $W_{22}$ is the covariance operator of $\hat{\mathcal{C}}_2$ and $W_{22} \in \mathcal{S}(\mathcal{H}_2)$, $W_{23}$ is the covariance between $\hat{\mathcal{C}}_2$ and $\hat{\mathcal{C}}$ and $W_{23} \in \mathcal{S}(\mathcal{H}_2) \otimes \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$, $W_{31}$ is the covariance between $\hat{\mathcal{C}}$ and $\hat{\mathcal{C}}_1$ and $W_{31} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{S}(\mathcal{H}_1)$, $W_{32}$ is the covariance between $\hat{\mathcal{C}}$ and $\hat{\mathcal{C}}_2$ and $W_{32} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2) \otimes \mathcal{S}(\mathcal{H}_2)$ and finally $W_{33}$ is the covariance operator of $\hat{\mathcal{C}}$ and $W_{33} \in \mathcal{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. 
3.6.2 Proof of Theorem 3.3

Since we have the joint asymptotic distribution of $\hat{C}_1, \hat{C}_2, \hat{C}$, we can use the delta method again to find the asymptotic distribution of $\hat{C}_1 \otimes \hat{C}_2 - \hat{C}$ and in particular, we can find the form of $Q$, the asymptotic covariance of $\hat{C}_1 \otimes \hat{C}_2 - \hat{C}$. Consider the function $g(C) = f_1(C) \otimes f_2(C) - f_3(C) = C_1 \otimes C_2 - C$. Using a one term Taylor expansion we have that

$$g(\hat{C}) \approx g(C) + \frac{\partial (g(C))}{\partial C_1} (\hat{C}_1 - C_1) + \frac{\partial (g(C))}{\partial C_2} (\hat{C}_2 - C_2) + \frac{\partial (g(C))}{\partial C} (\hat{C} - C),$$

or equivalently

$$g(\hat{C}) = g(C) + \nabla g(C) \cdot \begin{pmatrix} \hat{C}_1 - C_1 \\ \hat{C}_2 - C_2 \\ \hat{C} - C \end{pmatrix} + O_p(N^{-1}),$$

$$\nabla g(C) = \begin{pmatrix} \frac{\partial (g(C))}{\partial C_1} \\ \frac{\partial (g(C))}{\partial C_2} \\ \frac{\partial (g(C))}{\partial C} \end{pmatrix},$$

which implies that the variance of $\hat{C}_1 \otimes \hat{C}_2 - \hat{C}$ is approximately:

$$\operatorname{Var}(\hat{C}_1 \otimes \hat{C}_2 - \hat{C}) \approx \nabla g(C)^* \circ W \circ \nabla g(C) := Q,$$

and therefore the delta method implies that the asymptotic distribution of $\hat{C}_1 \otimes \hat{C}_2 - \hat{C}$ is given by:

$$\sqrt{N}(g(\hat{C}) - g(C)) = \sqrt{N}((\hat{C}_1 \otimes \hat{C}_2 - \hat{C}) - (C_1 \otimes C_2 - C)) \xrightarrow{L} N(0, Q).$$

To complete the proof we need to find the partial derivatives. Taking the derivative with respect to $C_1$ yields

$$\frac{\partial (g(C))}{\partial C_1} = \frac{\partial (C_1 \otimes C_2 - C)}{\partial C_1} = \mathcal{I}_4 \otimes C_2 = G_1,$$
with respect to $C_2$

$$\frac{\partial (g(C))}{\partial C_2} = \frac{\partial (C_1 \tilde{\otimes} C_2 - C)}{\partial C_2} = C_1 \tilde{\otimes} I_4 = G_2,$$

and with respect to $C$

$$\frac{\partial (g(C))}{\partial C} = \frac{\partial (C_1 \tilde{\otimes} C_2 - C)}{\partial C} = -I_8$$

where $I_4$ and $I_8$ are the fourth and eighth order identity tensors. Therefore by using the above partial derivatives we obtain the desired asymptotic distribution which is:

$$\sqrt{N}((\hat{C}_1 \tilde{\otimes} \hat{C}_2 - \hat{C}) - (C_1 \tilde{\otimes} C_2 - C)) \xrightarrow{d} N(0, Q),$$

where $Q$ is given by:

$$Q = \begin{pmatrix} G_1 \\ G_2 \\ -I_8 \end{pmatrix} \star W \begin{pmatrix} G_1 \\ G_2 \\ -I_8 \end{pmatrix}. \quad (3.14)$$

3.6.3 Details on how $W^*$ and $Q^*$ are actually computed in practice

Remark 2. Here we provide more details on how $W^*$ and $Q^*$ are actually computed in practice. Recall that $\hat{C}^*$ is a $K \times J \times K \times J$ array, $\hat{C}_1^*$ is a $K \times K$ matrix, and $\hat{C}_2^*$ is a $J \times J$ matrix. Also $\hat{\Gamma}^*$ is a $K \times J \times K \times J \times K \times J \times J$ array, and can be computed as described in Procedure 4. To find $W^*$ we need to find first the array analogs of $M_1$, $M_2$, $M_3$, which we denote by $M_1^*$, $M_2^*$, $M_3^*$. $M_2^*$ is a $J \times J \times K \times K \times J \times J$ array with entries $M_2^*[j, l, i, j', k, l'] = \delta_{jj'}\delta_{ll'}\delta_{ii'}$, that is, it’s an array with entries zeros and ones. $M_3^*$ is a $K \times J \times K \times J \times K \times J \times J$ array with entries $M_3^*[i, j, k, l, i', j', k', l'] = \delta_{ii'}\delta_{jj'}\delta_{kk'}\delta_{ll'}$, which is the eighth order
identity array. Finally, $M^*_1$ is a $K \times K \times K \times J \times K \times J$ array. To find the entries of $M^*_1$ we first have to calculate $\Delta$ and $I_4$. $\Delta$ is a $K \times K \times K \times J \times K \times J$ array with entries $\Delta[i,k,i',j,k',l] = \delta_{ii'} \delta_{kk'} \delta_{jj'}$, that is, it's an array with entries zeros and ones. $I_4$ is a $K \times J \times K \times J$ identity array, that is an array with entries $I_4[i,j,k,l] = \delta_{ik} \delta_{jl}$. Then, we need to find $\text{Tr}_2(\hat{\mathcal{C}}^*)$, which is a $K \times K$ matrix with entries $\text{Tr}_2(\hat{\mathcal{C}}^*)[i,k] = \sum_{j',j''} \hat{\mathcal{C}}^*_{i,j'j''}kj''$ and the scalar $\text{Tr}(\hat{\mathcal{C}}^*) = \sum_{i',j',j''} \hat{\mathcal{C}}^*_{i',j'j''}i'j'j''$. Combining $\Delta$, $I_4$, $\text{Tr}_2(\hat{\mathcal{C}}^*)$ and $\text{Tr}(\hat{\mathcal{C}}^*)$ we can compute $M^*_1$ by:

$$
M^*_1 = \frac{\Delta \text{Tr}(\hat{\mathcal{C}}^*) - \text{Tr}_2(\hat{\mathcal{C}}^*) \otimes I_4}{(\text{Tr}(\hat{\mathcal{C}}^*))^2},
$$

where the tensor product can be easily implemented by using the R package "tensorA" by van den Boogaart (2007).

Since we have $M^*_1$, $M^*_2$, $M^*_3$ we can compute $W^*$. Notice that $W^*$ has a block structure of the following form:

$$
W^* = 
\begin{pmatrix}
W^*_{11} & W^*_{12} & W^*_{13} \\
W^*_{21} & W^*_{22} & W^*_{23} \\
W^*_{31} & W^*_{32} & W^*_{33}
\end{pmatrix}, \quad (3.15)
$$

where

$$
W^*_{11} = M_1^* \Gamma^* M_1^*, \quad W^*_{12} = M_1^* \Gamma^* M_2^*, \quad W^*_{13} = M_1^* \Gamma^* M_3^* \\
W^*_{21} = M_2^* \Gamma^* M_1^*, \quad W^*_{22} = M_2^* \Gamma^* M_2^*, \quad W^*_{23} = M_2^* \Gamma^* M_3^* \\
W^*_{31} = M_3^* \Gamma^* M_1^*, \quad W^*_{32} = M_3^* \Gamma^* M_2^*, \quad W^*_{33} = M_3^* \Gamma^* M_3^* = \Gamma^*.
$$

$W^*_{11}$ is the variance–covariance array of $\hat{C}_1^*$ with dimensions $K \times K \times K \times K$, $W^*_{12}$ is the covariance between $\hat{C}_1^*$ and $\hat{C}_2^*$ with dimensions $K \times K \times J \times J$, $W^*_{13}$ is the covariance between $\hat{C}_1^*$ and $\hat{C}^*$ with dimensions $K \times K \times J \times K \times J$,
\( W_{21} \) is the covariance between \( \hat{C}_2^\star \) and \( \hat{C}_1^\star \) with dimensions \( J \times J \times K \times K \), \( W_{22}^\star \) is the variance–covariance array of \( \hat{C}_2^\star \) with dimensions \( J \times J \times J \), \( W_{23}^\star \) is the covariance between \( \hat{C}_2^\star \) and \( \hat{C}_1^\star \) with dimensions \( J \times J \times K \times K \), \( W_{31}^\star \) is the covariance between \( \hat{C}_2^\star \) and \( \hat{C}_1^\star \) with dimensions \( K \times J \times J \times K \times K \), \( W_{32}^\star \) is the covariance between \( \hat{C}_2^\star \) and \( \hat{C}_1^\star \) with dimensions \( K \times J \times J \times J \times J \) and finally \( W_{33}^\star \) is the variance–covariance of \( \hat{C}_2^\star \) which is \( \Gamma^\star \).

To compute \( Q^\star \) we need to find the array analogs of the derivatives \( G_1^\star, G_2^\star \) and \( -I_8^\star \), which we denote by \( G_1^\star, G_2^\star \) and \( -I_8^\star \). First, notice that \( -I_8^\star = -M_3^\star \). \( G_1^\star \) is \( K \times K \times K \times J \times J \) array, which can be computed by the tensor product between the identity array of dimensions \( K \times K \times K \times K \) and \( \hat{C}_2^\star \). Similarly, \( G_2^\star \) is \( K \times J \times J \times J \times J \) array, which can be computed by the tensor product between \( \hat{C}_1^\star \) and the identity array of dimensions \( J \times J \times J \times J \).

Since we have \( G_1^\star, G_2^\star \) and \( -I_8^\star \) we can compute \( Q^\star \). Notice that \( Q^\star \) has the following form:

\[
Q^\star = Q_1^\star + Q_2^\star + Q_3^\star + Q_4^\star + Q_5^\star + Q_6^\star + Q_7^\star + Q_8^\star + Q_9^\star \tag{3.16}
\]

where

\[
\begin{align*}
Q_1^\star &= G_1^{\star*} W_{11}^\star G_1^\star, & Q_2^\star &= G_2^{\star*} W_{21}^\star G_2^\star, & Q_3^\star &= -I_8^\star W_{31}^\star G_1^\star; \\
Q_4^\star &= G_1^{\star*} W_{12}^\star G_2^\star, & Q_5^\star &= G_2^{\star*} W_{22}^\star G_2^\star, & Q_6^\star &= -I_8^\star W_{32}^\star G_2^\star; \\
Q_7^\star &= -G_1^{\star*} W_{13}^\star I_8^\star, & Q_8^\star &= -G_2^{\star*} W_{23}^\star I_8^\star, & Q_9^\star &= I_8^\star W_{33}^\star I_8^\star = W_{33}^\star,
\end{align*}
\]

where \( Q_i^\star, i=1,\ldots,9, \) are \( K \times J \times K \times J \times K \times J \times J \) arrays.

### 3.6.4 Proof of Theorem 3.6

**Proof.** Let \( T^\star = \tilde{C}_1 \tilde{C}_2^\star - \tilde{C} \) then we can write
\[ \hat{T} = N\|\hat{C}_1 \otimes \hat{C}_2 - \hat{C}\|^2 = N\langle \hat{C}_1 \otimes \hat{C}_2 - \hat{C}, \hat{C}_1 \otimes \hat{C}_2 - \hat{C} \rangle \]
\[ = N\langle T^* - \Delta + \Delta, T^* - \Delta + \Delta \rangle \]
\[ = N[\|\Delta\|^2 + 2\langle T^* - \Delta, \Delta \rangle + \langle T^* - \Delta, T^* - \Delta \rangle] \]
\[ = N\|\Delta\|^2 + 2N^{1/2}\langle N^{1/2}(T^* - \Delta), \Delta \rangle + \langle N^{1/2}(T^* - \Delta), N^{1/2}(T^* - \Delta) \rangle \]
\[ = N\|\Delta\|^2 + O_P(N^{1/2}) + O_P(1), \]

and the claim follows. \qed
<table>
<thead>
<tr>
<th>N</th>
<th>S = 4</th>
<th>S = 6</th>
<th>S = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>J = 2</td>
<td>6.4 (87%)</td>
<td>5.6 (85%)</td>
<td>3.4 (87%)</td>
</tr>
<tr>
<td>J = 3</td>
<td>5.0 (90%)</td>
<td>5.9 (93%)</td>
<td>5.0 (91%)</td>
</tr>
<tr>
<td>J = 4</td>
<td>4.0 (84%)</td>
<td>4.3 (92%)</td>
<td>5.0 (89%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>S = 4</th>
<th>S = 6</th>
<th>S = 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>J = 2</td>
<td>5.5 (92%)</td>
<td>6.5 (93%)</td>
<td>6.5 (93%)</td>
</tr>
<tr>
<td>J = 3</td>
<td>5.1 (90%)</td>
<td>5.6 (91%)</td>
<td>6.0 (91%)</td>
</tr>
<tr>
<td>J = 4</td>
<td>4.7 (92%)</td>
<td>4.7 (92%)</td>
<td>4.7 (92%)</td>
</tr>
</tbody>
</table>

Table 3.1: Rejection rates under $H_0$ ($c = 0$) at the nominal 5 percent level; $J$ the number of temporal PC’s. The explained variance of the temporal FPCA is given in parentheses.
Table 3.2: Empirical power ($c = 1$); $J$ the number of temporal PC’s. The explained variance of the temporal FPCA is given in parentheses.
Table 3.3: Rejection rates under $H_0$ ($c = 0$); $K$ is the reduced panel dimension and $J$ the number of temporal PC’s. The explained variance of the dimension reduction is given in parentheses.
<table>
<thead>
<tr>
<th>$S = 10$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$N = 100$</th>
<th>$N = 150$</th>
<th>$N = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J = 2$</td>
<td>31.3 (85%)</td>
<td>43.0 (83%)</td>
<td>42.2 (85%)</td>
<td>42.9 (83%)</td>
<td>43.3 (85%)</td>
<td>43.7 (87%)</td>
</tr>
<tr>
<td>$J = 3$</td>
<td>49.2 (87%)</td>
<td>73.3 (83%)</td>
<td>73.5 (85%)</td>
<td>72.5 (83%)</td>
<td>72.0 (85%)</td>
<td>70.4 (87%)</td>
</tr>
<tr>
<td>$J = 4$</td>
<td>60.8 (87%)</td>
<td>81.7 (83%)</td>
<td>84.7 (85%)</td>
<td>82.8 (83%)</td>
<td>82.9 (85%)</td>
<td>80.4 (87%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$S = 12$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$N = 100$</th>
<th>$N = 150$</th>
<th>$N = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J = 2$</td>
<td>30.8 (82%)</td>
<td>42.9 (83%)</td>
<td>43.3 (85%)</td>
<td>43.3 (85%)</td>
<td>43.7 (84%)</td>
<td>43.7 (87%)</td>
</tr>
<tr>
<td>$J = 3$</td>
<td>49.7 (82%)</td>
<td>72.5 (83%)</td>
<td>72.0 (85%)</td>
<td>72.5 (83%)</td>
<td>72.0 (85%)</td>
<td>70.4 (87%)</td>
</tr>
<tr>
<td>$J = 4$</td>
<td>57.4 (87%)</td>
<td>82.8 (83%)</td>
<td>82.9 (85%)</td>
<td>82.8 (83%)</td>
<td>82.9 (85%)</td>
<td>80.4 (87%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$S = 14$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
<th>$N = 100$</th>
<th>$N = 150$</th>
<th>$N = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J = 2$</td>
<td>27.7 (87%)</td>
<td>39.2 (84%)</td>
<td>43.3 (85%)</td>
<td>43.7 (85%)</td>
<td>43.7 (84%)</td>
<td>43.7 (87%)</td>
</tr>
<tr>
<td>$J = 3$</td>
<td>46.2 (87%)</td>
<td>66.6 (85%)</td>
<td>73.5 (85%)</td>
<td>73.5 (85%)</td>
<td>73.5 (85%)</td>
<td>70.4 (87%)</td>
</tr>
<tr>
<td>$J = 4$</td>
<td>55.0 (87%)</td>
<td>81.3 (84%)</td>
<td>84.7 (85%)</td>
<td>84.7 (85%)</td>
<td>84.7 (85%)</td>
<td>80.4 (87%)</td>
</tr>
</tbody>
</table>

Table 3.4: Empirical power ($c = 1$); $K$ and $J$ are as in Table 3.3. The explained variance of the dimension reduction is given in parentheses.
Bibliography


Vita
Panayiotis Constantinou

Education
• Ph.D. candidate in Statistics, The Pennsylvania State University, University Park, Pennsylvania USA
• B.Sc. in Mathematics and Statistics, June, 2012 University of Cyprus, Nicosia, Cyprus

Publications

Submitted Papers

Employment
• The Pennsylvania State University, University Park, Pennsylvania USA