Journal of Statistical Planning and Inference 211 (2021) 326–339

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Projection pursuit based tests of normality with functional data☆

Adam Kolkiewicz, Gregory Rice, Yijun Xie *

University of Waterloo Department of Statistics and Actuarial Science, Mathematics 3 (M3) University of Waterloo 200 University Avenue West Waterloo, ON, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 10 April 2019
Received in revised form 30 June 2020
Accepted 2 July 2020
Available online 11 July 2020

Keywords:
Functional data analysis
Projection pursuit
Goodness-of-fit

ABSTRACT

Methods for validating the assumption of normality of functional data have been only lightly developed to date, with existing methods based primarily on summarizing the data by their projections into random or principal component subspaces, and applying multivariate normality tests to the vectors of scores defining these projections. While this is effective in some cases, we show with both real and synthetic data examples some pitfalls of this approach, including their sensitivity to the basis used to smooth the raw data. We propose a new normality test for functional data based on a projection pursuit that overcomes some of these challenges. Asymptotic theory is developed for the proposed statistics, and we develop several new computational tools needed to implement the high-dimensional projection pursuit. As a by-product of evaluating the test statistic, our method furnishes a way of decomposing functional data into its approximately Gaussian and non-Gaussian components, which is useful for the purpose of data visualization and subsequent analyses. A simulation study and analysis of three data sets demonstrate the complimentary advantages of the proposed test to those currently available in the literature.

1. Introduction

Statistical methods based on the assumption of normality of the observations and/or model errors are ubiquitous in classical statistics, and are also widely used in more modern settings when the data to be analyzed are high-dimensional or functional in nature. To give some recent examples, Panaretos et al. (2010) and Cuevas et al. (2004) assume normality in order to perform two sample and analysis of variance tests with functional data, and in Kowal et al. (2017) and Kowal et al. (2019) normality of the data is used in performing Bayesian inference with complex functional data. We refer the reader to the text books Horváth and Kokoszka (2012) and Ramsay and Silverman (2002) for introductions to functional data analysis, and some further applications of normality in this setting can be found in Gromenko et al. (2017), Yao et al. (2005), and Constantinou et al. (2017), although this list is far from exhaustive. Given the usefulness of these procedures, it is important to have ways of measuring the validity of the assumption of normality for a given sample of functional data. At the least such a validation would lend further credibility to the conclusions of procedures in which normality is assumed, but evidence for normality of functional data may also be of independent interest.

The much related problem of testing for normality in multivariate data enjoys an enormous literature dating back at least to the 1960s. A myriad of techniques are now available, and, crudely, they can be categorized into four groups

☆ The three authors contribute to all aspects of this paper.

* Corresponding author.
E-mail address: yijun.xie@uwaterloo.ca (Y. Xie).

https://doi.org/10.1016/j.jspi.2020.07.001
0378-3758/

based on two characteristics. The first is how departures from normality in the data are measured, in which typically either moment based measures are used, such as the sample skewness and kurtosis, or goodness-of-fit tests involving the empirical distribution or characteristic function are employed. The second is how information is aggregated across the coordinates of the data, which usually amounts to either pooling/averaging the information across coordinates, or searching for linear combinations of the coordinates that maximize a given measure of non-Gaussianity. Approaches following the later paradigm are often termed "projection pursuit" methods, since finding such a linear combination can be framed as a classical projection pursuit problem as put forward in Kruskal (1972), and Friedman and Tukey (1974). Canonical test statistics based on moment methods of each type are Mardia's multivariate skewness (Mardia et al., 1979), which aggregates the skewness across coordinates, and the skewness measure of Malkovich and Afifi (1973), which is the maximal sample skewness among all linear combinations of the coordinates. One test is expected to be preferable to the other depending on how "sparse" the non-Gaussianity is in the data: data for which all linear combinations of the coordinates are non-Gaussian should be more apparently non-Gaussian by considering aggregation based methods, while non-Gaussianity that can be explained by only a few linear combination of the coordinates would typically be more easily detected using projection pursuit methods. Some examples of multivariate projection pursuit based normality tests can be found in Liang et al. (2000), Henze and Wagner (1997), Baringhaus and Henze (1991), Zhu et al. (1995a), Zhu et al. (1995b), and general reviews of tests for multivariate normality are given in Mecklin and Mundfrom (2004), Henze (2002), and Szekely and Rizzo (2005).

In contrast, testing for normality of functional data objects has received considerably less attention. Methods based on random projections and subsequent Carmér–von Mises and Kolmogorov–Smirnov type goodness-of-fit tests are proposed and reviewed in Cuesta-Albertos et al. (2006), Cuesta-Albertos et al. (2007), Bugni et al. (2009), and Cuevas (2014). To date and to the knowledge of the authors, the only test available for this purpose based on moment methods was put forward in Górecki et al. (2018), henceforth referred to as the GHHK test. Their approach involves projecting the functional data onto the span of the first several functional principal components estimated from the data, and then applying a test based on combining Mardia's skewness and kurtosis to the vectors of coefficients defining these projections, i.e. applying a multivariate Jarque–Bera test (Jarque and Bera, 1980) to the projection scores. They also extend their method to serially correlated functional data. While this method proves to be effective in many cases, it evidently might be improved upon in several others. One is if the non-Gaussian components of the data are sparse among the leading principal components, analogously to the multivariate setting, but another is if the non-Gaussian components of the data are orthogonal to the leading principal components, in which case the GHHK test would not be expected to have more than trivial power. As we see below, this latter situation might occur more often than one might think, as it can arise from simply misspecifying the basis used to smooth/generate functional data objects from raw data and/or estimate the functional principal components. Although one may argue that this situation could be avoided by including more principal components, as later shown in a data example in Section 4.3, increasing the number of principal components incorporated into the GHHK test may not solve the problem.

In this paper, we propose and study an alternative normality test for functional data based on projection pursuit that overcomes some of these challenges. We consider as test statistics the maximal sample skewness and sample kurtosis among all scalar projections of the data onto a user selected compact subset of the unit ball, and hence the proposed test can be thought of as a functional generalization of the tests of Malkovich and Afifi (1973) and Baringhaus and Henze (1991). We show that the compact subset selected can be taken to be relatively high dimensional, and can also be generated by the functional principal components, which gives the test complimentary strengths to the GHHK test. A complete asymptotic theory is developed for the proposed statistics, and computational tools are introduced to conduct the required high-dimensional projection pursuit. These computational tools may be of independent interest since they may be used for more general projection pursuit problems with functional data. In addition to providing a test for Gaussianity, this projection pursuit method also furnishes a way to decompose functional data into a direct sum of approximately Gaussian and non-Gaussian components useful for data visualization or subsequent analyses, which we demonstrate. This latter application builds upon some recent efforts to develop projection pursuit methods for functional data; see for example Bali et al. (2011). We study the proposed methods and compare them to the GHHK method in a simulation study, as well as in three applications to real data sets, which show the complimentary strengths of the two tests.

The rest of the paper is organized as follows: In Section 2, we define our projection pursuit-based test statistics, and present their asymptotic properties. In Section 3, we detail several computational methods useful for calculating the proposed statistics and their critical values, and also describe and present the results of a simulation study. The results of the data analyses are presented in Section 4. The proofs of all technical results are contained in the online supplementary material.

Below we use the following notation: We let $L^2([0,1],\mathbb{R})$ denote the space of real valued functions defined on the unit interval with finite squared integral, which is a Hilbert space when equipped with the inner product defined for $x, y \in L^2([0,1],\mathbb{R})$ by $\langle x, y\rangle = \int_0^1 x(t)y(t)dt$. The corresponding norm is defined by $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$.

2. Problem statement, definition of test statistics, and their asymptotic properties

Suppose that $X_1(t), \ldots, X_n(t)$, $t \in [0,1]$, is a simple random sample of size n of functional data sharing the same distribution as X. We assume throughout that each functional observation is then an independent stochastic process,

whose sample path is in $L^2([0,1],\mathbb{R})$. Since typically functional data are only observed at some discrete collection of points, here we assume that the sample of curves under consideration have been obtained after a pre-processsing step, such as basis smoothing or principal analysis by conditional expectation as described in Ramsay and Silverman (2002) or in Yao et al. (2005). More generally, we could consider a simple random sample of elements from a general, separable, Hilbert space, but because of the type of data applications we present in this paper we consider the space $L^2([0,1],\mathbb{R})$ for clarity of presentation. Given this data, we are interested in testing the null hypothesis

$$\mathcal{H}_0 : X \text{ is a Gaussian process in } L^2([0,1],\mathbb{R}).$$

By definition, $\mathcal{H}_0$ can be equivalently stated as

$$\mathcal{H}_0 : \text{ For each nonzero } v \in L^2([0,1],\mathbb{R}), \text{ the scalar random variable } \langle X, v\rangle \text{ is Gaussian.}$$

As discussed in the multivariate setting in Malkovich and Afifi (1973), the latter formulation motivates developing test statistics aiming to find the "least Gaussian" projection of X. Indeed, if the distribution of such a projection does not significantly deviate from normality, then the same apparently holds for the entire process. In order to evaluate the normality of the projection of the data onto the direction v, a natural measure is the squared skewness and/or the absolute kurtosis:

$$S_n(v) = \frac{1}{n^2\hat{\sigma}^6(v)}\left[\sum_{i=1}^{n}(\langle X_i, v\rangle - \langle \bar{X}, v\rangle)^3\right]^2,$$

and

$$K_n(v) = \left|\frac{1}{n\hat{\sigma}^4(v)}\sum_{i=1}^{n}(\langle X_i, v\rangle - \langle \bar{X}, v\rangle)^4 - 3\right|.$$

Above we use $\bar{X}(t) = (1/n)\sum_{i=1}^{n} X_i(t)$ to denote the sample mean function and $\hat{\sigma}^2(v)$ to denote the sample variance of the scalar observations $\langle X_1, v\rangle, \ldots, \langle X_n, v\rangle$. Though here we consider "Jarque–Bera" moment based evaluations of normality, one could also consider projection pursuit methods based on other measures, for instance those surveyed in Thadewald and Büning (2007). Some benefits of using such moment based measures in this setting stem from their affine invariance and asymptotic properties, which, as we shall see below, are crucial in deriving feasible computational techniques to carry out a projection pursuit test in high dimensions.

Letting $U^\infty = \{u \in L^2([0,1],\mathbb{R}) : \|u\| = 1\}$ denote the unit sphere in $L^2([0,1],\mathbb{R})$, the least Gaussian projection may be calculated by evaluating the test statistics

$$S_n = \sup_{v\in U^\infty} S_n(v), \text{ and } K_n = \sup_{v\in U^\infty} K_n(v).$$

An issue that presents itself here, in contrast with the multivariate setting, is that these statistics are not necessarily well defined (finite), owing to the fact that the unit sphere in $L^2([0,1],\mathbb{R})$ is not compact. An obvious way to fix this is to restrict the search for projections of the data to compact subsets of U^∞, which, as a result of Riesz's lemma (see e.g., Riesz and Sz.-Nagy (1990)), must be finite dimensional. Such a finite dimensional subset must be spanned by a finite collection of orthonormal basis functions, and hence a natural way to explore compact subsets of U^∞ is to consider those that intersect a k dimensional linear subspace of the form $L_k = \text{span}(\phi_1, \ldots, \phi_k)$, for some orthonormal basis elements $\phi_1, \ldots, \phi_k$ chosen by the practitioner. For a chosen subspace L_k, we then instead consider the statistics

$$S_n^{L_k} = \sup_{v\in U^\infty\cap L_k} S_n(v), \text{ and } K_n^{L_k} = \sup_{v\in U^\infty\cap L_k} K_n(v). \tag{2.1}$$

These statistics are well defined if the functions $S_n(v)$ and $K_n(v)$ are at least continuous (almost surely) over $U^\infty \cap L_k$, which holds under quite mild conditions in addition to $\mathcal{H}_0$, basically entailing that L_k is not orthogonal to the data. Effectively, these statistics are measuring for multivariate normality in the subspace L_k based on the third and fourth order moments.

One might choose the subspace L_k and its dimension based on a number of considerations. If the observed functional data have been obtained by smoothing over a particular basis, such as the Fourier basis or a spline basis, then that basis, with the dimension used for smoothing, is a natural choice for the subspace L_k. If departures from normality are sought or expected in a particular way, then this information can also be used to select the basis. For instance, if it is believed that the functional data exhibits non-normality on a subset of its domain, then a Haar basis could be used.

In the case when one would like a parsimonious finite dimensional representation of the observed functional data, functional principal component analysis is often employed. In order to introduce the principal component basis, we assume for the moment that $E\|X_i\|^2 < \infty$, and we let $C(t,s) = \text{cov}(X_i(t), X_i(s))$. The covariance function C defines a Hilbert–Schmidt integral operator of the form

$$c(f)(t) = \int_0^1 C(t,s)f(s)ds.$$

Let v_i be the orthonormal eigenfunctions of c with the corresponding eigenvalues λ_i satisfying

$$c(v_i)(t) = \lambda_i v_i(t), \quad \lambda_1 \geq \lambda_2 \geq \cdots. \tag{2.2}$$

The subspace $P_k = \text{span}(v_1, \ldots, v_k)$ is well known to be optimal in terms of representing the curves X_i in a k dimensional space, at least when the error is measured in the mean squared sense. These basis elements are generally unknown in practice but can be estimated via estimates of C. The natural estimator of C is

$$\hat{C}(t,s) = \frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \bar{X}(t))(X_i(s) - \bar{X}(s)),$$

which gives an estimator of c

$$\hat{c}(f)(t) = \int_0^1 \hat{C}(t,s)f(s)ds,$$

and subsequently estimators $\hat{v}_i$ and $\hat{\lambda}_i$ of v_i and λ_i satisfying

$$\hat{c}(\hat{v}_i)(t) = \hat{\lambda}_i \hat{v}_i(t), \quad i = 1, \ldots, n. \tag{2.3}$$

With this notation, we note that the test statistic proposed in Górecki et al. (2018) is of the form

$$\text{GHHK}_k = \sum_{i=1}^{k}[S_n(\hat{v}_i) + K_n^2(\hat{v}_i)],$$

which, under the condition that the first k eigenvalues in (2.2) are bounded away from zero and with suitable normalization, converges in distribution to a χ^2-random variable under $\mathcal{H}_0$. Letting $\hat{P}_k = \text{span}(\hat{v}_1, \ldots, \hat{v}_k)$, one might alternatively test for normality in the principal component subspace by considering the statistics

$$S_n^{\hat{P},k} = \sup_{v \in U^\infty \cap \hat{P}_k} S_n(v), \text{ and } K_n^{\hat{P},k} = \sup_{v \in U^\infty \cap \hat{P}_k} K_n(v), \tag{2.4}$$

or

$$M_n^{\hat{P},k} = \max_{1 \leq i \leq k} \frac{n}{6}\left(S_n(\hat{v}_i) + \frac{1}{4}K_n^2(\hat{v}_i)\right), \tag{2.5}$$

in which the maximal sum of the skewness and kurtosis is evaluated only over the first k principal component directions.

2.1. Large sample properties

The asymptotic properties of each of these statistics under $\mathcal{H}_0$ are detailed by the following two results.

Theorem 2.1. *Suppose $X_1, \ldots, X_n$ are independent and identically distributed elements of $L^2([0,1], \mathbb{R})$ such that*

1. *$\mathcal{H}_0$ holds, and*
2. *$\inf_{v \in U^\infty \cap L_k} E\langle X_i, v\rangle^2 > 0$.*

Then, with $S_n^{L_k}$ and $K_n^{L_k}$ defined in (2.1),

$$(nS_n^{L_k}, \sqrt{n}K_n^{L_k})^\top \xrightarrow{\mathcal{D}} (\sup_{v \in U^\infty \cap L_k} Z_1^2(v), \sup_{v \in U^\infty \cap L_k} |Z_2(v)|)^\top,$$

where Z_1 and Z_2 are independent mean zero Gaussian processes defined on $U^\infty \cap L_k$, whose covariance functions, defined in the supplementary material, depend only on k.

This result may be proven in a similar fashion to the main theorem of Baringhaus and Henze (1991). We also note here that an asymptotic result of this type can easily be established under the more general condition that the projections of X onto L_k are elliptically symmetric, but we do not pursue that here. The asymptotic distribution presented in Theorem 2.1 can be used to estimate valid critical values for each test statistic under $\mathcal{H}_0$ using simulation. Furthermore, the form of this distribution shows that the tests based on $S_n^{L_k}$ and $K_n^{L_k}$ are asymptotically independent, which is useful in calculating a p value for $\mathcal{H}_0$ using both statistics jointly.

In order to derive similar results when the subspace used to define the test statistics is random and generated from the principal component basis, we make the following assumption.

Assumption 2.1. The eigenvalues λ_i defined in (2.2) satisfy $\lambda_1 > \cdots > \lambda_k > \lambda_{k+1} \geq 0$.

Assumption 2.1 implies that the principal component subspaces are asymptotically one dimensional, and in particular it implies that the estimated principal components are consistent up to a sign. This assumption could likely be relaxed to the one that only requires $\lambda_k > \lambda_{k+1}$ at the expense of some simplicity in the proof.

Theorem 2.2. *Suppose Assumption 2.1 holds, and that $X_1, \ldots, X_n$ satisfy $\mathcal{H}_0$ and are independent and identically distributed. Then with $S_n^{\hat{P},k}$ and $K_n^{\hat{P},k}$ defined in (2.4),*

$$(nS_n^{\hat{P},k}, \sqrt{n}K_n^{\hat{P},k})^\top \xrightarrow{\mathcal{D}} (\sup_{v\in U^\infty\cap P_k} Z_1^2(v), \sup_{v\in U^\infty\cap P_k} |Z_2(v)|)^\top,$$

where Z_1 and Z_2 are independent mean zero Gaussian processes defined on $U^\infty \cap P_k$. Their covariance functions are defined in the supplementary material. Furthermore,

$$M_n^{\hat{P},k} \xrightarrow{\mathcal{D}} \max_{1\le i\le k} \chi_i^2(2),$$

where $\chi_i^2(2),\ i = 1, \ldots, k$, denote independent and identically distributed χ^2 random variables with two degrees of freedom.

Theorem 2.2 shows that at least when the principal component subspaces are fixed and one dimensional, the distribution of the maximal skewness and kurtosis is not asymptotically affected by the error in estimating the principal components. This comes basically as a result of the continuity of the functions $S_n(v)$ and $K_n(v)$. This result also shows that a test of asymptotic size α is obtained by rejecting $\mathcal{H}_0$ when $M_n^{\hat{P},k}$ exceeds $\chi^2([1-\alpha]^{1/k}, 2)$, where $\chi^2(\beta, 2)$ is the β^{th} quantile of the χ^2 distribution with two degrees of freedom.

3. Implementation and a simulation study

Practical evaluation of the estimates of the test statistics $\hat{S}_n^{L_k}$ and $\hat{K}_n^{L_k}$ defined in (2.1) requires maximizing the objective functions $S_n(v)$ and $K_n(v)$ over a potentially high-dimensional unit sphere, which presents a difficult optimization problem. In previous projection pursuit based tests for normality with multivariate data, this is often handled using number theoretic optimization tools that effectively aim to maximize $S_n(v)$ and $K_n(v)$ by evaluating each function at a dense collection of points on the unit sphere; see for example Zhu et al. (1995a). Due to the potentially high dimension of the unit sphere that may arise in the present application, traditional methods for generating dense collections of points have proven ineffective. One reason for the poor performance of such methods is the fact that the selected points may not cover the search area evenly. The method that we propose here to address this issue borrows from recent advances in the generation of low discrepancy sequences developed in the context of quasi Monte Carlo integration.

To explain the main idea behind these sequences, consider a p-dimensional unit hybercube $[0,1)^p$. Let $Z = \{\mathbf{x}_j \in [0,1)^p,\ j = 0, 1, 2, \ldots\}$ be a sequence of points in the cube, $[a, b) = \{x \in [0,1)^p : a_i \le x_i < b_i, i = 1, \ldots, p\}$ denote a sub-rectangular prism, and $A([\mathbf{a}, \mathbf{b}), N)$ be the number of the first N points from Z that lie in $[a, b)$.

A desirable property of the sequence Z is that

$$\lim_{N\to\infty} \frac{A([\mathbf{a}, \mathbf{b}), N)}{N} = \lambda_p([\mathbf{a}, \mathbf{b}))$$

for any selection of the rectangle $[a, b)$, where λ_p denotes the p-dimensional Lebesgue measure. In order to quantify the rate at which the fraction $A([\mathbf{a}, \mathbf{b}), N)/N$ converges to the limit, different measures of discrepancy have been proposed in the literature. Among them, the following star discrepancy has received perhaps the most attention to date:

$$D_N^*(S) = \sup_{\mathbf{b}\in[0,1]^p} \left| \frac{A([\mathbf{0}, \mathbf{b}), N)}{N} - \lambda_p([\mathbf{0}, \mathbf{b})) \right|.$$

In the context of numerical integration methods, the importance of star discrepancy stems from the Koksma–Hlawka inequality, which provides an upper bound for the error estimate for quasi Monte Carlo integration rules (see, for example, Niederreiter (1992) or Leobacher and Pillichshammer (2014)). This bound depends on the underlying integration nodes only through the star discrepancy, and this explains why sequences with low discrepancy are desirable.

As demonstrated by numerous authors, sequences with low discrepancy can also improve efficiency of some global optimization methods (for example, Kimura and Matsumura (2007), Pant et al. (2008), Georgieva and Jordanov (2009), and Monica et al. (2011)). In our problem, the goal is to generate a low discrepancy sequence on the unit sphere that could be used to efficiently explore $U^\infty \cap L_k$. The recent work by Brauchart et al. (2015) provides an algorithm for generating such a sequence, which we use to propose a two-step optimization method to estimate the test statistics. In the rest of this sub-section we describe how we evaluate only $\hat{S}_n^{L_k}$, since $\hat{K}_n^{L_k}$ can be evaluated similarly.

First, we generate a low discrepancy sequence of length J on $U^\infty \cap L_k$ as described in Brauchart et al. (2015). Denote these points by $\boldsymbol{\xi}_j = (\xi_{j,1}, \ldots, \xi_{j,k})$ for $j = 1, 2, \ldots, J$. For each $\boldsymbol{\xi}_j$, and selected basis functions $\phi_i(t),\ i = 1, \ldots, k$, spanning L_k, we construct functions of the form $u_j(t) = \sum_{k=1}^K \xi_{j,k}\phi_k(t)$. Then we calculate the skewness of the projection of our data onto $u_j(t)$ as

$$Sk_j = S_n(u_j), \quad j = 1, \ldots, J.$$

The Sk_j's may then be ranked, and we denote the largest M of them as $Sk_{(1)} \ge \cdots \ge Sk_{(M)}$. We denote the low-discrepancy points that produce $Sk_{(m)}$ by $\boldsymbol{\xi}_{(m)}$, where $m = 1, \ldots, M$.

In a second step, to maximize $S_n(v)$ we apply M times a local optimization procedure where as initial points we use $\boldsymbol{\xi}_{(m)}$, $m = 1, \ldots, M$. In our implementation of the method we have used the L-BFGS-B algorithm proposed by Byrd et al. (1995), which allows the user to specify constraints on the domain over which the objective function is optimized. We also tried other optimization techniques, such as particle swarm optimization (Clerc, 2010) and conjugate gradient descendant (Fletcher and Reeves, 1964), but the results were almost identical. Therefore, we only report results from L-BFGS-B method, since it is slightly faster. Let $\tilde{\boldsymbol{\xi}}_{(m)}$ denote the point at which S_n is optimized starting from the function on the unit sphere corresponding to the initial point $\boldsymbol{\xi}_{(m)}$, $m = 1, \ldots, M$. Then our final estimated vector of coefficients $\hat{\boldsymbol{\xi}}$ is determined as

$$\hat{\boldsymbol{\xi}} = \{\tilde{\boldsymbol{\xi}}_{(m)} : Sk(\tilde{\boldsymbol{\xi}}_{(m)}) = \max_{m=1,\ldots,M} Sk(\tilde{\boldsymbol{\xi}}_{(m)})\}.$$

For

$$\hat{u}(t) = \sum_{k=1}^{K} \hat{\xi}_k \hat{v}_k(t), \tag{3.1}$$

our estimated test statistic is then given by

$$\hat{S}_n^{L_k} = S_n(\hat{u}). \tag{3.2}$$

This procedure is similar to the coarse-to-fine optimization schemes popular in the machine learning community (see, for example, Pedersoli et al. (2015) and Charniak and Johnson (2005) for two applications in computer vision and natural language processing). Our algorithm is summarized in Algorithm 3.1:

Algorithm 3.1: Two-Step Approximation Algorithm for $\hat{S}_n^{L_k}$

Input: $x_1(t), \cdots, x_n(t)$, $\phi_1(t), \cdots, \phi_k(t)$
Result: $\hat{S}_n^{L_k}$
generate $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_J$;
for $j = 1$ **to** J **do**
  generate $u_j(t) = \sum_{l=1}^{k} \xi_{jl}\phi_l(t)$;
  calculate $Sk_j = S_n(u_j)$ where the requisite integration is approximated using a Riemann integral;
end
rank $Sk_1, \cdots, Sk_J$ in decreasing order as $Sk_{(1)}, \cdots, Sk_{(J)}$;
for $m = 1$ **to** M **do**
  find $\boldsymbol{\xi}_{(m)}$ corresponding to $Sk_{(m)}$;
  find the spherical coordinate $\{1, \theta_{(m),1}, \cdots, \theta_{(m),k-1}\}$ of $\boldsymbol{\xi}_{(m)}$;
  fix the $(k-1)$-dimension hypercube $[\theta_{(m),1} - 0.2\pi, \theta_{(m),1} + 0.2\pi] \times \cdots \times [\theta_{(m),k-1} - 0.2\pi, \theta_{(m),k-1} + 0.2\pi]$;
  find optimized $\tilde{\boldsymbol{\xi}}_{(m)}$ in this hypercube with the objective function equal to negative value of skewness;
end
let $\hat{\boldsymbol{\xi}} = \{\tilde{\boldsymbol{\xi}}_{(m)} : Sk(\tilde{\boldsymbol{\xi}}_{(m)}) = \max_{m=1,\cdots,M} Sk(\tilde{\boldsymbol{\xi}}_{(m)})\}$;
construct $\hat{u}(t) = \sum_{l=1}^{k} \hat{\xi}_l \phi_l(t)$;
calculate $\hat{S}_n^{L_k} = S_n(\hat{u})$.

This procedure necessitates the selection of two tuning parameters: the length of the low discrepancy sequence J and the number of initial points M. Our recommended procedure is to start from some initial values, like those we propose below, and stop as soon as we observe that the hypothesis testing decision and/or p-values are not sensitive to increasing values of these parameters. This is equivalent to checking that the statistic calculated and null quantiles estimated achieve stability as M and J increase. We have conducted a number of simulations to investigate what choices for these parameters are appropriate in practice. The results of some of these experiments are discussed and shown in Figure 2.1 in the supplementary material. In terms of stability in estimating the quantiles of the test statistics defined in Section 2.1, we have found that reasonable choices of these parameters in a dimension of 21 or less are $J = 3 \times 10^4$ and $M = 5$. We also illustrate here how one might choose the dimension of the subspace k in practice, with an additional discussion presented in the supplementary material. A natural idea is to perform the test for a range of values of k in order to further understand how any non-Gaussianity is manifested in the data, or if k should potentially be increased. For Gaussian data, one expects that, as a function of k, the p-values of the test applied for different choices of k will fluctuate as dependent uniform random variables on [0, 1], while for non-Gaussian data the p-values as a function of k should at some point become small.

In order to estimate the null distributions of $S_n^{L_k}$ and $K_n^{L_k}$, we utilize the fact that their limiting distributions are pivotal under $\mathcal{H}_0$ and estimate their critical values by simulation. In particular, letting q_α^S and q_α^K denote the α quantiles

of $S_n^{L_k}$ and $K_n^{L_k}$ respectively, these are approximated by generating n k-dimensional multivariate normally distributed random vectors, $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,k})^\top$, $i = 1, \ldots, n$, with mean zero and identity covariance matrix. A functional sample $Y_i(t) = \sum_{j=1}^k y_{i,j}\phi_j(t)$ can be constructed from these vectors, to which we apply Algorithm 3.1 to calculate the statistics $S_{n,1}^{L_k}$ and $K_{n,1}^{L_k}$. By repeating this simulation B times we obtain a sample from statistics $S_{n,j}^{L_k}$ and $K_{n,j}^{L_k}$, $j = 1, \ldots, B$, and then we take q_α^S and q_α^K to be the α empirical quantiles of these respective samples. Below we take $B = 2000$ for estimating critical values. We found that this number is suitable for estimating the 1% and 5% critical values, although we recommend that it be increased if one wishes to consider values even further in the tail of the distribution.

We can estimate a p-value for the test based on these statistics as follows. With $\hat{S}_n^{L_k}$ and $\hat{K}_n^{L_k}$ denoting the test statistics estimated from the data, we take

$$p = P(S_n^{L_k} > \hat{S}_n^{L_k} \cup K_n^{L_k} > \hat{K}_n^{L_k}) = 1 - P(S_n^{L_k} \le \hat{S}_n^{L_k} \cap K_n^{L_k} \le \hat{K}_n^{L_k}) \approx 1 - P(S_n^{L_k} \le \hat{S}_n^{L_k})P(u_n^{L_k} \le \hat{K}_n^{L_k}),$$

where the last approximation is justified by the asymptotic independence of $S_n^{L_k}$ and $K_n^{L_k}$. The probabilities $P(S_n^{L_k} \le \hat{S}_n^{L_k})$ and $P(K_n^{L_k} \le \hat{K}_n^{L_k})$ can be estimated from the empirical CDF estimated from the simulation described above.

3.1. Simulation study

In order to evaluate the performance of the tests and the numerical methods proposed above, we conducted a simulation study, the results of which we now present. The synthetic data that we considered was generated from the basic model

$$X_i(t) = \sum_{j=1}^{D} \epsilon_{i,j} f_j(t), \tag{3.3}$$

where $D = 101$, and $f_1, \ldots, f_{101}$ are the first 101 Fourier basis functions defined as $f_1 = 1$, $f_j(t) = \sqrt{2}\sin(\frac{j-1}{2}\pi t)$ for $j = 3, 5, \ldots, 101$, and $f_j(t) = \sqrt{2}\cos(\frac{j}{2}\pi t)$ for $j = 2, 4, \ldots, 100$. We also studied the case in which the basis elements f_j were non-smooth Haar basis elements. Our results, which for completeness are presented in Section 4 of the supplementary material, suggested that the performances of the tests were very similar to those in the smooth case.

We produced raw discrete data from the model (3.3) by evaluating $X_i(t)$ at 100 equally spaced points in the unit interval. To simulate data following $\mathcal{H}_0$, we generated the coefficient vectors $\epsilon_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,D})^\top$ from a multivariate normal distribution with mean zero and covariance matrix $\Sigma = \Sigma_{D\times D}$. We considered three different types of the covariance structure. In two cases, Σ was diagonal,

$$\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_D^2),$$

where we either took the diagonal elements to decay quickly, so that

$$\sigma_w^2 = \frac{1}{w^2}, \quad w = 1, \ldots, D,$$

and the resulting covariance matrix was labeled Σ_{fast}, or more slowly, in which case we took

$$\sigma_w^2 = \begin{cases} \frac{1}{\sqrt{w}} \text{ for } w = 1, 2, 3 \\ \frac{2.4065}{w^2} \text{ for } w \ge 4, \end{cases}$$

and then the resulting covariance matrix was labeled Σ_{slow}. The normalizing constant 2.4065 was computed so that for both covariance matrices

$$\frac{\sum_{i=1}^7 \sigma_i^2}{tr(\Sigma)} \approx 0.9,$$

which is a common threshold when using the total variance explained (TVE) in principal component analysis to select the number of components to retain. We should note that the TVE level could be arbitrary, for example in Górecki et al. (2018) the authors use 85%. In the rest of this paper, we use 90% as the threshold for TVE.

For these diagonal covariance matrices the first d population level principal components of the observations X_i, $i = 1, \ldots, n$, are the functions f_1,..., f_d. Since initial Fourier basis elements do not fluctuate too much, the first d principal components can be estimated quite accurately using most standard initial smoothing methods, like those based on B-splines. In order to investigate the situation in which the estimation of the principal components might be sensitive to the choice of the basis used to smooth the raw data, we also considered generating data having a randomly constructed covariance matrix Σ_{ran} in the following way: we represent $\Sigma = P\Lambda P^{-1}$, where P is a $D \times D$ matrix whose columns are orthonormal to each other, and Λ is a diagonal matrix. We generate P by applying a QR decomposition to a $D \times D$ matrix filled by independent and identically distributed normal random variables with zero mean and unit variance, and we take $\Lambda = diag(101, 100, \ldots, 1)$. In this case the leading principal components of X_i in (3.3) are equally likely to be any of the functions $f_1, \ldots, f_D$, or linear combinations of them, and the eigenvalues of the covariance matrix decay quite slowly.

In order to generate data under $\mathcal{H}_A$, we consider three alternatives, which we label as $L1$, $L3$, and $M10$. For the alternative $L1$, we assume that the leading error term $\epsilon_{i,1}$ in (3.3) follows a scaled t-distribution with 5 degrees of freedom, mean zero, and variance equal to $\Sigma(1, 1)$. In $L3$, the first three leading coefficients $\epsilon_{i,1}$, $\epsilon_{i,2}$, $\epsilon_{i,3}$ follow independently a scaled t-distribution with 5 degrees of freedom and variances $\Sigma(1, 1)$, $\Sigma(2, 2)$, and $\Sigma(3, 3)$, respectively. In the last case $M10$, we assume $\epsilon_{i,10}$ follows the scaled t-distribution with 5 degrees of freedom and variance equal to $\Sigma(10, 10)$. In both of the cases L_1 and L_3, the non-Gaussianity of the observations is contained in the leading principal components, and hence the methods based on PCA are expected to perform well. In contrast, for the alternative $M10$ the non-Gaussian component is orthogonal to the PCA subspaces of dimensions nine or less.

To conduct the simulations, for each setting we generated 1000 samples of lengths $n = 150$, 450 and 900. For each sample of curves, to estimate the test statistics $S_n^{L,21}$ and $K_n^{L,21}$ we applied the approximation method described in the above algorithm. For linear spaces L_{21}, we considered $F_{21} = \text{span}(f_1, \ldots, f_{21})$, where f_i are the Fourier basis elements described above, and $B_{21} = \text{span}(b_1, \ldots, b_{21})$, where b_i are ortho-normalized B-splines constructed from 75 equally spaced knots of order 4.

We considered the following tests:

1. **PP-F-21**: Projection pursuit test with the subspace spanned by F_{21}.
2. **PP-B-21**: Projection pursuit test with the subspace spanned by B_{21}.
3. **PP-PF-7**: Projection pursuit test with the subspace spanned by the first 7 functional principal components estimated by initially smoothing the raw data using the Fourier basis.
4. **PP-PB-7**: Projection pursuit test with the subspace spanned by the first 7 functional principal components estimated by initially smoothing the raw data using the B-spline basis.
5. **GHHK-F**: GHHK test where we smooth the data using the first 75 Fourier basis functions and then estimate the principal components from the coefficients. We use the 90% TVE criterion to select the number of principal components included.
6. **GHHK-B**: GHHK test where we smooth the data using 75 B-spline basis functions and then estimate the principal components from the coefficients. We use the 90% TVE criterion to select the number of principal components included.
7. **MAX-F**: MAX test defined in (2.5) with data smoothed by Fourier basis. We use the 90% TVE criterion to select the number of principal components included.
8. **MAX-B**: MAX test defined in (2.5) with data smoothed by B-spline basis. We use the 90% TVE criterion to select the number of principal components included.

The percentage of rejections from the 1000 simulations at levels 5% and 1% are presented in Tables 3.1–3.3 for each covariance structure. The numbers in the Null column show the test sizes for different methods, while the numbers in L1, L3, and M10 columns show the power of each test under these three scenarios. The results can be summarized as follows:

- Each test exhibited reasonable size. The GHHK test and the MAX type tests were a bit oversized for large n, while the projection pursuit based tests tended to be a bit undersized.
- For the covariance structures Σ_{fast} and Σ_{slow} and the alternatives $L1$ and $L3$, the GHHK and MAX type tests performed superiorly and worked well regardless of the basis used to smooth the data. The projection pursuit based tests exhibited good power and consistency in these cases. By comparing the results for **PP-PF-7** and **PP-F-21**, one can get a sense of the sacrifice in power that is made by increasing the dimension of the search space, which can be quite severe: when the dimension increased from 7 to 21, the power was roughly halved at the significance levels of 5% and 1%.
- As expected, in the case $M10$ the GHHK test, the MAX type test, and the projection pursuit tests based on functional principal components have no more than trivial power, while the power of the other projection pursuit tests is very similar to what was observed under the alternative $L1$.
- When the covariance matrix used to generate the data was Σ_{ran}, then the performance of the GHHK test was strongly affected by the choice of basis used to smooth the raw data. When the Fourier basis was used, the GHHK test still exhibited strong, although somewhat diminished, power. On the other hand, when orthogonal B-splines were used to smooth the data, then the power was strongly diminished. This can be explained by the fact that the non-Gaussian signal in these cases often ends up in the Fourier basis elements that cannot be well represented by the first seven principal components calculated after initially smoothing the raw data using the orthogonal B-splines. In this case, the projection pursuit type tests are essentially unaffected by the choice of the basis, since even when the non-Gaussian component of the data is not well represented in the early principal components, it remains present in some linear combinations of the coordinates of the full data and can be essentially recovered without loss by the projection pursuit optimization.

Table 3.1
Percentage of rejections under the fast decaying covariance matrix Σ_{fast}.

Level	Method	$\alpha = 5\%$				$\alpha = 1\%$			
		Null	L1	L3	M10	Null	L1	L3	M10
n = 150	PP-F-21	5.3	13.7	29.9	15.1	1.6	8.7	22.2	9.4
	PP-B-21	2.9	15.0	34.4	15.8	1.0	9.8	24.5	9.3
	PP-PF-7	3.1	34.5	65.1	3.8	0.8	23.6	48.9	1.5
	PP-PB-7	3.7	33.0	65.7	11.0	0.8	20.4	50.2	6.6
	GHHK-F	4.7	69.1	97.4	4. 9	1.5	59.4	95.8	1.4
	GHHK-B	4.0	66.0	96.9	5.9	1.6	57.7	94.3	3.2
	MAX-F	7.2	74.0	97.1	7.1	2.7	66.4	95.3	2.8
	MAX-B	7.7	73.7	96.9	9.1	3.3	65.8	95.0	4.9
n = 450	PP-F-21	4.9	34.1	70.8	35.5	1.1	25.4	60.9	26.0
	PP-B-21	3.9	38.3	75.6	38.6	0.7	28.7	64.5	29.0
	PP-PF-7	3.3	79.1	99.1	3.7	0.3	65.2	95.1	0.4
	PP-PB-7	4.4	80.4	99.5	10.5	0.5	66.1	95.8	6.2
	GHHK-F	5.5	98.3	100	5.5	2.0	96.3	100	2.0
	GHHK-B	6.0	97.6	100	6.9	2.4	95.7	100	2.5
	MAX-F	6.4	99.0	100	6.4	2.8	97.5	100	4.2
	MAX-B	8.1	98.8	100	8.9	3.9	96.9	100	3.2
n = 900	PP-F-21	4.4	65.0	94.7	66.4	0.7	50.2	85.7	50.7
	PP-B-21	4.7	72.5	97.6	71.8	0.6	56.1	89.3	55.1
	PP-PF-7	4.2	98.6	99.9	4.6	0.7	96.9	99.9	0.8
	PP-PB-7	4.4	98.5	100	7.9	0.9	97.2	99.9	4.0
	GHHK-F	6.8	100	100	6.8	1.5	100	100	1.5
	GHHK-B	6.3	100	100	7.5	1.7	99.9	100	2.2
	MAX-F	7.9	100	100	8.0	3.2	100	100	3.2
	MAX-B	8.1	100	100	8.8	3.5	100	100	4.4

Table 3.2
Percentage of rejections under the slow decaying covariance matrix Σ_{slow}.

Level	Method	$\alpha = 5\%$				$\alpha = 1\%$			
		Null	L1	L3	M10	Null	L1	L3	M10
n = 150	PP-F-21	6.4	15.0	31.6	17.2	1.4	9.7	22.7	9.2
	PP-B-21	2.8	15.0	34.4	16.5	1.0	9.8	24.8	9.4
	PP-PF-7	3.0	32.7	65.2	3.9	0.5	20.6	49.7	1.2
	PP-PB-7	3.3	32.8	66.3	4.2	0.7	21.5	51.3	1.5
	GHHK-F	4.4	66.4	92.6	4.8	2.0	57.9	88.9	2.1
	GHHK-B	4.8	64.1	91.6	10.6	2.0	55.6	87.3	6.8
	MAX-F	7.3	73.5	92.8	7.3	3.5	65.3	88.9	3.5
	MAX-B	8.1	72.5	92.4	14.8	4.1	64.4	88.6	10.2
n = 450	PP-F-21	5.5	37.6	71.6	35.0	1.5	29.0	61.0	28.1
	PP-B-21	4.1	38.3	76.3	38.5	0.8	28.7	64.6	29.0
	PP-PF-7	4.1	79.6	99.4	6.8	0.4	65.1	96.6	3.1
	PP-PB-7	4.5	80.3	99.5	7.6	0.4	65.6	97.0	3.4
	GHHK-F	6.4	98.1	100	6.9	2.3	95.5	100	2.4
	GHHK-B	5.9	97.0	100	12.9	2.6	94.6	100	8.1
	MAX-F	6.7	98.7	100	6.8	3.5	96.6	100	3.4
	MAX-B	8.8	98.6	100	15.6	4.3	96.6	100	10.8
n = 900	PP-F-21	4.7	67.7	95.1	67.4	1.1	52.3	85.6	51.2
	PP-B-21	4.6	72.5	98.0	70.6	0.5	56.1	89.6	54.9
	PP-PF-7	4.0	98.5	100	6.7	1.2	97.3	100	3.8
	PP-PB-7	4.6	98.5	100	7.2	1.5	97.4	100	4.0
	GHHK-F	6.6	100	100	6.2	1.5	99.9	100	1.3
	GHHK-B	7.1	100	100	10.7	1.7	99.9	100	5.0
	MAX-F	7.3	100	100	7.2	2.8	100	100	2.9
	MAX-B	8.4	100	100	12.9	3.3	100	100	7.3

4. Data analysis

In this section, we apply our proposed normality test to several real data sets, with the main objective of comparing its performance with that of some of the existing methods. While for some data sets all tests give similar results, in one case the proposed test leads to different conclusions than those implied by the existing methods. In addition, we also explain how the proposed projection pursuit method can be used for identifying and visualizing the non-Gaussian components of functional data.

Table 3.3
Percentage of rejections under the random covariance matrix Σ_{ran}.

level	Method	$\alpha = 5\%$				$\alpha = 1\%$			
		Null	L1	L3	M10	Null	L1	L3	M10
n = 150	PP-F-21	3.8	13.4	30.7	14.4	0.7	7.8	21.4	8.3
	PP-B-21	4.9	15.9	36.0	17.1	1.0	7.6	20.9	8.2
	PP-PF-7	3.1	30.9	66.0	23.6	0.8	19.8	48.9	15.6
	PP-PB-7	3.6	10.8	26.5	10.8	0.3	5.6	16.4	5.5
	GHHK-F	5.5	41.0	74.1	15.7	2.5	34.2	67.1	11.1
	GHHK-B	5.9	9.8	17.8	10.3	2.6	5.7	13.1	5.9
	MAX-F	13.3	51.4	80.6	26.5	6.8	44.3	75.0	18.9
	MAX-B	28.6	34.9	42.4	34.7	18.1	23.7	32.1	23.0
n = 450	PP-F-21	3.8	38.8	73.3	38.1	1.2	29.3	60.4	27.8
	PP-B-21	6.5	42.6	78.7	39.9	0.9	31.3	64.4	29.0
	PP-PF-7	3.3	77.1	99.2	56.0	0.6	63.0	95.5	44.1
	PP-PB-7	5.5	30.4	59.6	28.3	0.8	21.1	47.4	18.1
	GHHK-F	5.2	77.5	98.5	54.0	2.3	71.3	97.1	45.5
	GHHK-B	5.2	19.3	38.6	16.6	2.0	13.1	29.3	10.5
	MAX-F	12.0	89.8	99.1	52.5	4.6	86.1	98.3	24.8
	MAX-B	21.7	38.4	59.7	37.3	12.8	27.4	47.9	70.8
n = 900	PP-F-21	3.7	68.8	96.5	67.5	0.6	50.5	87.1	49.4
	PP-B-21	7.9	76.7	98.3	72.9	0.9	55.9	89.6	51.5
	PP-PF-7	5.6	97.8	100	70.0	1.1	95.4	100	64.5
	PP-PB-7	6.0	52.4	84.9	50.1	1.7	43.5	79.3	40.3
	GHHK-F	5.8	98.6	100	71.5	1.7	97.9	100	62.6
	GHHK-B	5.6	28.4	59.9	26.8	1.5	20.7	51.6	19.9
	MAX-F	9.2	99.3	100.0	77.1	3.6	98.6	99.9	70.8
	MAX-B	15.2	44.5	73.9	41.4	6.6	33.4	63.9	31.8

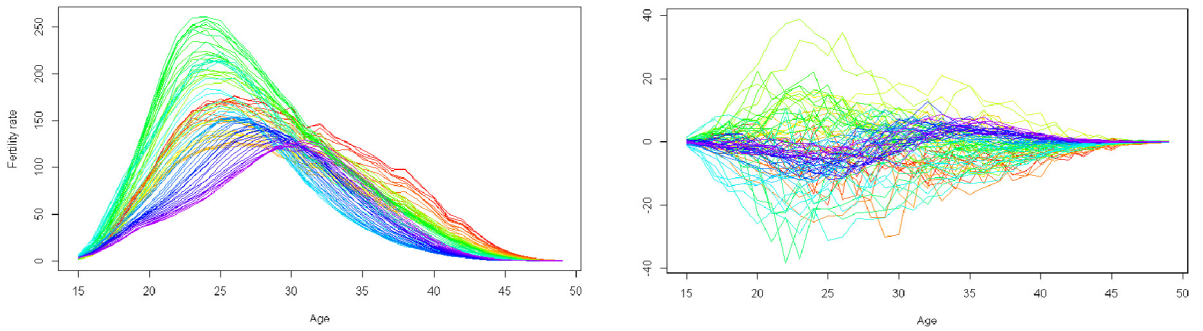

Fig. 4.1. Fertility rate by age in Australia from 1921 to 2006.

4.1. Fertility rate in Australia

We first consider Australian fertility rate data from 1921 to 2006 among women aged from 15 to 49. The data set has been collected by the Australian Bureau of Statistics and is available in the R package `rainbow` (Shang and Hyndman, 2016). In the left panel of Fig. 4.1 each curve represents the distribution of the number of births per 1000 females at each age. From the rainbow plot, and some further analysis, we have found that the second order differencing of the curves is sufficient to remove the prevalent trend in the sequence of curves. The detrended curves are depicted in the right panel of Fig. 4.1. After applying the GHHK-F test described in Section 3.1 to the detrended data, we have obtained a p-value equal to 0.826, which suggests that these curves are reasonably Gaussian. Using the proposed PP-F-21 we have obtained values of the test statistics $\hat{S}_n = 114.825$ and $\hat{K}_n = 23.895$, while the 95% level critical values are 132.915 and 32.792, respectively. The corresponding empirical p-value is 0.325, which is in apparent agreement with the GHHK test.

4.2. Conditional intra-day stock prices

In modern finance, Brownian bridges arise naturally as conditioned Brownian motions in the context of the Black–Scholes model for option pricing. But there are numerous other applications of Brownian bridges, and more generally conditioned diffusion processes. For example, in applications that involve modeling of the flow of information in the market, like in Brody et al. (2008), a Brownian bridge represents the noise in the information about a future market event. In Cartea et al. (2016) the authors utilize a randomized Brownian bridge to model the mid-price of an asset with a

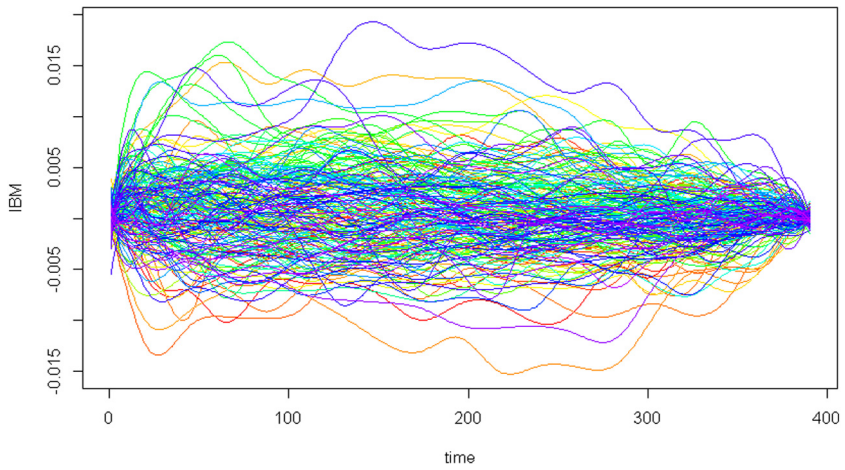

Fig. 4.2. Daily curves of the transformed IBM prices from 06/15/2006 to 04/02/2007.

random end-point that follows a distribution that is not necessary Gaussian. Such models can be justified, to some extent, by the fact that Brownian motion is not the only diffusion process that produces a Brownian bridge when conditioned on its terminal value (Benjamini and Lee, 1997).

In this example we test whether a conditioned log-price of a traded security follows a Gaussian process, which is a less stringent requirement than the assumption that the price follows a geometric Brownian motion. To this end, we utilize the intra-day stock prices of IBM from 06/15/2006 to 04/02/2007, which are available in the R package `fChange` (Sonmez et al., 2018). The closing prices of one share of IBM stock were recorded from 9 a.m. to 4:30 p.m. at a one-minute resolution, and hence there are 390 observations each day. By analogy to the well-known construction for the Brownian bridge (e.g., Karlin and Taylor (1981)), we have transformed the observed prices to conditioned prices in the following way. Suppose the observed intra-day prices on a given day i are denoted $X_i(t_1), \ldots, X_i(t_n)$, and $y_{i,j} = \log X_i(t_j), j = 1, \ldots, 390$. We denote the straight line connecting $y_{i,1}$ and $y_{i,390}$ as $L_i(t)$. Then the bridged log prices are defined as $Y_i(t_1) = (y_{i,1} - L_i(t_1)), \ldots, Y_i(t_{390}) = (y_{i,390} - L_i(t_{390}))$. The widely used Black–Scholes model assumes that log-prices follow a Brownian motion, and hence these transformed price curves should follow a Brownian bridge, which is a Gaussian process. The daily curves of the transformed prices are shown in Fig. 4.2. The p-value calculated from GHHK-F test is 3.78×10^{-8}, which suggests that these curves are non-Gaussian. Our PP-F-21 test generates test statistics $\hat{S}_n = 180.965$ and $\hat{K}_n = 42.401$, while the 95% level critical values are 101.635 and 29.381 respectively. The corresponding empirical p-value is 0, and hence it is in agreement with the GHHK test.

4.3. Yearly lower temperature profiles in Australia

In this final example we consider data comprised of the daily lowest temperature recorded in the Gayndah Post Office from 1893 to 2009, which is available both from the Australian Government Bureau of Meteorology and the R package `fChange` (Sonmez et al., 2018). Gayndah is a small town in Queensland, Australia, which is approximately 200km northwest of Brisbane. The settlement was established in 1849, and the Post Office was established at Gayndah in 1850. We analyze temperature records from 1894 to 2008, as the records prior to 1894 are not complete. In this case each functional observation $X_i(t)$ is defined to be the daily lowest temperature recorded in the Post Office for day $t = 1, 2, \ldots, 365$, in year $i = 1894, \ldots, 2008$. For leap years a 366th data point is added. Since these yearly records have different lengths, we scale the data to the unit interval and smooth the curves using 21 Fourier basis. We than evaluate these curves on 365 equally spaced points in the unit interval. Fig. 4.3 shows a rainbow plot of the data.

The p-value of the GHHK test applied to this data is 0.928, which suggests that these temperature curves are plausibly realizations of a Gaussian process. However, in this case our projection pursuit based method suggests that these curves have components that are both skewed and heavy-tailed. The estimated test statistics for our PP-F-21 test are $\hat{S}_n^{L_k} = 185.49$ and $\hat{K}_n^{L_k} = 47.32$, which both exceed the corresponding estimated 95% critical values (130.59 and 33.97 respectively). The empirical p-value has been estimated as 0.002.

Letting $p_1(t)$ denote the function that maximizes the skewness (or kurtosis) defined in (3.1), we can estimate the skewed (or leptokurtic) direction of each curve $X_i(t)$ as

$$g_{i1}(t) = \langle X_i, p_1 \rangle p_1(t).$$

One can further remove this non-Gaussian component by point-wise subtraction to obtain the residual

$$X_i^{\text{new}}(t) = X_i(t) - g_{i1}(t).$$

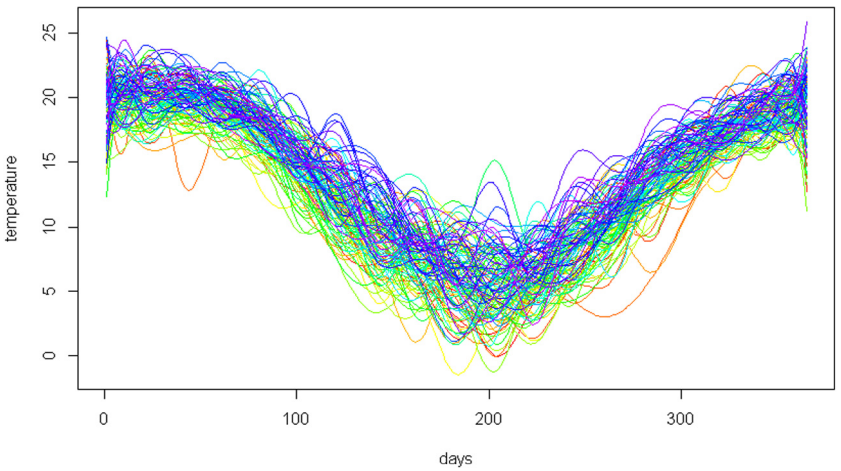

Fig. 4.3. Daily lowest temperature at Gayndah Australia from 1894 to 2008.

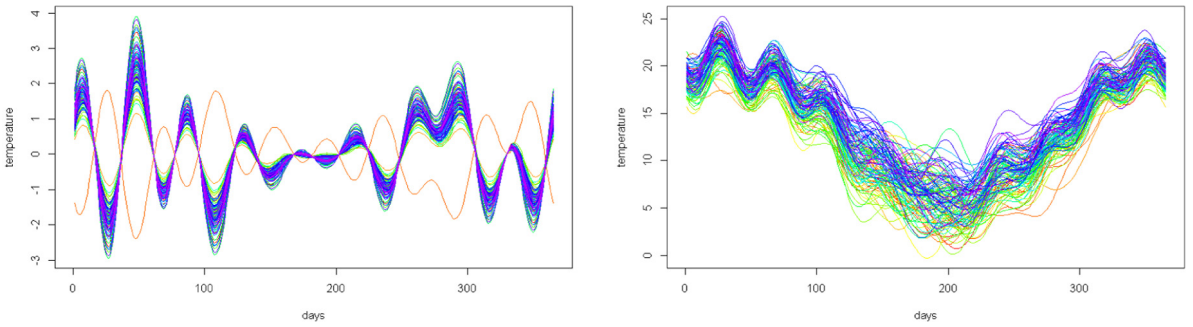

Fig. 4.4. The left panel shows the non-Gaussian components we found when running the projection pursuit based test. The right panel are the residuals of the daily low temperature profile after removing these non-Gaussian components.

Subsequent tests for Gaussianity may be applied to the sample $X_1^{\text{new}}(t), \ldots, X_n^{\text{new}}(t)$ to find further directions $p_2(t)$, $p_3(t)$, ... that will maximize the kurtosis or skewness. Suppose after m steps we are no longer able to reject the null hypothesis that the residuals are Gaussian processes. Then the curve $X_i(t)$ can be decomposed into two parts: an approximate non-Gaussian component $g_i(t) = g_{i1}(t) + \cdots + g_{im}(t)$, and an approximate Gaussian component $r_i(t) = X_i(t) - g_i(t)$.

These two components for the Gayndah temperature curves are presented in Fig. 4.4, where we find 2 directions with excessive kurtosis and 1 direction with excessive skewness. We notice that most variants of the direction tend to vary more prominently at either end of the function, which corresponds to the summer in Australia.

We ran the GHHK-F test again on both the estimated non-Gaussian components and the residuals. The p-values were 0.000 and 0.678, respectively. The total variance explained (TVE) of the non-Gaussian components was around 3%, which is quite small relative to the usual TVE thresholds used to select the number of FPCs. While one could in general try to increase the TVE used to select the number of components in the GHHK test with the aim of discovering such a sparse non-Gaussian component, we point out that intuitively the GHHK method is a joint Jarque–Bera test applied to the projections onto FPCs. Therefore, increasing the number of FPCs will typically lead to a loss of overall testing power. A case in point is this example, in which after increasing the TVE threshold to 99% for the GHHK test, the test still fails to reject the Gaussianity of the curves at the 0.05% level with a p-value equals to 0.723.

P-values of the test applied to the temperature data as a function of k are displayed in Fig. 4.5, and show that strong non-Gaussianity is evident in the data after projecting onto the first 3 Fourier basis elements, and then becomes essentially zero for k greater than 7.

Acknowledgments

This study is supported by Natural Science and Engineering Research Council of Canada Discovery Grant, RGPIN-50503-10477 and RGPIN-04059-2015.

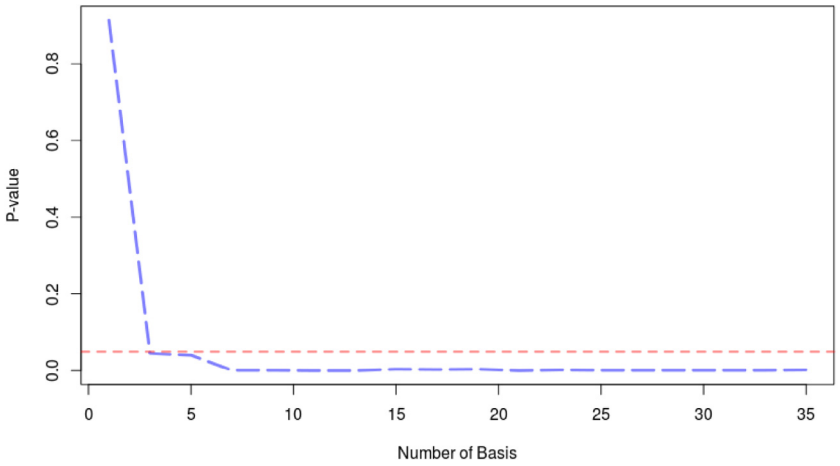

Fig. 4.5. The p-values of proposed normality test under different number basis functions to construct the subspace with daily low temperature at Gayndah, Australia. The red horizontal dash-line is positioned at p = 0.05.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2020.07.001. Online supplementary materials include proofs of our theorem and futher discussion about tuning parameters.

References

Bali, J.L., Boente, G., Tyler, D.E., Wang, J.-L., et al., 2011. Robust functional principal components: A projection-pursuit approach. Ann. Statist. 39 (6), 2852–2882.
Baringhaus, L., Henze, N., 1991. Limit distributions for measures of multivariate skewness and kurtosis based on projections. J. Multivariate Anal. 38 (1), 51–69.
Benjamini, I., Lee, S., 1997. Conditioned diffusions which are brownian bridges. J. Theoret. Probab. 10 (3), 733–736.
Brauchart, J.S., Dick, J., Fang, L., 2015. Spatial low-discrepancy sequences, spherical cone discrepancy, and applications in financial modeling. J. Comput. Appl. Math. 286, 28–53.
Brody, D.C., Hughston, L.P., Macrina, A., 2008. Information-based asset pricing. Int. J. Theor. Appl. Finance 11 (01), 107–142.
Bugni, F.A., Hall, P., Horowitz, J.L., Neumann, G.R., 2009. Goodness-of-fit tests for functional data. Econom. J. 12 (s1), 1–18.
Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208.
Cartea, Á., Jaimungal, S., Kinzebulatov, D., 2016. Algorithmic trading with learning. Int. J. Theor. Appl. Finance 19 (04), 1650028.
Charniak, E., Johnson, M., 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, pp. 173–180.
Clerc, M., 2010. Particle Swarm Optimization, Vol. 93. John Wiley & Sons.
Constantinou, P., Kokoszka, P., Reimherr, M., 2017. Testing separability of space-time functional processes. Biometrika 104 (2), 425–437.
Cuesta-Albertos, J., del Barrio, E., Fraiman, R., Matrán, C., 2007. The random projection method in goodness of fit for functional data. Comput. Statist. Data Anal. 51 (10), 4814–4831.
Cuesta-Albertos, J.A., Fraiman, R., Ransford, T., 2006. Random projections and goodness-of-fit tests in infinite-dimensional spaces. Bull. Braz. Math. Soc. 37 (4), 477–501.
Cuevas, A., 2014. A partial overview of the theory of statistics with functional data. J. Statist. Plann. Inference 147, 1–23.
Cuevas, A., Febrero, M., Fraiman, R., 2004. An ANOVA test for functional data. Comput. Statist. Data Anal. 47, 111–122.
Fletcher, R., Reeves, C.M., 1964. Function minimization by conjugate gradients. Comput. J. 7 (2), 149–154.
Friedman, J.H., Tukey, J.W., 1974. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Comput. 100 (9), 881–890.
Georgieva, A., Jordanov, I., 2009. Global optimization based on novel heuristics, low-discrepancy sequences and genetic algorithms. European J. Oper. Res. 196 (2), 413–422.
Górecki, T., Hörmann, S., Horváth, L., Kokoszka, P., 2018. Testing normality of functional time series. J. Time Ser. Anal. 39 (4), 471–487.
Gromenko, O., Kokoszka, P., Sojka, J., et al., 2017. Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. Ann. Appl. Stat. 11 (2), 898–918.
Henze, N., 2002. Invariant tests for multivariate normality: a critical review. Statist. Papers 43 (4), 467–506.
Henze, N., Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivariate Anal. 62 (1), 1–23.
Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer.
Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econom. Lett. 6 (3), 255–259.
Karlin, S., Taylor, H.E., 1981. A Second Course in Stochastic Processes. Academic Press.
Kimura, S., Matsumura, K., 2007. Improvement of the performances of genetic algorithms by using low-discrepancy sequences. Trans. Soc. Instrum. Control Eng. E-6 (1), 16–25.
Kowal, D.R., Matteson, D.S., Ruppert, D., 2017. A Bayesian multivariate functional dynamic linear model. J. Amer. Statist. Assoc. 112 (518), 733–744.
Kowal, D.R., Matteson, D.S., Ruppert, D., 2019. Functional autoregression for sparsely sampled data. J. Bus. Econom. Statist. 37 (1), 97–109.
Kruskal, J.B., 1972. Linear transformation of multivariate data to reveal clustering. Multidimens. Scaling 1, 101–115.
Leobacher, G., Pillichshammer, F., 2014. Introduction to Quasi-Monte Carlo Integration and Applications. Springer.

Liang, J., Li, R., Fang, H., Fang, K.-T., 2000. Testing multinormality based on low-dimensional projection. J. Statist. Plann. Inference 86 (1), 129–141.
Malkovich, J.F., Afifi, A., 1973. On tests for multivariate normality. J Amer. Statist. Assoc. 68 (341), 176–179.
Mardia, K., Kent, J., Bibby, J., 1979. Multivariate Analysis. Academic Press.
Mecklin, C.J., Mundfrom, D.J., 2004. An appraisal and bibliography of tests for multivariate normality. Internat. Statist. Rev. 72 (1), 123–138.
Monica, T., Rajasekhar, A., Pant, M., Abraham, A., 2011. Enhancing the local exploration capabilities of artificial bee colony using low discrepancy sobol sequence. In: International Conference on Contemporary Computing. Springer, pp. 158–168.
Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. In: CBMS-NSF Series in Applied Mathematics, vol. 63, SIAM, Philadelphia.
Panaretos, V.M., Kraus, D., Maddocks, J.H., 2010. Second-order comparison of gaussian random functions and the geometry of DNA minicircles. J. Amer. Statist. Assoc. 105, 670–682.
Pant, M., Thangaraj, R., Grosan, C., Abraham, A., 2008. Improved particle swarm optimization with low-discrepancy sequences. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). IEEE, pp. 3011–3018.
Pedersoli, M., Vedaldi, A., Gonzalez, J., Roca, X., 2015. A coarse-to-fine approach for fast deformable object detection. Pattern Recognit. 48 (5), 1844–1853.
Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis. Springer, New York.
Riesz, F., Sz.-Nagy, B., 1990. Functional Analysis. Dover.
Shang, H.L., Hyndman, R.J., 2016. Rainbow: rainbow plots, bagplots and boxplots for functional data. R package version 3.4.
Sonmez, O., Aue, A., Rice, G., 2018. Fchange: Change point analysis in functional data. R package version 0.2.0.
Szekely, G.J., Rizzo, M.L., 2005. A new test for multivariate normality. J. Multivariate Anal. 93 (1), 58–80.
Thadewald, T., Büning, H., 2007. Jarque–bera test and its competitors for testing normality–a power comparison. J. Appl. Stat. 34 (1), 87–105.
Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100, 577–590.
Zhu, L., Fang, K., Zhang, J., 1995a. A projection NT-type test for spherical symmetry of a multivariate distribution. New Trends Probab. Statist. 3, 109–122.
Zhu, L.-X., Wong, H.L., Fang, K.-T., 1995b. A test for multivariate normality based on sample entropy and projection pursuit. J. Statist. Plan. Inference 45 (3), 373–385.