A new test for functional one-way ANOVA with applications to ischemic heart screening

Jin-Ting Zhang  
Department of Statistics and Applied Probability, National University of Singapore, Singapore

Ming-Yen Cheng  
Department of Mathematics, Hong Kong Baptist University, Hong Kong

Hau-Tieng Wu  
Department of Mathematics and Department of Statistical Science, Duke University, United States

Bu Zhou  
School of Statistics and Mathematics, Zhejiang Gongshang University, China

Follow this and additional works at: https://repository.hkbu.edu.hk/hkbu_staff_publication

Part of the Mathematics Commons

This document is the authors' final version of the published article.  
Link to published article: https://doi.org/10.1016/j.csda.2018.05.004

APA Citation  
https://doi.org/10.1016/j.csda.2018.05.004

This Journal Article is brought to you for free and open access by HKBU Institutional Repository. It has been accepted for inclusion in HKBU Staff Publication by an authorized administrator of HKBU Institutional Repository. For more information, please contact repository@hkbu.edu.hk.
A New Test for Functional One-Way ANOVA with Applications to Ischemic Heart Screening

Jin-Ting Zhang\textsuperscript{a}, Ming-Yen Cheng\textsuperscript{b},\textsuperscript{*}, Hau-Tieng Wu\textsuperscript{c}, Bu Zhou\textsuperscript{a}

\textsuperscript{a}Department of Statistics and Applied Probability, National University of Singapore
\textsuperscript{b}Department of Mathematics, Hong Kong Baptist University
\textsuperscript{c}Department of Mathematics and Department of Statistical Science, Duke University

Abstract

Motivated by an ischemic heart screening problem, a new global test for one-way ANOVA in functional data analysis is studied. The test statistic is taken as the maximum of the pointwise $F$-test statistic over the interval the functional responses are observed. Nonparametric bootstrap, which is applicable in more general situations and easier to implement than parametric bootstrap, is employed to approximate the null distribution and to obtain an approximate critical value. Under mild conditions, asymptotically our test has the correct level and is root-$n$ consistent in detecting local alternatives. Simulation studies show that the proposed test outperforms several existing tests in terms of both size control and power when the correlation between observations at any two different points is high or moderate, and it is comparable with the competitors otherwise. Application to an ischemic heart dataset suggests that resting electrocardiogram signals may contain enough information for ischemic heart screening at outpatient clinics, without the help of stress tests required by the current standard procedure.

Keywords: Functional hypothesis testing, functional data, local power, nonparametric bootstrap, smoothing and nonparametric regression

2010 MSC: 62G10, 62-07

1. Introduction

As data collection technology evolves, functional data are increasingly common in a wide range of fields, such as chemometrics, climatology, genomics, medicine, seismology, and so on. Compared with traditional data, which consist of scalar or vector observations, functional data might contain more detailed information about the underlying system. On the other hand, new challenges arise in the endeavor to extract any meaningful information hidden in the functional data at hand. In the past two decades, functional data
analysis has emerged as an important area and significant progress has been achieved (Ferraty, 2011; Ramsay & Silverman, 2005; Zhang, 2013). We focus on the one-way ANOVA problem with functional responses (Cuevas et al., 2004; Zhang, 2011; Zhang & Liang, 2013), which is a fundamental problem in the inference but has received much less attention in functional data analysis compared to other problems such as regression, clustering, and classification (Alonoso et al., 2012; Slaets et al., 2012; Zhu et al., 2012; Wu et al., 2010).

A motivating example is detection of stable ischemic hearts, i.e. ischemic hearts without symptoms, which occurs in outpatient clinics on a daily basis. In this problem, we ask if we can distinguish the normal group from the group of subjects with stable ischemic heart by reading their resting surface ECG (electrocardiogram) signals. Traditionally, based on the physiological facts, physicians “reduce the dimension” of an ECG signal by focusing on a specific subinterval, for example, the ST interval, and then try to distinguish between the two groups based on the deformation of the ECG signals on that subinterval. Obviously, we lose information about the underlying system by doing so. Furthermore, due to the inevitable variation between subjects and influence of other physiological activities, the available information might be masked. For example, in the group with stable ischemic hearts, the ECG signal deformation coming from ischemia might be negligible as compared with the noise. A direct consequence is that we are not able to distinguish between the two groups with the reduced data. Thus, in clinics, further stress tests are needed to assist the diagnosis. However, stress tests cannot be applied to patients who are vulnerable to acute attacks during the testing, which is associated with the stress itself. An immediate question we may ask is whether it is possible to differentiate between the two groups if we take into account all the information contained in the whole resting ECG signals during the diagnosis. Naturally, the first step is to test equality of the mean functions of the different groups of some manipulated resting ECG functions. We will treat this example in greater detail in Section 4.

Let \( y_{i1}(t), y_{i2}(t), \ldots, y_{in_i}(t), \) \( i = 1, 2, \ldots, k, \) denote \( k \) groups of random functions defined over a given finite interval \( T = [a, b], \) where \( n_i \in \mathbb{N} \) is the number of observed functions in the \( i \)-th group. Let \( \text{SP}(\mu, \gamma) \) denote a stochastic process defined on \( T \) with mean function \( \mu(t), t \in T, \) and within-function covariance function \( \gamma(s,t), s,t \in T. \) Throughout this article, we refer the within-function correlation to the correlation between observations on the function at any two different points i.e. \( \gamma(s, t)/\sqrt{\gamma(s,s)\gamma(t,t)}. \) Assuming that

\[
y_{i1}(t), y_{i2}(t), \ldots, y_{in_i}(t) \overset{i.i.d.}{\sim} \text{SP}(\mu, \gamma), \quad i = 1, 2, \ldots, k, \quad (1)
\]

the one-way ANOVA problem (or the \( k \)-sample testing problem) for the functional data is to test the equality of the \( k \) mean functions; that is,

\[
H_0 : \mu_1(t) = \mu_2(t) = \ldots = \mu_k(t), \quad \forall t \in T, \quad (2)
\]

against the alternative that at least two of the mean functions are not equal. In this problem, the \( k \) mean functions are often decomposed as \( \mu_i(t) = \mu_0(t) + \alpha_i(t), i = 1, 2, \ldots, k, \) where \( \mu_0(t) \) is the grand mean.
function and $\alpha_i(t), i = 1, 2, \ldots, k$, are the main-effect functions.

There are several existing methods for testing the one-way functional ANOVA problem (2). For example, the $L^2$-norm based test and the $F$-type test for linear models with functional responses (Cuevas et al., 2004; Fan & Lin, 1998; Faraway, 1997; Shen & Faraway, 2004; Zhang, 2011; Zhang & Chen, 2007) may be adopted for this purpose. In addition, the pointwise $F$-test was proposed by Ramsay & Silverman (2005), naturally extending the classical $F$-test to the current context; see also Zhang (2013). Its test statistic is defined as

$$ F_n(t) := \frac{\text{SSR}_n(t)}{\text{SSE}_n(t)} \frac{(k - 1)}{(n - k)} $$

where and throughout $n := \sum_{i=1}^k n_i$ denotes the total sample size,

$$ \text{SSR}_n(t) := \sum_{i=1}^k n_i [\bar{y}_i(t) - \bar{y}(t)]^2 \quad \text{and} \quad \text{SSE}_n(t) := \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij}(t) - \bar{y}_i(t)]^2 $$

(3)

denote the pointwise between-group and within-group variations, respectively, and

$$ \bar{y}(t) := \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}(t) \quad \text{and} \quad \bar{y}_i(t) := \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}(t), \ i = 1, 2, \ldots, k, $$

(4)

are the sample grand mean function and the sample group mean functions, respectively. Then, the null hypothesis (2) is rejected as long as the pointwise null hypothesis $H_{0t} : \mu_1(t) = \mu_2(t) = \ldots = \mu_k(t)$ is rejected at some point $t \in \mathcal{T}$.

There are a few advantages of using the above pointwise $F$-test. When the functional data (1) are realizations of Gaussian processes, under the null hypothesis (2), $F_n(t)$ follows the same $F$-distribution for any given $t \in \mathcal{T}$. Therefore, for any pre-determined significance level, we can test the null hypothesis (2) at all of the points in $\mathcal{T}$ using the same critical value. On the other hand, the pointwise $F$-test has some limitations too. In particular, for a given significance level, it is not guaranteed that alternative hypothesis is overall significant even when the pointwise $F$-test is significant for every point in $\mathcal{T}$ at the same significance level. To overcome this difficulty, Cox & Lee (2008) proposed to correct for the multiple testing via incorporating the Westfall-Young randomization method. In addition, Zhang & Liang (2013) proposed and studied a so-called Globalized Pointwise $F$-test, abbreviated as GPF test, whose test statistic is taken as the integral of the pointwise $F$-test statistic over $\mathcal{T}$:

$$ T_n = \int_{\mathcal{T}} F_n(t) dt. $$

(5)

Via some simulation studies, Zhang & Liang (2013) showed that the GPF test is in general comparable with the $L^2$-norm based and the $F$-type tests, in terms of both size control and power. Further, via extensive simulation studies, Górecki & Smaga (2015) presented an exhaustive comparison of a number of existing tests and concluded that the GPF test is one of the best global tests.
Alternatively, we can globalize the pointwise $F$-test statistic using the supremum over $T$:

$$F_{\text{max}} := \sup_{t \in T} F_n(t).$$

(6)

Ramsay & Silverman (2005) (p. 234) briefly mentioned the use of the squared-root of $F_{\text{max}}$ as the test statistic and a permutation-based critical value. By an application to a real dataset using a bootstrap-based critical value of the $F_{\text{max}}$ test statistic, Zhang & Liang (2013) realized that it may have higher power than the GPF test, particularly when the within-function correlation is high or moderate which is often the case for functional data. In addition, when applied to an ischemic heart ECG dataset as presented in Section 4, the $F_{\text{max}}$-test is significant while the GPF test is not. Therefore, it is worthwhile to study the asymptotic level and power of the $F_{\text{max}}$-test.

The contributions of this paper are as follows, which are not offered by Ramsay & Silverman (2005) (p. 234). First, we derive the asymptotic null distribution of the $F_{\text{max}}$ test statistic under very general conditions. This allows us to prescribe a parametric bootstrap (PB) method to approximate the asymptotic null distribution. However, this approach is reliable only when the sample sizes $n_1, n_2, \ldots, n_k$ are all large, and the computation is complicated and time-consuming. To overcome these difficulties, we propose a simple and computationally efficient nonparametric bootstrap (NPB) method to approximate the finite sample null distribution. Indeed, via some extensive simulation studies we found that the null probability density function (pdf) of $F_{\text{max}}$ is skewed when the within-function correlation is moderate or high, and the NPB approximation works reasonably well. Secondly, we show that the $F_{\text{max}}$-test based on the NPB has the correct asymptotic level. Thus we can rely on the significance of the $F_{\text{max}}$-test when it is applied to the ischemic heart data and ignore the fact that the GPF test is not significant for the same dataset; see Section 4 for the details. This result has important implications in ischemic heart screening in outpatient clinics. In other words, we can conclude that ECG signals may contain relevant features that are sufficient for the purpose of distinguishing between ischemic heart and normal groups, without the help of stress tests. Therefore this work calls for the need of some larger scale clinical studies to confirm the finding. Furthermore, we show that the $F_{\text{max}}$-test is root-$n$ consistent, i.e., its power tends to 1 under local alternatives that depart from the null hypothesis at the $n^{-1/2}$ order.

Thirdly, we conclude from some simulation studies that the $F_{\text{max}}$-test is preferred to the $F$-type test (Shen & Faraway, 2004), the $L^2$-norm based test (Zhang & Chen, 2007), and the GPF test (Zhang & Liang, 2013) in the following senses. In terms of size control, generally the $F_{\text{max}}$-test is comparable with the $F$-type test and it outperforms the other two competitors. In terms of power, the $F_{\text{max}}$-test is substantially better than the other three tests when the within-function correlation is moderate or high; otherwise it is only slightly worse. Intuitively speaking, when the within-function correlation is moderate or high, the three competitors tend to average down the information provided by the functional data and hence they have lower powers. And, when the within-function correlation is low, they tend to summarize more uncorrelated information.
than the $F_{\text{max}}$-test as the latter takes into account only the information available at the maximum of the process $F_n(t)$. Since the within-function correlation in functional data is usually moderate or high, the $F_{\text{max}}$-test is therefore preferred to the three competitors in general. This also explains why we have the aforementioned contradictory results given by the $F_{\text{max}}$ and GPF tests in the ischemic heart example, as the ECG signals at different points are highly correlated.

Lastly, we mention that it is straightforward to extend the $F_{\text{max}}$-test to other linear regression models for functional data, including higher-way functional ANOVA (Zhang, 2013, ch. 5) and functional linear models (Zhang, 2013, ch. 6). In addition, when the functional responses are observed on a grid of (dense) points, which is usually the case in practice, it is shown in the supplementary material that under mild conditions the effect of discretization on the $F_{\text{max}}$-test is negligible in terms of both size control and local power.

The methodology and the main theoretical results on the level accuracy and local power of the $F_{\text{max}}$-test are presented in Section 2. Results of extensive simulation studies and the real data example on ischemic heart screening are given in Sections 3 and 4, respectively. Proofs of the theoretical results are deferred to Section 5. The supplementary material contains a theoretical study on effect of discretization, and additional simulation studies to consider asymmetric error distributions, non-smooth functional data, complicated mean differences, and unequal group covariance functions, and to study the effect of discretization resolution.

2. Methodology and Theoretical Results

2.1. Asymptotic Null Distribution

We first derive the asymptotic random expression of the $F_{\text{max}}$ test statistic under the null hypothesis (2). This result will be helpful when we study the asymptotic level of the bootstrap-based $F_{\text{max}}$-test, which can then be used to justify the significance of the $F_{\text{max}}$-test when applied to the ischemic dataset. Notice that for any $t \in \mathcal{T}$, the pointwise between-subject variation $\text{SSR}_n(t)$ given in (3) can be expressed as

\[
\text{SSR}_n(t) = [z_n(t) + \mu_n(t)]^T (I_k - b_n b_n^T / n) [z_n(t) + \mu_n(t)],
\]

where $I_k$ is the $k \times k$ identity matrix,

\[
\begin{align*}
  z_n(t) &= [\sqrt{n_1} [y_1(t) - \mu_1(t)], \sqrt{n_2} [y_2(t) - \mu_2(t)], \ldots, \sqrt{n_k} [y_k(t) - \mu_k(t)]]^T \in \mathbb{R}^k, \\
  \mu_n(t) &= [\sqrt{n_1} \mu_1(t), \sqrt{n_2} \mu_2(t), \ldots, \sqrt{n_k} \mu_k(t)]^T \in \mathbb{R}^k, \text{ and} \\
  b_n &= [\sqrt{n_1}, \sqrt{n_2}, \ldots, \sqrt{n_k}]^T \in \mathbb{R}^k.
\end{align*}
\]

Since $b_n^T b_n / n = 1$, it is easy to verify that $I_k - b_n b_n^T / n$ is an idempotent matrix with rank $k - 1$. In addition, as $n \to \infty$, we have

\[
I_k - b_n b_n^T / n \to I_k - b b^T, \quad \text{with} \quad b = [\sqrt{n_1}, \sqrt{n_2}, \ldots, \sqrt{n_k}]^T,
\]
where \( \tau_i = \lim_{n \to \infty} n_i/n, \) \( i = 1, 2, \ldots, k, \) are as given in Condition A3 stated below. Note that \( I_k - \mathbf{b}\mathbf{b}^T \) is also an idempotent matrix of rank \( k - 1, \) and it has the following singular value decomposition:

\[
I_k - \mathbf{b}\mathbf{b}^T = U \begin{pmatrix} I_{k-1} & 0 \\ 0^T & 0 \end{pmatrix} U^T, \tag{10}
\]

where the columns of \( U \) are the eigenvectors. Based on the \( k \) functional samples (1), the pooled sample covariance function is given by

\[
\tilde{\gamma}(s, t) = (n - k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \|y_{ij}(s) - \bar{y}_i(s)\| \|y_{ij}(t) - \bar{y}_i(t)\|, \tag{11}
\]

where \( \bar{y}_i(t), i = 1, 2, \ldots, k, \) are the sample group mean functions given in (4).

Let \( \mathcal{L}^2(\mathcal{T}) \) denote the set of all square-integrable functions over \( \mathcal{T} \) and let \( C^\beta(\mathcal{T}) \), where \( 0 < \beta \leq 1, \) denote the set of functions \( f \) over \( \mathcal{T} \) which are Hölder continuous with exponent \( \beta \) and Hölder modulus \( \|f\|_{C^\beta}. \) Let \( \text{tr}(\gamma) = \int_{\mathcal{T}} \gamma(t, t) dt \) denote the trace of the covariance function \( \gamma(s, t) \). We impose the following regularity conditions.

**Condition A**

(A1) The \( k \) mean functions \( \mu_1(t), \mu_2(t), \ldots, \mu_k(t) \) in (1) all belong to \( \mathcal{L}^2(\mathcal{T}). \)

(A2) The subject-effect functions \( \nu_{ij}(t) := y_{ij}(t) - \mu_i(t), j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k, \) are i.i.d. \( \text{SP}(0, \gamma) \) where 0 denotes the zero function whenever there is no ambiguity.

(A3) As \( n \to \infty, \) the \( k \) sample sizes \( n_1, n_2, \ldots, n_k \) satisfy \( n_i/n \to \tau_i > 0, \) \( i = 1, 2, \ldots, k. \)

(A4) The subject-effect function \( \nu_{11}(t) \) satisfies \( \mathbb{E}\|\nu_{11}\|^4 = \mathbb{E} \left[ \int_{\mathcal{T}} \nu_{11}^2(t) dt \right]^2 < \infty. \)

(A5) The covariance function \( \gamma(s, t) \) satisfies \( \gamma \in C^{2\beta}(\mathcal{T} \times \mathcal{T}) \) with \( \text{tr}(\gamma) < \infty, \) where \( 0 < 2\beta \leq 1. \) For any \( t \in \mathcal{T} \), \( \gamma(t, t) > 0. \)

(A6) We have \( \mathbb{E} \left[ \nu_{11}^2(s) \nu_{11}^2(t) \right] < C < \infty \) for some constant \( C \) independent of \( (s, t) \in \mathcal{T} \times \mathcal{T}. \)

Assumptions A1 and A2 are standard ones. Condition A3 guarantees that, as \( n \to \infty, \) the sample group mean functions \( \bar{y}_i(t), i = 1, 2, \ldots, k, \) will converge to some Gaussian processes weakly. Assumptions A4, A5 and A6 are imposed so that the pointwise \( F \)-statistic \( F_n(t) \) is well defined at any \( t \in \mathcal{T} \) and \( \gamma(s, t) \) converges to \( \gamma(s, t) \) uniformly over \( \mathcal{T} \times \mathcal{T}. \)

Let \( \text{GP}_k(\mu, \Gamma) \) denote a \( k \)-dimensional Gaussian process with mean function vector \( \mu(t) \in \mathbb{R}^k, \) \( t \in \mathcal{T}, \) and covariance function matrix \( \Gamma(s, t) \in \mathbb{R}^{k \times k}, s, t \in \mathcal{T}. \) We write \( \Gamma(s, t) = \text{diag} [\gamma(s, t), \gamma(s, t), \ldots, \gamma(s, t)] \) as usual, \( \text{GP}(\mu, \gamma) \) denotes a Gaussian process with mean function \( \mu(t) \) and covariance function \( \gamma(s, t). \) Let \( \overset{d}{\to} \) denote convergence in distribution in the sense of van der Vaart & Wellner (1996) (p. 50–51). Furthermore, let \( X \overset{d}{=} Y \) denote that \( X \) and \( Y \) have the same distribution.

The following proposition is useful for deriving a PB method and an NPB method to approximate the null distribution of the \( F_{\text{max}} \) test statistic, detailed in the next section.
Proposition 1. Under Condition A and the null hypothesis (2), as \( n \to \infty \), we have \( F_{\text{max}} \xrightarrow{d} R_0 \) with

\[
R_0 = \sup_{t \in T} \left\{ (k - 1)^{-1} \sum_{i=1}^{k-1} w_i^2(t) \right\},
\]

where \( w_1(t), w_2(t), \ldots, w_{k-1}(t) \overset{i.i.d.}{\sim} \text{GP}(0, \gamma_w) \) with \( \gamma_w(s, t) = \gamma(s, t)/\sqrt{\gamma(s, s)\gamma(t, t)} \).

2.2. Approximating the Null Distribution

By Proposition 1, \( w_1(t), w_2(t), \ldots, w_{k-1}(t) \overset{i.i.d.}{\sim} \text{GP}(0, \gamma_w) \) which is known except for \( \gamma_w(s, t) \). The within-function correlation function \( \gamma_w(s, t) \) can be estimated by

\[
\hat{\gamma}_w(s, t) = \frac{\hat{\gamma}(s, t)}{\sqrt{\hat{\gamma}(s, s)\hat{\gamma}(t, t)}},
\]

where \( \hat{\gamma}(s, t) \) is the pooled sample covariance function given in (11). Similar to Cuevas et al. (2004), for any given significance level \( \alpha \), we can approximate the critical value of the \( F_{\text{max}} \) test statistic by the upper \( 100\alpha \)-percentile of \( R_0 \). We can employ PB to approximate the distribution of \( R_0 \). Based on (12), we can generate i.i.d. Gaussian processes \( w_i(t), i = 1, 2, \ldots, k - 1 \), from \( \text{GP}(0, \hat{\gamma}_w) \) a large number of times to obtain a large PB sample of \( R_0 \). Alternatively, the following NPB method is applicable for both large and finite sample sizes. Denote the estimated subject-effect functions as

\[
v^*_{ij}(t) = y_{ij}(t) - \overline{y}_i(t), \quad j = 1, 2, \ldots, n; i = 1, 2, \ldots, k,
\]

which can be regarded as estimators of the subject-effect functions defined in Condition A2. Let

\[
v^*_i(t), j = 1, 2, \ldots, n; i = 1, 2, \ldots, k,
\]

be bootstrapped \( k \) samples randomly drawn from the estimated subject-effect functions given in (14). The NPB \( F_{\text{max}} \) test statistic can then be obtained as

\[
F^*_{\text{max}} = \sup_{t \in T} F^*_n(t), \quad \text{where } F^*_n(t) = \frac{\text{SSR}^*_n(t)/(k - 1)}{\text{SSE}^*_n(t)/(n - k)},
\]

with \( \text{SSR}^*_n(t) \) and \( \text{SSE}^*_n(t) \) obtained from (3) but based on the bootstrapped \( k \) samples (15). Repeat the above bootstrapping process a large number of times. Then we can calculate the upper \( 100\alpha \)-percentile of the bootstrap sample on \( F^*_{\text{max}} \) and denote it by \( C^*_\alpha \). Then we can conduct the level-\( \alpha \) \( F_{\text{max}} \)-test using it as the critical value. Note that this NPB method is simple and fast to compute.

Let \( C_\alpha \) denote the upper \( 100\alpha \)-percentiles of the random variable \( R_0 \) defined in Proposition 1, which characterizes the asymptotic distribution of \( F_{\text{max}} \) under the null hypothesis \( H_0 \). The following proposition shows that the \( F_{\text{max}} \)-test based on the NPB critical value \( C^*_\alpha \) has the correct level \( \alpha \) asymptotically.

Proposition 2. Under Condition A, as \( n \to \infty \), we have \( F_{\text{max}} \xrightarrow{d} R_0 \) and \( C^*_\alpha \xrightarrow{d} C_\alpha \).
Notice that Proposition 2 holds under both the null and the alternative hypotheses. This is a desirable property because in practice either of the null and alternative hypotheses may be true. It is mainly due to the fact that the $k$ NPB samples (15) all have the same group mean function as 0 and hence the null hypothesis always holds for them. Based on this result, we can rely on the significance of the $F_{\text{max}}$-test on the ischemic dataset as shown in Section 4, and say that ECG signals contain sufficient information that can be used to detect stable ischemic hearts in outpatient clinics.

Proposition 2 shows that for large samples, the NPB and PB methods will yield similar approximate critical values for the $F_{\text{max}}$ test statistic. However, the PB method may involve much more computational efforts as it requires sampling from the Gaussian processes $w_i(t), i = 1, 2, \ldots, k-1$, from GP($0, \gamma_w$) repeatedly, which may not be an easy task and can be time-consuming. Besides, larger sample sizes are necessary for the PB method to work reasonably well, in particular when the distribution of $R_0$ is skewed, but that may not be available in some applications. In general, we prefer the NPB method to the PB method because it can be used under more general conditions and its implementation is much simpler and efficient.

2.3. Asymptotic Power

To study the asymptotic power of the $F_{\text{max}}$-test, we specify the following local alternative:

$$H_{1n}: \mu_i(t) = \mu_0(t) + n_i^{-1/2}d_i(t), \quad i = 1, 2, \ldots, k,$$

where $\mu_0(t)$ is the grand mean function, and $d_1(t), d_2(t), \ldots, d_k(t)$ are any fixed real functions, independent of $n$. By (17), we have

$$\mu(t) = [\mu_1(t), \mu_2(t), \ldots, \mu_k(t)]^T = \mu_0(t)1_k + [n_1^{-1/2}d_1(t), n_2^{-1/2}d_2(t), \ldots, n_k^{-1/2}d_k(t)]^T,$$

where $1_k$ denotes the $k$-vector of ones. It follows that $\mu_n(t) = \mu_0(t)\mathbf{b}_n + \mathbf{d}(t)$ where $\mu_n(t)$ is defined in (8) and $\mathbf{d}(t) = [d_1(t), d_2(t), \ldots, d_k(t)]^T$. Under Condition A3, as $n$ tends to $\infty$, the local alternative (17) will tend to the null at the root-$n$ rate. In this sense, a test is said to be root-$n$ consistent if it can detect the local alternative (17) with probability tending to 1 as the amount of information contained in $\mathbf{d}(t)$ diverges, specifically as $\delta$ defined later diverges. In the following, we show that the $F_{\text{max}}$-test possesses this desirable root-$n$ consistency property. First, since $(I_k - \mathbf{b}_n\mathbf{b}_n^T/n)\mathbf{b}_n = 0$, under the local alternative (17), we have

$$\text{SSR}_n(t) = [\mathbf{z}_n(t) + \mathbf{d}(t)]^T (I_k - \mathbf{b}_n\mathbf{b}_n^T/n) [\mathbf{z}_n(t) + \mathbf{d}(t)],$$

where $\mathbf{z}_n(t)$ is defined in (8). We then have the following proposition about the asymptotic distribution of the $F_{\text{max}}$ test statistic under the local alternative (17).

**Proposition 3.** Under Condition A and the local alternative (17), as $n \to \infty$, we have $F_{\text{max}} \xrightarrow{d} R_1$ with

$$R_1 \overset{d}{=} \sup_{t \in \mathcal{T}} \left\{ (k-1)^{-1} \sum_{i=1}^{k-1} [w_i(t) + \delta_i(t)]^2 \right\},$$

(19)
where \( w_1(t), w_2(t), \ldots, w_{k-1}(t) \) are i.i.d. \( \text{GP}(0, \gamma_w) \) as in Proposition 1 and \( \delta_i(t), i = 1, 2, \ldots, k-1 \), are the \( k-1 \) components of \( \delta(t) = (I_{k-1}, 0) \mathbf{U}^T \mathbf{d}(t)/\sqrt{\gamma(t,t)} \) with \( \mathbf{U} \) given in (10).

With some abuse of notation, we set \( \delta^2 = \sum_{i=1}^{k-1} \int_{\mathcal{T}} \delta_i^2(t) dt \) which is a summary of the information coming from \( \mathbf{d}(t) \) with respect to the one-way ANOVA problem (2), where \( \delta_i(t), i = 1, 2, \ldots, k-1 \), are as defined in Proposition 3. The following proposition shows that the \( F_{\max} \)-test is root \( n \)-consistent.

**Proposition 4.** Under Condition A and the local alternative (17), as \( n \to \infty \), the power of the \( F_{\max} \)-test \( P(F_{\max} \geq C_n^\alpha) \) tends to 1 as \( \delta \to \infty \).

In the proof of Proposition 4, we use the following relationship between the \( F_{\max} \)-test statistic defined in (6) and the GPF test statistic \( T_n \) defined in (5):

\[
T_n = \int_{\mathcal{T}} F_n(t) dt \leq (b-a)F_{\max},
\]

where we use the fact that \( \mathcal{T} = [a,b] \). It then follows that

\[
P(F_{\max} \geq C_n^\alpha) \geq P(T_n \geq (b-a)C_n^\alpha), \tag{20}
\]

However, \((b-a)C_n^\alpha\) may not be equal or smaller than the upper (100\(\alpha\))-percentile of the GPF test statistic \( T_n \). Thus, (20) does not guarantee that the \( F_{\max} \)-test has higher power than the GPF test. To compare the powers of the two tests, some simulation studies are then needed.

**3. Simulation Studies**

In this section, we present results of some simulation studies. The aims are to check if the bootstrapped null distribution of the \( F_{\max} \) test statistics approximates the true underlying null distribution well, and how the \( F_{\max} \)-test performs compared with three other competitors: the \( L^2 \)-norm based (Zhang & Chen, 2007), \( F \)-type (Shen & Faraway, 2004) and GPF (Zhang & Liang, 2013) tests, denoted as \( L^2 \), \( F \) and GPF, respectively. The \( F_{\max} \)-test was implemented based on discretization as described in Section S.1 of the supplementary material. We refer to Zhang (2013) (Section 4.5.5) for the implementation of the \( L^2 \)-norm and \( F \)-type tests, and to Zhang & Liang (2013) (Section 2.4) for the implementation of the GPF test.

The \( k \) functional samples (1) were generated from the following one-way ANOVA model:

\[
y_{ij}(t) = \mu_i(t) + v_{ij}(t), \quad \mu_i(t) = c_i^T [1, t, t^2, t^3]^T, \quad v_{ij}(t) = b_{ij}^T \Psi(t), \quad t \in [0,1],
\]

\[
b_{ij} = [b_{ij1}, b_{ij2}, \ldots, b_{ijq}]^T, \quad b_{ijr} \overset{d}{=} \sqrt{\lambda_r} z_{ijr}, \quad r = 1, 2, \ldots, q, \tag{21}
\]

for \( j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k \). Here the parameter vectors \( c_i = [c_{i1}, c_{i2}, c_{i3}, c_{i4}]^T, i = 1, 2, \ldots, k \), can be flexibly specified, and the random variables \( z_{ijr}, r = 1, 2, \ldots, q, j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k \) are i.i.d. with mean 0 and variance 1. Also, \( \Psi(t) = [\psi_2(t), \psi_2(t), \ldots, \psi_q(t)]^T \) is a vector of \( q \) orthonormal basis functions.
\( \psi_r(t), t \in [0,1], r = 1, 2, \ldots, q \), the variance components \( \lambda_r, r = 1, 2, \ldots, q \) are positive and decreasing in \( r \), and \( q \) is an odd positive integer. These parameters help specify the group mean functions \( \mu_i(t) = c_{i1} + c_{i2}t + c_{i3}t^2 + c_{i4}t^3, i = 1, 2, \ldots, k \) and the common covariance function \( \gamma(s,t) = \Psi(s)t^{\delta}(\lambda_1, \lambda_2, \ldots, \lambda_q)\Psi(t) = \sum_{r=1}^{q} \lambda_r \psi_r(s) \psi_r(t) \), of the subject-effect functions \( v_{ij}(t), j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k \). For simplicity, we took the design time points for all the functions \( y_{ij}(t), j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k \) as the same and specified them as \( t_j = j/(M+1), j = 1, 2, \ldots, M \), where \( M \) is some positive integer. In practice, the functions can be observed at different design time points. In that case, some smoothing technique, such as the one discussed in Zhang & Liang (2013), can be used to reconstruct the functions \( y_{ij}(t), j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k \), and then to evaluate them at a common grid of time points. This setup is time consuming to carry out and we did not explore it.

We now specify the model parameters in (21). To specify the group mean functions \( \mu_1(t), \mu_1(t), \ldots, \mu_k(t) \), we set \( c_1 = [1, 2, 3, 4, 1.5]^T \) and \( c_i = c_1 + (i-1)\delta u, i = 2, 3, \ldots, k \), where the parameter \( \delta \) specifies the differences \( \mu_i(t) - \mu_1(t), i = 2, 3, \ldots, k \), and the constant vector \( u \) specifies the direction of these differences. We set \( \delta \) properly as listed in Tables 1 and 2 so that the null hypothesis (when \( \delta = 0 \)) and the four alternatives (when \( \delta > 0 \)) are fulfilled. In addition, we set \( u = [1, 2, 3, 4]^T/\sqrt{30} \) so that it is a unit vector.

To specify the common covariance function \( \gamma(s,t) \), for simplicity, we set \( \lambda_r = a \rho^r, r = 1, 2, \ldots, q \), for some \( a > 0 \) and \( 0 < \rho < 1 \). Notice that the parameter \( \rho \) not only determines the decay rate of \( \lambda_1, \lambda_2, \ldots, \lambda_q \), but also determines how the simulated within-function observations are correlated. When \( \rho \) is close to 0 (or 1), \( \lambda_1, \lambda_2, \ldots, \lambda_q \) will decay very fast (or slowly), indicating that they are highly correlated (or nearly uncorrelated). For simplicity, we set the basis functions as \( \psi_1(t) = 1, \psi_2(t) = \sqrt{2} \sin(2\pi rt), \psi_{2r+1}(t) = \sqrt{2} \cos(2\pi rt), t \in [0,1], r = 1, 2, \ldots, (q-1)/2 \). In addition, we set \( a = 1.5, q = 11 \) and \( \rho = 0.1, 0.3, 0.5, 0.7 \) or 0.9 to consider the five cases when the simulated functional data have very high, high, moderate, low or very low correlations. We set the number of groups \( k = 3 \) and specified three cases of the sample size vector \( n = [n_1, n_2, n_3]: n_1 = [20, 30, 30], n_2 = [40, 30, 70] \) and \( n_3 = [80, 70, 100] \), representing the small, moderate and large sample size cases, respectively. For the number of design time points \( M \): we took \( M = 80 \) or \( M = 150 \). Finally, we considered two cases for the distribution of the i.i.d. random variables \( z_{ijr}, r = 1, 2, \ldots, q; j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k: z_{ijr} \overset{i.i.d.}{\sim} N(0,1) \) and \( z_{ijr} \overset{i.i.d.}{\sim} t_4/\sqrt{2} \), allowing us to generate Gaussian and non-Gaussian functional data, respectively.

For a given model configuration, the \( k \) samples were generated as in (21). Then the \( L^2, F, \) GPF test statistics and their P-values were computed. Meanwhile, the \( F_{\max} \) test statistic was computed, and 10000 bootstrap replicates were generated to compute the P-value of the \( F_{\max} \)-test. The above process was repeated \( N = 5000 \) times. The nominal significance level \( \alpha \) was set to be 5%.

We first check if the bootstrapped null pdf of the \( F_{\max} \) test statistic works well in approximating the underlying null pdf. To this end, each panel in Fig. 1 displays the simulated null pdf and the first 50 bootstrapped null pdfs of the \( F_{\max} \) test statistic when \( z_{ijr}, r = 1, 2, \ldots, q; j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k \) \( \overset{i.i.d.}{\sim} \).
Figure 1: Simulated null pdfs (wider solid) and the first 50 bootstrapped null pdfs (dashed) of $F_{\text{max}}$ when $z_{ijr}, r = 1, \ldots, q; j = 1, \ldots, n_i; i = 1, \ldots, k$, are i.i.d. $N(0,1)$ and $M = 80$. From left to right, the columns correspond to the sample size vectors $n_1, n_2,$ and $n_3$, and from top to bottom the rows correspond to $\rho = 0.10, 0.50$ and 0.90.

Figure 2: Same as in Fig. 1 except that now $z_{ijr}, r = 1, \ldots, q; j = 1, \ldots, n_i; i = 1, \ldots, k$ i.i.d. $\sim t_4/\sqrt{2}$. 

11
Each of the pdfs was computed using the usual kernel density estimator (Wand & Jones, 1995, Ch. 2) based on the 5000 simulated $F_{\text{max}}$ test statistics (6) or the 10000 bootstrapped $F_{\text{max}}$ test statistics when $\delta = 0$ and $M = 80$. From Fig. 1, it is seen that the bootstrapped null pdfs of the $F_{\text{max}}$ test statistic approximate the associated simulated null pdf rather well in all of the nine cases, showing that the NPB approximation does work reasonably well. Furthermore, it seems that the effect of the sample size is insignificant; but the shapes of the simulated and bootstrapped null pdfs are strongly affected by the decay rates of the variance components $\lambda_r, r = 1, 2, \ldots, q$, namely, stronger correlation ($\rho$ smaller) causes more skewness. In Fig. 2, we display the simulated and bootstrapped null pdfs of the $F_{\text{max}}$ test statistic when $z_{ijr}, r = 1, 2, \ldots, q; j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k \overset{i.i.d.}{\sim} t_4/\sqrt{2}$. We have similar observations from Fig. 2 as from Fig. 1.

We summarize in Tables 1 and 2 the empirical sizes and powers (%) of the $L^2$, $F$, GPF and $F_{\text{max}}$ tests when $z_{ijr}, r = 1, 2, \ldots, q; j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k$ are i.i.d. $N(0, 1)$ and $t_4/\sqrt{2}$, respectively. From the columns associated with $\delta = 0$ in both tables, we see that in terms of size control, in general, the $F_{\text{max}}$-test is comparable with the $F$-type test and they both outperform the $L^2$ and GPF tests. From the columns associated with $\delta > 0$ in both tables, we also see that in terms of power, in general the $F_{\text{max}}$-test has higher powers than the $L^2$, $F$, and GPF tests, except when the within-function observations are less correlated ($\rho = 0.9$), and the advantages of the $F_{\text{max}}$-test over the $L^2$, $F$, and GPF tests become more significant as the correlation increases.

We only present the simulation results for the case when $M = 80$ since those when $M = 150$ are similar. From the simulation results, we conclude that the NPB method works reasonably well in approximating the underlying null pdf of the $F_{\text{max}}$ test statistic. Also, in general the $F_{\text{max}}$ test is preferred to the $L^2$, $F$ and GPF tests in terms of both size control and power, in particular when the within-function observations are moderately or strongly correlated which is usually the case for functional data.

Additional simulation studies considering asymmetric error distributions, non-smooth functional data, complicated mean function differences, and unequal group covariance functions, as well as an investigation on the effect of discretization resolution $M$, are presented in the supplementary material. The main findings are (i) similar conclusions to those given in this section can be drawn, (ii) in terms of size control the considered tests are robust against asymmetric error distributions, non-smooth functional data, and complicated mean function differences, and in terms of power they are relatively less sensitive to asymmetric error distributions than to the latter two situations, (iii) the considered tests are sensitive to heteroscedastic covariance functions especially when the group sample sizes are different from each other, and the bootstrapping step in $F_{\text{max}}$ test can be modified to deal with this problem, and (iv) taking $M = 80 \sim 150$ is generally sufficient to avoid effect of discretization.
Table 1: Empirical sizes and powers (%) of the $L^2$, F, GPF and $F_{\text{max}}$ tests when the nominal level is 5%, $z_{ijr}, r = 1, \ldots, q; j = 1, \ldots, n_i; i = 1, \ldots, k$, are i.i.d. $N(0,1)$ and $M = 80$. The associated standard deviations (%) are given in parentheses.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$n$</th>
<th>$L^2$</th>
<th>F</th>
<th>GPF</th>
<th>$F_{\text{max}}$</th>
<th>$L^2$</th>
<th>F</th>
<th>GPF</th>
<th>$F_{\text{max}}$</th>
<th>$L^2$</th>
<th>F</th>
<th>GPF</th>
<th>$F_{\text{max}}$</th>
<th>$L^2$</th>
<th>F</th>
<th>GPF</th>
<th>$F_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>5.2</td>
<td>5.30</td>
<td>4.60</td>
<td>5.00</td>
<td>0.10</td>
<td>4.75</td>
<td>5.14</td>
<td>4.40</td>
<td>3.85</td>
<td>7.77</td>
<td>8.60</td>
<td>5.25</td>
<td>19.40</td>
<td>18.50</td>
<td>20.50</td>
<td>27.02</td>
<td>42.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
</tr>
<tr>
<td>0.5</td>
<td>5.2</td>
<td>4.80</td>
<td>4.50</td>
<td>4.40</td>
<td>0.20</td>
<td>4.60</td>
<td>4.84</td>
<td>4.40</td>
<td>1.66</td>
<td>12.26</td>
<td>13.03</td>
<td>15.93</td>
<td>41.40</td>
<td>40.58</td>
<td>42.83</td>
<td>64.38</td>
<td>78.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.30)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
</tr>
<tr>
<td>0.7</td>
<td>4.5</td>
<td>4.50</td>
<td>4.26</td>
<td>4.60</td>
<td>0.30</td>
<td>4.60</td>
<td>4.86</td>
<td>4.60</td>
<td>2.50</td>
<td>24.80</td>
<td>25.84</td>
<td>26.16</td>
<td>82.78</td>
<td>82.48</td>
<td>83.86</td>
<td>89.36</td>
<td>99.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td>(0.29)</td>
</tr>
<tr>
<td>0.9</td>
<td>4.5</td>
<td>4.70</td>
<td>4.50</td>
<td>4.70</td>
<td>0.40</td>
<td>4.70</td>
<td>4.90</td>
<td>4.70</td>
<td>3.70</td>
<td>12.74</td>
<td>14.34</td>
<td>12.24</td>
<td>45.80</td>
<td>44.70</td>
<td>47.80</td>
<td>48.84</td>
<td>85.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.32)</td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
<td></td>
<td></td>
<td></td>
<td>(0.33)</td>
</tr>
</tbody>
</table>

$n_i = [20, 30, 30], n_2 = [40, 30, 70], n_3 = [80, 70, 100].
Table 2: Same as in Table 1, except that \( z_{ijr}, r = 1, \ldots, q; j = 1, \ldots, n_i; i = 1, \ldots, k, \) are i.i.d. \( t_4/\sqrt{2}. \)

| \( \rho \) | \( n \) | \( L^2 \) | \( F \) | \( GPF \) | \( F_{\text{max}} \) | \( L^2 \) | \( F \) | \( GPF \) | \( F_{\text{max}} \) | \( L^2 \) | \( F \) | \( GPF \) | \( F_{\text{max}} \) | \( L^2 \) | \( F \) | \( GPF \) | \( F_{\text{max}} \) | \( L^2 \) | \( F \) | \( GPF \) | \( F_{\text{max}} \) |
| \( 0.1 \) | \( 2 \) | \( 3 \) | \( 0.56 \) | 5.46 | 5.38 | 4.82 | 8.74 | 8.20 | 8.74 | 15.64 | 20.20 | 19.64 | 20.46 | 60.32 | 52.86 | 51.65 | 53.70 | 98.26 | 81.58 | 80.74 | 81.64 | 100 |
| \( 0.3 \) | \( 2 \) | \( 3 \) | \( 0.56 \) | 5.46 | 5.38 | 4.82 | 8.74 | 8.20 | 8.74 | 15.64 | 20.20 | 19.64 | 20.46 | 60.32 | 52.86 | 51.65 | 53.70 | 98.26 | 81.58 | 80.74 | 81.64 | 100 |

| \( n_1 \) | \( 5.20 \) | \( 5.20 \) | \( 4.92 \) | 8.00 | 8.00 | 8.90 | 9.46 | 20.18 | 19.23 | 21.06 | 28.46 | 43.08 | 41.98 | 44.90 | 65.66 | 69.84 | 68.68 | 71.94 | 91.86 |
| \( 0.3 \) | \( 0.31 \) | \( 0.32 \) | \( 0.30 \) | \( 0.39 \) | \( 0.38 \) | \( 0.40 \) | \( 0.41 \) | \( 0.56 \) | \( 0.55 \) | \( 0.57 \) | \( 0.63 \) | \( 0.70 \) | \( 0.69 \) | \( 0.70 \) | \( 0.67 \) | \( 0.64 \) | \( 0.65 \) | \( 0.63 \) | \( 0.37 \) | \( 0.38 \) | \( 0.35 \) | \( 0.35 \) |

| \( \delta = 0 \) | \( \delta = 0.05 \) | \( \delta = 0.1 \) | \( \delta = 0.15 \) | \( \delta = 0.2 \) |

| \( n_1 \) | \( 4.78 \) | 4.51 | 5.04 | 4.92 | 12.90 | 11.92 | 13.68 | 11.94 | 46.00 | 45.10 | 48.04 | 49.92 | 84.54 | 82.02 | 85.26 | 90.90 | 98.82 | 98.48 | 98.66 | 98.46 |
| \( 0.3 \) | \( 0.31 \) | \( 0.32 \) | \( 0.30 \) | \( 0.45 \) | \( 0.45 \) | \( 0.45 \) | \( 0.45 \) | \( 0.69 \) | \( 0.69 \) | \( 0.69 \) | \( 0.69 \) | \( 0.67 \) | \( 0.57 \) | \( 0.57 \) | \( 0.55 \) | \( 0.55 \) | \( 0.57 \) | \( 0.57 \) | \( 0.55 \) | \( 0.55 \) |
| \( \delta = 0 \) | \( \delta = 0.05 \) | \( \delta = 0.1 \) | \( \delta = 0.15 \) | \( \delta = 0.2 \) |

| \( n_1 \) | 4.12 | 3.88 | 4.56 | 4.94 | 14.04 | 13.42 | 15.22 | 11.54 | 54.26 | 53.24 | 55.80 | 42.48 | 93.24 | 90.03 | 92.75 | 89.24 | 99.94 | 99.94 | 99.94 | 99.94 | 99.92 | 99.32 |
| \( 0.28 \) | \( 0.27 \) | \( 0.29 \) | \( 0.30 \) | \( 0.49 \) | \( 0.48 \) | \( 0.50 \) | \( 0.45 \) | \( 0.70 \) | \( 0.70 \) | \( 0.69 \) | \( 0.35 \) | \( 0.36 \) | \( 0.41 \) | \( 0.49 \) | \( 0.03 \) | \( 0.03 \) | \( 0.03 \) | \( 0.03 \) | \( 0.03 \) | \( 0.03 \) |
| \( \delta = 0 \) | \( \delta = 0.05 \) | \( \delta = 0.1 \) | \( \delta = 0.15 \) | \( \delta = 0.2 \) |

| \( n_1 \) | 4.92 | 4.82 | 5.24 | 5.34 | 50.86 | 50.62 | 51.25 | 40.38 | 99.78 | 99.78 | 99.80 | 99.16 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| \( 0.30 \) | \( 0.30 \) | \( 0.31 \) | \( 0.31 \) | \( 0.70 \) | \( 0.70 \) | \( 0.70 \) | \( 0.69 \) | \( 0.66 \) | \( 0.66 \) | \( 0.66 \) | \( 0.66 \) | \( 0.53 \) | \( 0.54 \) | \( 0.53 \) | \( 0.53 \) | \( 0.54 \) | \( 0.54 \) | \( 0.54 \) | \( 0.54 \) | \( 0.54 \) | \( 0.54 \) |

\( n_1 = \{20, 30, 30\}, n_2 = \{40, 30, 70\}, n_3 = \{80, 70, 100\}. \)
4. Screening Myocardial Ischemia

In this section we apply the $F_{\text{max}}$-test to an ischemia dataset to demonstrate that it is useful in discovering new methods for detecting stable myocardial ischemia (MI). Detecting myocardial ischemia (MI) is an important clinical mission, in particular in outpatient evaluations. Typical evaluations include resting surface electrocardiography (ECG) examination and non-invasive stress testing, for example, exercise ECG and the single-photon emission computed tomography (SPECT) thallium scan. A stress testing is performed to assess the heart’s response to stress via reading the ECG signal. Although resting ECG is a fundamental measure in the diagnosis, compared with stress tests, it is not accurate enough in detecting MI in a typical chest pain clinic (Gibbons et al., 2003). On the other hand, while stress tests are more accurate than resting ECG, they have limitations in clinics. Among the limitations, the most important one is the stress itself — it is associated with the risk of acute attacks during the testing. Thus, stress tests cannot be performed on patients who are extremely susceptible to the provocations used in the tests, and such patients have to directly undergo invasive tests (Gibbons et al., 2003).

Recent findings from biophysics and pathology interdisciplinary work (Swan, 1979), researches on the physical characteristics of myocardial strain in echocardiography (Urheim et al., 2000) and studies on electric signals (del Rio et al., 2005) have necessitated a reappraisal of the ECG information for detecting MI. For example, spectral analysis of the resting ECG signals from dogs revealed a shift from high- to low-frequency ranges in ischemia cases (Mor-Avi & Akselrod, 1990). Similar phenomena were associated with balloon inflations during percutaneous transcatheter angioplasty in MI patients (Abboud et al., 1987; Pettersson et al., 2000). Based on these findings, we hypothesize that the ischemic myocardium information is contained in the resting ECG signals with some proper manipulation, and the power spectrum of the manipulated ECG signal is shifted to the low frequency. To test this hypothesis, the proposed $F_{\text{max}}$-test was applied to a clinical dataset.

4.1. Data Description

The dataset was based on 393 consecutive individuals, who visited an outpatient clinic complaining of chest pain and were evaluated the possibility of MI according to published guidelines (Gibbons et al., 2003). The following three groups were studied: the ischemic group, the normal group and the atrial fibrillation (Af) group, defined by the procedure shown in Fig. 3. The ischemic group consisted of patients who directly underwent a coronary arteriography (CAG) and diagnosed as MI and patients having positive exercise ECG and/or SPECT thallium scan. If a patient’s exercise ECG was inconclusive, the result of SPECT thallium scan was used for classification. The normal group comprised of patients who had at least one stress test, and any of the test results was negative. Individuals fulfilling any of the following criteria were excluded: (i) non-cardiac chest pain and a very low risk of developing MI; (ii) non-NSR (normal sinus rhythm); (iii)
already underwent exercise ECG, but were non-diagnostic or intolerant of the SPECT thallium scan, and unwilling to undertake further evaluation; (iv) a heart rate of less than 40 beats per minute. The Af group consisted of subjects excluded from the above categories and had atrial fibrillation. In the end, 161 patients were eligible for the analysis — 71 in the ischemic group, 66 in the normal group, and 24 in the Af group.

For each of the 161 subjects, 85 seconds of 12-lead resting ECG signals were acquired during his/her first visit to the outpatient clinic. The signals were acquired at 500Hz and quantized at 12 bits across ±10 mV (Bailey et al., 1990). The signals were passed through a digital low-pass filter with a −3dB cutoff at 60 Hz, and stored in double precision. All possible physiological activities, except for the inherent amplitude deviations, such as power line noise, thermal noise, etc., were preserved. The recorded 12 lead ECG signals for the $i$-th subject is denoted as $e^{(i)}$ so that $e^{(i)}(l,j)$ is the $l$-th lead ECG signal sampled at time $j\tau$, where $\tau = 1/500$ is the sampling interval and $j = 1, 2, \ldots, N$. The following procedure was apply to process the ECG signals.

For the $i$-th subject, we first reduce the influence of the lead system by recovering the 3-dimensional vectocardiograph (VCG) data with Levkov’s algorithm (Levkov, 1987), which is denoted as a $3 \times N$ matrix $V^{(i)}$, where the $j$-th column is the dipole current direction in $\mathbb{R}^3$ at time $j\tau$. Then, we reduce the heart rate variability (HRV) effect. Denote $t^{(i,l)} \in \{1, 2, \ldots, N\}$ to be the indices of time stamps of the $l$-th R-peak of the $n$-th subject. The R-peaks are detected using the standard method (Clifford et al., 2006). We then extract the first $L$ sinus heartbeat signals, where $L \in \mathbb{N}$. Here the $l$-th sinus heartbeat signal is the VCG
Figure 4: Upper left (right): a 5-second Lead II (Lead III) ECG signal. Lower left: the VCG signal of an extracted sinus heartbeat $b^{(i,l)}$ (black curve) superimposed with the direction of $v^{(i,l)}$ pointing from the origin (i.e., the atrio-ventricular node) to the R-peak (gray curve). Lower right: the adaptive lead signal $s_0^{(i,l)}$ constructed by projecting $b^{(i,l)}$ onto $v^{(i,l)}$.

signal between time $t^{(i,l)}$ and $t^{(i,l+1)}$, denoted as $b_0^{(i,l)}$, where $b_0^{(i,l)}(j,m) = e^{(i)}(j, t^{(i,l)} + m - 1)$, $j = 1, 2, 3$ and $m = 1, 2, \ldots, t^{(i,l+1)} - t^{(i,l)}$. Next, we interpolate each heartbeat by the cubic spline interpolation to be of the uniform length 500 to eliminate the HRV influence, and denote the $l$-th interpolated heartbeat as $b^{(i,l)}$. Then we “co-axial project” each $b^{(i,l)}$ onto the R-peak direction, which by our construction is the first column of $b^{(i,l)}$. Denote this R-peak direction as a column vector $v^{(i,l)} \in \mathbb{R}^3$, and we project $b^{(i,l)}$ onto $v^{(i,l)}$: $s_0^{(i,l)} = [v^{(i,l)}]^T b^{(i,l)} \in \mathbb{R}^{1 \times N}$. We call $s_0^{(i,l)}$ an adaptive lead signal, where the adjective adaptive means that the cardiac axis and lead system effects are reduced. Finally, we eliminate other physiological effects, in particular the respiration, by normalizing each beat $s_0^{(i,l)}$ such that its $L^2$ norm is 1, which is denoted as another row vector $s^{(i,l)}$ of length 500. The adaptive ECG waveform for the $i$-th subject is defined as

$$s^{(i)} = \frac{1}{L} \sum_{l=1}^{L} s^{(i,l)}. \quad (22)$$

A typical ECG signal, the VCG signal and the adaptive ECG waveform representation are demonstrated in Fig. 4.

To confirm our hypothesis derived from the results given in Mor-Avi & Akselrod (1990), Abboud et al. (1987), and Pettersson et al. (2000), that is, the power spectrum of a beat of an ischemic heart is shifted to the low frequency region, we study the power spectrum of $s^{(i)}$, which is defined as $P^{(i)} = |\mathcal{F}s^{(i)}|^2$, where $\mathcal{F}$ is the discrete Fourier transform. Here $P^{(i)}$ is a row vector of length 500. We follow the convention
and say the first entry of $P^{(i)}$ is the DC term, the second to 250 entries are of the positive frequencies and the left are of the negative frequencies. Furthermore, we define the cumulative power spectrum of $s^{(i)}$ as $f^{(i)}(m) = \sum_{j=2}^{m+1} P^{(i)}(j)$, where $m = 1, 2, \ldots, 250$. Fig. 5 shows a set of adaptive ECG waveforms $s^{(i)}$, power spectra $P^{(i)}$ and the electrophysiological fingerprints $f^{(i)}$.

4.2. Analysis and Results

We considered the one-way ANOVA problem (2) for the stable cardiac ischemia data. Note that all of the tests we have discussed require a common covariance function for the different groups. This equal covariance function assumption can be checked by some tests proposed in the literature, e.g., Fremdt et al. (2013) and Paparoditis & Sapatinas (2016). We applied these tests and the results suggest that there is no significant evidence to reject this equal covariance function assumption. For example, the P-values of the test proposed by Paparoditis & Sapatinas (2016) for testing whether the adaptive ECG waveform and cumulative power spectrum of the normal and ischemic groups have the same covariance function are respectively 0.1562 and 0.1738.
The simulation studies presented in Tables 1 and 2 indicates that in terms of power, the GPF test of Zhang & Liang (2013) is comparable with or outperforms the $L^2$-norm based test and the F-type test. Therefore, in the real data analysis presented in this section, we only applied the GPF and $F_{\text{max}}$ tests to the adaptive lead signals $s_0^{(i,l)}$, the adaptive ECG waveforms $s^{(i)}$, and their power spectra $P^{(i)}$ and cumulative power spectra $f^{(i)}$. The associated P-values are displayed in Table 3. It shows that the difference between the normal and ischemia groups is significant, at the 0.05 level, only when the $F_{\text{max}}$-test is applied to the adaptive ECG waveforms, their power spectra, or their cumulative power spectra. Even when the Af group is taken into account as the third group, the $F_{\text{max}}$-test is still significant with these three kinds of data. The significance of the $F_{\text{max}}$-test based on the adaptive ECG waveform indicates that the 12 lead ECG waveform does contain information about ischemic myocardium. However, this information is masked by the other physiological facts, like the lead system effect, the HRV and the respiration. HRV is the phenomenon of non-linear and non-stationary rhythms of heartbeats, caused by autonomic function, respiration and other physiological activities, even in resting conditions (Clifford et al., 2006). Thus, if we directly perform a Fourier transform to the 12 lead ECG signal, HRV would distort the information. In addition, lead system effect is inevitable — although all leads are placed in fixed locations on the body surface according to the standard, the cardiac axes vary among individuals and for each beat. These result in undesirable signal variations which hamper the visual inspection. Furthermore, the impedance inside the chest wall varies proportional to the breathing cycle, that is, the impedance increases as we inhale, and decreases as we exhale. Thus, the respiratory pattern will further compromise the information. The influence of the respiratory pattern is clearly evidenced by the insignificance of the $F_{\text{max}}$-test based on the adaptive lead signal $s_0^{(i,l)}$, as shown in Table 3. Note that the P-value is less than 0.05 when we apply GPF to the Adaptive ECG waveforms of the three groups. This is not surprising since the ventricular activity is perturbed by the abnormal atrial activity in the Af subjects.

### Table 3: P-values of the GPF and $F_{\text{max}}$ tests for screening stable ischemia.

<table>
<thead>
<tr>
<th></th>
<th>Normal and Ischemic</th>
<th>Normal, Ischemic and Af</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{\text{max}}$</td>
<td>GPF</td>
</tr>
<tr>
<td>Adaptive lead signal</td>
<td>0.0764</td>
<td>0.1317</td>
</tr>
<tr>
<td>Adaptive ECG waveform</td>
<td>0.0006</td>
<td>0.0578</td>
</tr>
<tr>
<td>Power spectrum</td>
<td>0.0084</td>
<td>0.1566</td>
</tr>
<tr>
<td>Cumulative power</td>
<td>0.0029</td>
<td>0.1531</td>
</tr>
</tbody>
</table>

The significant results of the $F_{\text{max}}$-test based on the power spectrum $P^{(i)}$ and the cumulative power spectrum $f^{(i)}$ confirm our another hypothesis — the power spectrum of the manipulated ECG signal is shifted to the low frequency. In conclusion, with the $F_{\text{max}}$-test, we are able to prove the concept that
quantitative analysis of a normalized spectrum of the resting ECG signals contains profound information which is useful in screening myocardial ischemia among patients with NSR in outpatient settings. This finding supports our hypothesis stimulated by the findings given by Mor-Avi & Akselrod (1990), Abboud et al. (1987), and Pettersson et al. (2000). Note also that our analysis does not require any information obtained from any stress tests. Based on these findings, we believe that a safer, and more convenient, affordable and cost-effective method for diagnosis of myocardial ischemia using statistical analysis of power spectrum of ECG signals may be very promising. Research in this direction is ongoing and the result may be reported in the near future.

5. Proofs of the Theoretical Results

We first state the following lemma given by Zhang & Liang (2013).

**Lemma 1.** Under Condition A, as \( n \to \infty \), we have

\[
\mathbf{z}_n(t) \xrightarrow{d} \mathcal{G} \mathcal{P}_k(0, \gamma \mathbf{I}_k), \quad \sqrt{n} [\gamma(s, t) - \gamma(s, t)] \xrightarrow{d} \mathcal{G} \mathcal{P}(0, \varpi),
\]

where \( \varpi = \{[s_1, t_1], (s_2, t_2]\} = E[v_{11}(s_1)v_{11}(t_1)v_{11}(s_2)v_{11}(t_2)] - \gamma(s_1, t_1)\gamma(s_2, t_2) \). In addition, we have \( \hat{\gamma}(s, t) = \gamma(s, t) + O_U \left\{ n^{-1/2} \right\} \), where \( O_U \) means "bounded in probability uniformly."

**Proof of Proposition 1** Under the given conditions and by Lemma 1, as \( n \to \infty \), we have \( \text{SSE}_n(t)/(n - k) = \hat{\gamma}(t, t) \xrightarrow{d} \gamma(t, t) \) uniformly for all \( t \in \mathcal{T} \). In addition, under the null hypothesis we have \( \mathbf{\mu}_n(t) \equiv 0 \) and hence \( \mathbf{z}_n \xrightarrow{d} \mathbf{z} \sim \mathcal{G} \mathcal{P}_k(0, \gamma \mathbf{I}_k) \). By Slutsky’s theorem, as \( n \to \infty \), we have \( F_n(t) - R(t) \xrightarrow{d} 0 \) for all \( t \in \mathcal{T} \), where \( R(t) = (k - 1)^{-1} \mathbf{z}(t)^T(\mathbf{I}_k - \mathbf{b} \mathbf{b}^T)\mathbf{z}(t)/\gamma(t, t) \) and \( \mathbf{I}_k - \mathbf{b} \mathbf{b}^T \) is the limit matrix of \( \mathbf{I}_k - \mathbf{b}_n \mathbf{b}_n^T \) as given by (9). Notice that \( \mathbf{I}_k - \mathbf{b} \mathbf{b}^T \) has the singular value decomposition (10). Let

\[
\mathbf{w}(t) := \mathbf{(I}_{k-1} - \mathbf{0}) \mathbf{U}^T \mathbf{z}(t)/\sqrt{\gamma(t, t)} = [w_1(t), w_2(t), \ldots, w_{k-1}(t)]^T.
\]  

(23)

Then \( \mathbf{w}(t) \sim \mathcal{G} \mathcal{P}_{k-1}(0, \gamma w \mathbf{I}_{k-1}) \), and it follows that \( R(t) = (k - 1)^{-1} \sum_{i=1}^{k-1} w_i^2(t) \).

By Condition A5 on \( \gamma(s, t) \) and a direct calculation by the definition of Hölder’s continuity, the covariance function \( \gamma_w(s, t) \) of the Gaussian process \( w(t) \) defined in (23) belongs to \( C^\beta (\mathcal{T} \times \mathcal{T}) \). Thus, Kolmogorov’s theorem (Koralov & Sinai, 2007, Theorem 18.19) says that, for all \( i = 1, 2, \ldots, k - 1 \) and \( 0 < \bar{\beta} < \beta \), there exists a subspace \( \Omega' \subset \Omega \), where \( \Omega \) is the probability event space, so that \( \Omega' \) is of probability 1 and for all \( \omega \in \Omega' \), there exists a continuous modification \( W_i(t), t \in \mathcal{T} \) of \( w_i(t), t \in \mathcal{T} \) so that \( W_i(t) \) is Hölder continuous with exponent \( \bar{\beta} \). That is, with probability 1, we can modify \( R(t) \) so that it is Hölder continuous with exponent \( \bar{\beta} \). We use the same notation \( R(t) \) to denote the modified version. Similarly, note that for each \( n \in \mathbb{N} \), by Condition A5, there exists an event space \( \Omega_n \subset \Omega \) with probability 1 so that for all \( \omega \in \Omega_n \), \( \text{SSE}_n(t) \) and \( \text{SSR}_n(t) \) are Hölder continuous. Since we have countably many \( \Omega_n \) and \( \Omega' \), we can determine
a subspace $\Omega_\infty \subset \Omega$ with probability 1 so that for all $\omega \in \Omega_\infty$, $R(t)$, $\text{SSE}_n(t)$ and $\text{SSR}_n(t)$ are Hölder continuous. We now work on $\Omega_\infty$ and $F_n(t) \xrightarrow{d} R(t)$ pointwisely still holds. Since $F_n(t) \xrightarrow{d} R(t)$ pointwisely and $R(t)$ is Hölder continuous, there exists $n_0 \in \mathbb{N}$ so that $F_n(t)$ is Hölder continuous for any $n > n_0$. Thus, we have $\sup_{t \in \mathcal{T}} F_n(t) \xrightarrow{d} \sup_{t \in \mathcal{T}} R(t) =: R_0$.

**Proof of Proposition 2** First of all, notice that given the original $k$ samples (1), the bootstrapped $k$ samples $v_{ij}(t), j = 1, 2, \ldots, n_i; i = 1, 2, \ldots, k \sim_i d_i$ SP$(0, \hat{\gamma})$ where $\hat{\gamma}(s, t)$ is the pooled sample covariance function (11). That is to say, the bootstrapped $k$ samples satisfy the null hypothesis (2). By Lemma 1 and under Condition A, as $n \to \infty$, we have $\hat{\gamma}(s, t) \xrightarrow{d} \gamma(s, t)$ uniformly over $\mathcal{T}^2$. Applying Proposition 1 leads to the first claim of the proposition and the second claim of the proposition follows immediately.

**Proof of Proposition 3** In the proof of Proposition 1, we show that under Condition A, as $n \to \infty$, $\text{SSE}_n(t)/(n - k) \xrightarrow{P} \gamma(t, t)$ uniformly in $t \in \mathcal{T}$ and $\mathbf{z}_n(t) \xrightarrow{d} \mathbf{z}(t) \sim \text{GP}_k(0, \gamma_{\mathbf{I}})$. Similarly, since $\mathcal{T}$ is a finite interval by Slutsky’s theorem, and (18), we can show that as $n \to \infty$, we have $F_{\max} \xrightarrow{d} R_1$ with $R_1 = \sup_{t \in \mathcal{T}} \{(k - 1)^{-1}[\mathbf{z}(t) + \mathbf{d}(t)]^T(\mathbf{I}_k - \mathbf{b}\mathbf{b}^T)[\mathbf{z}(t) + \mathbf{d}(t)]/\gamma(t, t)\}$, where $\mathbf{I}_k - \mathbf{b}\mathbf{b}^T$ has the singular value decomposition (10). Let $w(t)$ be defined as in the proof of Proposition 1 and let $\delta(t) = (\mathbf{I}_{k-1}, 0)\mathbf{U}^T \mathbf{d}(t)/\sqrt{\gamma(t, t)} = [\delta_1(t), \delta_2(t), \ldots, \delta_{k-1}(t)]^T$. Then $w(t) \sim \text{GP}_{k-1}(0, \gamma_w \mathbf{I}_{k-1})$ with $\gamma_w(s, t) = \gamma(s, t)/\sqrt{\gamma(s, s)\gamma(t, t)}$ and $(\mathbf{I}_{k-1}, 0)\mathbf{U}^T[\mathbf{z}(t) + \mathbf{d}(t)] = \mathbf{w}(t) + \mathbf{d}(t)$. Therefore, we have $R_1 = \sup_{t \in \mathcal{T}} \{(k - 1)^{-1}[w(t) + \mathbf{d}(t)]^T[w(t) + \mathbf{d}(t)]\} = \sup_{t \in \mathcal{T}} \{(k - 1)^{-1}\sum_{i=1}^{k-1}[\delta_i(t)]^2\}$.

**Proof of Proposition 4** By (20), we first have $P(F_{\max} \geq C_\alpha) \geq P(T_n \geq (b - a)C_\alpha^*)$. Notice that under Condition A and by Proposition 2, we have $(b - a)C_\alpha^* \xrightarrow{d} (b - a)C_\alpha$ with $C_\alpha$ being the upper 100$\alpha$ percentile of $R_0$. Under Condition A and the local alternative (17), by the proof of Proposition 3 in Zhang & Liang (2013), we have $P(T_n \geq (b - a)C_\alpha^*) \to 1$ as $\delta \to \infty$.

**Acknowledgements**

Zhang and Zhou’s research was supported by the National University of Singapore research grant R-155-000-164-112. Cheng’s research was supported by the Ministry of Science and Technology grant 104-2118-M-002-005-MY3. Wu’s research was supported by AFOSR grant FA9550-09-1-0551, NSF grant CCF-0939370, and FRG grant DMS-1160319. We thank Dr Chi-Jen Tseng (Fooyn University Hospital) for providing us with the ischemic heart dataset. The authors thank two reviewers and the AE for their invaluable comments and suggestions.

**References**


