Statistics and Its Interface Volume 12 (2019) 167–180

An L^2-norm based ANOVA test for the equality of weakly dependent functional time series

Jia Guo and Ying Chen*

We propose an L^2-norm based test for testing the equality of the mean functions of k groups of weakly dependent stationary functional time series. The proposed testing procedure is flexible and can be applied to both homoscedastic and heteroscedastic cases. Under the null hypothesis, the asymptotic random expression of the test statistic is a χ^2-type mixture, which is approximated by a two-cumulant and a three-cumulant matched χ^2 approximation methods, respectively. Under a local alternative hypothesis, the asymptotic random expression is also derived and the test is shown to be root-n consistent. Simulation studies are performed to compare the finite sample performance of the proposed test under various scenarios with alternatives e.g. an existing FPCA based test and some respective ANOVA tests. It is shown that the proposed test generally outperforms the alternative tests in terms of empirical sizes and powers. Two real data examples help to illustrate the implementation of our test based on the US yield curves and Google flu trends, respectively.

AMS 2000 subject classifications: Primary 62M10; secondary 62G10.
Keywords and phrases: χ^2-type mixture, Equality of the mean functions, Root-n consistency.

1. INTRODUCTION

Functional data nowadays has been available in many scientific fields including but not limited to economics, epidemiology, medicine and ergonomics. Statistical inference is useful in exploring the stochastic relationship between two or more data sets. We refer to Horváth and Kokoszka (2012) and Zhang (2013) and references therein for detailed discussions of functional data analysis. To evaluate the difference between the mean functions of functional data, Ramsay and Silverman (2005) extended the classical t-test and F-test to the functional domain. The test statistics are constructed pointwisely based on the finite discrete points of the functional observations and thus consider partial information. Faraway (1997) proposed an L^2-norm bootstrap based testing method integrating the pointwise test statistic over the continuous functional domain. The null distribution of the test statistic was however not provided rather a bootstrap method was applied. Zhang and Chen (2007) derived the asymptotic null distribution of the L^2-norm based test which was a χ^2-type mixture. With the help of functional principal component analysis (FPCA), Horváth et al. (2009) provided an alternative approach leading to a consistent testing procedure that projects the functional data onto a finite-dimensional space. Shen and Faraway (2004) and Zhang (2011) studied a global F-test by analogy to the usual F-test. For the functional two-sample problem, Zhang et al. (2010b) applied the L^2-norm based test to the two-sample Behrens–Fisher problem and Zhang et al. (2010a) advocated the use of the L^2-norm distance in two-sample mean testing problem. For the one-way ANOVA problem, Cuevas et al. (2004) studied the L^2-norm based test via bootstrap implementation and Zhang and Liang (2013) proposed another F-test statistic via globalizing the pointwise F-test statistic.

All the above works assume that the functional samples are independent. However, sometimes there exists serial dependence in functional observations which jeopardizes the tests under independence in terms of type I error. Hörmann and Kokoszka (2010) defined the m-dependence in functional time series and examined the effects of weak dependence on functional data analysis. Horváth et al. (2013) proposed a two-sample mean testing method for functional time series. Horváth and Rice (2015b) and Horváth and Rice (2015a) studied the one-way ANOVA problem for both independent functional samples and functional time series.

The inference for weakly dependent functional time series are usually based on FPCA. In particular, by means of projecting the functional samples into a finite d dimensional space, the functional curves are transformed into multivariate observations. A Wald type statistic is then used based on the projected samples whose null distribution is asymptotically a χ^2 distribution with the degrees of freedom being a function of d, where the value of d is generally chosen by cumulative variance approach (Section 3.3 in Horváth and Kokoszka 2012). It is worth noting the PCA based test statistic is constructed on a reduced multivariate domain that involves information loss. The cumulative variance approach chooses d based on the covariance only which has the potential to overlook the PCs though with small variance essential on interpreting difference on mean functions. Moreover, the test statistic depends on the inverse of the covariance matrix of the projected functional samples, which

*Corresponding author.

is not guaranteed to be invertible especially if the selected number of functional PCs is larger than the total sample size.

We propose a one-way ANOVA test for weakly dependent functional time series based on the L^2-norm. Instead of employing FPCA to convert functional data from infinite space to multivariate space, the test statistic is defined in infinite space and involves no information loss and also no need to select the number of PCs. The null distribution of the proposed test is well approximated by a two-cumulant or a three-cumulant matched χ^2 approximation method. Comparing with the existing L^2-norm tests which only consider the covariance function, we take the autocovariance function into account. Thus the proposed test is appropriate for both independent functional samples and weakly dependent functional time series. We study two cases of the one-way ANOVA problem. A homoscedastic case means dependent curves in different groups share the same covariance and autocovariance functions, while a heteroscedastic case considers various covariances or autocovariances in different groups. We show that the proposed test is root-n consistent. Simulation studies demonstrate the finite sample performance of the proposed test compared with the FPCA test proposed in Horváth and Rice (2015b) and the existing tests proposed for independent samples, see Zhang (2013) and Horváth and Rice (2015a). It shows the proposed test outperforms these alternatives under dependent data models and under independent data models performs comparably to the tests developed particularly for independent samples. As an illustration, we implement the proposed test to macroeconomics and public health by analyzing the US yield curves and the Google flu trends across four countries, respectively.

The rest of the paper is organized as follows. We present the main results in Section 2 and simulation studies in Section 3, respectively. Two real data examples are reported in Section 4. Section 5 concludes. Appendix provides the technical proofs of the main results.

2. MAIN RESULTS

We consider k groups of functional time series $\{y_{ij}(t),\ 1 \le i \le k,\ j \in \mathbb{Z},\ t \in \mathcal{T}\}$ defined on a continuous interval $\mathcal{T} = [a, b]$, where $\mathbb{Z}$ denotes the set of integers. We assume the functional time series is weakly stationary, i.e. at any time point t, the mean, covariance and serial dependence of the time series are constant within each group. On the other hand, the associated statistics may not be the same for different groups. Without loss of generality, we have

(1) $y_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t),\ t \in \mathcal{T},\ 1 \le j \le n_i$ and $1 \le i \le k$,

where $\mu_i(t)$ is the mean function of the i-th group with $\mu_i(t) \in L^2(\mathcal{T})$ and there are n_i observations in group i. Throughout this paper, $L^2(\mathcal{T})$ denotes the Hilbert space formed by all the squared integrable functions over $\mathcal{T}$ with the inner-product defined as $< f, g >= \int_{\mathcal{T}} f(t)g(t)\mathrm{d}t$, $f(t),\ g(t) \in L^2(\mathcal{T})$. And $\{\epsilon_{ij}(t),\ 1 \le j \le n_i,\ 1 \le i \le k,\ t \in \mathcal{T}\}$ denote the k sequences of random error functions exhibiting serial dependence with $\mathrm{E}\{\epsilon_{ij}(t)\} = 0$ and $\int_{\mathcal{T}} \mathrm{E}\{\epsilon_{ij}^2(t)\}\mathrm{d}t < \infty$. Within each group i, there exists serial dependence between $\epsilon_{ij}(t),\ 1 \le j \le n_i$. Across different groups, we assume that $\{\epsilon_{ij}(t),\ 1 \le j \le n_i,\ t \in \mathcal{T}\}$ and $\{\epsilon_{i'j}(t),\ 1 \le j \le n_{i'},\ t \in \mathcal{T}\}$, $i' \ne i$, are independent.

Our interest is to test the equality of the k mean functions

$$H_0 : \mu_1(t) = \cdots = \mu_k(t) = \mu(t).$$

This is an ANOVA (analysis of variance) problem for functional time series. If the k groups of samples share a common covariance function (exhibit homoscedasticity), under the null hypothesis, all the k samples can be pooled together which simplifies the testing problem. Otherwise if the underlying group covariance functions are different from each other (exhibit heteroscedasticity), the problem becomes more involved because the difference of covariance must be taken into consideration. We study the ANOVA problem under both homoscedasticity and heteroscedasticity with dependent functional time series characterized by the following assumptions:

Assumption 1. We assume that $\{\epsilon_{ij}(t),\ 1 \le j \le n_i,\ 1 \le i \le k\}$ are L^2 $\mathtt{m}$-approximable functional processes.

The meaning of "L^2 $\mathtt{m}$-approximable functional process" is as follows. We firstly suppose $\epsilon_{ij}(t)$ is a function of a sequence of Bernoulli shifts, i.e., $\epsilon_{ij}(t) = f(\theta_{ij}, \theta_{i(j-1)}, \dots)$, where θ_{ij} are i.i.d. random elements in a measurable space $\mathcal{H}$ and the measurable function f is a mapping from $\mathcal{H}^\infty$ to $L^2(\mathcal{T})$. In our study, we choose $\mathcal{H} = L^2(\mathcal{T})$. Let $\{\tilde{\theta}_{ij},\ 1 \le i \le k,\ j \in \mathbb{Z}\}$ be independent copies of $\{\theta_{ij},\ 1 \le i \le k,\ j \in \mathbb{Z}\}$ and we assume the condition that $\sum_{\mathtt{m}=1}^{\infty} \{\mathrm{E}||\epsilon_{ij}(t) - \epsilon_{ij}^{(\mathtt{m})}(t)||^2\}^{1/2} < \infty$ where $\epsilon_{ij}^{(\mathtt{m})}(t) = f(\theta_{ij}, \dots, \theta_{i(j-\mathtt{m}+1)}, \tilde{\theta}_{i(j-\mathtt{m})}, \tilde{\theta}_{i(j-\mathtt{m}-1)}, \dots)$. This assumption is used to approximate a stationary time series using finite dependent random variables (see Ibragimov 1962, Hörmann and Kokoszka 2010 and Aue et al. 2012).

It is important to consider weak dependence in time series analysis. Rosenblatt (1956) introduced strong mixing conditions, also see Doukhan (1994) and Bradley (2005) for the derived properties. Doukhan and Louhichi (1999) proposed a unified weak dependence condition based on the covariance structure. Wu (2005) proposed the physical and predictive dependence measures. However, m-dependence is a more direct relaxation of independence, see Section 21 of Billingsley (1968) and Pötscher and Prucha (1997). In the functional time series literature, Bosq (2012) developed the functional auto-regressive (FAR) model. Hörmann and Kokoszka (2010) proposed the $\mathtt{m}$-approximable functional processes, which is a general moment-based notion of dependence and thus appropriate for functional time series that are not necessarily to be linearly dependent. Aue et al. (2012) gave various examples and applications of the moment-based dependence structure.

Throughout this paper, we assume the group sample sizes tend to infinity proportionally, as in the following assumption:

Assumption 2. As $n \to \infty$, the k sample sizes satisfy $n_i/n \to \tau_i$, $i = 1, \dots, k$ such that $\tau_1, \dots, \tau_k \in (0,1)$ where $n = \sum_{i=1}^k n_i$ denotes the total sample size.

We denote the estimators of the group mean functions and the group covariance functions as

$$\hat{\mu}_i(t) = \bar{y}_i(t) = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}(t),\ 1 \le i \le k,$$

$$\hat{\gamma}_{i0}(s,t) = \frac{1}{n_i}\sum_{j=1}^{n_i}\{y_{ij}(s) - \bar{y}_i(s)\}\{y_{ij}(t) - \bar{y}_i(t)\},\ 1 \le i \le k.$$

Under the null hypothesis, the common mean function $\mu(t)$ of the k samples is estimated by the pooled sample mean function

$$\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}(t) = \frac{1}{n}\sum_{i=1}^{k} n_i \hat{\mu}_i(t).$$

While for independent functional data, the only focus is on the covariance functions $\gamma_{i0}(s,t)$. For weakly dependent functional time series, the autocovariance function is considered as well

$$\gamma_{ip}(s,t) = \mathrm{E}\{\epsilon_{i0}(s)\epsilon_{ip}(t)\},\ p \in \mathbb{Z}^+.$$

By taking account of the serial dependence between the random error functions $\{\epsilon_{ij}(t),\ j \in \mathbb{Z},\ t \in \mathcal{T}\}$, the long run covariance function is defined as

$$\begin{aligned} c_i(s,t) &= \mathrm{E}\{\epsilon_{i0}(s)\epsilon_{i0}(t)\} \\ &\quad + \sum_{p=1}^{\infty}[\mathrm{E}\{\epsilon_{i0}(s)\epsilon_{ip}(t)\} + \mathrm{E}\{\epsilon_{i0}(t)\epsilon_{ip}(s)\}] \\ &= \gamma_{i0}(s,t) + \sum_{p=1}^{\infty}\{\gamma_{ip}(s,t) + \gamma_{ip}(t,s)\}. \end{aligned}$$

The estimators of the autocovariance function and the long run covariance function are respectively as follows

$$\begin{aligned} (2)\quad \hat{\gamma}_{ip}(s,t) &= (n_i - p)^{-1} \\ &\quad \times \sum_{j=p+1}^{n_i}\{y_{ij}(s) - \bar{y}_i(s)\}\{y_{i(j-p)}(t) - \bar{y}_i(t)\}, \\ &\quad 0 \le p \le n_i - 1, \end{aligned}$$

$$(3)\quad \hat{c}_i(s,t) = \hat{\gamma}_{i0}(s,t) + \sum_{p=1}^{n_i-1} K(\frac{p}{h})\{\hat{\gamma}_{ip}(s,t) + \hat{\gamma}_{ip}(t,s)\},$$

where $K(\cdot)$ is a kernel function and $h = h(n)$ is the smoothing bandwidth. The estimator of the long run covariance function (3) is inspired by the spectral analysis of scalar time series. For univariate scalar time series, Grenander and Rosenblatt (1953) and Parzen (1957) constructed estimators of the spectral density function based on periodogram. Although asymptotically unbiased, the periodogram estimator of the spectral density function is not consistent. Kernel based smoothed estimators overcome the inconsistency problem of the periodogram estimator (Taniguchi and Kakizawa 2012), which is also applicable to the long run covariance estimator (3).

To ensure the consistency of the estimators (2) and (3), we need the following assumptions on the kernel function and bandwidth function, respectively, as imposed by Horváth et al. (2013).

Assumption 3. The kernel function $K(\cdot)$ satisfies that $K(0) = 1$, $K(\cdot)$ is continuous and bounded, and if $g > c$ for some $c > 0$, $K(g) = 0$.

Assumption 4. The bandwidth function h satisfies that $h(n) \to \infty$ and $h(n)/n \to 0$ as the total sample size $n \to \infty$.

We refer to Horváth et al. (2013), Horváth and Rice (2015a) and Horváth et al. (2016) for the choice of the kernel function and the bandwidth h.

2.1 Test statistic

It is obvious that $\hat{\mu}_i(t) - \hat{\mu}(t)$ estimates the difference between the individual mean function and pooled mean function. We use the L^2-norm test statistic for the one-way ANOVA problem of the weakly dependent functional data:

$$T_n = \sum_{i=1}^{k} n_i \int_{\mathcal{T}}\{\hat{\mu}_i(t) - \hat{\mu}(t)\}^2 \mathrm{d}t.$$

Note that T_n can be expressed as

$$T_n = \int_{\mathcal{T}} \mathbf{z}_n(t)^\top \boldsymbol{M}_n \mathbf{z}_n(t)\mathrm{d}t$$

where $\mathbf{z}_n(t) = [z_{n1}(t), \dots, z_{nk}(t)]^\top$ with $z_{ni}(t) = \sqrt{n_i}\{\hat{\mu}_i(t) - \mu(t)\}$, $i = 1, \dots, k$ and $\mathbf{M}_n = \mathbf{I}_k - \mathbf{b}_n\mathbf{b}_n^\top/n$ with $\mathbf{b}_n = (n_1^{1/2}, \dots, n_k^{1/2})^\top$. It is easy to verify that $\mathbf{M}_n$ is an idempotent matrix with rank $k-1$. The (i,j)-th entry of $\mathbf{M}_n$ is

$$m_{n,ij} = \begin{cases} 1 - n_i/n, & \text{if } i = j; \\ -\sqrt{n_i n_j}/n, & \text{if } i \neq j. \end{cases}$$

As $n \to \infty$, we have $\mathbf{M}_n \to \mathbf{M} = (m_{ij}) :\ k \times k$, where $m_{ij} = \begin{cases} 1 - \tau_i, & \text{if } i = j; \\ -\sqrt{\tau_i\tau_j}, & \text{if } i \neq j. \end{cases}$

To better understand the test statistic, we will derive the mean and variance of T_n under the null hypothesis. Throughout this paper, denote $\xrightarrow{d}$ and $\xrightarrow{P}$ as convergence in distribution and in probability, respectively, and let $\mathrm{GP}\{\mu(t), c(s,t)\}$ and $\mathcal{N}(\mu, \sigma^2)$ represent a Gaussian process with mean function $\mu(t)$ and covariance function $c(s,t)$, and a normal distribution with mean μ and variance σ^2, respectively. Define the trace of a bivariate function $c(s,t)$ as

$$\mathrm{tr}(c) = \int_{\mathcal{T}} c(t,t)\mathrm{d}t,$$

and the r-th power of $c(s,t)$ as

$$c^{\otimes r} = \int_{\mathcal{T}} \cdots \int_{\mathcal{T}} c(s,u_1)c(u_1,u_2)\cdots c(u_{r-1},t)\mathrm{d}u_1\cdots \mathrm{d}u_{r-1}.$$

Lemma 1 derives the asymptotic mean and variance of T_n under H_0 which helps to study the asymptotic null distribution of the test statistic for heteroscedastic populations.

Lemma 1. *Under Assumptions 1, 2 and the null hypothesis H_0, as $n \to \infty$, we have $T_n \xrightarrow{d} T_0$, where*

$$E(T_0) = \sum_{i=1}^{k} m_{ii} tr(c_i),$$
$$Var(T_0) = 2\sum_{\alpha=1}^{k}\sum_{\beta=1}^{k} m_{\alpha\beta}^2 tr(c_\alpha \otimes c_\beta).$$

The null distribution of T_n can be approximated based on the first two moments of the test statistic. Next we will show how to conduct various ANOVA inferences for functional time series with the homogeneous and heterogeneous long-run covariances.

2.2 Main results for homoscedastic populations

Throughout this subsection, we assume the k samples share the same long-run covariance function, that is, the following homoscedasticity assumption holds.

Assumption 5. We assume that all the long-run covariance functions are the same, i.e., $c_1(s,t) = \cdots = c_k(s,t) = c(s,t)$.

If Assumption 5 is satisfied, the common long run covariance function can be estimated by

$$\hat{c}(s,t) = \sum_{i=1}^{k} n_i/n\hat{c}_i(s,t). \tag{4}$$

2.2.1 Approximate and asymptotic null distributions for homoscedastic populations

Next we will show that for the homoscedastic one-way ANOVA problem, the test statistic asymptotically follows a χ^2-mixture. Let $\chi^2_{d,\delta}$ denote a χ^2 random variable with degrees of freedom d and noncentral parameter δ, and when $\delta = 0$, we simply write χ^2_d.

Theorem 1. *Under Assumptions 1, 2, 5 and the null hypothesis H_0, as $n \to \infty$, we have $T_n \xrightarrow{d} T_0$,*

$$T_0 \stackrel{d}{=} \int_{\mathcal{T}} ||\mathbf{z}^*(t)||^2 dt \stackrel{d}{=} \sum_{r=1}^{q} \lambda_r A_r,$$

where $\mathbf{z}^*(t) \sim GP_{k-1}(\mathbf{0}, c\boldsymbol{I}_{k-1})$, $A_r \sim \chi^2_{k-1}$, $r = 1, \dots, q$ are independent, λ_r, $r = 1, \dots, \infty$ are the decreasing-ordered eigenvalues of the common long run covariance function $c(s,t)$ with q being the number of all the positive eigenvalues so that $\lambda_q > 0$ and $\lambda_r = 0$, $r > q$.

Remark 1. q can be finite or infinite. Infinite q means that the common long run covariance function has infinite positive eigenvalues and the above theorem turns to $T_0 \stackrel{d}{=} \sum_{r=1}^{\infty} \lambda_r A_r$. Later in Theorem 4 we have $A_r \sim \chi^2_{k-1,(\lambda_r^{-1}\delta_r^2)}$, where q is used to distinguish positive eigenvalues and 0. It is worth noting that q is the number of positive eigenvalues rather a hyperparameter to be chosen. As mentioned, the proposed testing method involves no information loss and also no need to choose hyperparameter e.g. the number of PCs in the FPCA based methods.

Given that the asymptotic null distribution of T_n is the same as that of a χ^2-type mixture according to Theorem 1, we can approximate its distribution using the well-known Welch–Satterthwaite χ^2-approximation. That is, we approximate T_0 using that of a random variable

$$R_1 \stackrel{d}{=} \beta_1 \chi^2_{d_1}$$

via matching the first two cumulants of T_0 and R_1. By some simple algebra, we have

$$\beta_1 = \frac{\mathrm{tr}(c^{\otimes 2})}{\mathrm{tr}(c)},\ d_1 = (k-1)\frac{\mathrm{tr}^2(c)}{\mathrm{tr}(c^{\otimes 2})}. \tag{5}$$

Alternatively, we can also approximate the distribution of T_0 by that of a random variable of form $R_2 = \beta_2\chi^2_{d_2} + \beta_0$. The parameters β_2, d_2 and β_0 are determined via matching the first three cumulants of T_0 and R_2. Then we get

$$\begin{aligned} \beta_2 =& \frac{\mathrm{tr}(c^{\otimes 3})}{\mathrm{tr}(c^{\otimes 2})},\ d_2 = (k-1)\frac{\mathrm{tr}^3(c^{\otimes 2})}{\mathrm{tr}^2(c^{\otimes 3})} \\ \beta_0 =& (k-1)\{\mathrm{tr}(c) - \frac{\mathrm{tr}^2(c^{\otimes 2})}{\mathrm{tr}(c^{\otimes 3})}\} \end{aligned} \tag{6}$$

The Welch–Satterthwaite χ^2-approximation method has been shown to be able to deliver good accuracy with fast computation speed in the literature, including Satterthwaite (1946), Welch (1947), Zhang (2005) and recently by Zhang et al. (2015) in high-dimensional settings. Zhang et al. (2015) provided the error bound of the two-cumulant (2-c) Welch–Satterthwaite χ^2-approximation method which is $O(M) + O(d_1^{-1}) + O\{(d^*)^{-1/2} - d_1^{1/2}\}$, where $M = \frac{\mathrm{tr}(c^{\otimes 4})}{(k-1)\mathrm{tr}^2(c^{\otimes 2})}$ and $d^* = \frac{(k-1)\mathrm{tr}^2(c^{\otimes 3})}{\mathrm{tr}^3(c^{\otimes 2})}$ and

Zhang (2005) offered the error bound of the three-cumulant (3-c) matched Welch–Satterthwaite χ^2-approximation, i.e., $O(M) + O\{(d^*)^{-1}\}$.

Horváth et al. (2013), Horváth and Rice (2015b) and Horváth and Rice (2015a) developed FPCA based tests where an asymptotic χ^2-approximation is also applied to the null distribution. It is worth noting that the degrees of freedom of the asymptotic χ^2-distribution depends on the number of dimensions selected. This limitation may cause not only loss of information but also low power problem of their test as we will demonstrate in the simulation studies. In addition, the selected number of dimensions cannot be larger than the sample size due to the requirement of invertibility of sample covariance matrix in the test statistic.

2.2.2 Implementation for homoscedastic populations

For the given samples, we obtain the following naive estimators of β_1 and d_1 via replacing $c(s,t)$ with its estimator $\hat{c}(s,t)$ as given in (4) in the expression (5):

$$\hat{\beta}_1 = \frac{\mathrm{tr}(\hat{c}^{\otimes 2})}{\mathrm{tr}(\hat{c})},\ \hat{d}_1 = \frac{\mathrm{tr}^2(\hat{c})}{\mathrm{tr}(\hat{c}^{\otimes 2})}. \tag{7}$$

Similarly, for the three-cumulant matched χ^2-approximation, we get the estimators of β_2, d_2 and β_0 via replacing $c(s,t)$ in (6) with $\hat{c}(s,t)$ given in (4), and we denote the estimators as $\hat{\beta}_2$, $\hat{d}_2$ and $\hat{\beta}_0$, respectively. Let $\chi^2_d(\alpha)$ denote the upper α-quantile of a central χ^2 distribution with degrees of freedom d, we show that these naive estimators converge in probability to their underlying true values, in the sense that the estimated upper α-quantile converges to the theoretical upper α-quantile. The following assumption is used in Horváth et al. (2013) to prove the consistency of the long run covariance functions.

Assumption 6. $\lim_{\mathtt{m}\to\infty}\mathtt{m}\{\mathrm{E}||\epsilon_{ij}(t) - \epsilon_{ij}^{(\mathtt{m})}(t)||^2\}^{1/2} = 0$.

The relationship between Assumption 6 and Assumption 1 can be found in Horváth et al. (2013). There are many examples of functional time series satisfying both assumptions, such as the functional autoregressive process presented in Hörmann and Kokoszka (2010).

Theorem 2. *Under Assumptions 1–6, as $n \to \infty$, we have $\hat{\beta}_1 \xrightarrow{P} \beta_1$, $\hat{d}_1 \xrightarrow{P} d_1$, $\hat{\beta}_2 \xrightarrow{P} \beta_2$, $\hat{d}_2 \xrightarrow{P} d_2$ and $\hat{\beta}_0 \xrightarrow{P} \beta_0$. In addition, we have $\hat{C}_\alpha \xrightarrow{P} C_\alpha$ where $\hat{C}_\alpha = \hat{\beta}_1\chi^2_{\hat{d}_1}(\alpha)$ (or $\hat{\beta}_2\chi^2_{\hat{d}_2}(\alpha)+\hat{\beta}_0$) is the estimated critical value of T_n and $C_\alpha = \beta_1\chi^2_{d_1}(\alpha)$ (or $\beta_2\chi^2_{d_2}(\alpha) + \beta_0$) is the approximate theoretical critical value of T_n.*

2.3 Extension to heteroscedastic populations

In the previous section, we study the homoscedastic ANOVA for functional time series. In this section, we continue the ANOVA problem when the underlying population long-run covariance functions are heterogeneous.

2.3.1 Two-sample heteroscedastic case

We firstly consider a two-sample case. Given group number $k = 2$, we show the χ^2-mixture representation of T_0 under heteroscedasticity. Note the test statistic $T_n = \sum_{i=1}^2 n_i \int_{\mathcal{T}}\{\hat{\mu}_i(t) - \hat{\mu}(t)\}^2\mathrm{d}t = \frac{n_1n_2}{n}\int_{\mathcal{T}}\{\hat{\mu}_1(t) - \hat{\mu}_2(t)\}^2\mathrm{d}t = \int_{\mathcal{T}} P_n^2(t)\mathrm{d}t$ where $P_n(t) = \sqrt{\frac{n_1n_2}{n}}\{\hat{\mu}_1(t) - \hat{\mu}_2(t)\}$ is the pivotal test function for this two-sample problem that measures the mean difference between the two groups. We calculate the mean and covariance of $P_n(t)$,

$$\mathrm{E}\{P_n(t)\} = \sqrt{\frac{n_1n_2}{n}}\{\mu_1(t) - \mu_2(t)\},$$

$$\mathrm{Cov}\{P_n(s), P_n(t)\} = \frac{n_2}{n}c_1(s,t) + \frac{n_1}{n}c_2(s,t). \tag{8}$$

Theorem 3. *Under Assumptions 1, 2 and the null hypothesis H_0, when $k = 2$, as $n \to \infty$, we have $T_n \xrightarrow{d} T_0$, where $T_0 \overset{d}{=} \sum_{r=1}^q \lambda_r A_r$ with $A_r \sim \chi^2_1$ and λ_r, $r = 1, \ldots, \infty$ are the decreasing-ordered eigenvalues of Cov$\{P_n(s), P_n(t)\}$ with q being the number of all the positive eigenvalues so that $\lambda_q > 0$ and $\lambda_r = 0$, $r > q$.*

Similar to the k-sample homoscedastic case we studied in Section 2.2, we approximate the distribution of T_0 by that of a random variable of form $R_1 = \beta_1\chi^2_{d_1}$ or $R_2 = \beta_2\chi^2_{d_2} + \beta_0$ in this two-sample case. The associated parameters are determined via matching the first two cumulants of T_0 and R_1 or first three cumulants of T_0 and R_2. Then the parameters β_1, d_1, β_2, d_2, β_0 can be obtained by (5) and (6) via replacing the common long run covariance with $\mathrm{Cov}\{P_n(s), P_n(t)\}$ given in (8). Similarly, we obtain the parameter estimators via replacing $c(s,t)$ in (5) and (6) with $\widehat{\mathrm{Cov}}\{P_n(s), P_n(t)\} = \frac{n_2}{n}\hat{c}_1(s,t) + \frac{n_1}{n}\hat{c}_2(s,t)$ where $\hat{c}_i$ is defined in (3). Furthermore, the consistency can also be obtained by similar argument as the proof of Theorem 2 under Assumptions 1–4 and 6.

2.3.2 k-sample heteroscedastic case

For a general case when $k > 2$, under heteroscedasticity, it is noted the first two cumulants of T_0 given in Lemma 1 holds for $k > 2$ under heteroscedasticity. Inspired by the approximate χ^2-mixture representation derived in the k-sample case under homoscedasticity and heteroscedastic two-sample case, and also the two-cumulant (2-c) matched χ^2-approximation can be implemented as long as the first two cumulants of T_0 are available, we can similarly approximate the distribution of T_0 by that of a random variable of form $R_1 = \beta_1\chi^2_{d_1}$. The associated parameters are determined via matching the first two cumulants of T_0 and R_1. Based on Lemma 1, we have

$$\beta_1 = \frac{\mathrm{Var}(T_0)}{2\mathrm{E}(T_0)},\ d_1 = \frac{2\mathrm{E}^2(T_0)}{\mathrm{Var}(T_0)},$$

where $\mathrm{E}(T_0)$ and $\mathrm{Var}(T_0)$ are given in Lemma 1. Estimators of β_1 and d_1 can be obtained by plugging $\hat{c}_i$ defined in (3) into $\mathrm{E}(T_0)$ and $\mathrm{Var}(T_0)$. The consistency results of the associated parameters and estimated critical value are straightforward under Assumptions 1–4 and 6.

2.4 Approximate and asymptotic powers of a local alternative hypothesis

To study the asymptotic power of the proposed test, we specify the following local alternative

$$(9) \qquad H_{1n}: \ \mu_i(t) = \mu(t) + n_i^{-1/2} d_i(t), \ i = 1, \ldots, k,$$

where $\mu(t)$ is the common mean function and $\mathbf{d}(t) = [d_1(t), \ldots, d_k(t)]$ is a vector of fixed functions, which is independent from n. The local alternative is of particular interest as it will tend to the null hypothesis H_0 in a root-n rate and hence it is difficult to detect. In this section, we derive the alternative distribution of the proposed test under the homoscedastic case. Next we will discuss the root-n consistency property for the test statistic under both homoscedastic and heteroscedastic cases.

Theorem 4. *Under Assumptions 1, 2, 5 and the alternative hypothesis, as $n \to \infty$, we have $T_n \xrightarrow{d} T_1$,*

$$T_1 \stackrel{d}{=} \int_{\mathcal{T}} ||\mathbf{z}_1^*(t)||^2 \, dt \stackrel{d}{=} \sum_{r=1}^{q} \lambda_r A_r + \sum_{r=q+1}^{\infty} \delta_r^2,$$

where $\boldsymbol{z}_1^*(t) \sim GP_{k-1}\{(\mathbf{I}_{k-1}, \mathbf{0})\mathbf{U}^\top \mathbf{d}(t), c\mathbf{I}_{k-1}\}$, $A_r \sim \chi^2_{k-1, \lambda_r^{-1}\delta_r^2}$, $r = 1, \ldots, q$ are independent, λ_r, $r = 1, \ldots, \infty$ are the decreasing-ordered eigenvalues of the common long run covariance $c(s,t)$, $\phi_r(t)$, $r = 1, \ldots, \infty$ are the associated eigenfunctions, $\delta_r^2 = || \int_{\mathcal{T}} (\mathbf{I}_{k-1}, \mathbf{0})\mathbf{U}^\top \mathbf{d}(t) \phi_r(t) dt||^2$, $r = 1, \ldots, q$, with q being the number of all the positive eigenvalues so that $\lambda_q > 0$ and $\lambda_r = 0$, $r > q$.

Theorem 5. *Under Assumptions 1–4 and the alternative hypothesis, as $\delta^2 = \sum_{i=1}^k \int_{\mathcal{T}} \tilde{d}_i^2(t) dt \to \infty$, where $\tilde{d}_i$ is the i-th component of $\tilde{\mathbf{d}}(t) = \mathbf{M}_n \mathbf{d}(t)$, the proposed L^2-norm based test has asymptotic power 1, i.e., $\Pr(T_n > \hat{C}_\alpha) \to 1$ where $\hat{C}_\alpha$ can be $\hat{\beta}_1 \chi^2_{\hat{d}_1}(\alpha)$ or $\hat{\beta}_2 \chi^2_{\hat{d}_2}(\alpha) + \hat{\beta}_0$.*

Theorem 4 derives the asymptotic alternative distribution of the test statistic under homoscedasticity. When $\mathbf{d}(t) = \mathbf{0}$, we have $\delta_r^2 = 0$ and Theorem 4 then reduces to Theorem 1. And Theorem 5 shows that the asymptotic power of the proposed test will tend to 1 as long as the information provided by $\mathbf{d}(t)$ diverges. Thus, the proposed test enjoys root-n consistency property under both homoscedasticity and heteroscedasticity.

In summary, we present a new testing procedure for the ANOVA problem for functional time series based on squared L^2-norm. When the underlying population long run covariance functions are homogeneous, the asymptotic null distribution of the test statistics is a χ^2-type mixture. The proposed test is conducted by means of the 2-c and 3-c matched Welch–Satterthwaite χ^2 approximation methods. The method also applies for the two-sample case when the underlying population long run covariance functions are heterogeneous. Although the approximate χ^2-mixture representation is not easy to derive for multi-sample ($k > 2$) heteroscedastic case, the 2-c χ^2 approximation method is still applicable and thus the proposed test can be conducted. In terms of power study, the asymptotic alternative distribution under homoscedasticity is illustrated. Regardless whether under homoscedasticity and heteroscedasticity, the root-n consistency property of the test statistics is provided. Compared with the existing tests proposed for independent samples, our proposed test is applicable to both independent and dependent functional data. In contrast to the FPCA based test proposed in Horváth and Rice (2015a), the proposed test involves no information loss and also no need to specify the number of PCs.

3. NUMERICAL SIMULATIONS

In this section, we perform simulations to investigate the finite sample performance of the proposed tests under known data generating process. We compare the empirical sizes and powers of the proposed tests with the FPCA based methods proposed in Horváth and Rice (2015b) under various scenarios with both homoscedastic and heteroscedastic serially dependent functional data. To demonstrate the advantage of utilizing appropriate tests for weakly dependent functional data, we also generate independent functional samples and investigate the performance of the L^2-norm based tests and FPCA based tests; see references Zhang (2013) and Horváth and Rice (2015a), respectively. For easy reference, the considered tests are named as follows:

- L2OM2d and L2OM3d: the L^2-norm based tests with the two-cumulant (2-c) or three-cumulant (3-c) matched Welch–Satterthwaite χ^2-approximation for the homoscedastic populations developed in our study where "d" stands for weakly dependent functional samples,
- L2TM2d: the L^2-norm based test with the 2-c matched χ^2-approximation for the heteroscedastic populations developed in this study,
- HRO3d, HRO4d and HRO5d: the FPCA based methods for the homoscedastic populations with the first three, four and five functional principal components selected, proposed by Horváth and Rice (2015b),
- HRO.9d: the FPCA based method with some d so that 90% of the sample variance can be explained by the first d principal components, proposed by Horváth and Rice (2015b),
- HRT3d, HRT4d, HRT5d and HRT.9d: similar FPCA based methods for the heteroscedastic populations proposed by Horváth and Rice (2015b).

For the tests proposed for independent functional data, we change the last letter "d" of the above notations to "i", i.e., L2OM2i, L2OM3i and L2TM2i proposed by Zhang (2013) and HRO3i, HRO4i, HRO5i, HRO.9i, HRT3i, HRT4i, HRT5i and HRT.9i proposed by Horváth and Rice (2015a). We also study the finite sample performance of the result stated in Theorem 5 to show how the power function of the proposed test changes as the difference of the sample mean functions of the k groups changes.

In the simulation studies, in total $k = 3$ groups of functional time series are generated following the dynamics $y_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t)$ as specified in (1). The mean functions are specified as $\mu_1(t) = c_1 + c_2 t + c_3 t^2 + c_4 t^3$ with $\mathbf{c} = (1, 2.3, 3.4, 1.5)^\top$ and $\mu_i(t) = \mu_1(t) + (i-1)\omega\Delta\mu(t)$, $i = 2, \ldots, k$ with $\Delta\mu(t) = \sum_{r=1}^{L} u_i \phi_i(t)$ where $\mathbf{u} = (u_1, \ldots, u_L) = (1, \ldots, L)/||(1, \ldots, L)||$, $\phi_1(t), \ldots, \phi_L(t)$ are the generated Fourier basis functions with $L = 21$. The parameter ω controls the differences between the means of different groups and $\Delta\mu(t)$ controls the direction of these differences. The functional time series are sampled discretely at $J = 100$ evenly spaced design time points within $\mathcal{T} = [0, 1]$. In particular, we use 49 Fourier bases to smooth the data and convert discretely simulated data into functional objects, before applying the FPCA based method proposed in Horváth and Rice (2015b).

Note here we only considered an ideal situation that the functional data are observed densely without any measurement errors. However, in practice, data curves may be observed with missing values and measurement errors. In such case, the observed curves should be reconstructed before conducting the tests; see, e.g., Ramsay and Silverman (2005), Ch.4, Zhang and Chen (2007), and Zhang (2013), Ch.3, for discussions of methods for reconstructing functional data. Additional simulations with a modified data generating model by taking missing values and measurement errors into consideration, and adopting a regression spline reconstruction of the curves, are presented in the supplementary material (http://intlpress.com/site/pub/pages/journals/items/SII/content/vols/0012/0001/s003).

To conduct the L^2-norm based test via Welch–Satterthwaite χ^2-approximation, we need to calculate the estimator of the long run covariance $\hat{c}$. Its value depends on the choice of the kernel function $K(\cdot)$ and the bandwidth function h. For simplicity, we choose the flat top kernel

$$K^*(t) = \begin{cases} 1, & 0 \leq t < 0.1; \\ 1.1 - |t|, & 0.1 \leq t < 1.1; \\ 0, & |t| \geq 1.1; \end{cases}$$

recommended by Politis and Romano (1996) and $h = n_i^{1/3}$.

3.1 Simulation HOM

In Simulation HOM (HOMoscedasticity), we compare the proposed L^2-norm based tests with the above-mentioned existing methods under homoscedasticity. The random error functions $\epsilon_{ij}(t)$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ are generated from the following three models:

- Model IID: $\epsilon_{ij}(t) = \eta_{ij}(t)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, where $\eta_{ij}(t)$'s are i.i.d. Brownian bridges;
- Model AR: $\epsilon_{ij}(t)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$ is a functional AR(1) process, i.e.,

$$\epsilon_{ij}(t) = \int_0^1 K(t,s)\epsilon_{i(j-1)}(s)\mathrm{d}s + \eta_{ij}(t),$$

  where $K(t,s) = \frac{\exp\{-(t^2+s^2)/2\}}{4\int_0^1 \exp(-x^2)\mathrm{d}x}$ is the kernel function and $\eta_{ij}(t)$'s are i.i.d. Brownian bridges.
- Model MA: $\epsilon_{ij}(t)$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ is a functional MA(1) process, i.e.,

$$\epsilon_{ij}(t) = \int_0^1 K(t,s)\eta_{i(j-1)}(s)\mathrm{d}s + \eta_{ij}(t),$$

  where $K(t,s)$ and $\eta_{ij}(t)$'s are the same as in Model AR.

All above models generate homoscedastic populations with Model IID generating independent functional samples and Models AR and MA generating dependent functional samples.

We study the impact of sample size on the tests' performance by specifying four cases of sample sizes: $\mathbf{n}_1 = (50, 60, 70)$, $\mathbf{n}_2 = (100, 120, 110)$, $\mathbf{n}_3 = (200, 220, 240)$ and $\mathbf{n}_4 = (300, 360, 320)$. The null hypothesis is rejected if the calculated p-value of a testing procedure is smaller than the nominal significance level $\alpha = 5\%$. We repeat the above process $10,000$ times. The empirical sizes or powers of the testing procedures are computed as the percentages of rejection in the $10,000$ runs. To assess the performance of a test in maintaining the type I error, we define the average relative error as $\text{ARE} = 100 S^{-1} \sum_{i=1}^{S} |\hat{\alpha}_i - \alpha|/\alpha$, where α is the significance level (5% here) and $\hat{\alpha}_i$, $i = 1, \ldots, S$ denote the empirical sizes under consideration. A smaller ARE value indicates better overall performance of the associated test in terms of maintaining the nominal size.

Table 1 gives the empirical sizes (in percentages) of our L^2-norm based tests and the existing tests under different settings of sample sizes and data models. For easy recognition, the names of our L^2-norm based tests are marked in bold in the table. In addition, the best (smallest) ARE value is marked in bold and the second best ARE value is underlined. In general, the L2OM3d test, i.e., the proposed L^2-norm based test with 3-c matched Welch–Satterthwaite χ^2-approximation for the homoscedastic populations, has the smallest ARE value of 38.18 i.e. it outperforms the other tests. Among the tests constructed for dependent functional data, the sizes of the tests designed for homoscedastic case are comparable. As sample size increases, their empirical sizes become better. Note that the FPCA based tests are less stable, for example, under Model AR with sample size $\mathbf{n}_3$, when increasing the number of functional principal components selected, the HRO4d test improves the empirical size of HRO3d test from 7.05 to 6.85 while the HRO5d test worsens from 6.85 to 7.04. It implies choosing a larger d cannot always improve the results of the FPCA based tests. For the tests designed for heteroscedastic case, the L2TM2d test, i.e., the proposed L^2-norm based test with the 2-c matched χ^2-approximation for the heteroscedastic populations, is superior with the second smallest ARE value of 40.72 while the empirical sizes of the heteroscedastic FPCA based tests are quite inflated. For the tests constructed for independent functional data, we can see that under the independent scenario, i.e., Model IID, the empirical sizes of the respective ANOVA tests constructed for independent samples are around the nominal size 5%. However, for the AR and MA models, their empirical sizes are too inflated and unacceptable. In general, those tests have better performance under the independent scenario but perform quite poorly for the dependent data.

Table 2 compares empirical powers. The column of ω represents the difference of mean functions. It shows that L^2-norm based tests outperform the FPCA based methods regardless whether the methods are constructed for dependent or independent functional data. Most of the empirical powers of the L^2-norm based tests are above 70%, that are much higher than those of the FPCA based methods with values smaller than 20%. In Model IID, the empirical powers of the proposed L2OM2d, L2OM3d and L2TM2d tests are slightly higher than those of the L2OM2i, L2OM3i and L2TM2i tests given that the empirical sizes of L2OM2d, L2OM3d and L2TM2d tests are slightly inflated as shown in Table 1. Similarly for Models AR and MA, the slightly higher empirical powers of the L2OM2i, L2OM3i and L2TM2i tests are not so informative because of their inflated empirical sizes.

3.2 Simulation HET

Simulation HET (HETeroscedasticity) is conducted to demonstrate the effect of heteroscedasticity. In the heteroscedastic scenarios, $\eta_{ij}(t)$ are generated by the following model:

$$(10)\qquad \eta_{ij}(t) = \mathbf{b}_{ij}^\top \mathbf{\Psi}(t),\ t \in [0,1],$$
$$\mathbf{b}_{ij} = [b_{ij1}, \ldots, b_{ijq}]^\top,\ b_{ijr} \stackrel{d}{=} \sqrt{\lambda_{ir}} z_{ijr},\ r = 1, \ldots, q,$$

where $\mathbf{\Psi}(t) = [\phi_1(t), \ldots, \phi_q(t)]^\top$ is a vector of q orthonormal basis functions, $\lambda_{ir} = \rho_i^{r-1}$, $r = 1, \ldots, q$ for $0 < \rho < 1$ and $z_{ijr},\ r = 1, \ldots, q,\ j = 1, \ldots, n_i,\ i = 1, \ldots, k \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$. We let $\rho_1 = 0.1$, $\rho_2 = \rho_3 = 0.9$ and $q = 21$. We use the same three models considered in Simulation HOM via replacing the i.i.d. Brownian bridges $\eta_{ij}(t)$ by the newly constructed $\eta_{ij}(t)$ defined in (10). For simplicity, we only consider two cases of sample sizes $\mathbf{n}_1 = (50, 150, 150)$ and $\mathbf{n}_2 = (100, 300, 300)$.

Table 3 shows the empirical sizes of our L^2-norm based tests and the existing tests under the heteroscedastic scenario. The homoscedastic tests, including the L^2-norm based tests and the FPCA based tests constructed for both dependent and independent functional data, are rather conservative in this case, indicating that these tests are inappropriate. On the other hand, the proposed L2TM2d test reaches to the smallest ARE with slightly inflated sizes under all the three models. When increasing the sample sizes, its empirical size tends to the nominal significance level. On the contrary, the tests HRT3d, HRT4d and HRT5d based on FPCA have conservative sizes and HRT.9d test has quite inflated sizes especially under the small sample size. As expected, the tests L2TM2i, HRT3i, HRT4i, HRT5i and HRT.9i only have good performance under Model IID which is reasonable given that these tests are designed for independent functional data. Under Model IID, their empirical sizes are conservative especially for the HRT.9i test. Again it is found that the FPCA based tests are not stable, choosing larger d or increasing the sample size may not always lead to better results. In terms of powers, Table 4 indicates that L2TM2d and L2TM2i tests have comparably high powers and outperform the other tests. However, the high power of L2TM2i test may due to the inflated sizes. Similar to homoscedastic case in Simulation 1, our L^2-norm based tests outperform the FPCA based methods regardless whether the methods are constructed for dependent or independent functional data.

In general, when the generated functional data exhibit serial dependence and share the same long run covariance function, the L2OM3d test performs best under different settings of sample sizes and data models. And when the dependent data have different long run covariance functions, the L2TM2d test outperforms the other tests. In contrast, the L2OM3i test only has best performance when the generated functional data are independent and homoscedastic and the L2TM2i test only outperforms the other tests when the generated functional data are independent and heteroscedastic. If there exists serial dependence in the generated data, the empirical sizes of L2OM3i and L2TM2i are quite inflated and unacceptable. Although the FPCA based tests sometimes have good empirical sizes, they are not stable and quite powerless under the simulation settings presented in this paper. Additionally, the FPCA based tests need the selection of the functional principal components and they offer different results when choosing different numbers of FPCs.

4. REAL DATA ANALYSIS

We implement the proposed test to two real data examples of the US yield curves and Google flu data. The US yield curve data contains the monthly interest rates over 27 years with different maturities. We would like to test the equality of various time period to see whether there is a monetary policy change in terms of level. And the Google flu trends dataset records estimates of weekly influenza activity for more than 25 countries over 13 years. Similarly, we are curious about whether different time periods have

Table 1. Empirical sizes (in percentages) of our L^2-norm based tests and the existing tests under homoscedasticity

		Tests for dependent functional data											Tests for independent functional data										
		Homoscedastic						Heteroscedastic					Homoscedastic						Heteroscedastic				
Model	$\mathbf{n}$	**L2OM2d**	**L2OM3d**	HRO3d	HRO4d	HRO5d	HRO.9d	**L2TM2d**	HRT3d	HRT4d	HRT5d	HRT.9d	L2OM2i	L2OM3i	HRO3i	HRO4i	HRO5i	HRO.9i	L2TM2i	HRT3i	HRT4i	HRT5i	HRT.9i
IID	$\mathbf{n}_1$	7.83	7.47	8.35	8.93	8.98	8.86	7.51	12.04	14.24	17.15	16.89	5.65	5.41	6.29	6.15	6.10	6.07	5.58	7.04	7.33	7.82	8.07
	$\mathbf{n}_2$	6.66	6.28	7.14	7.78	7.99	7.77	6.42	8.90	10.44	11.77	11.81	5.41	5.06	5.57	5.79	5.59	5.40	5.35	5.93	6.56	6.29	6.26
	$\mathbf{n}_3$	5.73	5.42	6.31	6.26	6.20	6.32	5.53	7.33	8.02	8.44	8.59	4.76	4.40	5.23	5.13	5.08	5.02	4.71	5.31	5.60	5.41	5.37
	$\mathbf{n}_4$	5.89	5.58	6.14	6.35	6.27	6.10	5.82	6.78	7.47	8.24	8.08	5.27	4.91	5.35	5.33	5.51	5.40	5.23	5.35	5.65	5.77	5.71
	ARE	30.55	23.75	39.70	46.60	47.20	45.25	26.40	75.25	100.85	128.00	126.85	7.85	**5.80**	12.20	12.00	11.40	9.45	7.25	18.15	25.70	26.45	27.05
AR	$\mathbf{n}_1$	9.91	9.55	9.57	9.61	10.19	9.54	9.69	13.20	15.20	18.37	15.37	15.68	15.06	12.92	12.46	12.21	12.11	15.43	14.19	13.71	13.96	14.03
	$\mathbf{n}_2$	8.38	8.10	7.82	8.02	8.56	8.09	8.16	9.67	11.12	12.58	11.20	15.33	14.75	12.14	11.40	11.28	11.06	15.24	12.42	12.58	12.50	12.23
	$\mathbf{n}_3$	7.10	6.89	7.05	6.85	7.04	6.79	6.97	8.04	8.31	9.34	8.30	15.29	14.64	11.87	11.13	10.72	10.26	15.19	12.12	11.39	11.26	10.78
	$\mathbf{n}_4$	6.54	6.21	6.42	6.42	6.65	6.44	6.44	7.52	7.78	8.70	7.80	14.98	14.28	11.95	11.10	10.81	10.57	14.87	12.15	11.40	11.11	11.00
	ARE	59.65	**53.75**	54.30	54.50	62.20	54.30	56.30	92.15	112.05	144.95	113.35	206.40	193.65	144.40	130.45	125.10	120.00	203.65	154.40	145.40	144.15	140.20
MA	$\mathbf{n}_1$	8.73	8.35	8.96	9.22	9.37	9.01	8.43	12.51	14.72	17.39	15.25	12.42	11.87	10.48	10.06	9.48	9.23	12.25	11.55	11.53	11.42	11.46
	$\mathbf{n}_2$	7.35	6.98	7.73	7.78	8.10	7.72	7.11	9.82	10.93	12.40	11.45	12.02	11.46	9.94	9.81	9.14	8.94	11.94	10.65	10.38	10.08	9.96
	$\mathbf{n}_3$	6.67	6.30	7.07	7.24	7.08	6.96	6.44	8.13	8.65	9.12	8.66	12.83	12.09	10.13	9.56	9.00	8.51	12.69	10.27	9.84	9.45	9.01
	$\mathbf{n}_4$	6.05	5.78	6.42	6.68	6.40	6.37	5.91	7.35	7.86	8.11	7.73	12.03	11.57	9.53	9.15	8.46	8.27	11.99	9.63	9.34	8.86	8.60
	ARE	44.00	**37.05**	50.90	54.60	54.75	50.30	39.45	89.05	110.80	135.10	115.45	146.50	134.95	100.40	92.90	80.40	74.75	144.35	110.50	105.45	99.05	95.15
ARE		44.73	**38.18**	48.30	51.90	54.72	49.95	40.72	85.48	107.90	136.02	118.55	120.25	111.47	85.67	78.45	72.30	68.07	118.42	94.35	92.18	89.88	87.47

Table 2. Empirical powers (in percentages) of our L^2-norm based tests and the existing tests under homoscedasticity

			Tests for dependent functional data											Tests for independent functional data										
			Homoscedastic						Heteroscedastic					Homoscedastic						Heteroscedastic				
Model	$\mathbf{n}$	ω	**L2OM2d**	**L2OM3d**	HRO3d	HRO4d	HRO5d	HRO.9d	**L2TM2d**	HRT3d	HRT4d	HRT5d	HRT.9d	L2OM2i	L2OM3i	HRO3i	HRO4i	HRO5i	HRO.9i	L2TM2i	HRT3i	HRT4i	HRT5i	HRT.9i
IID	$\mathbf{n}_1$	0.07	90.67	89.22	9.01	10.89	13.32	13.87	89.00	12.59	17.00	22.82	23.80	87.81	85.53	6.14	8.06	9.02	11.61	86.83	7.34	9.56	10.93	13.86
	$\mathbf{n}_2$	0.05	82.25	80.13	8.11	9.78	11.36	12.69	80.67	10.28	12.98	16.21	18.03	76.87	73.86	6.15	7.56	7.85	10.78	76.40	6.60	8.22	8.78	11.97
	$\mathbf{n}_3$	0.04	99.24	98.92	6.85	9.59	10.46	12.88	99.11	8.55	11.63	13.69	16.57	99.87	99.77	5.91	8.09	7.85	12.91	99.88	6.23	8.52	8.64	13.87
	$\mathbf{n}_4$	0.03	87.60	85.29	6.61	7.91	8.81	11.00	86.86	7.49	9.35	10.95	13.80	86.80	83.58	5.45	6.94	7.13	11.05	86.56	5.63	7.26	7.47	11.56
AR	$\mathbf{n}_1$	0.08	83.26	81.87	10.07	12.95	15.53	12.94	81.82	14.25	19.30	24.96	19.67	99.69	99.53	13.29	15.31	15.41	18.84	99.63	14.40	17.20	18.05	22.04
	$\mathbf{n}_2$	0.06	78.85	77.06	8.53	10.99	12.88	11.21	77.19	10.68	14.85	18.07	15.18	99.90	99.83	12.81	14.55	14.36	18.60	99.88	13.37	15.63	15.80	20.06
	$\mathbf{n}_3$	0.04	65.11	62.78	7.41	9.94	10.69	9.83	64.07	8.74	11.88	13.77	11.94	99.68	99.48	12.59	14.58	13.85	17.86	99.67	13.04	15.04	14.49	18.51
	$\mathbf{n}_4$	0.04	97.38	96.85	7.19	10.42	11.35	10.28	97.15	8.12	11.76	13.52	11.67	100.00	100.00	12.88	15.32	14.46	20.84	100.00	12.92	15.65	14.81	21.31
MA	$\mathbf{n}_1$	0.09	98.21	97.86	9.99	13.25	17.30	14.14	97.61	13.55	19.87	27.41	21.54	100.00	100.00	11.10	14.48	14.94	19.46	100.00	12.31	16.36	17.06	22.17
	$\mathbf{n}_2$	0.07	99.42	99.28	8.34	12.18	15.03	12.86	99.27	10.83	15.64	20.85	17.19	100.00	100.00	10.70	13.76	13.98	19.94	100.00	11.29	14.85	15.05	21.56
	$\mathbf{n}_3$	0.04	73.69	71.40	7.14	9.32	10.33	9.45	72.68	8.22	11.15	13.20	11.65	99.61	99.33	10.05	12.23	11.58	15.71	99.54	10.26	12.71	11.93	16.15
	$\mathbf{n}_4$	0.04	99.47	99.29	6.76	9.89	11.21	10.24	99.39	7.75	11.54	13.42	12.18	100.00	100.00	10.02	12.96	12.47	18.73	100.00	10.15	13.17	12.99	19.34

Table 3. Empirical sizes (in percentages) of our L^2-norm based tests and the existing tests under heteroscedasticity

		Tests for dependent functional data											Tests for independent functional data										
		Homoscedastic						Heteroscedastic					Homoscedastic						Heteroscedastic				
Model	**n**	**L2OM2d**	**L2OM3d**	HRO3d	HRO4d	HRO5d	HRO.9d	**L2TM2d**	HRT3d	HRT4d	HRT5d	HRT.9d	L2OM2i	L2OM3i	HRO3i	HRO4i	HRO5i	HRO.9i	L2TM2i	HRT3i	HRT4i	HRT5i	HRT.9i
IID	$\mathbf{n}_1$	0.67	0.64	0.56	0.43	0.34	0.20	6.13	2.95	3.13	3.54	16.36	0.46	0.43	1.67	1.45	1.04	0.00	5.57	4.30	4.10	4.22	0.77
	$\mathbf{n}_2$	0.54	0.49	0.75	0.64	0.44	0.09	5.30	2.78	2.76	3.09	5.51	0.50	0.50	1.74	1.36	1.11	0.01	4.95	4.10	3.99	4.20	0.34
	ARE	87.90	88.70	86.90	89.30	92.20	97.10	14.30	42.70	41.10	33.70	118.70	90.40	90.70	65.90	71.90	78.50	99.90	**6.20**	16.00	19.10	15.80	88.90
AR	$\mathbf{n}_1$	1.65	1.53	0.96	0.79	0.59	0.33	7.21	3.68	3.75	4.09	15.14	2.93	2.79	5.44	4.20	3.32	0.10	11.75	9.68	8.79	8.45	1.82
	$\mathbf{n}_2$	1.52	1.46	1.23	0.85	0.71	0.11	6.29	3.57	3.50	3.57	6.00	2.92	2.71	6.53	4.75	3.77	0.05	11.97	10.58	9.56	8.97	0.81
	ARE	68.30	70.10	78.10	83.60	87.00	95.60	35.00	27.50	27.50	23.40	111.40	41.50	45.00	19.70	**10.50**	29.10	98.50	137.20	102.60	83.50	74.20	73.70
MA	$\mathbf{n}_1$	1.60	1.43	1.09	0.67	0.59	0.38	6.44	3.73	3.90	4.12	15.06	2.13	2.02	3.95	3.03	2.60	0.08	9.68	7.90	7.26	7.11	1.32
	$\mathbf{n}_2$	1.30	1.21	1.32	1.00	0.66	0.23	6.10	3.43	3.48	3.42	5.93	2.13	2.03	5.04	3.60	2.87	0.07	9.81	8.69	7.77	7.22	0.73
	ARE	71.00	73.60	75.90	83.30	87.50	93.90	25.40	28.40	26.20	**24.60**	109.90	57.40	59.50	10.90	33.70	45.30	98.50	94.90	65.90	50.30	43.30	79.50
ARE		75.73	77.47	80.30	85.40	88.90	95.53	**24.90**	32.87	31.60	27.23	113.33	63.10	65.07	32.17	38.70	50.97	98.97	79.43	61.50	50.97	44.43	80.70

Table 4. Empirical powers (in percentages) of our L^2-norm based tests and the existing tests under heteroscedasticity

			Tests for dependent functional data											Tests for independent functional data										
			Homoscedastic						Heteroscedastic					Homoscedastic						Heteroscedastic				
Model	**n**	ω	**L2OM2d**	**L2OM3d**	HRO3d	HRO4d	HRO5d	HRO.9d	**L2TM2d**	HRT3d	HRT4d	HRT5d	HRT.9d	L2OM2i	L2OM3i	HRO3i	HRO4i	HRO5i	HRO.9i	L2TM2i	HRT3i	HRT4i	HRT5i	HRT.9i
IID	$\mathbf{n}_1$	0.26	55.69	54.22	0.53	0.54	0.60	1.83	92.70	5.24	7.36	11.26	82.56	52.02	51.06	2.00	1.73	1.58	0.44	92.76	6.34	7.89	11.09	61.20
	$\mathbf{n}_2$	0.20	74.56	73.35	1.15	1.04	0.87	1.42	98.26	6.03	8.63	12.62	83.66	73.19	72.22	2.35	1.89	1.82	0.46	98.38	7.43	9.09	13.07	70.63
AR	$\mathbf{n}_1$	0.28	60.06	57.80	1.49	1.19	0.98	2.27	91.57	7.68	10.02	14.12	86.30	75.65	74.58	6.12	4.89	4.21	1.26	98.15	13.92	15.83	19.31	77.58
	$\mathbf{n}_2$	0.20	60.04	57.41	1.78	1.37	1.18	1.67	91.38	7.11	10.21	13.98	83.14	78.29	77.34	7.22	5.74	5.22	0.92	98.55	14.76	15.69	19.35	75.08
MA	$\mathbf{n}_1$	0.28	62.49	60.21	1.07	1.02	0.90	2.21	92.80	6.66	9.33	13.92	86.49	74.11	73.06	4.62	3.58	3.26	1.10	98.06	11.83	13.07	17.24	76.35
	$\mathbf{n}_2$	0.20	63.40	61.42	1.46	1.19	1.09	1.66	94.00	6.98	9.55	13.50	84.10	77.25	76.17	5.41	4.31	4.11	0.78	98.55	12.68	14.11	17.71	74.71

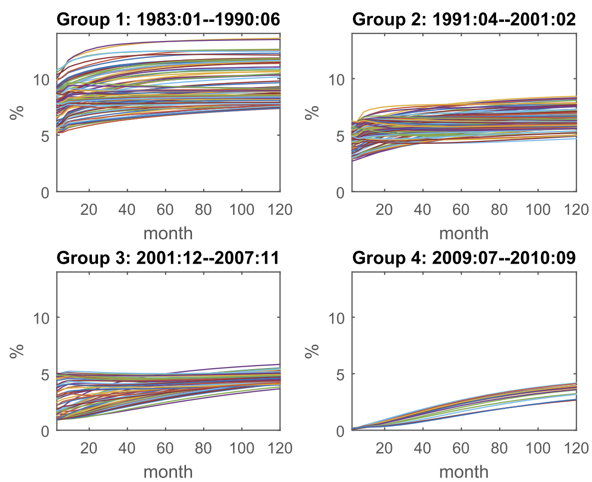

Figure 1. The monthly interest rates of different maturities for four groups.

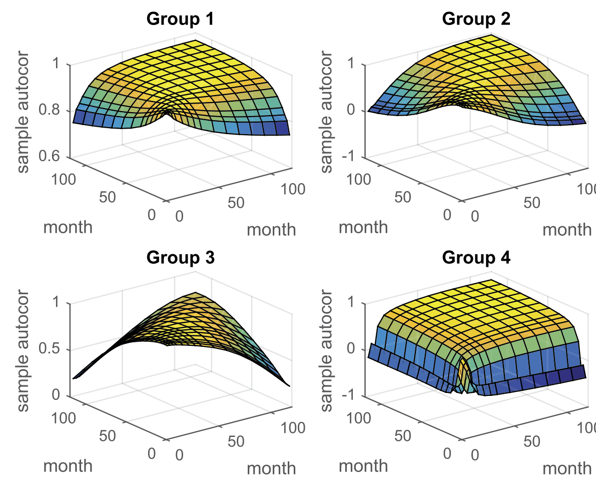

Figure 2. The sample autocorrelation functions of lag(1) for four groups.

equality in terms of public health (influenza). The observational curves are constructed using linear interpolation. Those curves within each population can be seen as functional time series. Testing whether there is a level shift in multiple groups is a commonly encountered problem. To solve this problem, our proposed test and the FPCA based test proposed in Horváth and Rice (2015b) are applied to the two real data examples and provide convincing statistical results. It is shown that our proposed test is preferred in practice due to its consistency and easy implementation.

4.1 US yield curve data

In this section, we consider a real data example of the US yield curve data which was previously studied in Chen and Niu (2014). This dataset records the monthly interest rates from January 1983 to September 2010 with maturities of 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. According to the National Bureau of Economic Research (NBER), there were three recessions during the sample period: namely 1990:7–1991:3, 2001:3–2001:11 and 2007:12–2009:6. We thus divide the whole dataset into 4 groups. Samples within the months of the recessions are discarded. The sample sizes of the 4 groups are $n_1 = 90$, $n_2 = 119$, $n_3 = 72$, $n_4 = 15$, respectively. Our interest is to compare the equality of the level of interest rates in the four periods. Figure 1 displays the monthly yield curve of interest rates against different maturities for all the 4 groups. It implies that the mean functions of the four groups are not identical. The functional samples within each group are obtained by separating a long continuous time record into monthly observations. So it is natural to assume that there exists serial weak dependence in the monthly yield curves. The sample autocorrelation functions of lag(1) for the four groups are displayed in Figure 2. Clearly, samples within each group are not independent, so the tests developed for independent functional data are not appropriate. In our study, we use the proposed L2OM2d, L2OM3d, L2TM2d, L2TM3d tests and the FPCA based tests proposed in Horváth and Rice (2015b).

Table 5. P-values of our L^2-norm based tests and the FPCA based tests applied to the US yield curve data

Methods	All	1 vs 2	2 vs3	3 vs 4
L2OM2d	1.99e-61	6.93e-20	6.31e-16	3.75e-04
L2OM3d	1.04e-55	2.06e-18	1.65e-14	4.46e-04
L2TM2d	4.26e-46	1.66e-16	1.09e-15	1.09e-15
L2TM3d	-	1.60e-15	2.33e-14	2.33e-14
HRO3d	2.23e-04	2.21e-02	1.09e-02	2.32e-01
HRO4d	4.09e-05	2.07e-02	1.82e-02	2.84e-01
HRO5d	1.22e-08	4.08e-02	1.41e-02	1.42e-03
HRO.9d	5.60e-05	1.43e-02	1.09e-02	2.60e-01
HRT3d	6.33e-33	4.79e-02	4.55e-03	3.22e-02
HRT4d	4.95e-52	4.60e-02	7.43e-03	4.54e-02
HRT5d	4.79e-101	8.45e-02	3.52e-03	2.63e-07
HRT.9d	1.91e-29	2.82e-02	4.55e-03	1.57e-02

The number of Fourier bases for the FPCA based tests, the kernel function and the bandwidth function for estimating the long run covariance function are the same as those used in simulation studies. Table 5 shows the p-values of our L^2-norm based tests and the existing tests applied to the US yield curve data. For space saving, we conduct four different group comparisons, one is all the four group, one is comparing Group 1 and Group 2, one is comparing Group 2 and Group 3, and another one is comparing Group 3 and Group 4. The results shown in Table 5 are consistent with the three recessions identified by NBER and Figure 1. Most of the p-values are less than 5% showing that the associated tests reject the null hypothesis. On the other hand, some FPCA based methods have p-values larger than the significance level 5%. In addition to the economic aspect

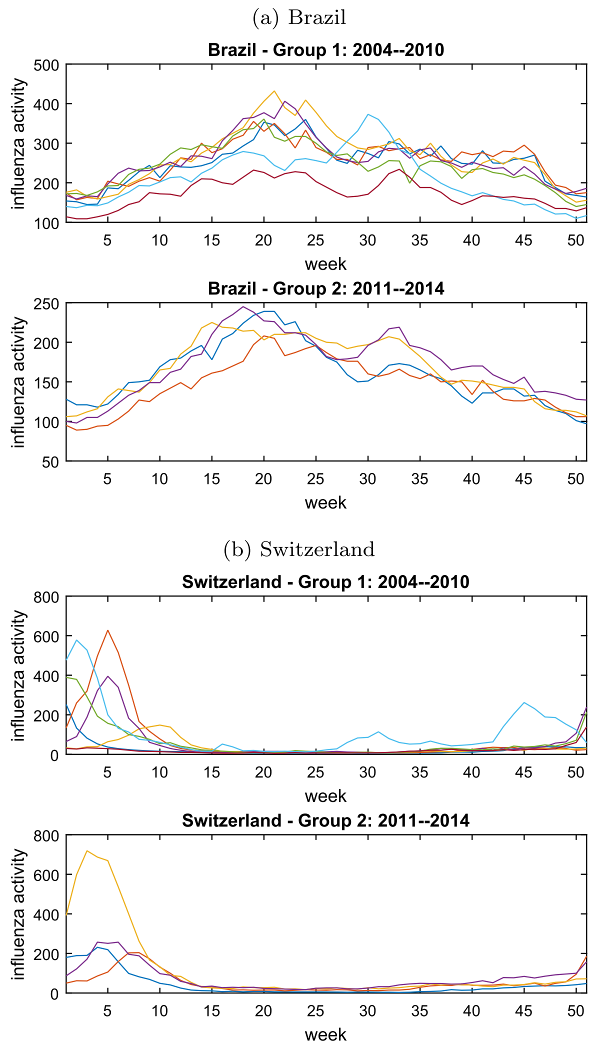

Figure 3. The yearly influenza activity estimates for two groups of Brazil and Switzerland.

of defining the three recessions, we provide an evidence of the three recessions via the significant results given by the proposed tests.

On the other hand, the FPCA based tests give us a misleading result of not rejecting the null hypothesis (with significance level 5%) if a wrong number of functional principal components is chosen. However, there is no standard method for selecting that number. Compared with their method, our proposed tests give consistent results and do not need any FPCA selecting procedure which is a big advantage.

4.2 Google flu trends data

The google flu trends dataset contains estimates of weekly influenza activity for more than 25 countries from the year 2002 to the year 2015 based on Google search queries. The data are available at http://www.google.org/flutrends. For space saving, only the countries Argentina, Brazil, Germany

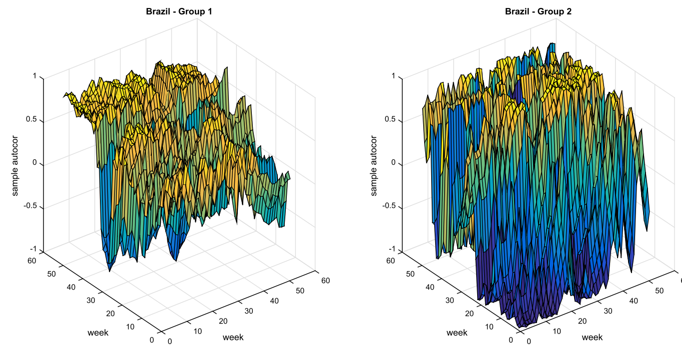

Figure 4. The sample autocorrelation functions of lag(1) for two groups of Brazil.

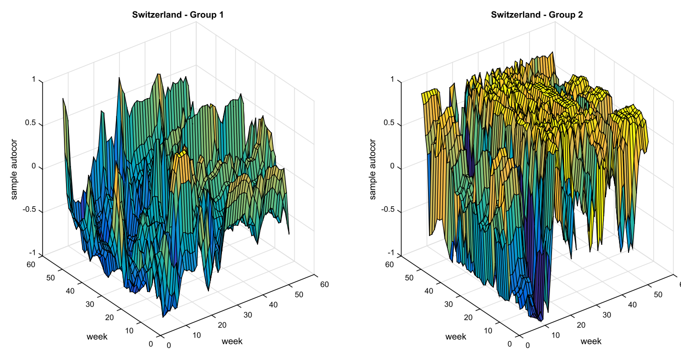

Figure 5. The sample autocorrelation functions of lag(1) for two groups of Switzerland.

and Switzerland are considered in our study. A primary examination of the raw data shows that the mean influenza activity from 2011 to 2014 is lower than that from 2004 to 2010 for some countries. Our interest is testing the mean influenza activity differences of this two periods. We consider yearly influenza activity curves as samples, and the curves of the years 2004–2010 as the first group, the curves of the years 2011–2014 as the second group, then the statistical problem of interest is to test the mean difference of these two groups. Take Brazil and Switzerland as an illustration, Figure 3 shows the yearly influenza activity estimates of the two countries. It is seen that there may exist mean difference of the two groups of Brazil while the mean functions of the two groups of Switzerland may be the same. Since the samples of yearly influenza activity are obtained by dividing records over 11 years into yearly observations, the samples may exhibit serial dependence. Figures 4 and 5 give the sample autocorrelation functions of lag(1) for the two groups of Brazil and Switzerland. Clearly, samples within each group are dependent functional time series.

Since we have a quite small sample size for the second group, we only consider the FPCA based tests with the first one, two and three functional principal components selected. Table 6 gives the testing results for the four countries Ar-

Table 6. P-values of our L^2-norm based tests and the FPCA based tests applied to the Google flu trends data

Methods	Argentina	Brazil	Germany	Switzerland
L2OM2d	1.59e-03	2.86e-04	6.64e-02	3.45e-01
L2OM3d	1.90e-03	4.53e-04	6.58e-02	3.30e-01
L2TM2d	5.29e-05	7.36e-07	1.96e-02	3.13e-01
L2TM3d	9.99e-05	3.01e-06	2.29e-02	2.94e-01
HRO1d	5.91e-03	6.35e-04	4.59e-01	5.93e-01
HRO2d	6.12e-03	1.45e-03	5.55e-02	2.76e-01
HRO3d	8.02e-07	3.50e-07	1.15e-01	4.61e-01
HRO.9d	6.12e-03	1.45e-03	1.15e-01	2.76e-01
HRT1d	1.31e-03	9.21e-06	3.68e-01	5.48e-01
HRT2d	9.83e-04	1.58e-05	1.80e-02	3.91e-01
HRT3d	2.22e-09	4.90e-08	2.09e-02	5.74e-01
HRT.9d	9.83e-04	1.58e-05	2.09e-02	3.91e-01

gentina, Brazil, Germany and Switzerland. All tests suggest that there is significant influenza activity difference between the two periods for Argentina and Brazil and suggest that there is no significant difference for Switzerland. The results for Germany are not consistent. Tests without assuming homoscedasticity, i.e., L2TM2d, L2TM3d and HRT2d, HRT3d, HRT.9d think the difference is significant, while tests assuming homoscedasticity, i.e., L2OM2d, L2OM3d and HRO1d, HRO2d, HRO3d, HRO.9d are not so sure. However, the results without assuming homoscedasticity are more reliable because homoscedasticity is not a very realistic assumption in practice and is difficult to verify, especially when sample sizes are small as in this example. Besides, the p-values of L2OM2d, L2OM3d and HRO2d are close to the significance level 0.05, which also implies that there is at least some difference between the means of the two periods.

From this example, it also can be seen that FPCA based tests depend on the selection of d, which can affect the p-values a lot. When d is too small, these two tests are unable to detect the mean difference, and choosing a larger d may not always increase the power of these two tests.

5. CONCLUSION

For the mean testing problem of k groups of weakly dependent stationary functional time series, we propose an L^2-norm based test which can be applied to both homoscedastic and heteroscedastic cases and enjoys root-n consistency property. Simulation studies and real data analysis show that the proposed test generally outperforms the alternative tests in terms of empirical sizes and powers. One possible extension of this paper is to consider the covariance testing problem of two or multi groups of functional time series.

ACKNOWLEDGEMENT

This work is supported by Singapore Ministry of Education Academic Research Fund Tier 1 and Institute of Data Science at National University of Singapore. The first author also would like to thank Professor Wen-Lung Shiau, advanced data analysis center (PLS-SEM of Zhejiang University of Technology) for his help on this research.

Received 24 February 2018

REFERENCES

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(04):804–837. MR2959126

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons. MR0233396

Bosq, D. (2012). *Linear processes in function spaces: theory and applications*. Springer Science & Business Media. MR1783138

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2:107–144. MR2178042

Chen, Y. and Niu, L. (2014). Adaptive dynamic nelson–siegel term structure model with applications. *Journal of Econometrics*, 180(1):98–115. MR3188914

Cuevas, A., Febrero, M., and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122. MR2087932

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer. MR1312160

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342. MR1719345

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39(3):254–261. MR1462586

Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *The Annals of Mathematical Statistics*, 24(4):537–558. MR0058901

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3):1845–1884. MR2662361

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer Science & Business Media. MR2920735

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122. MR3008273

Horváth, L., Kokoszka, P., and Reimherr, M. (2009). Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37(4):571–591. MR2588950

Horváth, L. and Rice, G. (2015a). An introduction to functional data analysis and a principal component approach for testing the equality of mean curves. *Revista Matemática Complutense*, 28(3):505–548. MR3379038

Horváth, L. and Rice, G. (2015b). Testing equality of means when the observations are from functional time series. *Journal of Time Series Analysis*, 36(1):84–108. MR3300207

Horváth, L., Rice, G., and Whipple, S. (2016). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics & Data Analysis*, 100:676–693. MR3505826

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382. MR0148125

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, 28(2):329–348. MR0088833

Politis, D. N. and Romano, J. P. (1996). On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*, 51(1):41–53. MR1394143

Pötscher, B. M. and Prucha, I. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer Science & Business Media. MR1468737

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Spring, New York, 2nd edition. MR2168993
Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences*, 42(1):43–47. MR0074711
Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, 2(6):110–114.
Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica*, 14(4):1239–1258. MR2126351
Taniguchi, M. and Kakizawa, Y. (2012). *Asymptotic theory of statistical inference for time series*. Springer Science & Business Media. MR1785484
Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35. MR0019277
Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154. MR2172215
Zhang, C., Peng, H., and Zhang, J.-T. (2010a). Two samples tests for functional data. *Communications in Statistics – Theory and Methods*, 39(4):559–578. MR2745305
Zhang, J.-T. (2005). Approximate and asymptotic distributions of chi-squared–type mixtures with applications. *Journal of the American Statistical Association*, 100(469):273–285. MR2156837
Zhang, J.-T. (2011). Statistical inferences for linear models with functional responses. *Statistica Sinica*, 21(3):1431. MR2827530
Zhang, J.-T. (2013). *Analysis of variance for functional data*. CRC Press. MR3185072
Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079. MR2341698
Zhang, J.-T., Guo, J., Zhou, B., and Cheng, M. (2015). A simple and adaptive two-sample test in high-dimensions based on L^2 norm. Manuscript.
Zhang, J.-T. and Liang, X. (2013). *One-way ANOVA for functional data via globalizing the pointwise F-test*.
Zhang, J.-T., Liang, X., and Xiao, S. (2010b). On the two-sample Behrens–Fisher problem for functional data. *Journal of Statistical Theory and Practice*, 4(4):571–587. MR2758746

Jia Guo
College of Economics and Management
Zhejiang University of Technology
China
E-mail address: jia.guo@u.nus.edu

Ying Chen
Department of Statistics and Applied Probability
National University of Singapore
Singapore
E-mail address: stacheny@nus.edu.sg