Title
Test of Significance When Data Are Curves

Permalink
https://escholarship.org/uc/item/3f23v732

Authors
Fan, Jianqing
Lin, Sheng-Kuei

Publication Date
1998

Test of Significance when data are curves[*]

Jianqing Fan and Sheng-Kuei Lin

January 6, 1998

With modern technology, massive data can easily be collected in a form of multiple sets of curves. New statistical challenge includes testing whether there is any statistically significant difference among these sets of curves. In this paper, we propose some new tests for comparing two groups of curves based on the adaptive Neyman test and the wavelet thresholding techniques introduced in Fan (1996). We demonstrate that these tests inherit the properties outlined in Fan (1996) and they are simple and powerful for detecting differences between two sets of curves. We then further generalize the idea to compare multiple sets of curves, resulting in an adaptive high-dimensional analysis of variance, called HANOVA. These newly developed techniques are illustrated by using a dataset on pizza commercial where observations are curves and an analysis of cornea topography in ophthalmology where images of individuals are observed. A simulation example is also presented to illustrate the power of the adaptive Neyman test.

KEY WORDS: Adaptive Neyman test, adaptive ANOVA, repeated measurements, functional data, thresholding, wavelets.

SHORT TITLE: Testing sets of curves.

1   Introduction

With modern equipment, massive data can easily be scanned in a form of curves. More precisely, the observations of individuals at different time points are recorded. This kind of data is not uncommon. Examples include seismic recordings of earthquakes and nuclear explosions presented in Shumway (1988), a gait analysis in Rice and Silverman (1991), temperature-precipitation patterns in Ramsay and Dalzell (1991), and brain potentials evoked by flashes of light in Kneip and Gasser (1992). These

kinds of data are also called longitudinal data or functional data in the literature. See for example recent monographs by Jones (1993), Diggle, Liang and Zegger (1994), Hand and Crowder (1996), and Ramsay and Silverman (1997). The main difference between longitudinal data analysis and functional data analysis is in dimensionality of data vector, though such a distinction is not always clear. The dimensionality in functional data analysis is usually much higher and hence smoothing techniques are needed.

1.1 Business commercial data
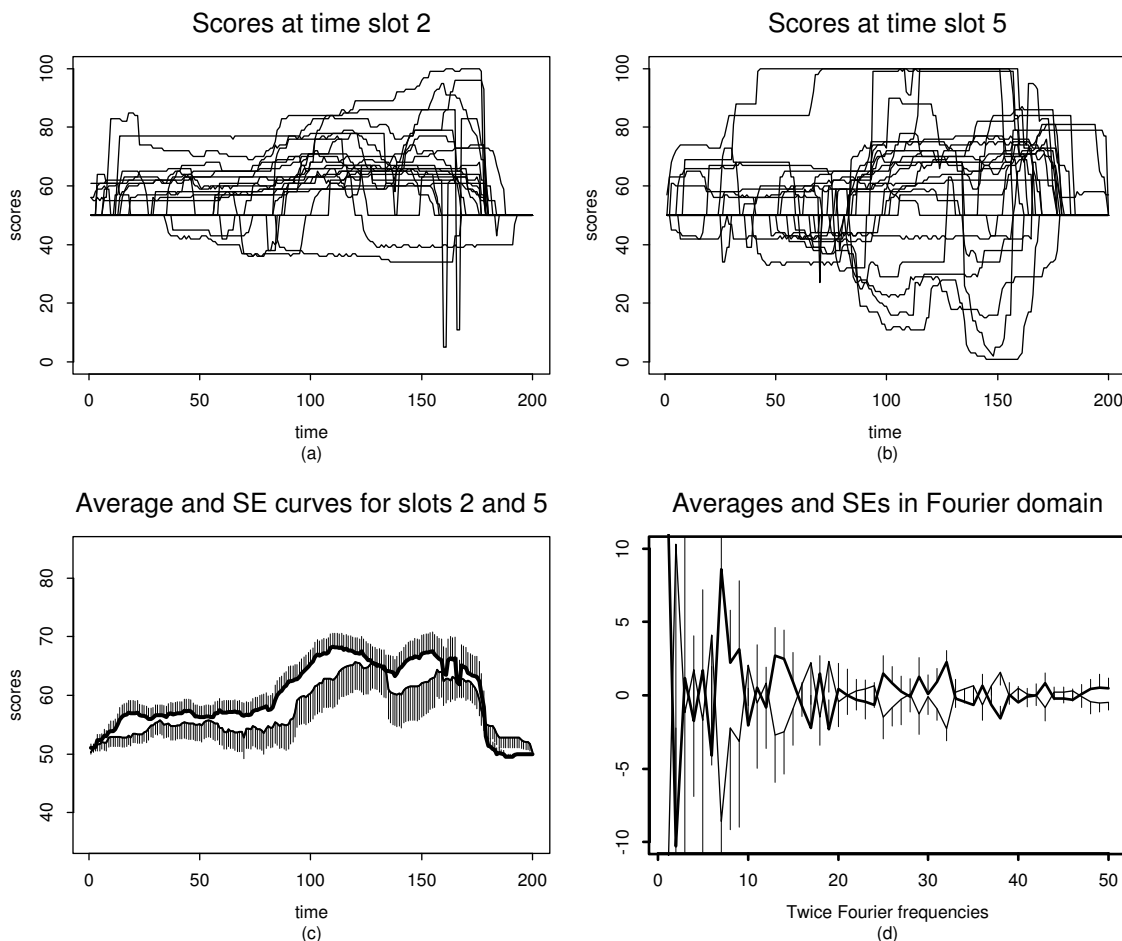
Figure 1: (a) Scores of a pizza commercial at time slot 2 with $n_1 = 21$. (b) Scores of the pizza commercial at time slot 5 with $n_2 = 24$. (c) The average and SD curves for time slots 2 and 5. Thick curve is the average curve at time slot 2 and thin curve is the average curve at time slot 5. The bars on the top or bottom indicate the estimated standard errors for the *average scores* at that time point. (d) A similar plot to (c), but now in the Fourier domain. Only a part of the coefficients are plotted here.

2

In evaluating business advertisements, evaluators are asked to dynamically assign scores to a commercial as they are watching. The resulting observations are a collection of curves: the score of the ith subject assigned at time t_j of the commercial advertisement. Figure 1 presents this kind of data on a pizza commercial. The commercial was played at studios at six different time slots and assessed by different evaluators. Of interest is to test if there is any significant time effect. For simplicity, we first consider the pairwise difference between the time slots 2 and 5, the scores at which are shown in Figures 1(a) and 1(b) respectively. The sample sizes are 21 for slot 2 and 24 for slot 5.

The observed curves are summarized in Figures 1(c). Presented in Figure 1(c) are the average and their standard error curves, which are respectively the sample average at each given time and the sample standard deviation at given time divided by the square root of sample size. There are several naive approaches to handling this kind of testing problem. Firstly, from Figure 1 (c) the time slot 2 is slightly preferred, though the evidence looks very weak since at each given time point the standardized difference is less than 0.5. This approach of pointwise two-sample t-test ignores completely the fact that at each given time point slot 2 is slightly preferred and when these evidences are combined properly the P-value should be much smaller. The challenge is how to combine these tests to yield a powerful overall test. The second naive approach is to treat each sample curve as a long multivariate vector and then use a multivariate technique such as Hotelling's T^2-test. This approach suffers from two serious drawbacks: It ignores completely the continuity of the scores assigned at neighboring time points and the dimensionality (200 for this example) is typically much larger than the sample sizes. The third naive approach is to compute the average score assigned by each individual and then apply a two-sample t-test. The average of the average scores assigned at slots 2 and 5 are respectively 59.1 and 57.6, with an SD of 6.5 and 11.7, respectively. The two-sample t-test statistic is 0.54 resulting in a P-value of 58.9%. This simple method ignores completely the fact that the average curve for time slot 2 is almost always somewhat larger than that for time slot 5.

The scores assigned by each subject are undoubtly correlated. A common device for decorrelating stationary data is to apply the discrete Fourier transform to each curve and obtain the transformed data in frequency domain, which are nearly independent and Gaussian. One can now summarize the transformed data in a similar way to Figure 1(c): At each given Fourier frequency,

3

compute the sample mean and its estimated standard error. These summary statistics in the frequency domain are presented in Figure 1(d). Shumway (1988) suggests to carry out a two-sample *t*-test at each given frequency and look for the frequencies that have significant differences between the two samples. This approach, though useful as an exploratory tool, involves what is so-called "data snooping" and does not combine the evidence at each given frequency to yield a more powerful overall testing procedure.

An objective of this paper is to propose a simple and powerful approach to combine properly the test statistics at different time points or different frequencies to obtain an overall test. This is then extended to compare multiple groups of curves. These techniques will be illustrated, in Sections 3.5 and 5.2, by using the pizza commercial data.

1.2   Other related work

There is large literature on longitudinal data analysis where various useful testing procedures have been developped. See for example Jones (1993), Diggle, Liang and Zeger (1994), Hand and Crowder (1996), Schimid (1996), among others. The procedures can also be applicable to our functional data analysis setting. They usually treat longitudinal data as a multivariate vector and do not incorporate a dimensionality reduction technique. For functional data analysis, the dimensionality is high and hence dimensionality reduction techniques are required. While they are powerful for analyzing longitudinal data, traditional tests for high-dimensional problems need some tuning. Faraway (1997) proposes to conduct smoothing on the functional data first and then use traditional ANOVA type of analysis.

As noted above, the dimensionality of curve testing problems is very high and some dimensionality reduction techniques are needed. The techniques used in this paper are based on orthogonal transforms such as the Fourier transform and the wavelet transforms to compress signals. This requires that data be sampled at equispaced design points such as those in Figure 1. For irregularly spaced designs, one can use either binning or interpolation methods to preprocess the data. See for example Cai (1996) and Hall and Turlach (1997) in a different context. Once the data are transformed by Fourier or wavelet techniques, useful coefficients in the transformed domain are then adaptively chosen to yield a test statistic. It remains to be seen how efficient these simple and crude methods are.

Another possible technique is to use a smoothed principal component analysis (Besse and Ramsay 1986 and Rice and Silverman 1991) and to project data on the first few important principal directions. The question is then how many principal directions should be chosen. Our adaptive Neyman test in Section 2 can be used for this purpose. The smoothed principal component approach involves choosing smoothing parameters and studying the effect of projected data on the estimated principal axes. Further, the principal axes are not necessary an efficient basis for compressing mean functions and can be quite different for two given sets of curves under alternative hypotheses. The idea along this has not yet been developed and its power remains to be seen.

Several recent papers study statistical issues when data are curves. Hart and Wehrly (1986) use kernel regression approach to estimate the mean curve. The bandwidth can be chosen by a cross-validation method, which is shown to be consistency by Hart and Wehrly (1993) even when curves are sampled from a dependent Gaussian process. See Hart (1996) for an overview of smoothing for dependent data. Besse and Ramsay (1986) investigate the principal component analysis of sampled functions. Rice and Silverman (1991) and Ramsay and Dalzell (1991) study estimation of mean curves and use principal component analysis to extract salient features of curves. Pezzulli and Silverman (1993) show the weak consistency of the estimated eigenvalues and eigenfunctions in the smoothed principal component analysis and study the effect of smoothing. Statistical problems associated with curves sampled from a population with individual variations were investigated by Kneip and Gasser (1992), Silverman (1995) and Capra and Müller (1996). Leurgans, Moyeed and Silverman (1993) extend the canonical correlation analysis to random functions and show that smoothing is needed in order to give sensible analyses. Further developments on functional data analysis can be found in the book by Ramsay and Silverman (1997) and the references therein.

The problems on hypothesis testing for curves were considered in the time series literature (see for example Shumway 1988). However, nonparametric treatments of the problems appear relatively new. The literature on nonparametric goodness-of-fit provides useful insights to our study. Closely related idea to ours is the recent paper by Fan (1996) who introduces a few powerful tests for high-dimensional testing problems. Hall and Hart (1990) propose a bootstrap test for detecting a difference between two mean functions in a nonparametric regression setting. Eubank and Hart (1992) and Eubank and LaRiccia (1992) consider the goodness-of-fit test problem based on cross-validation. Bickel and Ritov (1992) provide illuminating insights into nonparametric tests.

Data-driven methods for smoothed test were studied by Inglot, Kallenberg and Ledwina (1994), Ledwina (1994), Inglot and Ledwina (1996) and references therein.

1.3 Outline

We begin with summarizing the key ingredient of high-dimensional hypothesis testing techniques developed by Fan (1996). The key ideas on the adaptive Neyman test and wavelet thresholding tests are discussed in Section 2. In Section 3, we show how these procedures can be applied to detect differences between two sets of curves. We then develop in Section 4 the analysis of variance when the dimensionality is high, leading to an adaptive high-dimensional ANOVA, called HANOVA. The HANOVA is applied in Section 5 to test differences among multiple sets of curves. Some concluding remarks are noted in Section 6. Technical proofs are given in the Appendix.

2 A review of the adaptive Neyman test and thresholding tests

This section is mainly excerpted from Fan (1996) except Table 1. Let $\mathbf{X} \sim N(\theta, I_n)$ be an n-dimensional normal random vector. We wish to test

$$H_0 : \theta = 0 \quad \longleftrightarrow \quad H_1 : \theta \neq 0. \tag{2.1}$$

The maximum likelihood ratio test statistic for problem (2.1) is $\|\mathbf{X}\|^2$, which tests all components of $\mathbf{X}$. This test has approximate power

$$1 - \Phi\left(\frac{z_{1-\alpha} - \|\theta_1\|^2/\sqrt{2n}}{\sqrt{1 + 2\|\theta_1\|^2/n}}\right) \approx 1 - \Phi\left(z_{1-\alpha} - \|\theta_1\|^2/\sqrt{2n}\right), \tag{2.2}$$

at the alternative $\theta = \theta_1$, provided that $\|\theta_1\|^2 = o(n)$, where α is the significant level and $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$. The power tends to α even though $\|\theta_1\| \to \infty$ with $\|\theta_1\|^2 = o(\sqrt{n})$. The moral is that testing too many dimensions accumulates large stochastic noise and hence decreases the discrimability power of the test. The price is reflected in the factor $1/\sqrt{n}$ at the right hand side of (2.2).

2.1 Adaptive Neyman's test

Expression (2.2) suggests that it is not a good idea to test all components of the vector θ. If there is vague prior that large absolute values of θ are mainly located on the first m components, then

one would only test the first m-dimensional subproblem, leading to the test statistic $\sum_{j=1}^{m} X_j^2$, or equivalently its standardized form $(2m)^{-1/2} \sum_{j=1}^{m} (X_j^2 - 1)$. The parameter m has to be determined. Based on the power consideration, Fan (1996) proposes to use

$$\hat{m} = \text{argmax}_{m:1 \leq m \leq n} \left\{ m^{-1/2} \sum_{j=1}^{m} (X_j^2 - 1) \right\}.$$

This leads to the adaptive Neyman test statistic

$$T_{AN}^* = (\sqrt{2\hat{m}})^{-1} \sum_{j=1}^{\hat{m}} (X_j^2 - 1) = \max_{1 \leq m \leq n} \left\{ (\sqrt{2m})^{-1} \sum_{j=1}^{m} (X_j^2 - 1) \right\}. \tag{2.3}$$

This test is equivalent to rejecting H_0 when

$$T_{AN} = \sqrt{2 \log \log n} \, T_{AN}^* - \{ 2 \log \log n + 0.5 \log \log \log n - 0.5 \log(4\pi) \} \tag{2.4}$$

is too large. Let us call the distribution of T_{AN} under H_0 as J_n. The asymptotic distribution of J_n is given by

$$P\{T_{AN} \leq x\} \rightarrow \exp\{-\exp(-x)\}, \tag{2.5}$$

which is a corollary of a result in Darlin and Erdös (1956). As noted in Fan (1996), the convergence of (2.5) is very slow. Table 1 gives the finite sample distribution of J_n based on one million simulations.

Note that when $X_i's$ are not normally distributed, the constant factor 2 in (2.3) should be replaced by $\text{var}(X_i^2)$ under H_0. In particular, when the model is replaced by $X_i \sim t_d(\theta_i)$, a noncentral t distribution, one adjusts T_{AN}^* as

$$T_{AN}^* = \max_{1 \leq m \leq n} \left\{ \sqrt{\frac{(d-2)^2(d-4)}{2md^2(d-1)}} \sum_{j=1}^{m} (X_j^2 - 1) \right\}. \tag{2.6}$$

To examine the robustness of the distribution of T_{AN} with respect to the underlying distribution of X_i, we simulated the distribution of T_{AN} when X_i follows a Student t-distribution. When degrees of freedom is 20, the results are close to Table 1 for $0.5 \geq \alpha \geq 0.01$

Fan (1996) has also shown that having to estimate m, the adaptive Neyman test performs at least as good as the ideal Neyman test, which uses the knowledge of the optimal m, within a factor of $(\log \log n)^{1/2}$. See Theorem 2.2 of that paper. To obtain the required vague prior, one can apply the discrete Fourier transform to the observations $\mathbf{X}$ before implementing the adaptive Neyman test (2.3).

Table 1: α upper quantile of the distribution $J_n^{\dagger}$

$n\backslash\alpha$	0.00001	0.00005	0.0001	0.00025	0.0005	0.001	0.0025	0.005	0.01	0.025	0.05	0.10	0.25	0.50
5	12.45	11.39	10.55	9.42	8.65	7.80	6.74	5.97	5.21	4.23	3.50	2.77	1.77	0.96
10	16.53	13.91	12.79	11.53	10.47	9.13	7.73	6.77	5.78	4.57	3.67	2.74	1.49	0.40
20	16.82	14.88	13.95	12.49	11.27	9.83	8.26	7.16	6.07	4.75	3.77	2.78	1.41	0.18
30	17.58	15.47	14.62	12.90	11.65	10.11	8.47	7.29	6.18	4.82	3.83	2.81	1.39	0.11
40	18.71	16.11	15.14	13.34	11.92	10.34	8.65	7.41	6.22	4.87	3.85	2.82	1.39	0.08
50	19.85	16.08	15.15	13.35	11.89	10.32	8.67	7.43	6.28	4.89	3.86	2.84	1.39	0.07
60	18.45	16.25	15.26	13.53	12.16	10.56	8.80	7.51	6.32	4.91	3.88	2.85	1.39	0.07
70	20.37	17.04	15.40	13.65	12.27	10.59	8.81	7.55	6.34	4.92	3.88	2.85	1.40	0.06
80	20.46	16.24	15.17	13.68	12.24	10.54	8.81	7.57	6.37	4.93	3.89	2.85	1.40	0.06
90	21.84	17.67	15.78	14.06	12.62	10.79	8.95	7.65	6.40	4.94	3.90	2.86	1.40	0.06
100	20.73	17.37	15.87	14.12	12.59	10.80	8.95	7.65	6.41	4.95	3.90	2.86	1.41	0.06
120	20.02	16.99	15.87	14.19	12.77	10.87	8.96	7.65	6.41	4.95	3.90	2.87	1.41	0.05
140	21.08	17.14	15.98	14.08	12.43	10.80	9.00	7.66	6.42	4.95	3.90	2.86	1.41	0.05
160	19.42	17.07	16.19	14.40	12.97	10.88	8.95	7.69	6.42	4.95	3.91	2.87	1.41	0.06
180	21.33	17.71	16.14	14.50	12.97	11.02	9.10	7.77	6.47	4.95	3.90	2.87	1.41	0.06
200	20.50	17.51	16.34	14.31	12.78	11.10	9.08	7.72	6.43	4.95	3.89	2.86	1.42	0.06
∞	11.51	9.90	9.21	8.29	7.60	6.91	5.99	5.30	4.60	3.68	2.97	2.25	1.25	0.37

†The results are based on 1,000,000 simulations. The relative errors are expected to be around .3% – 3%.

2.2 Thresholding tests

If we have vague prior that large coefficients of the vector θ lie only on a few components, then a simple method for finding these informative components is thresholding. Such a vague prior can be obtained by a wavelet transform of the observation vector $\mathbf{X}$. See for example Donoho and Johnstone (1994).

A thresholding test statistic is defined by

$$\hat{T}_H^* = \sum_{j=1}^{n} X_j^2 I(|X_j| > \delta),$$

where $\delta > 0$ is called a thresholding parameter. It is shown in Fan (1996) that an approximate level α test for problem (2.1) is to reject H_0 when

$$\hat{T}_H = \sigma_{n,H}^{-1}\left(\hat{T}_H^* - \mu_{n,H}\right) > z_{1-\alpha}, \tag{2.7}$$

where with $\delta = \sqrt{2\log n\, a_n}$ for some sequence a_n tending to zero at a logarithmic rate,

$$\mu_{n,H} = \sqrt{2/\pi}\, a_n^{-1}\delta(1 + \delta^{-2}), \qquad \sigma_{n,H}^2 = \sqrt{2/\pi}\, a_n^{-1}\delta^3(1 + 3\delta^{-2}).$$

It is recommended to take $\delta_n = \sqrt{2\log(n/\log^2 n)}$ based on the considerations of both the power of the procedure and the accuracy of the approximation (2.7). It is also shown there that under some mild conditions, the testing procedure (2.7) performs within a logarithmic factor to the ideal thresholding estimator. Minimax properties of the thresholding test can be found in Spokoiny (1996).

Fan (1996) observes that the power of the above thresholding test can be improved upon if δ is replaced by the hard-thresholding parameter:

$$\delta_H = \sqrt{2\log(n\hat{a}_n)}, \quad \hat{a}_n = \min(4(X_{\max})^{-4}, \log^{-2} n),$$

where $X_{\max} = \max_{1 \leq i \leq n} |X_i|$. The intuition for this is that under H_0, $X_{\max} \approx (2\log n)^{1/2}$, and it is expected to be larger under the alternative hypothesis. A similar intuition leads to soft-thresholding parameter:

$$\delta_S = \sqrt{2\log(n\hat{a}_n)}, \quad \hat{a}_n = \min(\{\log(\sum_{i=1}^{n} X_i^2)\}^{-2}, \log^{-2} n).$$

A specific version of the thresholding test is $\hat{T}_H^* = \max_{1 \leq i \leq n} X_i^2$. This corresponds to a data-dependent thresholding parameter $\hat{\delta}$, which is the second largest order statistic among $\{X_1^2, X_2^2, \cdots, X_n^2\}$. It is closely related to the Fisher (1929) g-test for harmonic analysis. The test is expected to be powerful for alternatives with energy concentrated mostly on one dimension.

3 Comparing two sets of curves

We assume that the observed curves in the first group are a random sample from the model

$$X_j(t) = f_1(t) + \varepsilon_j(t), \qquad t = 1, \cdots, T, \quad j = 1, 2, 3, ..., n_1, \tag{3.1}$$

where the random variables $\varepsilon_j(t)$ have mean zero. Similarly, the second group of curves are a random sample from the model

$$Y_j(t) = f_2(t) + \varepsilon_j'(t), \qquad t = 1, \cdots, T, \quad j = 1, 2, 3, ..., n_2, \tag{3.2}$$

with the random variables $\varepsilon_j'(t)$ having mean zero. Of interest is to test

$$H_0 : f_1(t) = f_2(t) \qquad \longleftrightarrow \qquad H_1 : f_1(t) \neq f_2(t), \tag{3.3}$$

based on the above two sets of curves.

3.1 Independent heteroscedastic errors

We assume further that the random variables $\varepsilon_j(t) \sim N(0, \sigma_1^2(t))$ and $\varepsilon_j'(t) \sim N(0, \sigma_2^2(t))$ are independent for all j and t. Consider the summarized curves:

$$\bar{X}(t) = n_1^{-1} \sum_{j=1}^{n_1} X_j(t), \quad \bar{Y}(t) = n_2^{-1} \sum_{j=1}^{n_2} Y_j(t), \tag{3.4}$$

9

and

$$\hat{\sigma}_1^2(t) = \frac{1}{(n_1 - 1)} \sum_{j=1}^{n_1} \{X_j(t) - \bar{X}(t)\}^2, \quad \text{and } \hat{\sigma}_2^2(t) = \frac{1}{(n_2 - 1)} \sum_{j=1}^{n_2} \{Y_j(t) - \bar{Y}(t)\}^2. \tag{3.5}$$

Denote the standardized difference by

$$Z(t) = \{n_1^{-1}\hat{\sigma}_1^2(t) + n_2^{-1}\hat{\sigma}_2^2(t)\}^{-1/2}\{\bar{X}(t) - \bar{Y}(t)\} \tag{3.6}$$

and put $\mathbf{Z} = (Z(t), \cdots, Z(T))^T$. When n_1 and n_2 are reasonably large,

$$Z(t) \stackrel{a}{\sim} N\{d(t), 1\},$$

where $d(t) \approx \{n_1^{-1}\sigma_1^2(t) + n_2^{-1/2}\sigma_2^2(t)\}^{-1/2}\{f_1(t) - f_2(t)\}$, and "$\stackrel{a}{\sim}$" means "distributed approximately". Note that when $\sigma_1^2(t) = \sigma_2^2(t)$ is imposed, one can use the pooled variance estimate

$$\hat{\sigma}^2(t) = \frac{n_1 - 1}{n_1 + n_2 - 2}\hat{\sigma}_1^2(t) + \frac{n_2 - 1}{n_1 + n_2 - 2}\hat{\sigma}_2^2(t)$$

in (3.6). This technique is particularly useful when n_1 and n_2 are small.

Now one can apply the Fourier transform to the standardized difference vector $\mathbf{Z}$ to compress useful signals into low frequencies. Let $\mathbf{Z}^*$ be the resulting vector. One then applies the adaptive Neyman test statistic to the vector $\mathbf{Z}^*$ and consults Table 1 for the P-value of the test. As noted in Section 2.1, when the degrees of freedom are moderately large, with the adjustment (2.6) with $d = n_1 + n_2 - 2$, the null distribution of the adaptive Neyman test is reasonably close to that for normal errors. The adaptive Neyman test is useful when the population standardized difference function $d(t)$ is smooth.

In practice, one does not have to apply the adaptive Neyman test to the whole vector of $\mathbf{Z}^*$. Indeed, when T is large (e.g. $T = 200$), the high frequency components of $\mathbf{Z}^*$ are usually noises (e.g. $Z^*(k)$ with $k > 100$ are noises). Using this prior, the power of the test can somewhat be gained by only applying the adaptive Neyman test to the first T^* components of $\mathbf{Z}^*$ (e.g. $T^* = 100$ and looking up Table 1 with $n = 100$ instead of 200). If there is no such a prior, then one needs to use the whole vector $\mathbf{Z}^*$, i.e. $T^* = T$.

When the population standardized difference function exhibits discontinuities and sharp aberrations, wavelet transforms can be a good family of orthogonal transforms for signal compression. Once the wavelet transform is conducted on the vector $\mathbf{Z}$, we can apply the thresholding test outlined in Section 2.

3.2 Impact of variance substitution on the null distributions

The aim of this section is to show that the impact of variance substitution on null distribution is negligible when variances are estimated with good precision. For simplicity of technical arguments, we assume further that $\sigma_1^2(t) = \sigma_1^2$ and $\sigma_2^2(t) = \sigma_2^2$ for $t = 1, \cdots, T$. As outlined in the appendix, even for this simple case, the technical arguments are not trivial.

Under the homoscedastic error model, natural estimators for σ_1^2 and σ_2^2 are the pooled estimators:

$$\hat{\sigma}_1^2 = \{T(n_1 - 1)\}^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n_1} (X_j(t) - \bar{X}(t))^2,$$

and

$$\hat{\sigma}_2^2 = \{T(n_2 - 1)\}^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n_2} (Y_j(t) - \bar{Y}(t))^2.$$

Then, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are consistent estimators of σ_1^2 and σ_2^2 (as $T \to \infty$). Denote the standardized difference by

$$Z(t) = \{n_1^{-1}\hat{\sigma}_1^2 + n_2^{-1}\hat{\sigma}_2^2\}^{-1/2}\{\bar{X}(t) - \bar{Y}(t)\}. \tag{3.7}$$

Let Γ be an orthonormal transform and $\mathbf{Z}^* = \Gamma\mathbf{Z}$. This can be in one case the discrete Fourier transform and in another case a discrete wavelet transform. Let $c_T \leq T$ be a sequence of constants, tending to infinity. Define

$$T_{AN}^* = \max_{1 \leq m \leq c_T} \left\{ (\sqrt{2m})^{-1} \sum_{k=1}^{m} (Z^*(k)^2 - 1) \right\}. \tag{3.8}$$

This adaptive Neyman test uses the information contained in the first $T^* = c_T$ coordinates. Following the discussions at the end of Section 3.1, this is hardly a restriction for practical purposes if c_T is large enough. Now, normalize T_{AN}^* as in (2.4) except replacing n by c_T. We have the following results:

Theorem 3.1 *The asymptotic distribution of T_{AN} under H_0 is given by*

$$P(T_{AN} < x) \to \exp\{-\exp(-x)\} \quad as \ T \to \infty,$$

provided that $c_T = O(T/\log^a T)$, for any $a > 0$.

The proofs of this and next result are given in the appendix. For the thresholding estimator (2.7) with respect to the transformed data $\mathbf{Z}^*$, we have the following result.

11

Theorem 3.2 *Under H_0, the asymptotic distribution of $\hat{T}_H$ is given by*

$$\hat{T}_H \xrightarrow{\mathcal{L}} N(0,1), \quad as \quad T \to \infty,$$

provided that the thresholding parameter $\delta_T = \sqrt{2\log(Ta_T)}$ with $a_T = c\log^{-d} T$ and $d > 0.5$.

3.3 Stationary Errors

When the error processes $\varepsilon(t)$ and $\varepsilon'(t)$ are stationary, the adaptive Neyman test is still applicable. One needs to preprocess the data via the Fourier transform of each curve. The Fourier transform serves two purposes here: Transform stationary errors into approximately independent Gaussian errors (see e.g. Chapter 10 of Brockwell and Davis 1991) and compress useful signals into low frequencies. Let $\{X_j^*(k)\}_{k=1}^T$ and $\{Y_j^*(k)\}_{k=1}^T$ be the sets of the transformed data. More precisely, $X_j^*(1)$ is the real part of the Fourier transform of the vector $\{X_j(t)\}_{t=1}^T$ at the zeroth Fourier frequency, $X_j^*(2)$ is the real part at the first Fourier frequency, $X_j^*(3)$ is the imaginary part at the first Fourier frequency, and so on. An advantage of considering separately real and imaginary parts is that we use fully the information contained in both amplitudes and phases. This marks a difference between our approach and the technique outlined in Section 4.5 of Shumway (1988). As shown in (3.15) in the next subsection, $\{X_j^*(k)\}$ and $\{Y_j^*(k)\}$ satisfy approximately the models given in (3.1) and (3.2). One can now apply the recipe given in Section 3.1 to $\{X_j^*(k)\}$ and $\{Y_j^*(k)\}$ to complete the analysis. The only difference is that one does not need to apply the Fourier transform again to the standardized vector **Z**. Details of the procedure can be found in the next subsection, where we will demonstrate the impact of stationary errors and of the variance substitution on the null distribution is negligible.

For repeated measurements with stationary stochastic errors, Shumway (1988) outlined an approach to test differences among multiple sets of measurements. Namely, for each given frequency, a two-sample t-test statistic or more generally an ANOVA statistic is used to detect if the mean curves have the same power spectrum at that given frequency. This results in a test statistic at each given frequency. The innovation of our procedure is that it combines properly the separate test statistics at different frequencies and uses information from both amplitudes and phases to yield a powerful overall test.

3.4 Null distribution under stationarity errors

We now show that the impact of stationarity and of variance estimates on the null distribution is asymptotically negligible. The conclusion is also supported by our simulation study presented in Example 2 of next subsection. To ease technical arguments, we assume that the stochastic errors in (3.1) and (3.2) are stationary linear Gaussian process:

$$\begin{aligned} \varepsilon_j(t) &= \textstyle\sum_{l=-\infty}^{\infty} a_l Z_j(t-l), &\text{with } \{Z_j(t)\} \text{ i.i.d. } N(0,1) \\ \varepsilon_j'(t) &= \textstyle\sum_{l=-\infty}^{\infty} b_l Z_j'(t-l), &\text{with } \{Z_j'(t)\} \text{ i.i.d. } N(0,1). \end{aligned} \tag{3.9}$$

Assume further

$$\sum_{l=-\infty}^{\infty} (|a_l| + |b_l|)|l|^{1/2} < \infty. \tag{3.10}$$

Let the discrete Fourier transform of a vector $\{X_t\}_{t=1}^{T}$ be

$$J_X(\lambda) = T^{-1/2} \sum_{t=1}^{T} X_t \exp\left(-i\lambda(t-1)\right).$$

This transform will be evaluated at the Fourier frequencies $\omega_j = 2\pi j/T$ for $j = 0, 1, \cdots, [T/2]$. Applying the Fourier transform to each of the sample curves in (3.1), we obtain

$$J_{X_j}(\omega_k) = J_{f_1}(\omega_k) + J_{\varepsilon_j}(\omega_k). \tag{3.11}$$

By Theorem 10.3.1 of Brockwell and Davis (1987),

$$J_{\varepsilon_j}(\omega_k) = A(\exp(-i\omega_k))J_{Z_j}(\omega_k) + O_p\{T^{-1/2}\}, \quad \text{with } A(\lambda) = \sum_{j=-\infty}^{\infty} a_j \lambda^j. \tag{3.12}$$

The uniformity of the O_p-terms in k will be considered in the proof of Theorem 3.3. The sequence $\{J_{Z_j}(\omega_k)\}_{k=0}^{[T/2]}$ is a Gaussian white noise. Therefore, the transformed data become nearly independent and approximately Gaussian. Analogously, by applying the Fourier transform to the curves in (3.2), we have

$$J_{Y_j}(\omega_k) = J_{f_2}(\omega_k) + J_{\varepsilon_j'}(\omega_k) \tag{3.13}$$

with

$$J_{\varepsilon_j'}(\omega_k) = B(\exp(-i\omega_k))J_{Z_j'}(\omega_k) + O_p\{T^{-1/2}\}, \quad \text{with } B(\lambda) = \sum_{j=-\infty}^{\infty} b_j \lambda^j. \tag{3.14}$$

Let $\{X_j^*(k)\}_{k=1}^{T}$ and $\{Y_j^*(k)\}_{k=1}^{T}$ be respectively the real and imaginary parts of the discrete Fourier transforms $\{J_{X_j}(\omega_k)\}$ and $\{J_{Y_j}(\omega_k)\}$, arranged as in Section 3.3. Then, from $(3.11)-(3.14)$,

we have

$$
\begin{aligned}
X_j^*(k) &= f_1^*(k) + \varepsilon_j^*(k) + O_p\{T^{-1/2}\}, \\
Y_j^*(k) &= f_2^*(k) + \varepsilon_j^{**}(k) + O_p\{T^{-1/2}\},
\end{aligned} \tag{3.15}
$$

where $f_1^*(k)$ and $f_2^*(k)$ contains respectively the real and imaginary parts of the Fourier transforms $J_{f_1}(\omega_k)$ and $J_{f_2}(\omega_k)$, and $\varepsilon_j^*(k) \sim N(0, \sigma_1^2(k))$ and $\varepsilon_j^{**}(k) \sim N(0, \sigma_2^2(k))$ are independent for all k. Here the variance functions $\sigma_1^2(k)$ and $\sigma_2^2(k)$ can easily be derived respectively from $A(\exp(-i\omega_k))$ and $B(\exp(-i\omega_k))$.

Consider the difference of the sample means in the frequency domain. Denote by

$$
D^*(k) = n_1^{-1} \sum_{j=1}^{n_1} X_j^*(k) - n_2^{-1} \sum_{j=1}^{n_2} Y_j^*(k), \quad \text{for } k = 1, \cdots, T, \tag{3.16}
$$

and let the standardized difference be

$$
Z(k) = D^*(k) / \{\hat\sigma_1^2(k)/n_1 + \hat\sigma_2^2(k)/n_2\}^{1/2}, \tag{3.17}
$$

where $\hat\sigma_1^2(k)$ and $\hat\sigma_2^2(k)$ are estimators of $\sigma_1^2(k)$ and $\sigma_2^2(k)$, respectively. A natural choice would be the sample variance of $\{X_j^*(k)\}_{j=1}^{n_1}$ and $\{Y_j^*(k)\}_{j=1}^{n_2}$. For theoretical considerations, we wish to determine how good the estimators of $\sigma_1^2(k)$ and $\sigma_2^2(k)$ should be in order not to affect asymptotic analyses. For this purpose, we define the sum of the maximum relative errors by

$$
e_T = \max_{1 \le k \le c_T} |\hat\sigma_1^2(k)/\sigma_1^2(k) - 1| + \max_{1 \le k \le c_T} |\hat\sigma_2^2(k)/\sigma_2^2(k) - 1|, \tag{3.18}
$$

for some constant c_T tending to infinity with $c_T \le T$.

Let the adaptive Neyman test be

$$
T_{AN}^* = \max_{1 \le m \le c_T} (2m)^{-1/2} \sum_{k=1}^{m} \{|Z(k)|^2 - 1\}. \tag{3.19}
$$

Define the standardized adaptive Neyman test as

$$
T_{AN} = \sqrt{2 \log \log c_T}\, T_{AN}^* - \{2 \log \log c_T + 0.5 \log \log \log c_T - 0.5 \log(4\pi)\}. \tag{3.20}
$$

Compare with (2.4).

In many applications, T is usually much larger than n_1 and n_2. For example, in the business advertisement data, T is in an order of 200, while n_1 and n_2 are in an order of 20. Thus, our asymptotic analysis is based on the assumption that T tend to ∞ and n_1 and n_2 tends to ∞ (at a possibly slower rate than T) as $T \to \infty$. The following theorem describes the asymptotic result.

14

Theorem 3.3 *Assume that $c_T = O(T^{a_0})$ for some $a_0 < 1$, and $e_T = O_P(c_T^{-1/2}/\log T)$. If (3.9) and (3.10) hold, then the asymptotic distribution of T_{AN} under H_0 is given by*

$$P(T_{AN} < x) \to \exp\{-\exp(-x)\} \quad as\ T \to \infty,$$

provided that n_1 and n_2 are of the same order.

The implications of Theorem 3.3 are that the impact of stationary errors on the null distribution of the adaptive Neyman test is asymptotically negligible and that the impact of variance estimate on the null distribution is also negligible.

To construct variance estimators having the prescribed accuracy, we can first compute the sample variance in the frequency domain as in Section 3.3. Then, conduct a kernel smoothing on the sample variances to improve the rates of convergence. The resulting variance estimators can achieve the required accuracy when n_1 and n_2 are moderately large comparing with T. Since the arguments are very technically involved, we omit the details here. See Section 4.8 of Lin (1997) for details.

For the wavelet thresholding test, it is shown in Section 4.7 of Lin (1997) that the impact of stationarity assumption on the thresholding test is also negligible. We do not reproduce the proof here since it is lengthy and complicated.

3.5 Examples

The sets of curves discussed in this section are relatively smooth. Thus, the adaptive Neyman test is employed throughout this section. Further, the Fourier transform are always employed to preprocess the data, since the stochastic errors here are expected to be correlated.

Example 1 (TV commercial). The data collected are about a business commercial on pizza (courtesy of Professors D. Hudge and N.M. Didow of the Kenan Flag Business School, University of North Carolina at Chapel Hill). The aim of this study is to examine whether there is any time effect on the commercial. For example, one wishes to know if the pizza commercial should be better aired near the lunch/dinner hours. The advertisement was played at studios at 6 different time slots with different evaluators. As they are watching, the assessors turn the knob of a device to assign scores at a rate of one per second. The scale of the score ranges from 1 to 100 and the device was originally set at 50.
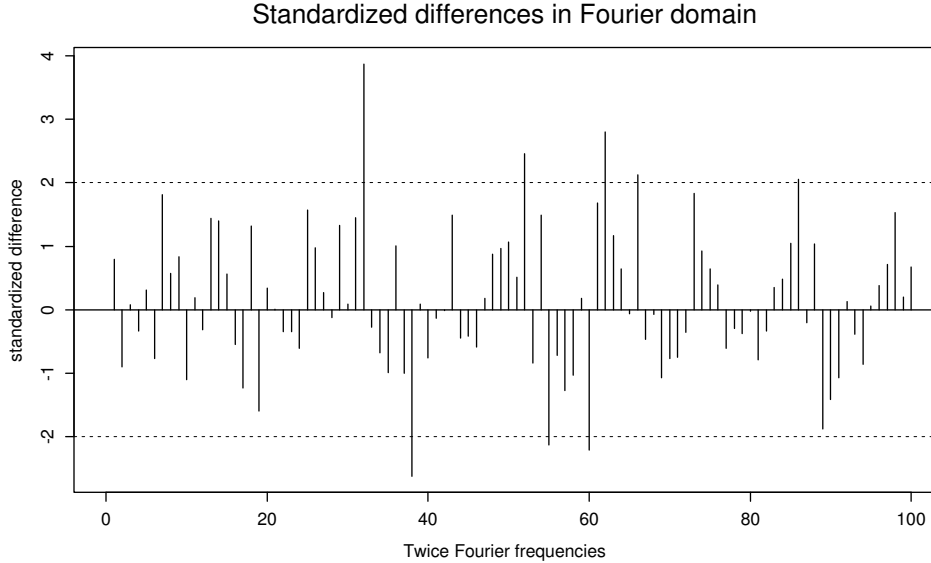
Figure 2: The standardized difference between slot 2 and 5 at each given frequency.

We now apply our new method to test whether there is any significant time effect between slot 2 and 5 for the pizza commercial. Since the neighboring scores are expected to be corrected, we apply the recipes outlined in Section 3.3 (Admittedly, the model here is only an idealization of the data). Figure 1(d) depicts the averages and SDs of the Fourier transformed data. The transformed coefficient at the zeroth Fourier frequency is labeled as 0 at the x-axis of Figure 1 (d). The real and imaginary part of the transformed data at the first Fourier frequency are labeled respectively as 1 and 2 on the x-axis, and so on. We took the first 100 coefficients for analysis. Namely, we regard the data beyond the 50th Fourier frequency as noise. See also remarks at the end of Section 3.1. The standardized difference $Z(k)$ in the Fourier domain, computed based on Figure 1(d), is presented in Figure 2. The adaptive Neyman test is employed, resulting in the observed test statistic $T_{AN} = 2.70$ with $\hat{m} = 66$, and a P-value $\approx 10.10\%$ (From Table 1, it is about 10%). This indicates in turn that there is some evidence for the time effect on the commercial, but the evidence is somewhat weak given the available data. We will analyze this problem further in Example 4.

Example 2 (Simulation). This example is similar to example 1, except data are now simulated. Let $g_1(t)$ and $g_2(t)$ be respectively the mean curve of the first (time slot 2) and second group (time slot 5) given in Figure 1(c). We then generate a random sample of size $n_1 = 21$ from model (3.1) and of size $n_2 = 24$ from model (3.2) with

$$f_1(t) = g_1(t) \quad \text{and} \quad f_2(t) = g_1(t) + \theta(g_2(t) - g_1(t)) \quad (0 \le \theta \le 1.2).$$
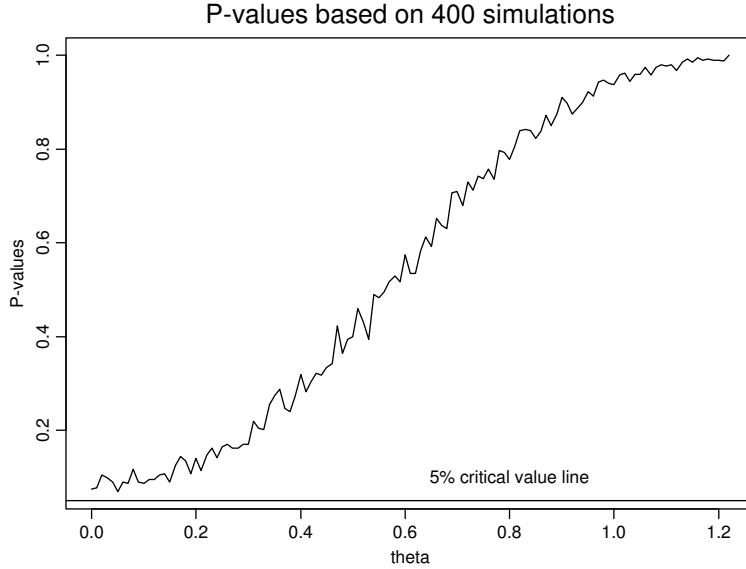
16

Figure 3: Plot of power function for the simulated models in Example 2.

The stochastic error is generated from an AR(1) process:

$$\varepsilon(t) = 0.7\varepsilon(t-1) + \eta(t),$$

where $\{\eta(t)\}$ are i.i.d. Gaussian white noise with variance chosen so that variance of ε is approximately the same as the sample variance in Example 1. More precisely, for the first group the standard deviation of $\eta(t)$ was taken to be 7.26 and for the second group, the standard deviation of $\eta(t)$ was chosen to be 11.65. Figure 3 depicts the power function at the significant level 5% based on 400 simulations. As expected, when $\theta = 0$ the power function is approximately the same as the significant level 5%. This supports our theoretical result that impact of stationarity on the adaptive Neyman test is nearly negligible. When $\theta = 1$, the power is about 94%, which demonstrates the excellent discriminability power of the adaptive Neyman test even when the two mean curves are as close as those in Figure 1(c). This gives us a rough idea about the chance of making Type II error if the data curves were generated from our models.

Example 3 (Cornea Topography). We briefly describe measurements obtained by a keratoscopic device called "Keratron". See Cohen *et al* (1994) and Tripolli *et al* (1995) for more details. The measurements used here are the heights of a cornea measured from the tangent plane of the keratoscope's optical axis intersected with the cornea surface. These height measurements reflect the cornea profile and are measured at each of intersections of 26 rings with 256 radial directions
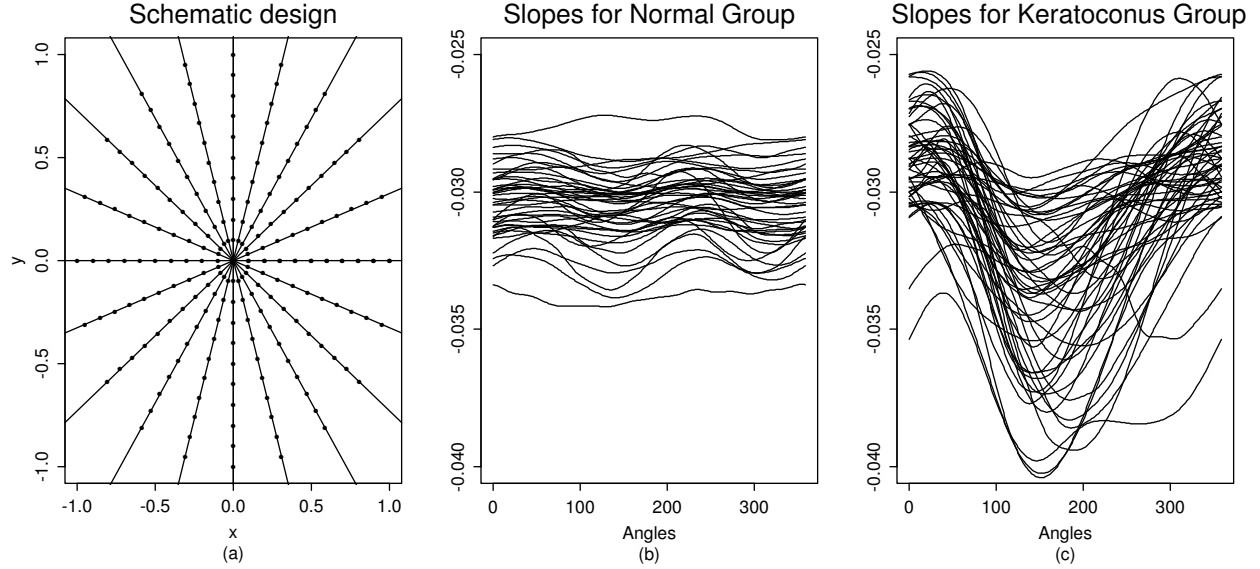
17

Figure 4: (a) Schematic design of height measurements of a cornea surface. (b) and (c): Plot of slope coefficients along 256 radial directions for the normal and keratoconus group. The presented curves are the smoothed version of the original data via a truncated Fourier method — the Fourier coefficients at frequencies higher than $20\pi/256$ are set to zero.

on a keratograph. A schematic representation of the locations where the height measurements are obtained is presented in Figure 4(a). For this kind of images as data, the first step is to extract salient features. For each radial direction, a Legendre polynomial of order 7 was fitted to the 26 data points along this radial direction and 8 least-squares coefficients were obtained. This process is repeated for all 256 radial directions, resulting in a 256×8 matrix. For each Legendre order (i.e. each column), a Fourier transform is obtained and the Fourier coefficients of the first 10 frequencies are kept, resulting in a $(2 * 10 - 1) \times 8$ matrix, called the feature matrix of a cornea topograph. Let L_i denote the ith order Legendre polynomial. Note that the coefficients of L_1 and L_2 in the least squares fit reflect the slope and curvature of a cornea surface along a radial direction. For illustration purpose, we consider only the slope here. Of interest is to test whether there is any statistical difference between two given clinical groups. In the study presented here, 42 subjects of the normal cornea group and 52 of the keratoconus group were available for our analysis. The data were kindly provided by Nancy K. Tripoli (M.A.) and Kenneth L. Cohen (M.D.) of Department of Ophthalmology, University of North Carolina at Chapel Hill and analyzed with computing assistance of Mr. Jin-Ting Zhang (M.S.) and advice of Ms. Nancy K. Tripoli. Figures 4(b) and (c) show the smoothed version (via the Fourier inversion of the first 19 Fourier coefficients in the feature

18

matrices) of the slope coefficients along the 256 radial directions. The adaptive Neyman test in Section 3.2 is applied to this set of data. The result is highly statistically significant, corresponding to a P-value .06%. This in turn suggests that there is a statistical difference of cornea slopes between the normal and keratoconus group. The same statistical analysis was applied to compare the difference between the right and left eyes of normal corneas. As anticipated, there is no statistical difference between the left and right eyes. The corresponding P-value is about 14.41%.

4   High-dimensional Analysis of Variance

In the business commercial data, evaluations were conducted at six different time slots. The collected data are of form: $\{X_{ij}(t)\}$. Here, i denotes the index of groups $(1 \leq i \leq I)$, j represents the membership of each individual $(1 \leq j \leq n_i)$ and t is the index of the grid point that a value of a curve is observed $(1 \leq t \leq T)$. Of interest is to detect if there is any significant time effect. This is clearly a generalization of the two-sample problem discussed in Section 3, and requires us to generalize the adaptive Neyman test to testing multiple groups of curves.

We will assume throughout this paper that for a given j and t, $X_{ij}(t)$ are independently and identically distributed. Namely, observations within each group at a given time are a random sample from an unknown population.

4.1   Connections with high-dimensional ANOVA

Let $\{X_{ij}^*(k)\}$ be the discrete Fourier transform of the vector $\{X_{ij}(t)\}$ for given i and j. As mentioned before, this transform serves two purposes: compress signals into low frequencies and transform stationary errors into nearly independent Gaussian errors. For simplicity of discussions, we first assume the following model:

$$X_{ij}^*(k) = f_i^*(k) + \varepsilon_{ij}^*(k), \tag{4.1}$$

where $\varepsilon_{ij}^*(k)$ are all independent and $\varepsilon_{ij}^*(k) \sim N(0, \sigma_i^2(k))$. Compare with (3.15). Of interest is to examine whether there is any intergroup difference:

$$H_0 : f_i^*(k) = f^*(k), \text{ for } i = 1, \cdots, I \text{ and } k = 1, \cdots, T, \tag{4.2}$$

19

for some unspecified function $f^*(\cdot)$. Let $\bar{X}_i^*(k)$ be the average curve of the ith group in the frequency domain, namely

$$\bar{X}_i^*(k) = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}^*(k), \qquad (4.3)$$

and $\hat{\sigma}_i^*(k)$ be the SD curve

$$\hat{\sigma}_i^{*2}(k) = (n_i - 1)^{-1} \sum_{j=1}^{n_i} \{X_{ij}^*(k) - \bar{X}_i^*(k)\}^2. \qquad (4.4)$$

Then,

$$\bar{X}_i^*(k) \sim N\{f_i^*(k), \sigma_i^{*2}(k)/n_i\}.$$

Note that $\bar{X}_i^*(k)$ and $\hat{\sigma}_i^{*2}(k)$ are sufficient statistics under the ideal model (4.1). Our curve testing problem becomes a high-dimensional (because T is large) ANOVA problem: testing (4.2) based on the observations $\bar{X}_i^*(k) \sim N\{f_i^*(k), \sigma_i^{*2}(k)/n_i\}$ and $\hat{\sigma}_i^{*2}(k)$. Similar to discussions in Section 2, a direct use of the ANOVA technique in such a high dimensional problem will accumulate large stochastic errors and results in low discriminability power. Next subsection offers a method to overcome this difficulty. The resulting test statistic is a generalization of the Adaptive Neyman test.

4.2 Adaptive ANOVA

For simplicity of notation, we restate the high-dimensional ANOVA as follows: Let $X_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$ be independent random variables. Of interest is to test

$$H_0 : \mu_{ij} = \mu_j \text{ for } i = 1, \cdots, I, j = 1, \cdots, n, \qquad (4.5)$$

where n is large. Note that when $\{\sigma_{ij}\}$ are unknown, the problem is a generalization of the famous Behrens-Fisher problem and the maximum likelihood ratio test can not be explicitly determined.

Throughout this section, we assume that $\{\sigma_{ij}^2\}$ are known and $\{\mu_j\}$ in (4.5) are unknown. Because of high dimensionality, testing all dimensions accumulates large stochastic noise and hence results in low discriminability power. Thus, we need to select a few useful cells to test. Suppose that we have vague prior knowledge that useful information is concentrated on the first m cells. We then consider the subproblem

$$H_0 : \mu_{ij} = \mu_j \text{ for } j = 1, \cdots, m. \qquad (4.6)$$

20

The maximum likelihood ratio tatistic for problem (4.6) is

$$X^2 = \sum_{j=1}^{m} \sum_{i=1}^{I} \sigma_{ij}^{-2}(X_{ij} - \bar{X}_{\cdot j})^2, \quad \text{with} \quad \bar{X}_{\cdot j} = \sum_{i=1}^{I} \sigma_{ij}^{-2} X_{ij} / \sum_{i=1}^{I} \sigma_{ij}^{-2}.$$

It can easily be shown that

$$X^2 \sim \chi^2_{(I-1)m}(\delta_m^2),$$

where $\delta_m^2 = \sum_{j=1}^{m} \sum_{i=1}^{I} \sigma_{ij}^{-2}(\mu_{ij} - \bar{\mu}_{\cdot j})^2$ with $\bar{\mu}_{\cdot j} = \sum_{i=1}^{I} \sigma_{ij}^{-2} \mu_{ij} / \sum_{i=1}^{I} \sigma_{ij}^{-2}$. Thus, a level α test is to reject H_0 when

$$
\begin{aligned}
F_m &= \frac{1}{\sqrt{2(I-1)m}} \left\{ \sum_{j=1}^{m} \sum_{i=1}^{I} \sigma_{ij}^{-2}(X_{ij} - \bar{X}_{\cdot j})^2 - (I-1)m \right\} \\
&\geq \frac{1}{\sqrt{2(I-1)m}} \{ \chi^2_{(I-1)m}(1-\alpha) - (I-1)m \}.
\end{aligned}
\tag{4.7}
$$

Note that when the degree of freedom $(I - 1)m$ is large, the test statistic F_m is approximately normally distributed with mean $\delta_m^{*2} = \delta_m^2 / \sqrt{2(I-1)m}$ and variance 1. Thus, the power is approximately an increasing function of δ_m^{*2}.

In practice, the parameter m needs to be determined. Since the power depends monotonously on δ_m^{*2}, following the idea of Fan (1996), we choose m_0 that maximizes δ_m^{*2}. Namely,

$$m_0 = \mathrm{argmax}_{1 \leq m \leq n} \delta_m^2 / \sqrt{2(I-1)m}. \tag{4.8}$$

In practice, such an ideal m_0 is not available to us. Since F_m is an unbiased estimate of δ_m^{*2}, a natural candidate is

$$\hat{m} = \mathrm{argmax}_{1 \leq m \leq n} F_m,$$

leading to the adaptive testing statistic

$$F_{\hat{m}} = \max_{1 \leq m \leq n} \frac{1}{\sqrt{2(I-1)m}} \left\{ \sum_{j=1}^{m} \sum_{i=1}^{I} \sigma_{ij}^{-2}(X_{ij} - \bar{X}_{\cdot j})^2 - (I-1)m \right\}. \tag{4.9}$$

For convenience, this ANOVA type of test statistic is called high-dimensional ANOVA (HANOVA). Specifically, when $I = 2$, the test statistic reduces to

$$F_{\hat{m}} = \max_{1 \leq m \leq n} \frac{1}{\sqrt{2m}} \left\{ \sum_{j=1}^{m} \frac{(X_{1j} - X_{2j})^2}{\sigma_{1j}^2 + \sigma_{2j}^2} - m \right\},$$

which is the adaptive-Neyman test for comparing two sets of curves. See (2.3) and (3.6).

4.3 Null distribution and power

We now derive the distribution of the HANOVA test statistic $F_{\hat{m}}$ under the null hypothesis. As in (2.4), we consider the following normalization:

$$F_{I,n} = \sqrt{2\log\log n}\, F_{\hat{m}} - \{2\log\log n + 0.5\log\log\log n - 0.5\log(4\pi)\}. \qquad (4.10)$$

Recall that the distribution J_n is defined as that of T_{AN} in (2.4) and is tabulated in Table 1.

Theorem 4.1 *Under the null hypothesis (4.5), when n is large, the statistic $F_{I,n}$ is distributed approximately as J_n and*

$$\lim_{n\to\infty} P(F_{I,n} < x) \to \exp(-\exp(-x)). \qquad (4.11)$$

In the proof of Theorem 4.1, we will show that the statistic $F_{I,n}$ has exactly the same form as the definition of J_n (see (A.14)). The only difference is that in the definition of J_n the random variables $V'_j = 2^{-1/2}(Z^2_{2j} - 1)$ in (2.3) instead of $V_j = \{2(I-1)\}^{-1/2}\sum_{i=2}^{I}(Z^2_{ij} - 1)$ are used, where $\{Z_{ij}\}$ are standard Gaussian random variables. Note that both V'_j and V_j have mean zero and variance one, and when the number of groups is not too large, the shapes of the densities of V'_j and V_j are similar. Thus, it is expected that J_n is a closer approximation to the distribution of $F_{I,n}$ than the limiting distribution given at the right hand side of (4.11).

We now consider the asymptotic power of the HANOVA test. Recall that the ideal m_0 is given by (4.8). By (4.7), the power of this ideal test is given by

$$P_\mu\{F_{m_0} - \delta^{*2}_{m_0} \geq C_{m_0}(1-\alpha) - \delta^{*2}_{m_0}\},$$

where

$$C_{m_0}(1-\alpha) = \frac{1}{\sqrt{2(I-1)m_0}}\{\chi^2_{(I-1)m_0}(1-\alpha) - (I-1)m_0\}.$$

Note that because of the asymptotic normality of $F_m - \delta^{*2}_m$ as $m \to \infty$, the sequence of random variables $\{F_m - \delta^{*2}_m\}$ is stochastically bounded (tight) and hence $\{C_m(1-\alpha)\}$ is also a bounded sequence. The following theorem shows that having to estimate m the HANOVA test pays at most a price of $\log\log n$ comparing with the ideal test.

Theorem 4.2 *The power of HANOVA is at least*

$$P_\mu[F_{m_0} - \delta^{*2}_{m_0} \geq \sqrt{2\log\log n}\{1 + o(1)\} - \delta^{*2}_{m_0}].$$

In particular, when $\max_{1\leq m\leq n} \delta^2_m/\sqrt{2(I-1)m} \to \infty$, the HANOVA has the asymptotic power one.

5 Hypothesis testing for multiple groups of curves

5.1 Testing differences among multiple groups of curves

Consider the observed curves from I different groups: $\{X_{ij}(t), i = 1, \cdots, I, j = 1, \cdots, n_i, t = 1, \cdots, T\}$. We assume that

$$X_{ij}(t) = f_i(t) + \varepsilon_{ij}(t), \tag{5.1}$$

where $\{\varepsilon_{ij}(t), t = 1, \cdots, T\}$ are stationary time series with mean zero. Of interest is to test

$$H_0 : f_i(t) = f(t), \text{ for } i = 1, \cdots, I \text{ and } t = 1, \cdots, T. \tag{5.2}$$

Let $\{X_{ij}^*(k)\}$ be the discrete Fourier transform of the vector $\{X_{ij}(t)\}$ for given i and j. Then, $\{X_{ij}^*(k)\}$ satisfy approximately the ideal model (4.1) and the problem (5.2) is equivalent to (4.2). Thus, the HANOVA technique in Section 4 can be used.

Suppose that the maximum number of dimensions to be tested is T^*. This can usually be a convenient number tabulated in Table 1 (e.g. $T^* = T/2$ or simply $= T$). As discussed before, this choice does not alter the result very much, as long as T^* is large enough so that high frequency cells are basically noise. Applying the HANOVA and replacing the unknown variance by the sample variance, we obtain the test statistic

$$F_{\hat{m}} = \max_{1 \le m \le T^*} \frac{1}{\sqrt{2(I-1)m}} \left[\sum_{k=1}^{m} \sum_{i=1}^{I} n_i \hat{\sigma}_i^*(k)^{-2} \{\bar{X}_i^*(k) - \bar{X}^*(k)\}^2 - (I-1)m \right], \tag{5.3}$$

where $\bar{X}_i^*(k)$ and $\hat{\sigma}_i^{*2}(k)$ were defined respectively in (4.3) and (4.4), and

$$\bar{X}^*(k) = \sum_{i=1}^{I} n_i \hat{\sigma}_i^*(k)^{-2} \bar{X}_i^*(k) / \sum_{i=1}^{I} n_i \hat{\sigma}_i^*(k)^{-2}.$$

Finally, one can normalize the test statistic as in (4.10), leading to the HANOVA test statistic:

$$T_{\text{HANOVA}} = \sqrt{2 \log \log T^*} F_{\hat{m}} - \{2 \log \log T^* + 0.5 \log \log \log T^* - 0.5 \log(4\pi)\}. \tag{5.4}$$

One can use Table 1 with $n = T^*$ to find P-values.

5.2 An application

In this section, we use the business advertisement data in Example 1 to illustrate our HANOVA technique.
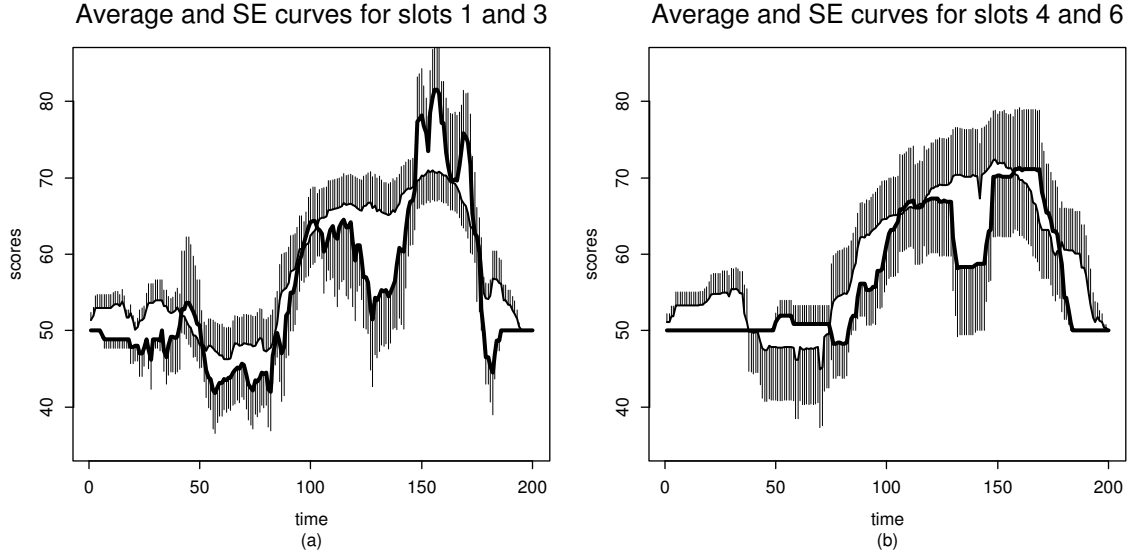
Figure 5: The average and SE curves of pizza commercial at different time slots. (a) Thick curve for time slot 1 and thin curve for time slot 3; (b) thick curve for time slot 4 and thin curve for time slot 6. The bars on the top or bottom indicate the estimated standard errors of the average score at that time point.

Example 4. We now use the data collected in all six time slots to test whether there is any time effect. The sample sizes for six different time slots are respectively 6, 21, 15, 7, 24, and 10. Figure 5 depicts the average and standard error curves for time slots 1, 3, 4, and 6 (see Figure 1(c) for slots 2 and 5). The HANOVA statistic is computed to be 8.12 with $\hat{m} = 80$ for $T^* = 100$. This in turns gives a P-value .355% (Table 1 gives a P-value between .25% − .5%)

To better understand the time effect, we apply the adaptive Neyman test as in Example 1 to obtain pairwise comparisons between any two given time slots. The results are summarized in Table 2. This example illustrates that even if the P-values are not very significant for pairwise comparisons, when the information is pulled, it can yield a highly statistical significance result.

Table 2: P-values (in percents) for pairwise comparisons
of time effect of a pizza commercial

time slots	time slots	1	2	3	4	5	6
	sample size	6	21	15	7	24	10
1			3.15	35.6	34.1	6.10	4.60
2				22.2	64.1	11.1	0.24
3					21.6	66.3	29.8
4						69.8	4.16
5							6.10

24

5.3 Extensions

The ideas outlined in Sections 4 and 5 can easily be generalized to more general ANOVA settings. We consider a three-factor ANOVA problem as an illustration.

Suppose after preprocessing data as in Section 4.1 which includes using Fourier transform to compress information into low frequencies, we have the cell averages

$$\{\bar{X}_{ij}^*(k), i = 1, \cdots, I; j = 1, \cdots, J \text{ and } k = 1, \cdots, T\}.$$

Assume that $\bar{X}_{ij}^*(k) \sim N\{f_{ij}^*(k), \sigma_{ij}^2(k)/n_{ij}\}$ with known $\sigma_{ij}^2(k)$, where n_{ij} is the sample size at the (i, j)-cell. In practice, the variances $\{\sigma_{ij}^2(k)\}$ can be estimated from the sample. Suppose that we are interested in testing the hypothesis that

$$H_0 : f_{ij}^*(k) = f_j^*(k), \text{ for } i = 1, \cdots, I, j = 1, \cdots, J, k = 1, \cdots, T. \tag{5.5}$$

Following the derivations in Section 4, we reject H_0 when

$$F_{\hat{m}} = \max_{1 \le m \le T^*} \frac{1}{\sqrt{2(I-1)Jm}} \left[\sum_{k=1}^{m} \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} \sigma_{ij}(k)^{-2} \{\bar{X}_{ij}^*(k) - \bar{X}_{\cdot j}^*(k)\}^2 - (I-1)Jm \right], \tag{5.6}$$

where T^* is the highest dimension (frequency) to be tested and

$$\bar{X}_{\cdot j}^*(k) = \sum_{i=1}^{I} \frac{n_{ij}}{\sigma_{ij}^2(k)} \bar{X}_{ij}^*(k) \bigg/ \sum_{i=1}^{I} \frac{n_{ij}}{\sigma_{ij}^2(k)}.$$

The normalized statistic is

$$T_{\text{HANOVA}} = \sqrt{2 \log \log T^*} F_{\hat{m}} - \{2 \log \log T^* + 0.5 \log \log \log T^* - 0.5 \log(4\pi)\},$$

which is distributed approximately as J_{T^*} tabulated in Table 1.

6 Concluding remarks

We have proposed the adaptive Neyman test and the wavelet thresholding tests for comparing two sets of curves. When the underlying mean curves are reasonably smooth, one would employ the adaptive Neyman test; otherwise the wavelet thresholding tests. We then extend the idea of the adaptive Neyman test to detect the inter-group difference among multiple sets of curves, resulting in an adaptive ANOVA for high dimensional data, called HANOVA. Extensions of wavelet thresholding

tests to this setup are less straightforward, because the information in each set of curves is not necessarily compressed into the same cells by a wavelet transform and hence a useful thresholding rule is hard to define.

While the adaptive Neyman test and the HANOVA are simple and powerful, their practical use depends on the distribution of J_n. Table 1 is provided for convenience, but the explicit form of the distribution J_n is hard to find. With computers exist virtually everywhere, this issue is no longer so crucial. The distribution of J_n can easily be simulated. For example, based on 100,000 simulations, it took us less than 1 minute cpu time on an Ultra Sparc Station 1 or a Pentium 166 to complete Table 2.

We have C-code available for computing the distribution of J_n and Splus code for computing the adaptive ANOVA test statistic, which includes the adaptive Neyman test as a specific example. We would be happy to provide these codes upon request.

The dimensionality reduction techniques used in this paper are based on orthogonal transforms. Other useful techniques include the functional data analysis methods in Ramsay and Silverman (1997). Studies of these approaches are needed.

7   Appendix: Proofs

Proof of Theorem 3.1. Let $\sigma_n^2 = \sigma_1^2/n_1 + \sigma_2^2/n_2$ and $\mathbf{U} = (U_1, U_2, ..., U_T)^T = \Gamma(\bar{X} - \bar{Y})/\sigma_n$. Then, under H_0, $\mathbf{U} \sim N(0, I_T)$ and

$$\mathbf{Z}^* = (\sigma_n/\hat{\sigma}_n)\mathbf{U}, \quad \text{with} \quad \hat{\sigma}_n^2 = \hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2.$$

By (3.8), we have

$$T_{AN}^* = \max_{1 \le m \le c_T} \sum_{j=1}^m \frac{\sigma_n^2/\hat{\sigma}_n^2 U_j^2 - 1}{\sqrt{2m}} = \max_{1 \le m \le c_T} \sum_{j=1}^m \left\{ \frac{U_j^2 - 1}{\sqrt{2m}} + \frac{(\sigma_n^2/\hat{\sigma}_n^2 - 1) U_j^2}{\sqrt{2m}} \right\}. \qquad (A.1)$$

Note that $\sigma_n^2/\hat{\sigma}_n^2 - 1 = O_p(T^{-1/2})$ and

$$\sum_{j=1}^m \frac{(\sigma_n^2/\hat{\sigma}_n^2 - 1)U_j^2}{\sqrt{2m}} = \left(\frac{\sigma_n^2}{\hat{\sigma}_n^2} - 1 \right) \left(\sum_{j=1}^m \frac{U_j^2 - 1}{\sqrt{2m}} + \sqrt{\frac{m}{2}} \right).$$

By a result of Darling and Erdös (1956), we have

$$\max_{1 \le m \le c_T} \sum_{j=1}^m \frac{U_j^2 - 1}{\sqrt{2m}} = O_p(\sqrt{\log \log c_T}).$$

Therefore, we have

$$\max_{1 \le m \le c_T} \sum_{j=1}^{m} \frac{(\sigma_n^2/\hat{\sigma}_n^2 - 1) \, U_j^2}{\sqrt{2m}} = O_p\{(c_T/T)^{1/2}\}$$

and it follows from (A.1) that

$$T_{AN}^* = \max_{1 \le m \le c_T} \sum_{j=1}^{m} \frac{U_j^2 - 1}{\sqrt{2m}} + O_p\left((c_T/T)^{1/2}\right).$$

Thus, by Slutsky's theorem and the result of Darling and Erdös (1956), the desired conclusion follows.   ■

Proof of Theorem 3.2. We continue to use the notation introduced in the proof of Theorem 3.1. Let $\hat{c}_T = \sigma_n/\hat{\sigma}_n$. Then $\hat{c}_T^2 = 1 + O_p(T^{-1/2})$. Now decompose $\hat{T}_H^*$ as the following:

$$\begin{aligned}
\hat{T}_H^* &= \sum_{j=1}^{T} |\hat{c}_T|^2 |U_j|^2 1(|\hat{c}_T U_j| > \delta) \\
&= (1 + O_p(T^{-1/2})) \left[\sum_{j=1}^{T} |U_j|^2 1(|U_j| > \delta) + \sum_{j=1}^{T} |U_j|^2 \{1(|\hat{c}_T U_j| > \delta) - 1(|U_j| > \delta)\} \right] \\
&= \sum_{j=1}^{T} |U_j|^2 1(|U_j| > \delta) + O_p(T^{-1/2})\alpha_T + \{1 + O_p(T^{-1/2})\}\beta_T, \quad (A.2)
\end{aligned}$$

where $\alpha_T = \sum_{j=1}^{T} |U_j|^2 1(|U_j| > \delta)$ and $\beta_T = \sum_{j=1}^{T} |U_j|^2 (1(|\hat{c}_T U_j| > \delta) - 1(|U_j| > \delta))$. By Theorem 2.3 of Fan (1996), we have

$$\alpha_T = O_p[T E\{|U_j|^2 1(|U_j| > \delta)\}] = O_p\{(\log T)^{1/2+d}\},$$

and we will show that

$$\beta_T = O_p((\log T)^{-k}) \qquad \text{for any } k > 0. \quad (A.3)$$

Therefore, by (A.2), we obtain

$$\hat{T}_H^* = \sum_{j=1}^{T} |U_j|^2 1(|U_j| > \delta) + O_p(T^{-1/2}(\log T)^{1/2+d}) + O_p((\log T)^{-k}).$$

By Slutsky's theorem and Theorem 2.3 of Fan (1996), the conclusion of Theorem 3.2 follows. It remains to establish (A.3).

Note that

$$\begin{aligned}
\beta_T &\le \max_{1 \le j \le T} |U_j|^2 \sum_{j=1}^{T} \left| 1(|\hat{c}_T U_j| > \delta) - 1(|U_j| > \delta) \right| \\
&= O_p(\log T) O_p[T E\{|1(|\hat{c}_T U_j| > \delta) - 1(|U_j| > \delta)|\}]. \quad (A.4)
\end{aligned}$$

27

Let $\beta_{T1} = E\{|1(|\hat{c}_T U_j| > \delta) - 1(|U_j| > \delta)|\}$. Then,

$$\beta_{T1} = P(|\hat{c}_T U_j| > \delta, |U_j| \leq \delta) + P(|\hat{c}_T U_j| \leq \delta, |U_j| > \delta) \equiv \beta_{T2} + \beta_{T3}. \tag{A.5}$$

We first deal with the term β_{T2}. Observe that

$$\beta_{T2} \leq P(\delta \geq |U_j| > \delta/(1+b)) + P(|\hat{c}_T| > 1 + b). \tag{A.6}$$

Taking $b = (\log T)^{-f}$ for some large f, it follows from the direct calculation that

$$
\begin{aligned}
P\{\delta/(1+b) < |U_j| \leq \delta\} &= \frac{2}{\sqrt{2\pi}} \int_{\delta/(1+b)}^{\delta} e^{-x^2/2} dx \\
&\leq \frac{2(1+b)}{\delta\sqrt{2\pi}} \int_{\delta/(1+b)}^{\delta} x e^{-x^2/2} dx \\
&= O(\delta^{-1} e^{-\delta^2/2} b\delta^2) \\
&= O\{\frac{1}{T}(\log T)^{-f+1/2+d}\}.
\end{aligned} \tag{A.7}
$$

By Chebyshev's inequality,

$$
\begin{aligned}
P(|\hat{c}_T| > 1 + b) &= P\Big(\exp\{\sqrt{T}(1 - |\hat{c}_T|^{-2})\} \geq \exp[\sqrt{T}\{1 - (1+b)^{-2}\}]\Big) \\
&\leq \frac{E\exp\{\sqrt{T}(1 - |\hat{c}_T|^{-2})\}}{\exp\{\sqrt{T}(1 - (1+b)^{-2})\}}.
\end{aligned}
$$

Note that

$$|\hat{c}_T|^{-2} = \frac{\sigma_1^2/n_1}{\sigma_n^2} \frac{\hat{\sigma}_1^2}{\sigma_1^2} + \frac{\sigma_2^2/n_2}{\sigma_n^2} \frac{\hat{\sigma}_2^2}{\sigma_2^2}$$

is a mixture of χ^2 distribution. By using the moment generating function of the χ^2-distribution, it follows from tedious calculations that

$$P(|\hat{c}_n| > 1 + b) = O\{\exp(-2\sqrt{T}b)\}.$$

This and (A.7) show that $\beta_{T2} = O_P(T^{-1}\log^{-k}T)$ for any $k > 0$. Using a similar argument, we can show that $\beta_{T3} = O_P(T^{-1}\log^{-k}T)$ for any $k > 0$. By (A.5), we establish (A.3). The proof is completed. ∎

Proof of Theorem 3.3. Denote the remainder O_p-terms in (3.15) respectively by $R_{Xj}(k)$ and $R_{Yj}(k)$. Then,

$$E R_{Xj}(k) = R_{Yj}(k) = 0.$$

By the assumptions (3.9) and (3.10), it can be shown (see Section 4.3 of Lin 1997) that for any integer m,

$$ER_{Xj}(k)^{2m} = O(T^{-m}), \qquad ER_{Yj}(k)^{2m} = O(T^{-m}). \tag{A.8}$$

Using this, we now demonstrate

$$\max_{1 \le k \le T} n_1^{-1} \sum_{j=1}^{n_1} R_{Xj}(k) = o_P(n_1^{-1/2}T^{-a}), \quad \text{for any } a < 1/2. \tag{A.9}$$

To see this, we note that for any $\delta > 0$,

$$P\{\max_{1 \le k \le T} n_1^{-1} \sum_{j=1}^{n_1} R_{Xj}(k) > \delta n^{-1/2}T^{-a}\}$$

$$\le \sum_{k=1}^{T} P\{n_1^{-1/2} \sum_{j=1}^{n_1} R_{Xj}(k) > \delta T^{-a}\}$$

$$\le \sum_{k=1}^{T} \delta^{-2m} T^{2am} E\{n_1^{-1/2} \sum_{j=1}^{n_1} R_{Xj}(k)\}^{2m}.$$

By (A.8), the last term is bounded by

$$O(TT^{2am}T^{-m}),$$

which tends to zero if m is chosen largely enough. This establishes (A.9).

Let $\sigma_n^2(k) = n_1^{-1}\sigma_1^2(k) + n_2^{-1}\sigma_2^2(k)$. Then, by using (A.9) and a similar expression for $\{R_{Yj}(k)\}$, we have

$$D^*(k) = \varepsilon_k^* + R_k^*,$$

where

$$\varepsilon_k^* = n_1^{-1} \sum_{j=1}^{n_1} \varepsilon_j^*(k) - n_2^{-1} \sum_{j=1}^{n_2} \varepsilon_j^{**}(k) \sim N(0, \sigma_n^2(k)),$$

and

$$\max_{1 \le k \le T} R_k^* = o_P(n_1^{-1/2}T^{-a}), \quad \text{for any } a < 1/2. \tag{A.10}$$

Thus, by (3.15), we have

$$\begin{aligned} Z(k) &= \frac{D^*(k)}{\sigma_n(k)}\{1 + O_p(e_T)\} \\ &= \varepsilon_k + R_k, \quad \varepsilon_k \sim N(0,1), \max_{1 \le k \le T} R_k = o_P(T^{-a}) + O_P(e_T \log^{1/2} T), \end{aligned} \tag{A.11}$$

29

where the factor $\log^{1/2} T$ comes from the extreme value of the T independent Gaussian white noises. Consequently,

$$
\begin{aligned}
T_{AN}^* &= \max_{1 \le m \le c_T} (2m)^{-1/2} \sum_{k=1}^m \{(\varepsilon_k + R_k)^2 - 1\} \\
&= \max_{1 \le m \le c_T} (2m)^{-1/2} \{\sum_{k=1}^m (\varepsilon_k^2 - 1) + 2 \sum_{k=1}^m \varepsilon_k R_k + \sum_{k=1}^m R_k^2\}.
\end{aligned} \tag{A.12}
$$

By (A.11), the last term of (A.12) is bounded by

$$
\max_{1 \le m \le c_T} (2m)^{-1/2} \sum_{k=1}^m R_k^2 = o(c_T^{1/2} T^{-2a}) + O_P(c_T^{1/2} e_T^2 \log T)
$$

The second term of (A.12) is bounded by

$$
\max_{1 \le m \le c_T} m^{-1/2} \sum_{k=1}^m \varepsilon_k R_k \le \{o_P(T^{-a}) + O_P(e_T \log^{1/2} T)\} \max_{1 \le m \le c_T} m^{-1/2} \sum_{k=1}^m |\varepsilon_k|.
$$

Observe that

$$
\begin{aligned}
\max_{1 \le m \le c_T} m^{-1/2} \sum_{k=1}^m |\varepsilon_k| &\le \max_{1 \le m \le c_T} (\sum_{k=1}^m \varepsilon_k^2)^{1/2} \\
&\le [c_T^{1/2} \max_{1 \le m \le c_T} m^{-1/2} \{\sum_{k=1}^m (\varepsilon_k^2 - 1) + m\}]^{1/2} \\
&= [c_T^{1/2} O_p\{(\log \log c_T)^{1/2}\} + c_T]^{1/2} \\
&= O_p(c_T^{1/2}),
\end{aligned}
$$

where we used (2.4) and (2.5) to obtain the $O_p\{(\log \log c_T)^{1/2}\}$ term.

Combining all we have shown so far, we have

$$
\begin{aligned}
T_{AN}^* &= \max_{1 \le m \le c_T} (2m)^{-1/2} \sum_{k=1}^m (\varepsilon_k^2 - 1) + o_P(c_T^{1/2} T^{-a}) + O_P(c_T^{1/2} e_T \log^{1/2} T) \\
&= \max_{1 \le m \le c_T} (2m)^{-1/2} \sum_{k=1}^m (\varepsilon_k^2 - 1) + o_P(T^{-(a-a_0)/2}) + \log^{-1/2} T.
\end{aligned}
$$

By choosing $a > a_0$ and applying the result in (2.4) and (2.5) to the main term, we obtain the desired result. This completes the proof. ∎

Proof of Theorem 4.1. Note that under the null hypothesis, X^2 defined after (4.6) is a central χ^2 distribution with degrees of freedom $(I - 1)m$. Without loss of generality, assume $\mu_{ij} = 0$. Set

$Y_{ij} = X_{ij}/\sigma_{ij} \sim N(0, 1)$ and $a_j^2 = \sum_{i=1}^{I} \sigma_{ij}^{-2}$. Observe that

$$
\begin{aligned}
X^2 &= \sum_{j=1}^{m} \{\sum_{i=1}^{I} \sigma_{ij}^{-2} X_{ij}^2 - a_j^2 \bar{X}_{\cdot j}^2\} \\
&= \sum_{j=1}^{m} \{\sum_{i=1}^{I} Y_{ij}^2 - (\sum_{i=1}^{I} \sigma_{ij}^{-1} Y_{ij}/a_j)^2\} \quad (A.13)
\end{aligned}
$$

Denote by $\mathbf{Y}_j = (Y_{1j}, \cdots, Y_{Ij})^T$. Let Γ_j be an orthogonal matrix whose first row is $(\sigma_{1j}^{-1}, \cdots, \sigma_{Ij}^{-1})/a_j$ and $\mathbf{Z}_j \equiv (Z_{1j}, \cdots, Z_{Ij})^T = \Gamma_j \mathbf{Y}_j$. Then $\{Z_{ij}\}$ are independent and normally distributed with mean 0 and variance 1. By (A.13), we have

$$
X^2 = \sum_{j=1}^{m} \sum_{i=2}^{I} Z_{ij}^2.
$$

This together with (4.9) and (4.10) lead to

$$
\begin{aligned}
F_{\hat{m}} &= \max_{1 \le m \le n} \frac{1}{\sqrt{2(I-1)m}} \left\{ \sum_{j=1}^{m} \sum_{i=2}^{I} Z_{ij}^2 - (I-1)m \right\} \\
&= \max_{1 \le m \le n} m^{-1/2} \sum_{j=1}^{m} V_j, \quad (A.14)
\end{aligned}
$$

where $V_j = \{2(I-1)\}^{-1/2} \sum_{i=2}^{I} (Z_{ij}^2 - 1)$ has mean zero and variance 1. By a result of Darlin and Erdös (1956), we obtain (4.11). Since J_n has also the same asymptotic distribution as $F_{I,n}$, the first conclusion follows. ∎

Proof of Theorem 4.2. Let $c_\alpha = -\log(-\log(1-\alpha))$. Then, the power of the HANOVA test is given by

$$
P_\mu(F_{I,n} > c_\alpha) = P[F_{\hat{m}} \ge \sqrt{2 \log \log n} \{1 + o(1)\}].
$$

Since $F_{\hat{m}} \ge F_{m_0}$, it follows that

$$
P_\mu(F_{I,n} > c_\alpha) \ge P_\mu[F_{m_0} \ge \sqrt{2 \log \log n} \{1 + o(1)\}].
$$

This proves the first conclusion. The second conclusion follows from the stochastic boundedness of the random sequence $\{F_m - \delta_m^{*2}\}$. ∎

REFERENCES

Bickel, P.J. and Ritov, Y. (1992), "Testing for goodness of fit: a new approach," *Nonparametric Statistics and Related Topics* (A.K.Md.E. Saleh, ed.), Elsevier Science Publishers B.V., 51-57.

Besse, P. and Ramsay, J.O. (1986). "Principal components analysis of sampled functions". *Psychometrika*, **51**, 285-311.

Brockwell, P.J. and Davis, R.A. (1991), *Times Series : Theory and Methods*, Second edition, New York: Springer.

Cai, T. (1996). *Non-parametric function estimation via wavelets*, Ph.D. dissertation, Cornell University.

Capra, W.B. and Müller, H.G. (1996). An accelerated time model for response curves. *Manuscript.*

Cohen, K.L., Tripoli, N.K., Holmgren, D.E. and Coggins, J.M. (1994), "Assessment of the power and height of radial aspheres reported by a computer-assisted keratoscope", *American Journal of Ophthalmology*, 119, 723-732.

Darling, D. A. and P. Erdös (1956), "A limit theorem for the maximum of normalized sums of independent random variables," *Duke Mathematical Journal*, 23, 143-155.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994), *Analysis of Longitudinal Data*, Oxford: Clarendon Press.

Donoho, D.L. and Johnstone, I.M. (1994), "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, 81, 425 - 455.

Eubank, R.L. and Hart, J.D. (1992), "Testing goodness-of-fit in regression via order selection criteria," *The Annals of Statistics*, 20, 1412-1425.

Eubank, R.L. and LaRiccia, V.N. (1992), "Asymptotic comparison of Cramér-von Mises and non-parametric function estimation techniques for testing goodness-of-fit," *The Annals of Statistics*, 20, 2071-2086.

Fan, J. (1996), "Test of significance based on wavelet thresholding and Neyman's truncation", *Journal of American Statistical Association*, 91, 674–688.

Faraway, J.J. (1997), "Regression analysis for a functional response", *Technometrics*, **39**, 254-261.

Fisher, R.A. (1929), "Tests of significance in harmonic analysis", *Proceedings of the Royal Society*, A, **125**, 54-59.

Hall, P. and Hart, J.D. (1990), "Bootstrap test for difference between means in nonparametric regression", *Journal of American Statistical Association*, 85, 1039–1049.

Hall, P. and Turlach, B. (1997), "Interpolation methods for nonlinear wavelet regression with irregularly spaced design", *The Annals of Statistics*, 1912-1925.

Hand, D. and Crowder, M. (1996), *Practical Longitudinal Data Analysis*, Chapman and Hall, London.

Hart, J. D. (1996). Some automated methods of smoothing time-dependent data, *Nonparametric Statistics*, 6, 115-142.

Hart, J.D. and Wehrly, T.E. (1986), "Kernel regression estimation using repeated measurements data", *Journal of the American Statistical Association*, 81, 1080-1088.

Hart, J.D. and Wehrly, T.E. (1993), "Consistency of cross-validation when the data are curves", *Stochastic Process and their Applications*, 45, 351-361.

Inglot, T., Kallenberg, W.C.M. and Ledwina, T. (1994), "Power approximations to and power comparison of smooth goodness-of-fit tests", *Scandinavian Journal of Statistics.*, 21, 131-145.

Inglot, T., and Ledwina, T. (1996), "Asymptotic optimality of data-driven Neyman's tests for uniformity", *The Annals of Statistics*, 24, 1982-2019.

Jones, R.H. (1993), "Longitudinal Data With Serial Correlation: A State-space Approach", Chapman and Hall, London.

Kneip, A. and Gasser, T. (1992). "Statistical tools to analyze data representing a sample of curves", *The Annals of Statistics*, 20, 1266-1305.

Ledwina, T. (1994), "Data-driven version of Neyman's smooth test of fit", *Journal of the American Statistics Association*, 89, 1000-1005.

Leurgans, S.E., Moyeed, R.A. and Silverman, B.W. (1993), "Canonical correlation analysis when the data are curves", 55, 725-740.

Lin, S.K. (1997), *Test of Significance When the Data are Curves*, Ph.D. dissertation, University of North Carolina at Chapel Hill.

Pezzulli, S. and Silverman, B. W. (1993), "Some properties of smoothed principal components analysis for the functional data," *Computational Statistics*, 8, 1-16.

Ramsay, J.O. and Dalzzel, C.J. (1991). "Some tools for functional data analysis (with discussions)", *Journal of the Royal Statistics Society B*, 53, 539-572.

Ramsay, J.O. and Silverman, B.W. (1997), "The Analysis of Functional Data", Springer-Verlag, Berlin.

Rice, J. A. and Silverman, B. W. (1991). "Estimating the mean and covariance structure nonparametrically when the data are curves," *Journal of the Royal Statistics Society B* 53, 233-243.

Schmid, C.H. (1996), "An EM algorithm fitting first-order autoregressive models to longitudinal data", *Journal of American Statistical Association*, **91**, 1322-1330.

Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*. Prentice-Hall.

Silverman, B.W. (1995). "Incorporating parametric effects into functional components analysis", *Journal of the Royal Statistical Society B*, 57, 673-689.

Spokoiny, V. (1996), "Adaptive hypothesis testing using wavelets", *The Annals of Statistics*, to appear.

Tripoli, N.K., Cohen, K.L., Holmgren, D.E. and Coggins, J.M. (1995), "Assessment of radial aspheres by the arc-step algorithm as implemented by the keratron keratoscope", *American Journal of Ophthalmology*, 120, 658-664.