TWO-SAMPLE TESTING IN HIGH DIMENSION AND A SMOOTH BLOCK BOOTSTRAP FOR TIME SERIES

A Dissertation

by

KARL BRUCE GREGORY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Raymond Carroll
Co-Chair of Committee,	Soumendra Lahiri
Committee Members,	Mohsen Pourahmadi
	Bruce Lowe
Head of Department,	Simon Sheather

August 2014

Major Subject: Statistics

Copyright 2014 Karl Bruce Gregory

ABSTRACT

This document contains three sections. The first two present new methods for two-sample testing where there are many variables of interest and the third presents a new methodology for time series bootstrapping.

In the first section we develop a test statistic for testing the equality of two population mean vectors in the "large-p-small-n" setting. Such a test must surmount the rank-deficiency of the sample covariance matrix, which breaks down the classic Hotelling T^2 test. The proposed procedure, called the generalized component test, avoids full estimation of the covariance matrix by assuming that the p components admit a logical ordering such that the dependence between components is related to their displacement. The test is shown to be competitive with other recently developed methods under ARMA and long-range dependence structures and to achieve superior power for heavy-tailed data. The test does not assume equality of covariance matrices between the two populations, is robust to heteroscedasticity in the component variances, and requires very little computation time, which allows its use in settings with very large p. An analysis of mitochondrial calcium concentration in mouse cardiac muscles over time and of copy number variations in a glioblastoma multiforme data set from The Cancer Genome Atlas are carried out to illustrate the test.

In the second section we present a theorem establishing a power improvement to the Benjamini–Hochberg procedure for controlling the false discovery rate when it is applied to test statistics which have been adjusted for the effects of latent factors. We extend recently published methodology to the context of serially dependent test statistics by presenting a frequency-domain adaptation of their procedure. We show that our harmonic factor adjustment to the test statistics improves the power of the Benjamini–Hochberg procedure without compromising its control of the false discovery rate when the test statistics are affected by latent periodic components. An illustration of our methodology

is given in an analysis of copy number variations, which are measured along a chromosome and tend to exhibit serial dependence; power gains from our harmonic factor adjustment are demonstrated.

In the third section we present a smoothed bootstrap procedure for time series data. Unlike with independent data, smoothed boostraps have received little consideration for time series. However, as evidenced in the iid smooth bootstrap, additional data smoothing steps within resampling can improve bootstrap approximations of the distributions of statistics, especially when such sampling distributions depend critically on unknown and smooth (e.g., infinite-dimensional) population quantities, such as marginal densities. To broaden the effectiveness of the bootstrap for time series, we propose a smooth bootstrap based on modifying a state-of-the-art block resampling approach for dependent data based on tapering windows. The resulting smooth (extended) tapered block bootstrap (TBB) is shown to provide valid variance and distributional approximations over a broad class of parameters and statistics for stationary time series, formulated in terms of statistical functionals (e.g., smooth function model statistics, L- and M-estimators, rank statistics). Our treatment goes beyond statistics as smooth functions of sample averages, showing that the smooth TBB has applicability in inference cases which have not been formally established for other TBB versions. Some finite-sample simulations also provide evidence that smoothing steps enhance the performance of the block bootstrap for various statistical functionals.

ACKNOWLEDGEMENTS

Thanks to my parents for taking all of my career ideas seriously, for encouraging my imagination, for teaching me that "the job is the boss", and for paying my rent in college. Thanks to my advisors for their direction and trust. Thanks to Dr. Felix Famoye for telling me I should study statistics. Thanks to my roommates for turning the TV down.

NOMENCLATURE

Ch-Q	Refers to the test from Chen & Qin (2010)
CLX	Refers to the test from Cai et al. (2014)
SK	Refers to the test from Srivastava & Kubokawa (2013)
GCT	Generalized Component Test
ARMA	Autoregressive Moving Average
IND	Independence
LR	Long-range dependence
CBS	Circular Binary Segmentation
FDR	False Discovery Rate
BH	Refers to the procedure from Benjamini & Hochberg (1995)
FHG	Refers to the procedure from Fan et al. (2012)
MSE	Mean Squared Error
MBB	Moving Blocks Bootstrap
SMBB	Smooth Moving Blocks Bootstrap
ETBB	Extended Tapered Blocks Bootstrap
SEBB	Smooth Extended Tapered Blocks Bootstrap
HHJ	Refers to the block size selection method in Hall et al. (1995)

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

This dissertation is composed of three projects. The first two address two-sample testing problems when large numbers of variables are concerned and the third presents a smooth block bootstrap method for time series.

In Section 2 a test statistic for testing equality of mean vectors between two populations is proposed and its null distribution is derived. The test is developed under the setting in which the variables of interest admit an ordering in some index such as time or space from which the test statistics inherit a serial dependence structure. The proposed test demonstrates superior power over competitors in the literature in some simulation studies and in an analysis of copy number data from two groups of cancer patients.

Section 3 develops a dependence-adjusted multiple testing procedure for differences in means for a large number of variables, continuing under the assumption of serial dependence. Power gains from the dependence adjustment are established theoretically and demonstrated in simulations as well as on real data.

Section 4 develops a time-series bootstrap methodology involving block resampling with tapered block weights and smoothing by adding normal perturbations to the data values. The gains from smoothing in time series bootstrap methods have not been well explored in the literature, and this work demonstrates that estimates of the sampling distributions of some statistics such as the quantile and the trimmed mean can be substantially improved by smoothing. We prove the consistency of the recently introduced extended tapered block bootstrap and our smoothed version of it under a broader class of statistics than that originally considered.

Appendices for the three sections appear at the end of this document which contain most of the proofs and some supplementary material.

2. A TWO-SAMPLE TEST FOR EQUALITY OF MEANS IN HIGH DIMENSION

2.1 Introduction

In many applications it is desirable to test whether the means of high-dimensional random vectors are the same in two populations. Often, the number of components in the random vectors exceeds the number of sampled observations, the so-called "large-p-small-n" problem, and conventional test statistics become unviable. Given the steadily growing availability and interest in high-dimensional data, particularly in biological applications, test statistics that are viable for high-dimensional data are in increasing demand.

The challenge when $p \gg n$ is to model the structure of dependence among the p components without estimating each of the $p(p+1)/2$ unique entries in the full covariance matrix. The classical test for equal mean vectors between two populations is Hotelling's T^2 test, but the test statistic is undefined when p is larger than the sum of the sample sizes (minus 2), because it involves inverting the $p \times p$ sample covariance matrix. Several procedures are available which circumvent full covariance matrix estimation. We acheive this in the important case in which the p components admit an ordering in time, space, or in another index, such that the dependence between two components is related to their displacement. When measurements are taken along a chromosome, for example, the location of each measurement is recorded, providing an index over which dependence may be modeled, affording gains in power. For concreteness, it is here assumed that the components admit a unidirectional ordering.

To fix notation, let $X_1, X_2, \ldots, X_n \in \mathbb{R}^p$ and $Y_1, Y_2, \ldots, Y_m \in \mathbb{R}^p$ be independent identically distributed random samples from two populations having $p \times 1$ mean vectors μ_1 and μ_2 and $p \times p$ covariance matrices Σ_1 and Σ_2, respectively. The hypotheses of interest become $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$.

There are some methods available for testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ in the

"large-p-small-n" setting. Srivastava (2007) presented a modification of Hotelling's T^2 statistic which handles the singularity of the sample covariance matrix by replacing its inverse with the Moore-Penrose inverse. Wu et al. (2006) proposed the pooled component test, for which the test statistic is the sum of the squared univariate pooled two-sample t-statistics for all p vector components, which they assumed to follow a scaled chi-square distribution. Bai & Saranadasa (1996) presented a test statistic which uses only the trace of the sample covariance matrix and performs well when the random vectors of each population can be expressed as linear transformations of zero-mean i.i.d. random vectors with identity covariance matrices. Each of these methods assumes a common covariance matrix between the two populations, that is that $\Sigma_1 = \Sigma_2$.

More recently, under a setup similar to that of Bai & Saranadasa (1996), but which accommodates unequal covariances, Chen & Qin (2010) introduced a method (hereafter called the Ch-Q test), which allows $\Sigma_1 \neq \Sigma_2$ and sidesteps covariance matrix estimation altogether. Srivastava & Kubokawa (2013) proposed a method (hereafter called the SK test) for multivariate analysis of variance in the large-p-small-n setting, of which the high-dimensional two-sample problem is an instance. Cai et al. (2014) presented a test (hereafter called the CLX test) based upon the supremum of standardized differences between the observed mean vectors, and offer an illuminating discussion about the conditions under which supremum-based tests are likely to outperform sum-of-squares-based tests, which include the Ch-Q and SK tests as well as the test we introduce in this paper. If the differences between μ_1 and μ_2 are rare, but large where they occur, i.e. the signals are sparse but strong, a supremum-based test should have greater power than a sum-of-squares-based test. The reason is that tests which sum the differences across a large number of indices will not be greatly influenced by a very small number of large differences. If, however, there are many differences between μ_1 and μ_2, but these differences are small, i.e. the signals are dense but weak, the supremum of the differences across all the indices will not likely be extreme enough to arouse suspicion of the null. A sum-of-

squares based test statistic, however, will represent an accumulation of the large number of weak signals, and will have more power. Dense-but-weak signal settings do exist, for example in the analysis of copy number variations, where mildly elevated or reduced numbers of DNA segment copies in cancer patients are believed to occur over regions of the chromosome rather than at isolated points (Olshen et al. (2004), Baladandayuthapani et al. (2010)). It is for such cases that our test is designed.

Section 2.2 describes the GCT test statistic and Section 2.3 gives its asymptotic distribution. Section 2.4 presents a simulation study of the GCT, comparing its performance with that of the Ch-Q, SK, and CLX tests in terms of power and maintenance of nominal size. Section 2.5 implements the GCT as well as the Ch-Q, SK, and CLX tests on a copy number data set and a time series data set. Concluding remarks appear in Section 2.7 and Appendix A provides proofs of the main results. Full details for the proofs may be found in Appendix A.

2.2 Test Statistic

The GCT statistic is computed as follows. Let $T_n = p^{-1}(t_{n1}^2 + t_{n2}^2 + \cdots + t_{np}^2)$, where

$$t_{nj}^2 = (\overline{X}_{nj} - \overline{Y}_{mj})^2 (s_{nj}^2/n + \vartheta_{mj}^2/m)^{-1} \tag{2.1}$$

for $j = 1, \ldots, p$, where $\overline{X}_{nj}$ and $\overline{Y}_{mj}$ are the sample means of the j^{th} vector component and s_{nj}^2 and ϑ_{mj}^2 are the sample variances of the j^{th} vector component for the X and Y samples, respectively. Thus T_n is the mean of the squared unpooled univariate two-sample t-statistics t_{nj}^2 over all components $j = 1, \ldots, p$.

The GCT statistic is a centered and scaled version of T_n defined as $G_n \equiv p^{1/2}(T_n - \widehat{\xi}_n)/\widehat{\zeta}_n$, where $\widehat{\xi}_n$ and $p^{1/2}/\widehat{\zeta}_n$ are described below. The equal means hypothesis is rejected at level α when $|G_n| > \Phi^{-1}(1 - \alpha/2)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

In what shall be called the *moderate-p* version of the test, $\widehat{\xi}_n \equiv 1$, so that $G_n^{(M)} \equiv$

$p^{1/2}(T_n - 1)/\widehat{\zeta}_n$. For the *large-p* version, higher-order expansions suggest a centering of the form $\widehat{\xi}_n \equiv 1 + n^{-1}\widehat{a}_n + n^{-2}\widehat{b}_n$, so that

$$G_n^{(L)} \equiv p^{1/2}\{T_n - (1 + n^{-1}\widehat{a}_n + n^{-2}\widehat{b}_n)\}/\widehat{\zeta}_n. \tag{2.2}$$

The quantities $\widehat{a}_n$ and $\widehat{b}_n$ are defined as $\widehat{a}_n \equiv (\widehat{c}_{n1}+\cdots+\widehat{c}_{np})/p$ and $\widehat{b}_n \equiv (\widehat{d}_{n1}+\cdots+\widehat{d}_{np})/p$, where $\widehat{c}_{nj}$ and $\widehat{d}_{nj}$ are obtained by plugging sample moments into the expressions given in Lemma 1 for c_{nj} and d_{nj} for each of the components $j = 1, \ldots, p$.

Though T_n is a mean of squared marginal two-sample t-statistics, the construction of the scaling will account for the dependence among them. In both the moderate- and large-p versions of the test statistic, the scaling $p^{1/2}/\widehat{\zeta}_n$ is the same. Let

$$\widehat{\gamma}(k) = (p-k)^{-1}\textstyle\sum_{j=1}^{p-k}(t_{nj}^2 - T_n)(t_{n(j+k)}^2 - T_n), \tag{2.3}$$

which is the sample autocovariance function of the squared t-statistics. Then the scaling $\widehat{\zeta}_n$ is defined such that

$$\widehat{\zeta}_n^2 \equiv \textstyle\sum_{|k|<L} w(k/L)\widehat{\gamma}(k), \tag{2.4}$$

where $w(x)$ is an even, piecewise function of x such that $w(0) = 1$, $|w(x)| \leq 1$ for all x, and $w(x) = 0$ for $|x| > 1$, and L is a user-selected lag window size.

The choices of the lag window $w(\cdot)$ considered here are the Parzen window

$$w_p(x) \;=\; \begin{cases} 1 - 6|x|^2 + 6|x|^3, & |x| < 1/2 \\ 2(1-|x|)^3, & 1/2 \leq x \leq 1 \\ 0, & |x| > 1 \end{cases}$$

found in Brockwell & Davis (2009) and the trapezoid window

$$w_T(k/r) = \begin{cases} 1, & |k| < [L/2] \\ 1 - \left(\frac{k-[L/2]}{r-[L/2]}\right), & [L/2] \le k \le L \\ 0, & |k| > L \end{cases}$$

from Politis & Romano (1995), where $[x]$ denotes the largest integer not exceeding x.

2.3 Main Results

Let $\alpha(r) = \sup\{\alpha(\mathcal{F}_1^k, \mathcal{F}_{k+r}^p) : 1 \le k \le p-r\}$, where $\mathcal{F}_a^b \equiv \mathcal{F}_{a,n}^b = \sigma\langle\{t_{nj} : a \le j \le b\}\rangle$ and where for any σ-fields, $\mathcal{F}$ and $\mathcal{G}$,

$$\alpha(\mathcal{F}, \mathcal{G}) = \sup\{|P(A \cap B) - P(A)P(B)| \colon A \in \mathcal{F}, B \in \mathcal{G}\}$$

denotes the strong mixing coefficient between $\mathcal{F}$ and $\mathcal{G}$. Then the following conditions are assumed in deriving the asymptotic distribution of the test statistic T_n.

(C.1) For some $\delta \in (0, \infty)$, (i) $\sum_{r=1}^{\infty} \alpha(r)^{\delta/(2+\delta)} < \infty$, and (ii) $E|t_{nj}^2|^{2r+\delta} < c < \infty$ for all $j = 1, \ldots, p$ for some integer $r \ge 1$.

(C.2) The limit $\lim_{n\to\infty} \frac{1}{p-k} \sum_{j=1}^{p-k} \mathrm{Cov}(t_{nj}^2, t_{n(j+k)}^2) = \gamma(k)$ exists for all $k > 0$.

(C.3) (i) $\max\{E|X_{1j}|^{16}, E|Y_{1j}|^{16}, j = 1, \ldots, p\} = O(1)$.

(ii) $\min\{\mathrm{Var}(X_{1j}), \mathrm{Var}(Y_{1j})\} > c > 0$.

The following theorem establishes the asymptotic normality of the test statistic under the appropriate centering and scaling.

Theorem 1 *Suppose that $p \equiv p_n = o(n^6)$ and (C.1)–(C.3) hold with $r = 1$ in (C.1). Then*

$$\sup_{x \in \mathbb{R}} |P(T_n - 1 < x) - \Phi\{\sqrt{p}(x - n^{-1}a_n - n^{-2}b_n)/\tau_\infty\}| = o(1),$$

where $\tau_\infty^2 = \gamma(0) + 2\sum_{k=1}^{\infty}\gamma(k)$ *and* $a_n = (c_{n1} + \cdots + c_{np})/p$ *and* $b_n = (d_{n1} + \cdots + d_{np})/p$, *where* c_{nj} *and* d_{nj} *for* $j = 1, \ldots, p$ *are as in Lemma 1 in Appendix A.*

Remark 1 *Theorem 1 shows that* $G_n \equiv p^{1/2}(T_n - \widehat{\xi}_n)/\widehat{\zeta}_n \to^d Normal(0,1)$ *as* $n \to \infty$.

2.3.1 Technical Details

The choice of the centering quantity $\widehat{\xi}_n$ comes from noting that $ET_n = 1 + O(n^{-1})$ as $n \to \infty$. This follows from the fact that t_{nj} converges in distribution to Z, where $Z \sim \text{Normal}(0,1)$, for all $j = 1, \ldots, p$, and $EZ^2 = 1$. Thus $E\{\sqrt{p}(T_n - 1)\} = \sqrt{p}O(n^{-1})$, so that when $\widehat{\xi}_n \equiv 1$, the expectation of the test statistic differs from zero by $\sqrt{p}O(n^{-1})$, restricting p to grow at a rate such that $p = o(n^2)$. When $\widehat{\xi}_n \equiv 1 + n^{-1}\widehat{a}_n + n^{-2}\widehat{b}_n$, the expectation of the test statistic is $\sqrt{p}O(n^{-3})$, allowing $p = o(n^6)$. Hence the "moderate-" and "large-p" designations. One may also consider an intermediate-p version of the test which uses only $n^{-1}\widehat{a}_n$ in the centering correction, allowing $p = o(n^4)$, but its performance is not investigated here.

While the large-p test allows for $p = o(n^6)$, an advantage of the moderate-p test is its robustness to outliers. The centering correction in the large-p test involves high-order sample moments which are volatile when the data come from a very heavy-tailed distribution, in which case the centering value of 1 is preferable.

The formulation of $\widehat{\zeta}_n$ rests on the assumption that the p components admit a logical ordering such that their dependence is autocovarying and diminishing as components are further removed—that is, that the covariance between components may be described with an autocovariance function that decays sufficiently fast. In the proof of Theorem 1, the asymptotic variance of $p^{1/2}T_n$ under some regularity conditions is shown to be $\sum_{h=-\infty}^{\infty}\gamma(h)$, which is equal to 2π times the spectral density $f(\cdot)$ of the sequence $(t_{n1}^2, t_{n2}^2, \ldots)$ evaluated at 0. Thus $\widehat{f}(0) = (2\pi)^{-1}\sum_{|k|<L} w(k/L)\widehat{\gamma}(k)$ provides the scaling in (2.4).

2.3.2 Power of the Generalized Component Test

In order to compute the asymptotic power of the GCT, the expected value of $T_n = p^{-1}(t_{n1}^2, \ldots, t_{np}^2)$ must be computed under the alternative $H_1 : \mu_{1j} - \mu_{2j} = \delta_j$ for $j = 1, \ldots, p$ when $\delta_j \neq 0$ for at least one j. Let $\xi_n^{(1)}$ denote $E(T_n|H_1 \text{ true})$. Then the power of the GCT, which is $P(|p^{1/2}(T_n - \widehat{\xi}_n)/\widehat{\zeta}_n)| > z_{\alpha/2}|H_1 \text{ true})$, is equal to

$$1 - P(-z_{\alpha/2} - p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\widehat{\zeta}_n < p^{1/2}(T_n - \xi_n^{(1)})/\widehat{\zeta}_n) < z_{\alpha/2} - p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\widehat{\zeta}_n|H_1 \text{ true}).$$

Under conditions (C.1)–(C.3) we can invoke the asymptotic normality of $p^{1/2}(T_n - \xi_n^{(1)})/\widehat{\zeta}_n$ and the consistency of $\widehat{\zeta}_n$ for ζ and approximate the power with

$$1 - \{\Phi(z_{\alpha/2} - p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\zeta) - \Phi(-z_{\alpha/2} - p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\zeta)\}$$

so that it is a function of $p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\zeta$.

Given the tedium of computing $\xi_n^{(1)} = E\{p^{-1}\sum_{j=1}^p t_{nj}^2|H_1 \text{ true}\} = np^{-1}\sum_{j=1}^p E[(\overline{X}_{nj} - \overline{Y}_{mj})^2/\{s_{nj}^2 + (n/m)\vartheta_{mj}^2\}|H_1 \text{ true}]$ to within $O(n^{-3})$ of its true value as was done for $\widehat{\xi}_n$ under the null hypothesis (cf. Lemma 1), we replace s_{nj}^2 and ϑ_{mj}^2 with their population values σ_{1j}^2 and σ_{2j}^2 and get $\xi_n^{(1)} \approx 1 + n(\mu_{1j} - \mu_{2j})^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}$.

If we may replace n, $p^{1/2}(\xi_n^{(1)} - \widehat{\xi}_n)/\zeta$ with $np^{-1/2}\sum_{j=1}^p \delta_j^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}/\zeta$, then the power may be expressed

$$\begin{aligned} 1 - (&\Phi[z_{\alpha/2} - np^{-1/2}\textstyle\sum_{j=1}^p \delta_j^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}/\zeta] \\ &- \Phi[-z_{\alpha/2} - np^{-1/2}\textstyle\sum_{j=1}^p \delta_j^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}/\zeta]). \end{aligned}$$

From this expression we note that under $p = o(n^2)$

$$\text{Power} \quad \rightarrow \quad \begin{cases} 1, & p^{1/2}n^{-1} = o(\sum_{j=1}^p \delta_j^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}) \\ \alpha, & \sum_{j=1}^p \delta_j^2/\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}) = o(p^{1/2}n^{-1}) \end{cases}$$

For example, if $\delta_j = \delta p^{-1/2}$ for $j = 1, \ldots, p$ for some $\delta > 0$ then the power will converge to 1, but if $\delta_j = \delta p^{-(1/2+\epsilon)}$ for $j = 1, \ldots, p$ the test will have "nonpower" above the significance level as $n, p \to \infty$.

2.4 Simulation Studies

The performances of the GCT, Ch-Q, SK, and CLX tests were compared in terms of size control and power under various settings. For the sample sizes $(n, m) = (45, 60)$ and $(n, m) = (90, 120)$ with $p = 300$, two-sample data were generated such that for each subject the p components were (i) independent (IND), (ii) ARMA dependent, or (iii) long-range (LR) dependent. For each dependence structure, the innovations used to generate each subject series were (a) Normal(0,1), (b) skewed innovations, coming from a gamma$(4, 2)$ distribution centered at zero, thus having mean zero and variance $4(2)^2 = 16$, and (c) heavy-tailed innovations from a Pareto(a, b) distibution with distribution function $F(x) = 1 - (1 + x/b)^{-1/a}$ where the density was shifted to the origin and reflected across the vertical axis to form a "double" Pareto distribution. Under this double Pareto distribution,

$$E|X|^r = \begin{cases} \infty, & r \geq a \\ b^r \Gamma(a-r)\Gamma(1+r)/\Gamma(a), & r < a. \end{cases}$$

Once a zero-mean series was generated for each subject, it was added to the $p \times 1$ mean vector μ_1 or μ_2, depending on the population to which the subject belonged. Under IND, the zero-mean series consisted of p independent identically distributed innovations from the chosen innovation distribution. For the ARMA dependence structure, p-length series from an ARMA process with AR parameters $\phi_1 = \{0.4, -0.1\}$ and MA parameters $\theta_1 = \{0.2, 0.3\}$ were used for both populations. Under the LR structure, realizations of zero-mean, long-range-dependent processes with self-similarity parameter $H_1 = (1/2)(2 - 0.75) = 0.625$ were used. The algorithm used for generating vectors of

long-range dependent random variables is found in Hall et al. (1998).

At each sample size, dependence structure, and innovation distribution combination, a simulation was run in which $\Sigma_1 = \Sigma_2$ and in which $\Sigma_2 = 2\Sigma_1$, where the unequal covariance setting was imposed by scaling the zero-mean series for the population 2 subjects by $\sqrt{2}$.

For the CLX test, which features an equal-covariances and an unequal-covariances version, Cai et al. (2014) suggest first testing $H_0 : \Sigma_1 = \Sigma_2$ using a test from Cai et al. (2013a) and then choosing the version of the CLX test accordingly. Since in practice it is generally not known whether $\Sigma_1 = \Sigma_2$ holds, the test of $H_0 : \Sigma_1 = \Sigma_2$ was performed in each simulation run to determine which version of the CLX test would be used. The CLX test requires an estimate for the precision matrix $\Omega = \Sigma_1^{-1} (= \Sigma_2^{-1})$ or $\Omega = \{\Sigma_1 + (n/m)\Sigma_2\}^{-1}$ for the unequal-covariances version. Of the two methods the authors suggest for estimating Ω, that which is presented in Cai et al. (2011) and provided in the `R` package `fastclime` (Pang et al. (2013)) was chosen and implemented under default settings.

For power simulations, the alternate hypotheses were that $\mu_1 = 0$ and $\mu_2 = [\delta 1'_{\beta p}, (0)1'_{(1-\beta)p}]'$, where 1_k was a $k \times 1$ vector of ones, p was the number of components, and $\beta \in [0, 1]$ was the proportion of the p components for which the difference in means was nonzero. The number of components p was fixed at 300 and the power was simulated for $\beta \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1\}$. The difference or signal δ was chosen such that the signal to noise ratio δ/σ was equal to $1/8$, where σ was the standard deviation of the innovations used to construct each series (each p-variate observation); thus $\delta = \sigma/8$ was used.

Full factorial simulation results for $\{(45, 60), (90, 120)\} \times \{$IND, ARMA, LR$\} \times \{$Normal, Skewed, Heavy-tailed$\} \times \{\ \Sigma_1 = \Sigma_2\ ,\ \Sigma_2 = 2\Sigma_1\ \}$ were run, but only selected results are highlighted here. In addition to the factorial simulation, the tests were evaluated under heteroscedastic component variances and ultra-heavy tailed (infinite-variance) innova-

tions.

2.4.1 Performance Under Normality

Table 2.1 displays the simulated Type I error rates of the four tests under the sample sizes $(n, m) = (45, 60), (90, 120)$ across the three dependence structures under Normal$(0, 1)$ innovations and for $\Sigma_1 = \Sigma_2$. For the GCT, results are given for the Parzen and trapezoid lag windows at lag window sizes $L = 10, 15, 20$ for the moderate-p (upper panel) and the large-p (lower panel) choice of the centering. The Ch-Q, SK, and CLX Type I error rates are duplicated in the upper and lower panels as the moderate- and large-p versions of the GCT were applied to the same 500 simulated data sets.

$p = 300$, $\Sigma_1 = \Sigma_2$			Normal$(0, 1)$ Innovations								
$\widehat{\xi_n} \equiv 1$						Parzen Window			Trapezoid Window		
n	m	Cov	Ch-Q	SK	CLX	L = 10	L = 15	L = 20	L = 10	L = 15	L = 20
45	60	IND	0.07	0.04	0.09	0.06	0.07	0.07	0.06	0.08	0.07
		ARMA	0.06	0.04	0.08	0.06	0.07	0.07	0.07	0.08	0.07
		LR	0.05	0.04	0.10	0.06	0.06	0.07	0.08	0.09	0.07
90	120	IND	0.05	0.04	0.07	0.06	0.06	0.06	0.07	0.08	0.06
		ARMA	0.05	0.04	0.06	0.07	0.07	0.08	0.08	0.09	0.08
		LR	0.03	0.03	0.07	0.05	0.05	0.07	0.06	0.08	0.07
$\widehat{\xi_n} \equiv 1 + a_n/n + b_n/n^2$						Parzen Window			Trapezoid Window		
n	m	Cov	Ch-Q	SK	CLX	L = 10	L = 15	L = 20	L = 10	L = 15	L = 20
45	60	IND	0.07	0.04	0.09	0.07	0.07	0.07	0.07	0.08	0.07
		ARMA	0.06	0.04	0.08	0.07	0.07	0.07	0.07	0.08	0.07
		LR	0.05	0.04	0.10	0.07	0.07	0.08	0.08	0.09	0.08
90	120	IND	0.05	0.04	0.07	0.06	0.06	0.07	0.07	0.08	0.07
		ARMA	0.05	0.04	0.06	0.08	0.08	0.08	0.08	0.09	0.08
		LR	0.03	0.03	0.07	0.06	0.06	0.06	0.07	0.09	0.06

Table 2.1: Type I error rates over $S = 500$ simulations with nominal size $\alpha = .05$ for the moderate- and large-p GCT under the Parzen and trapezoid lag windows at lengths $L = 10, 15, 20$ and for the Ch-Q, SK, CLX tests under Normal$(0, 1)$ innovations with $\Sigma_1 = \Sigma_2$.

The Ch-Q and SK tests maintained very close-to-nominal Type I error rates. The

CLX test exhibited slightly inflated Type I error rates under the IND and LR dependence structures for the smaller sample sizes $(n, m) = (45, 60)$, but maintained close-to-nominal rates for $(n, m) = (90, 120)$. For the GCT, the Parzen window appeared to control the Type I error rate slightly better than the trapezoid window, and the Type I error rates were similar for the three choices of the lag window size.

2.4.2 Effect of Skewness

The results of the Type I error simulation with skewed innovations were similar to those in the Normal$(0, 1)$ case. For the power simulation, Figure 2.1 plots the proportion of rejections across 500 simulation runs against the proportion β of the $p = 300$ components in which μ_1 and μ_2 differed, where $\beta \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1\}$. The three panels show the power curves of the four tests under the IND, ARMA, and LR dependence structures, respectively, when the innovations came from the centered gamma$(4, 2)$ distribution and when the sample sizes were $(n, m) = (90, 120)$. The four tests exhibited similar performance under these settings, though under independence the size of the CLX test was somewhat inflated, yet its power increased more rapidly in β than that of the other tests under ARMA dependence.

2.4.3 Effect of Heavy-Tailedness

The results for the heavy-tailed simulation with innovations coming from the double Pareto$(16.5, 8)$ distribution did not differ greatly from those of the normal- and skewed-innovations simulations. In order to assess the robustness of the GCT to violations of its moment conditions, ultra-heavy tailed data were simulated using innovations from a double Pareto$(1.5, 1)$ distribution, which has infinite variance. Since the centering corrections $\widehat{a}_n$ and $\widehat{b}_n$ in the large-p GCT are computed using higher order sample moments, only the moderate-p GCT was here considered, as its centering of 1 gives it stability. Under these settings, the signal, which was set to $\delta = .5$, is very weak relative to the noise, such that as the proportion β of non-null mean differences goes to 1, a dense-but-

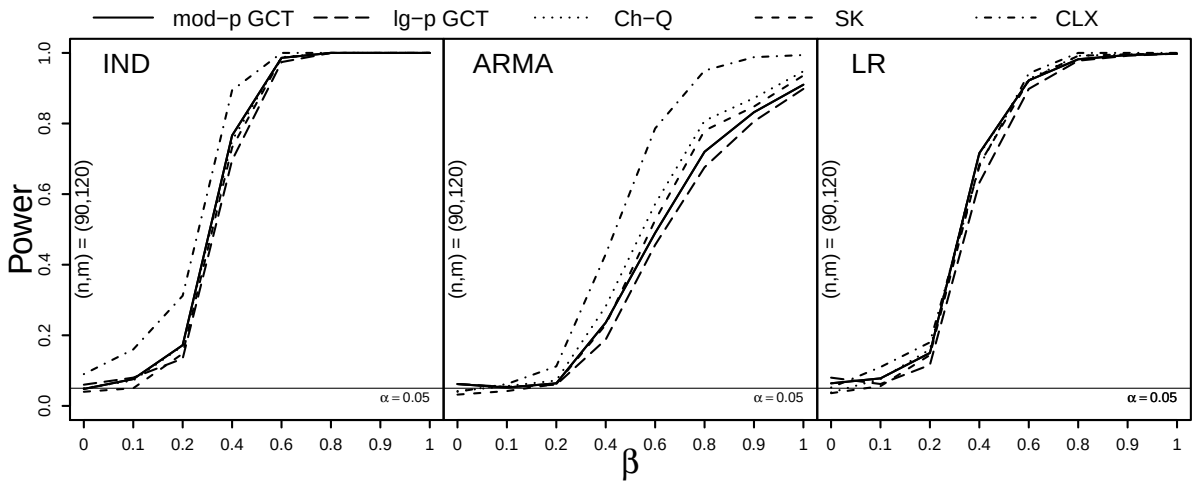

Figure 2.1: Power curves at sample sizes $(n, m) = (90, 120)$ for the moderate- and large-p GCT, Ch-Q, SK, and CLX tests against the proportion of nonzero mean differences β under IND, ARMA, and LR dependence (left to right) with centered gamma$(4, 2)$ innovations and $\Sigma_1 = \Sigma_2$. Based on $S = 500$ simulations.

weak signal structure is simulated. The resulting power curves are shown in Figure 2.2, in which the Ch-Q test is seen to have much less power than the others; the CLX also suffers a reduction in power under ARMA and LR dependence. Under LR dependence, the size of the GCT was somwhat inflated, but it was very close to nominal for the IND and ARMA cases. In the ARMA case, the GCT exhibited greater power than the other tests across the range of alternatives.

2.4.4 Effect of Heteroscedasticity

The effect of heteroscedasticity on the GCT may be anticipated by noting that t_{nj}^2 from (2.1) can be expressed

$$t_{nj}^2 = \left[\frac{\sqrt{n}\{(\overline{X}_{nj} - \mu_{1j}) - (\overline{Y}_{mj} - \mu_{2j})\}}{\sqrt{s_{nj}^2 + (n/m)\vartheta_{mj}^2}} + \frac{\sqrt{n}\delta_j}{\sqrt{s_{nj}^2 + (n/m)\vartheta_{mj}^2}} \right]^2 \tag{2.5}$$

where $\delta_j = \mu_{1j} - \mu_{2j}$, for $j = 1, \ldots, p$. The second term is equal to zero under H_0. Under H_1, for a fixed difference δ_j, the variances σ_{1j}^2 and σ_{2j}^2 affect the magnitude of t_{nj}^2 such

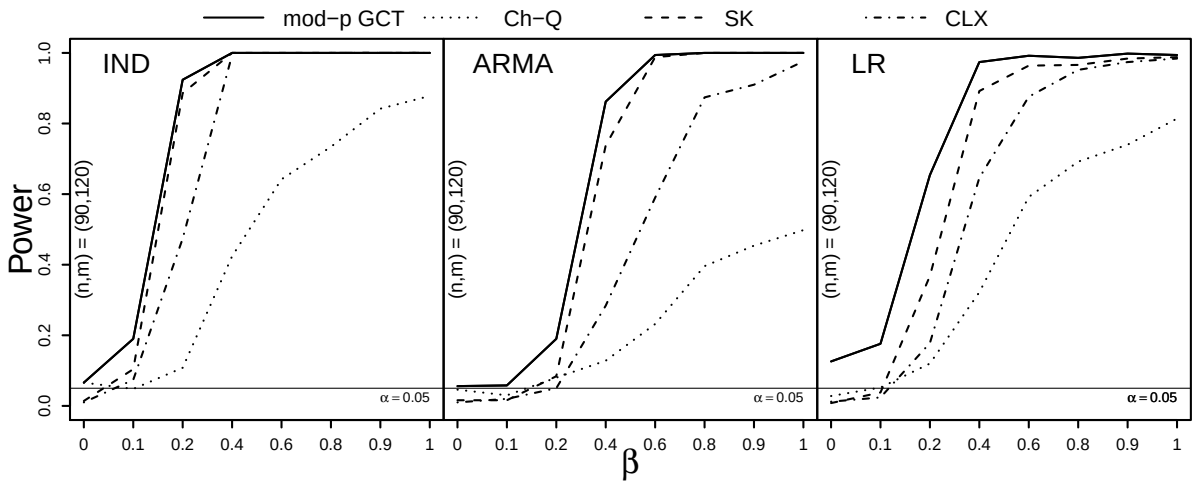

Figure 2.2: Power curves at sample sizes $(n, m) = (90, 120)$ for the large-p GCT, Ch-Q, SK, and CLX tests against the proportion of nonzero mean differences β under IND, ARMA, and LR dependence (left to right) with double Pareto(1.5,1) innovations and $\Sigma_1 = \Sigma_2$. Based on $S = 500$ simulations.

that very small values for σ_{1j}^2 and σ_{2j}^2 promote very large values of t_{nj}^2. Since the scaling $\widehat{\zeta}_n$ for T_n is a function of $\widehat{\gamma}(\cdot)$, the estimated autocovariance function of $t_{n1}^2, t_{n2}^2, \ldots, t_{np}^2$, as seen from (2.3) and (2.4), extreme values of t_{nj}^2 will pull $\widehat{\zeta}_n$ upward, shrinking T_n toward zero. Extreme values of t_{nj}^2 will tend to occur if σ_{1j}^2 and σ_{2j}^2 are very small when $\delta_j \neq 0$. Although smaller variances ought to ensure a greater likelihood of rejecting H_0, if $\widehat{\zeta}_n$ is inflated by extreme values of t_{nj}^2, the GCT statistic will be close to zero, and the test will fail to reject, hence condition (C.3) (ii). Large values of σ_{1j}^2 and σ_{2j}^2 when $\delta_j \neq 0$ will tend to reduce t_{nj}^2, but since it is bounded below by zero, extreme values will not occur. The size of the test should be robust to any scaling of the variances, as the second term in (2.5) will be zero when H_0 is true.

To investigate the impact of heterscedasticity on the performance of the four tests, the standard deviations of the components were each scaled by a realization from the exponential distribution with mean 1/2 shifted to the right by 1/2 such that the average scaling was 1 and so that the scaled variances were bounded away from 0. The power simulation with centered gamma$(4, 2)$ innovations was repeated under these heterosce-

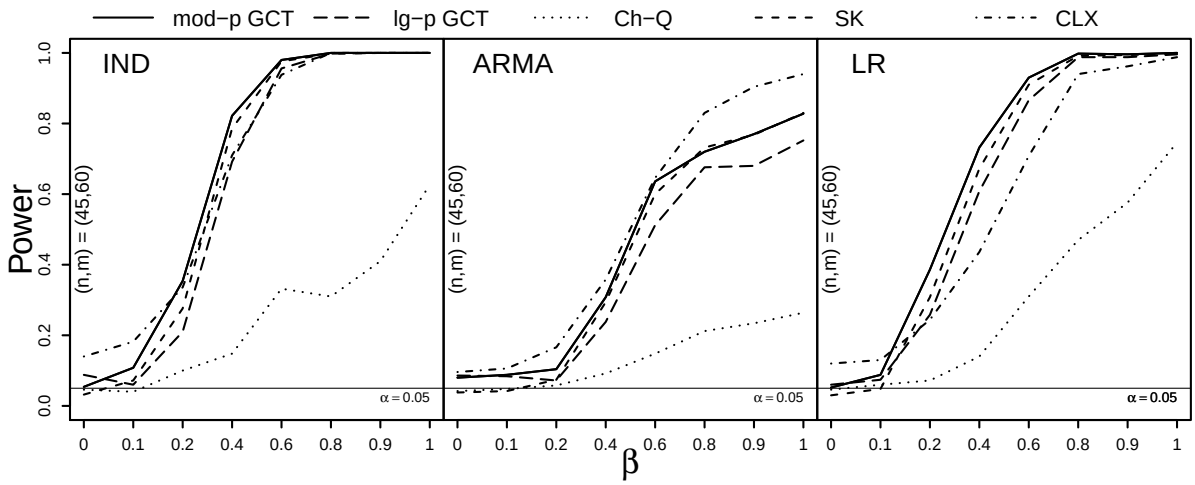

Figure 2.3: Power curves at sample sizes $(n, m) = (45, 60)$ for the moderate- and large-p GCT, Ch-Q, SK, and CLX tests against the proportion of nonzero mean differences β under IND, ARMA, and LR dependence (left to right) with heteroscedastic centered gamma$(4, 2)$ innovations and $\Sigma_1 = \Sigma_2$. Based on $S = 500$ simulations.

castic conditions with $(n, m) = (45, 60)$. Figure 2.3 exhibits a dramatic reduction in the power of the Ch-Q test due to heteroscedasticity. The CLX test exhibited somewhat inflated size under the IND and LR dependence structures, while the SK test and the GCT demonstrated robustness to the heteroscedstic variance scalings.

2.4.5 Effect of Unequal Covariance Matrices

Of the four tests, the SK test is the only one which assumes a common covariance matrix for the two populations. Cai et al. (2014) suggest first testing $H_0 : \Sigma_1 = \Sigma_2$ with a test from Cai et al. (2013a) and implementing the equal or unequal covariances version of the CLX test accordingly. The Ch-Q and the GCT do not require any assumption or testing of equality between the covariance matrices. The SK is thus anticipated to perform more poorly than the others when the covariance matrices are unequal.

To impose inequality between Σ_1 and Σ_2, the zero-mean sequences for each subject from population two were scaled by $\sqrt{2}$ before the signal μ_2 was added. This imposed the condition that $\Sigma_2 = 2\Sigma_1$.

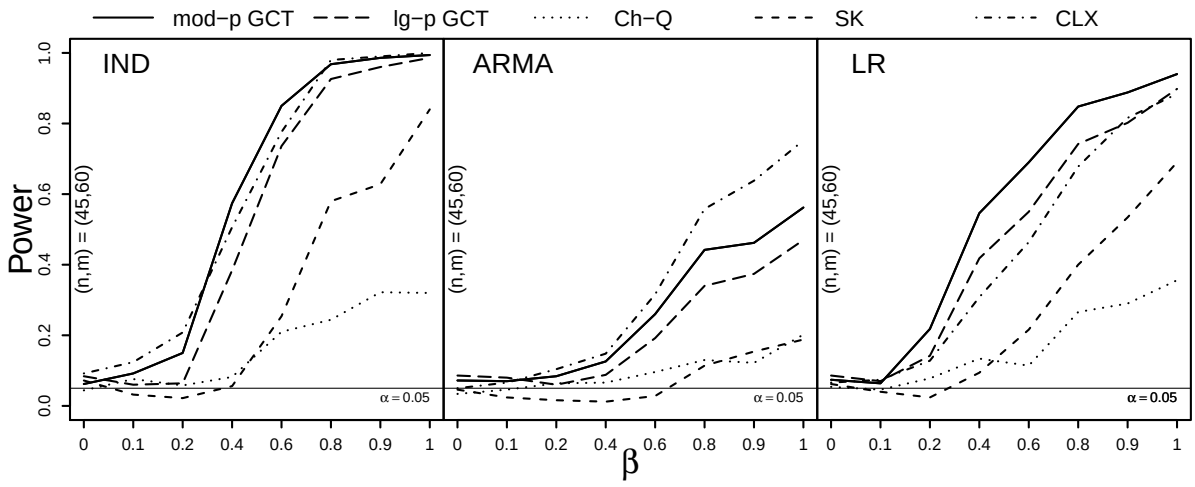

Figure 2.4: Power curves at sample sizes $(n, m) = (45, 60)$ for the moderate- and large-p GCT, Ch-Q, SK, and CLX tests against the proportion of nonzero mean differences β under IND, ARMA, and LR dependence (left to right) with heteroscedastic centered gamma$(4, 2)$ innovations and $\Sigma_2 = 2\Sigma_1$. Based on $S = 500$ simulations.

Figure 2.4 displays results for a simulation in which the variances of the second population were scaled by two and in which the variances in both populations were heteroscedastic. The SK lost much of its power under these settings, which was expected given its assumption of a common covariance matrix in the two populations. The Ch-Q test exhibited low power as before owing to the heteroscedasticity, but performed none the worse for the unequally scaled variances. The GCT achieved the greatest power under the LR dependence structure, having less power than the CLX test in the ARMA case.

Lastly, under the ultra heavy-tailed innovation distribution with unequally scaled covariances between the two populations, the GCT exhibited superior power to the Ch-Q, SK, and CLX tests under all three dependence stuctures at the $(n, m) = (90, 120)$ sample sizes. Although the size of the GCT was somewhat inflated under the LR dependence structure, it maintained the nominal Type I error rate in the ARMA case, under which it achieved roughly 60% power when $\beta = 0.4$ while the CLX test achieved only about 10% power.

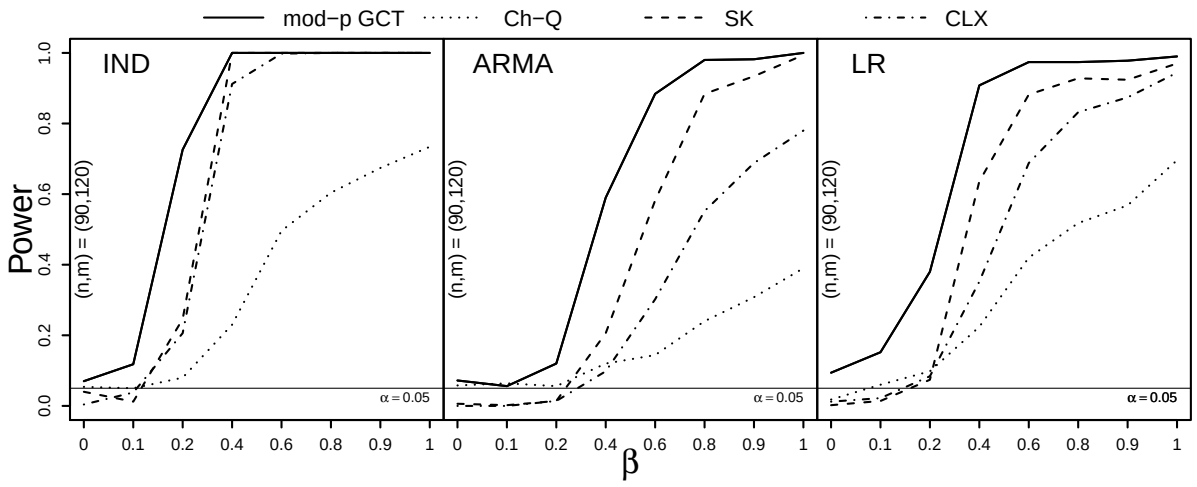

Figure 2.5: Power curves at sample sizes $(n, m) = (90, 120)$ for the moderate-p GCT, Ch-Q, SK, and CLX tests against the proportion of nonzero mean differences β under IND, ARMA, and LR dependence (left to right) with double Pareto(1.5,1) innovations and $\Sigma_2 = 2\Sigma_1$. Based on $S = 500$ simulations.

2.5 Copy Number Variation Example

The GCT, Ch-Q, SK, and CLX tests were each applied to a data set from The Cancer Genome Atlas containing copy number measurements at chromosomal copy number locations in 92 long-term-surviving patients, who survived for more than two years after their initial diagnosis and 138 short-term-surviving patients, who survived for fewer than 2 years after their initial diagnosis of a brain cancer called glioblastoma multiforme. Pinkel & Albertson (2005) suggest that the numbers of copies of certain DNA segments within a cell may be associated with cancer development and spread. It is thus of interest to identify regions along the genome in which high numbers of copies are associated with the incidence or severity of cancer, as such regions may harbor cancer-causing or tumor-suppressor genes. In studies having relatively few patients, several thousand copy number measurements are taken along each arm of each chromosome, which makes identifying regions for which two patient groups have different mean copy number profiles a high-dimensional problem. Additionally, it is believed that copy number variations between

patient groups will occur over stretches of the chromosome (spanning multiple probes) rather than at isolated points (singleton probe locations) (Olshen et al. (2004), Baladandayuthapani et al. (2010)), suggesting a serial dependence over the chromosome as well as the presence of a dense-but-weak rather than a sparse-but-strong signal structure.

We restricted our analysis to the q arm of chromosome 1, the longest chromosome, on which there are 8,895 copy number measurements. Each measurement is a log-ratio of the number of copies at each location over 2, where 2 is the number of copies found in normal DNA. Positive measurements thus indicate duplications and negative measurements indicate deletions. The measurements, in conformity with the assumption of the GCT that the components of interest admit a logical ordering, are recorded along with their locations given in the number of base pairs from the end of the DNA strand. For many of the 8,895 components, there are a few missing values in either or both of the samples such that the average proportion of missing values per component is 0.0276 for the long-term survivors and 0.0273 for the short-term survivors. Prior to analysis, each missing value was replaced with the mean of the non-missing values for the same component in the same sample.

Although a test may reject $H_0 : \mu_1 = \mu_2$ when μ_j is the 8895×1 vector of copy number means for $j = 1, 2$, a wholesale conclusion for the entire arm of the chromosome is of little use if it is desired to identify particular regions in which copy number differences lie. In order to break the chromosome arm into meaningful regions in which the equal means hypothesis is of interest, we performed a method of segmentation called circular binary segmentation (CBS) from Olshen et al. (2004). This procedure locates change points in the copy number sequence for a single sequence of copy number values, and is implemented in the R package `DNAcopy` (Seshan & Olshen (2013)). In order to segment the q arm of chromosome 1 for equal means hypothesis testing when multiple patients are observed, the CBS procedure was applied to the 8895×1 vector of differences in means $\bar{X} - \bar{Y}$ using weights equal to $s_j^2/n + \vartheta_j^2/m$ for $j = 1, \ldots, 8895$. Before computing

$\bar{X}$, $\bar{Y}$, and s_j^2 and ϑ_j^2 for $j = 1, \ldots, p$, each series was smoothed using the function `smooth.CNA()` from the `DNAcopy` package. The CBS procedure provided 26 segments of varying lengths at the edges of which change points were detected in the vector of differences in means. As a set of 7 contiguous segments contained small numbers of markers $(44, 14, 26, 39, 26, 21, 27)$ they were collapsed into a single segment having 197 markers, which left 20 regions within which the number of probes p ranged from 73 to 1811. Such splitting of the chromosome into windows or segments has been widely used in genome-wide association studies in searching for chromosomal regions in which genetic variants are associated with a continuous or dichotomous clinical outcome, as in Wu et al. (2011).

The large- and moderate-p GCT with lag window size $L = (2/3)p^{1/2}$ and the Ch-Q, SK, and CLX tests were applied to each of the twenty segments identified by the CBS procedure to test $H_{0k} : \mu_{1k} = \mu_{2k}$ for $k = 1, \ldots, 20$ (Though smoothing was used in identifying the segments, the analysis was carried out on the raw, unsmoothed data). Since the equal-means hypothesis was tested for twenty different regions simultaneously, the sets of p-values which each of the four tests generated were compared with the Benjamini & Hochberg (1995) discovery rate (FDR) threshold. For m tests of hypotheses, the m p-values are ordered $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$ and then the hypothesis to which $p_{(i)}$ corresponds is rejected if $i \leq k$, where $k = \max\{j : p_{(j)} \leq (j/m)q\}$. This procedure was originally shown to control the FDR at q for m independent hypothesis tests, though Benjamini & Yekutieli (2001) showed that for many common types of positive dependence among the m test statistics, the same procedure still adequately controls the FDR. The procedure was therefore applied to the twenty p-values computed from each test.

Figure 2.6 summarizes the analysis. The left panel displays the univariate two-sample t-statistics, which are the t_{nj} values for $j = 1, \ldots, 8895$, against their locations in base pairs along the q arm of chromosome 1. The vertical line at zero marks the value around which the t-statistics would be centered under the null hypotheses, and the horizontal

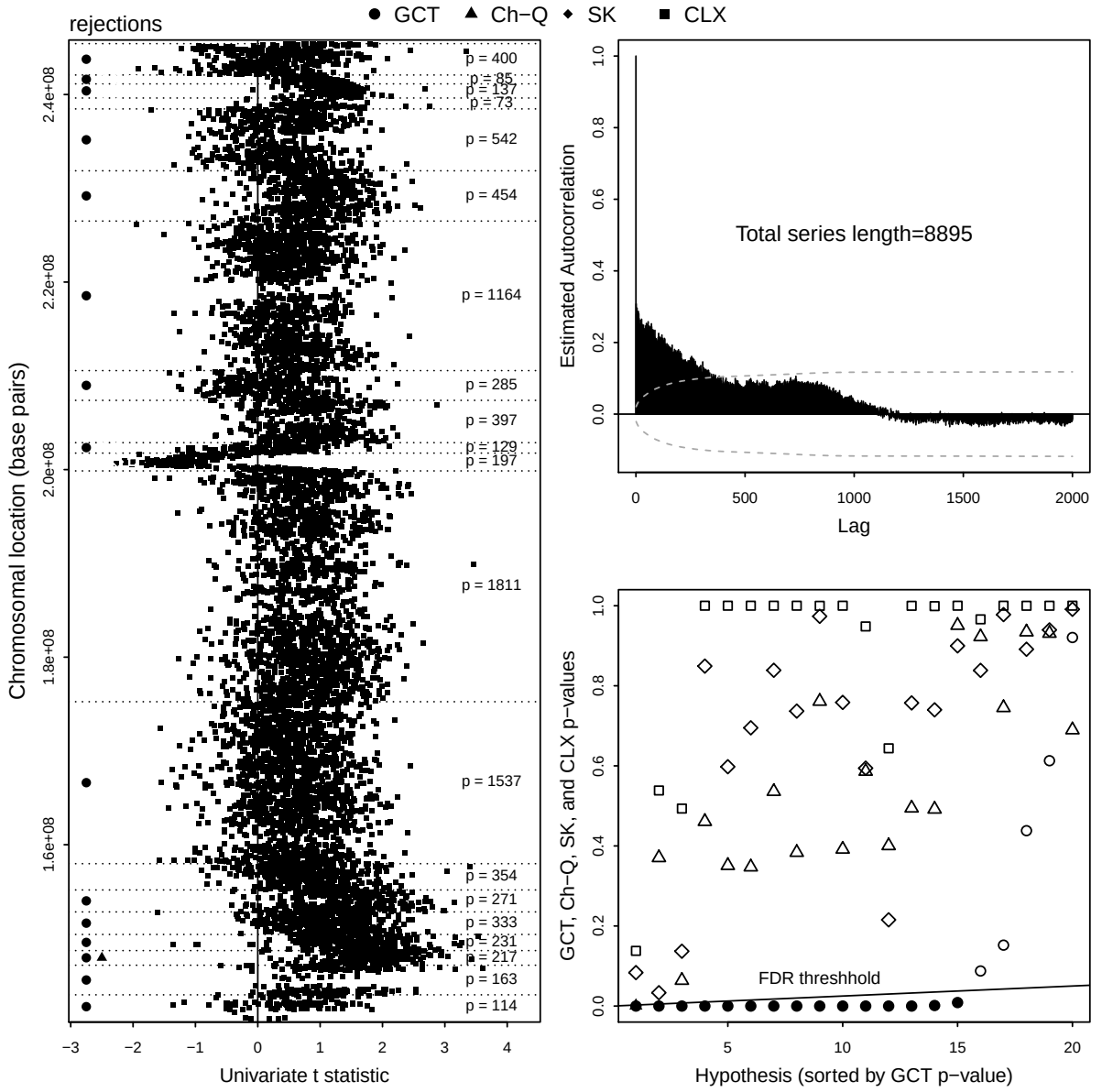

Figure 2.6: (Left) Univariate t-statistics (t_{nj}) plotted against base-pair location on q arm of chromosome 1. Filled symbols denote rejections from FDR procedure for the GCT, Ch-Q, SK, and CLX tests. The number of components p within each CBS-selected chromosomal region is shown. (Upper right) Estimated autocorrelation function for squared univariate t-statistics along q arm of chromosome 1 with large-lag confidence bands. (Lower right) FDR results, hypotheses sorted by GCT p-values. FDR rejection threshold shown with filled symbols denoting rejections.

dotted lines delineate the CBS-selected segments of the chromosome arm. The numbers of copy number markers p within each segment appear on the right. Rejections acheived by the tests are marked with symbols appearing on the left, where rejections for each test are determined by the Benjamini & Hochberg (1995) FDR procedure.

The upper right panel of Figure 2.6 displays the estimated autocorrelation function of the squared two-sample univariate t-statistics, the t_{nj}^2 values for $j = 1, \ldots, 8895$, along the q arm of chromosome 1. The 95% confidence bounds using the large-lag standard error described in Anderson (1977) are shown, which suggest that dependence decays in conformity with (C.1) (i).

The lower right panel of Figure 2.6 shows the results of the FDR procedure. The upward sloping line is given by $y = (x/m)q$, which is the Benjamini & Hochberg (1995) FDR rejection threshold. The p-values for all four tests are shown, but are ordered according to the ranking of the large-p GCT p-values (The rejection decisions were the same for the moderate- and large-p versions of the GCT). The SK and CLX tests did not achieve any rejections; the Ch-Q test achieved one rejection, and the GCT rejected equal means for fifteen of the twenty regions.

Figure 2.7 offers an explanation of the additional power demonstrated by the GCT. The upper and lower panels show the estimated standard deviation at each of the 8,895 copy number locations across the q arm of chromosome 1 for the 92 long-term and 138 short-term survivors, respectively. Both panels exhibit spikes at shared locations as well as prominent humps around 2.0×10^8 Mbps, suggesting that the variances are not constant across the chromosome; nor are the humps at equal heights for the two groups of patients. The boxplots of the 8,895 standard deviations for each group reveal significant right skewness, suggesting heavy-tailedness of some of the component distributions. The minimum estimated standard deviations for the long- and short-term survivors were 0.1314 and 0.1123, respectively, indicating that the component variances are bounded sufficiently away from zero. The severe heteroscedasticity as well as the inequality of

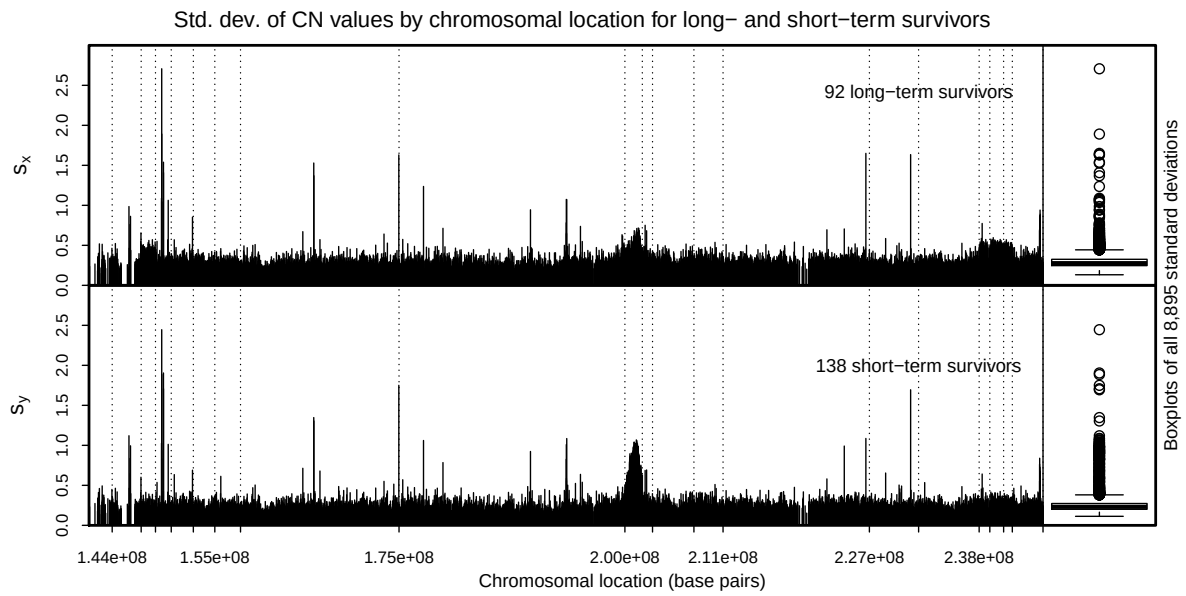

Figure 2.7: Sample standard deviations of copy number at all 8,894 copy number locations for long- and short-term survivors with boxplots at right. Gaps occur at chromosomal locations where no copy number measurements were taken. Vertical dashed lines delineate the twenty CBS-selected regions in which the equal means hypothesis was tested.

variances between the two samples appear to have attenuated the power of the Ch-Q and SK tests just as in the simulation.

None of the univariate two-sample t-statistics in the lefthand panel of Figure 2.6 are very extreme, the largest of their magnitudes being 3.607. This suggests that the difference between the copy number profiles of short- and long-term survivors consists of smaller differences distributed over a larger number of components rather than larger differences over a smaller number of components. That is, the signals appear to be dense but weak rather than sparse but strong. In such a setting the CLX test will likely have low power.

It is worth discussing the computation time of the four tests. For this analysis, in which each test was implemented twenty times at various values of the dimension p, the moderate-p GCT finished in 1.75 seconds and the large-p GCT finished in 6.60 seconds. The Ch-Q and SK tests finished in 2.32 and 2.68 minutes, respectively, and

the CLX took 2.79 hours to run on a MacBook Air with a 1.86 GHz Intel Core 2 Duo processor with 4 GB of memory. The SK procedure involves matrix operations which can be quite slow for large p, and the Ch-Q test involves a cross-validation type sum of inner products which becomes slow for large sample sizes. The CLX method must first test whether $\Sigma_1 = \Sigma_2$ and then directly estimate Σ^{-1} or $\{\Sigma_1 + (n/m)\Sigma_2\}^{-1}$ under sparsity assumptions. Estimating these large matrices quickly becomes computationally burdensome. The GCT requires only a summation over p components and computation of the sample autocovariance function of a p-length series, making it very fast to compute.

2.6 Mitochondrial Calcium Concentration

Ruiz-Meana et al. (2003) subjected cells from cardiac muscles in mice to conditions which simulated reduced blood flow for a period of one hour. To a treatment group, a dose of cariporide was administered, which is believed to inhibit cell death due to oxidative stress. The investigators measured the mitochondrial concentration of Ca^{2+} every ten seconds during the hour. The experiment was run twice, once on intact cells and once on cells with permeabilized membranes. The data have been made available by Febrero-Bande & Oviedo de la Fuente (2012) in the R package `fda.usc`.

The mean percent increase of the calcium concentration over its initial value for the treatment and control in both the experiments is plotted against time in Figure 2.8, where the sample sizes for each curve are shown. The first 180 seconds of the data are removed, given the erratic behavior of the curves, leaving $p = 342$ time points. The four tests were applied to both the intact and permeabilized data to test for equality between the true treatment and control mean curves. The p-values for the four tests are given in Table 2.2.

For the intact cells, the Ch-Q test and the GCT strongly rejected the null, while the CLX test, after failing to reject equality of the covariance matrices, produced a p-value of 0.086 under the equal covariances assumption, and the SK test failed to reject. For the permeabilized experiment the Ch-Q test and the GCT again strongly rejected the

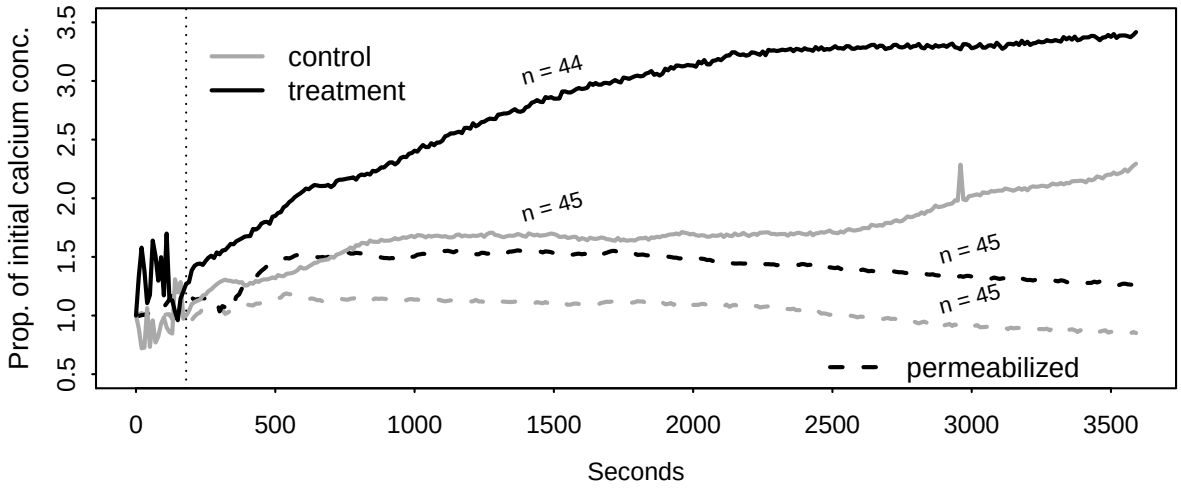

Figure 2.8: Mean curves of the proportional increase in calcium concentration over initial value in intact and permeabilized cells from cardiac muscles in mice over one hour with and without cariporide treatment. First 180 seconds removed from analysis.

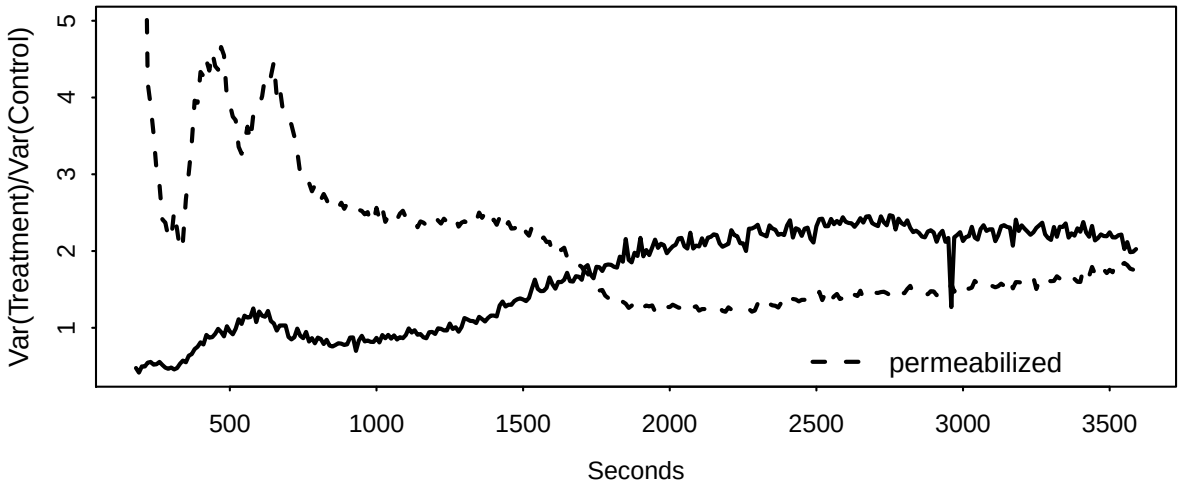

Figure 2.9: Ratios of the variances of the proportional increase in calcium concentration for the treatment versus control group plotted against time for the intact and permeabilized data sets.

	Ch-Q	SK	CLX	mod-p GCT	lg-p GCT
Intact	0.000	0.118	0.086	0.000	0.000
Permeabilized	0.001	0.358	0.817	0.000	0.000

Table 2.2: The p-values produced by the four tests for equality between the treatment and control calcium concentration curves in the intact and permeabilized experiments.

null. The CLX test again failed to reject equality of the covariance matrices, which is a dubious assumption for either the intact or permeabilized experiments given the plot in Figure 2.9 of s_j^2/ϑ_j^2 for $j = 1, \ldots, 342$ for each set of data. In this plot the variance of the treatment group measurements for the intact cells is well over twice as high as in the control group for the first ten minutes (fluctuating wildly), and for the permeabilized cells the variance of the treatment group measurements remains at roughly twice that of the control group measurements after half an hour has elapsed. The low power of the SK test apparently owes to the variance inequality depicted here.

The inability of the CLX test to reject what appears to be an implausible null hypothesis likely owes to a difference in mean functions which is characterized by gradual separation rather than by spikes in one function or the other. The large number of small differences are unable to produce a maximum which will exceed the CLX rejection threshold. However, the Ch-Q test and the GCT are able to register the large number of small differences cumulatively and reject the equal means hypothesis.

This example illustrates the applicability of our test in functional data contexts, in which each observation consists of a function observed at points over some domain. When it is of interest to compare the mean functions in two populations, the assumptions of the GCT are likely to apply.

2.7 Conclusions

The test we present for $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, called the *generalized component test*, was shown to be competitive in the $p \gg n$ setting when the p components admit an ordering allowing the dependence between two components to be modeled according to their displacement. Moderate- and large-p versions of the test were given for $p = o(n^2)$ and $p = o(n^6)$, respectively. The test requires very little computation time and is easily scalable to very-large p settings.

The moderate-p version of our test is robust to ultra heavy-tailedness, and both the moderate- and large-p versions are robust to heteroscedasticity and highly unequal

covariance matrices. The Chen and Qin (Ch-Q) test lost most of its power in the presence of heavy-tailedness or heteroscedasticity; the Srivastava and Kubokawa (SK) test lost much of its power when the covariance matrices were unequally scaled. The Cai, Liu, and Xia (CLX) test performed well under a variety of settings, proving to be robust to heteroscedasticity and to unequally scaled covariance matrices; however, when the data were very heavy-tailed, which rendered the signals very weak, the CLX lost considerable power. Also, since the CLX test requires estimating the $p \times p$ precision matrix, it is computationally much slower than the other tests, requiring over 2.5 hours to complete the copy number data analysis which the SK and Ch-Q tests completed in under 3 minutes and the GCT in under 10 seconds.

For the copy number analysis, the GCT exhibited superior power over the other three tests. This was likely due to heteroscedasticity in the component variances, under which the Ch-Q would lose power, unequally scaled variances between the two populations, under which the SK test would lose power, and likely to the presence of a dense-but-weak rather than a sparse-but-strong signal structure, under which the CLX test would have low power.

For the mitochondrial calcium concentration data set, only the Ch-Q test and the GCT were able to reject the equal means hypothesis. The SK test appears to have lost power due to unequal variances and the CLX supremum-based test was unable to detect the smooth separation of the two mean functions over time, which was characterized by small differences in many components rather than by large differences in a few.

2.7.1 Software

We created the R package `highD2pop` for implementing the GCT as well as the Ch-Q, SK, and CLX tests. A source version, `http://www.stat.tamu.edu/~kbgregory/Research/highD2pop/highD2pop.zip`, is available for download. The package includes copy number data for the CBS-selected segment of the q arm of chromosome 1 having $p = 400$ copy number probes. See package documentation in `http://www.stat.tamu.`

edu/~kbgregory/Research/highD2pop/highD2pop-manual.pdf.

3. FALSE DISCOVERY RATE CONTROL FOR SERIALLY DEPENDENT TEST STATISTICS

3.1 Introduction

Suppose it is of interest to test each of the hypotheses $H_1, \ldots, H_N$ with the test statistics $Z_1, \ldots, Z_N$. Let "null" refer to the state in which a hypothesis is true and "non-null" to the state in which it is false, and let the distribution of Z_i when H_i is null be known. Then if $Z_1, \ldots, Z_N$ are independent, the false discovery rate (FDR), the rate at which null hypotheses are rejected, can be controlled by choosing the critical region with the Benjamini & Hochberg (1995) procedure, hereafter called the BH procedure. The BH procedure was quickly adopted because of its simplicity and the cogency of its authors' arguments for controlling the FDR rather than the familywise error rate—the probability that *any* null hypotheses will be rejected—in large multiple testing scenarios. Reservations arose, however, around the independence assumption under which the BH procedure was developed. Benjamini & Yekutieli (2001) allayed some of this concern by showing that the BH procedure still controlled the FDR if the dependence among $Z_1, \ldots, Z_N$ satisfied the conditions of positive regression dependence, which they argued would be true in many settings. Nevertheless, the problem of accounting for dependence in multiple testing has received unwavering attention between then and now. Dependence among $Z_1, \ldots, Z_N$, it is reasoned, may have such an effect that the extremity of a test statistic will be significantly different when viewed conditionally rather than marginally. These effects may substantially reduce power, even in cases where the dependence structure does not threaten FDR control.

In this paper we shall be concerned with accounting for dependence among a set of test statistics $Z_1, \ldots, Z_N$ when they admit an ordering in some index such as time, from which they inherit a serial dependence structure. For the setting in which the sequence

$\{Z_t\}_{t\geq 1}$ is influenced by latent periodic components, we propose a procedure adapted from the Fan et al. (2012) method (FHG method) for removing the periodic components from $Z_1, \ldots, Z_N$ prior to carrying out the BH procedure. We also further develop the theory for the FHG dependence-adjusted procedure, proving that it increases the power of the BH procedure under some conditions. This result applies readily to the time series context with which we are concerned.

Section 3.2 introduces the FHG method and our frequency domain adaptation for the time series context. Section 3.3 offers a characterization of power for multiple tests of hypotheses and a reformulation of the BH critical region which is useful for power calculations. A heuristic explanation for the increased power of the BH procedure under factor-adjusted test statistics is also given as a prelude to the main result. Section 3.4 gives a theoretical result relating the gains in power from factor adjustment to the variances of the latent factors and the loadings of non-null test statistics upon them. Section 3.5 describes two simulation studies which support the main result of improved power from factor adjustment. Section 3.6 applies our adaptation of the FHG method to an analysis of differences in mean copy numbers along a chromosome between two groups of patients. Section 3.7 offers concluding remarks.

3.2 Methods

As in the introduction, suppose we are interested in testing each of the hypotheses $H_1, \ldots, H_N$, for which we observe the test statistics $Z_1, \ldots, Z_N$, and let "null" refer to the state in which a hypothesis is true and "non-null" to the state in which it is false. In many settings it may be reasonable to assume that

$$Z_i = \begin{cases} e_i & H_i \text{ null} \\ \delta_i + e_i & H_i \text{ non-null} \end{cases}$$

for $i = 1, \ldots, N$, where $\mathbf{e} = (e_1, \ldots, e_N)'$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_e$ with unit diagonal entries. If $\mathbf{\Sigma}_e$ is not a diagonal matrix, i.e. if $e_1, \ldots, e_N$ are correlated, a large value of Z_i may result from a dependence-induced larger-than-usual value of e_i rather than from a nonzero value of δ_i (non-nullity of H_i). A small value of Z_i may likewise result from the dependence among $e_1, \ldots, e_N$ rather than from the nullity of H_i.

However, if new test statistics $\tilde{Z}_i$, $i = 1, \ldots, N$ could be defined such that

$$\tilde{Z}_i \;=\; \begin{cases} K_i & H_i \text{ null} \\ \delta_i + K_i & H_i \text{ non-null,} \end{cases}$$

where $(K_1, \ldots, K_N)' \sim \text{Normal}(\mathbf{0}, \mathbf{D})$, where $\mathbf{D}$ is a diagonal matrix, standard FDR procedures (for independent tests of hypotheses) could be carried out on $\tilde{Z}_1, \ldots, \tilde{Z}_N$, the magnitudes of which would carry direct information concerning $H_1, \ldots, H_N$.

3.2.1 The Fan et al. (2012) Factor Model Approach

Fan et al. (2012) assumed a known covariance matrix for $\mathbf{e} = (e_1, \ldots, e_N)'$ with unit diagonals. If a decomposition $\mathbf{\Sigma}_e = \mathbf{L}\mathbf{\Delta}_m\mathbf{L}' + \mathbf{D}$ exists for which the matrix $\mathbf{L}$ has dimension $N \times m$ with $m \ll N$, $\mathbf{\Delta}_m$ is a $m \times m$ diagonal matrix, and $\mathbf{D}$ is diagonal, then each error term e_i can be expressed in the form of the factor model

$$e_i = \ell_{i1} f_1 + \cdots + \ell_{im} f_m + K_i = \boldsymbol{\ell}_i' \boldsymbol{f} + K_i, \tag{3.1}$$

where $\boldsymbol{\ell}_i'$ is the ith row of the matrix $\mathbf{L}$, and $\boldsymbol{f} = (f_1, \ldots, f_m)' \sim \text{Normal}(\mathbf{0}, \mathbf{\Delta}_m)$ independently of $(K_1, \ldots, K_n)' \sim \text{Normal}(\mathbf{0}, \mathbf{D})$. Then the test statistics $Z_1, \ldots, Z_N$ can be written as

$$Z_i \;=\; \begin{cases} \boldsymbol{\ell}_i' \boldsymbol{f} + K_i & H_i \text{ null} \\ \delta_i + \boldsymbol{\ell}_i' \boldsymbol{f} + K_i & H_i \text{ non-null.} \end{cases}$$

Suppose we can identify a subset $I_0 \subset \{1, \ldots, N\}$ of indices such that we are reasonably confident that H_i is null for $i \in I_0$ (One choice of I_0 could be the set of indices corresponding to the smallest 80%, say, of $Z_1, \ldots, Z_N$). Then it may be assumed that

$$Z_i = \boldsymbol{\ell}_i' \boldsymbol{f} + K_i \text{ for } i \in I_0. \tag{3.2}$$

The supposition that we observe $Z_i = \boldsymbol{\ell}_i' \boldsymbol{f} + K_i$ for $i \in I_0$ allows us to estimate the realized values of the latent factors $f_1, \ldots, f_m$ which have given rise to $Z_1, \ldots, Z_N$. Fan et al. (2012) obtain $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_N$ through spectral decomposition of the (assumed-to-be) known covariance matrix $\boldsymbol{\Sigma}_e$ and $\hat{\boldsymbol{f}} = (\hat{f}_1, \ldots, \hat{f}_m)'$ with regression, recommending a robust method of regression which will be less sensitive to a poor choice of I_0. Then the new uncorrelated test statistics are defined as

$$\tilde{Z}_i = (Z_i - \boldsymbol{\ell}_i' \hat{\boldsymbol{f}})(1 - \boldsymbol{\ell}_i' \boldsymbol{\Delta}_m \boldsymbol{\ell}_i)^{-1/2}, \tag{3.3}$$

where the rescaling comes from the fact that $\mathrm{Var}(Z_i) = \mathrm{Var}(\boldsymbol{\ell}_i' \boldsymbol{f} + K_i) = 1$, so that $\mathrm{Var}(K_i) = 1 - \boldsymbol{\ell}_i' \boldsymbol{\Delta}_m \boldsymbol{\ell}_i$.

3.2.2 A Remark on Strategy

In order to replace the dependent $e_1, \ldots, e_N$ random variables with uncorrelated random variables $K_1, \ldots, K_N$, we must remove from each e_i the parts which it has in common with the others, leaving only the innovative component. The principal obstacle to parsing each e_i into an innovation and a non-innovation is that we only observe e_i directly where H_i is null. For H_i non-null, we observe $Z_i = \delta_i + e_i$. If $\mathbf{e} = (\mathbf{e}_0', \mathbf{e}_1')'$, where $\mathbf{e}_0$ contains e_i for H_i null and $\mathbf{e}_1$ contains e_i for H_i non-null, we must estimate the non-innovative component of each entry in $\mathbf{e}_1$ using only the entries in $\mathbf{e}_0$, which we can only observe inasmuch as we can identify a set of indices I_0 such that $Z_i = e_i$ for $i \in I_0$.

3.2.3 Decomposition of Serially Dependent Errors

If the test statistics $Z_1, \ldots, Z_N$ are serially observed, as along a chromosome or in a time series, the error terms $e_1, \ldots, e_N$ may admit of a decomposition into sums of sinusoids such that for some choice of $m \ll N$,

$$e_t = \sum_{j=1}^{m} \{a_j \cos(\omega_j t) + b_j \sin(\omega_j t)\} + K_t,$$

where a_j and b_j are independent Normal$(0, \sigma_j^2)$ random variables for $j = 1, \ldots, m$, independent of $(K_1, \ldots, K_N) \sim$ Normal$(0, \mathbf{D})$, where $\mathbf{D}$ is a diagonal matrix, as before. If the error terms $e_1, \ldots, e_N$ possess such a structure, the test statistics will rise and fall artificially according to the activity of these latent periodic components, and the signals δ_t will be harder to detect with accuracy.

Assuming that the spectral density of $\{e_t\}_{t \geq 1}$ is known and that dominant frequencies $\omega_1, \ldots, \omega_m$ are readily identified, we may write

$$Z_t \quad = \quad \begin{cases} \sum_{j=1}^{m} \{a_j \cos(\omega_j t) + b_j \sin(\omega_j t)\} + K_t & H_t \text{ null} \\ \delta_t + \sum_{j=1}^{m} \{a_j \cos(\omega_j t) + b_j \sin(\omega_j t)\} + K_t & H_t \text{ non-null.} \end{cases}$$

Then the realized values of the random coefficients $a_1, \ldots, a_m$ and $b_1, \ldots, b_m$ may be estimated through fitting the regression

$$Z_t = \boldsymbol{x}_t' \boldsymbol{\beta} + K_t \text{ for } t \in I_0, \tag{3.4}$$

where $\boldsymbol{x}_t = \{\cos(\omega_1 t), \ldots, \cos(\omega_m t), \sin(\omega_1 t), \ldots, \sin(\omega_m t)\}'$, $\boldsymbol{\beta} = (a_1, \ldots, a_m, b_1, \ldots, b_m)'$. In agreement with Fan et al. (2012), an robust regression method is preferred for estimating the components of $\boldsymbol{\beta}$, as it will be less sensitive to poor choices of I_0.

Having obtained $\hat{\boldsymbol{\beta}} = (\hat{a}_1, \ldots, \hat{a}_m, \hat{b}_1, \ldots, \hat{b}_m)'$, new test statistics can be defined as

$$\tilde{Z}_t = (Z_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}})(1 - \textstyle\sum_{j=1}^{m} \sigma_j^2)^{-1/2}, \tag{3.5}$$

where σ_j^2, $j = 1, \ldots, m$ are the variances of the random coefficients of the m harmonic components retained. This scaling results from the fact that the variance of Z_t is equal to 1 for $t = 1, \ldots, N$, so that

$$\begin{aligned} 1 &= \mathrm{Var}(\boldsymbol{x}_t'\boldsymbol{\beta} + K_t) \\ &= \boldsymbol{x}_t'\mathrm{diag}(\sigma_1^2, \ldots \sigma_m^2, \sigma_1^2, \ldots, \sigma_m^2)\boldsymbol{x}_t + \mathrm{Var}(K_t) \\ &= \textstyle\sum_{j=1}^{m} \sigma_j^2 + \mathrm{Var}(K_t). \end{aligned}$$

3.2.4 Defining Factors from Data

In practice the covariance matrix or the spectral density of the test statistics will not be known, and must be estimated from data. How $Z_1, \ldots, Z_N$ inherit dependence from the data will depend on the context.

Fan et al. (2012) originally considered a regression setting in which $H_1, \ldots, H_N$ were zero-slope hypotheses for N candidate predictors. If $X_1, \ldots, X_N$ are the predictors and Y the response, then fitting N simple linear regression models according to $Y = \beta_i X_i + \epsilon_i$ results in the fitted values $\hat{\beta}_1, \ldots, \hat{\beta}_N$. A z-score for each $\hat{\beta}_i$ is $Z_i = \hat{\beta}_i\{\sigma/(\sqrt{n}s_{ii})\}^{-1}$, where n is the sample size, s_{ii} is the sample standard deviation of X_i, and σ is the standard deviation of ϵ_i. For N simple linear regressions, the covariance matrix of $(Z_1, \ldots, Z_N)'$ is equal to the correlation matrix of $(X_1, \ldots, X_N)'$.

Here we are interested in the two-sample problem in which $H_1, \ldots, H_N$ are equal-means hypotheses for N variables and $Z_1, \ldots, Z_N$ are two-sample t statistics (we assume that sample sizes are large enough to treat the two-sample t statistics as normal) where $t_i = (\bar{X}_{i\cdot} - \bar{Y}_{i\cdot})(s_i^2/n_1 + \vartheta_i^2/n_2)^{-1/2}$ for $i = 1, \ldots, N$. If $\boldsymbol{\Sigma}_X$ and $\boldsymbol{\Sigma}_Y$ are the covariance

matrices for the two populations, then

$$\mathbf{\Sigma}_Z \equiv \mathrm{Cov}\{(Z_1, \ldots, Z_N)'\} = D^{-1/2}(\mathbf{\Sigma}_X/n_1 + \mathbf{\Sigma}_Y/n_2)D^{-1/2},$$

where $D = \mathrm{diag}(\mathbf{\Sigma}_X/n_1 + \mathbf{\Sigma}_Y/n_2)$.

When the structure of $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$ is not known and $N \gg n_1, n_2$, estimates of $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$ are likely to be poor. However, in the context of serially dependent data, it may be reasonable to assume a Toeplitz structure for $\mathbf{\Sigma}_X$ and $\mathbf{\Sigma}_Y$. In this case, an unbiased estimator $\hat{\mathbf{\Sigma}}_X^{(T)}$ of $\mathbf{\Sigma}_X$ may be obtained by averaging the diagonals of each order of the sample covariance matrix $\mathbf{S}_X$ such that entry (i, j) of $\hat{\mathbf{\Sigma}}_X^{(T)}$ is given by $\hat{\mathbf{\Sigma}}_X^{(T)}(i, j) = (N - |i - j|)^{-1} \sum_{|l-k|=|i-j|} \mathbf{S}_X(l, k)$, as in Cai et al. (2013b).

A factor-adjustment of the test statistics $Z_1, \ldots, Z_N$ may now be carried out in two ways: By defining factors from the principal components of $\hat{\mathbf{\Sigma}}_Z^{(T)}$ and proceeding as in Fan et al. (2012), or by using $\hat{\gamma}_Z(k) \equiv \hat{\mathbf{\Sigma}}_Z^{(T)}(1, 1 + k)$, $k = 0, 1, \ldots, N - 1$, to estimate the spectral density of $\{Z_t\}_{t \geq 1}$ and then performing the harmonic factor adjustment described in Section 3.2.3.

3.2.5 Choosing the Number of Factors

Choosing the number of factors in a factor model or the number of frequencies in a harmonic decomposition of a time series are long-standing questions with which we are not primarily concerned here, though we give some guidelines. If using factors defined by the spectral decomposition of $\hat{\Sigma}_Z$, the appropriate number of factors to retain may be discerned from a plot of the eigenvalues ordered from largest to smallest. If using harmonic factors, the number of factors and the frequencies to which they correspond may be discerned from the periodogram or from a smoothed estimate of the spectral density. Factors corresponding to frequencies at which spikes occur in the spectral density should be retained.

If it is desired to retain a certain proportion ξ of the "total variability", one may

choose the number of factors m such that

$$m = \min\{k : (\hat{\lambda}_1^2 + \cdots + \hat{\lambda}_k^2)/\textstyle\sum_{j=1}^{N} \hat{\lambda}_j^2 \geq \xi\}, \tag{3.6}$$

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_N$ are the eigenvalues of the sample-covariance or Toeplitz estimate of $\mathbf{\Sigma}_Z$ ordered from largest to smallest, or in the harmonic case,

$$m = \min[k : \{\hat{f}^2(\omega_{(1)}) + \cdots + \hat{f}^2(\omega_{(k)})\}/\textstyle\sum_{j=1}^{N} \hat{f}(\omega_{(j)}) \geq \xi], \tag{3.7}$$

where $\hat{f}(\cdot)$ is the estimated spectral density of $\{Z_t\}_{t\geq 1}$ and $\omega_{(j)}$ is the frequency at which $\hat{f}(\cdot)$ is the jth largest.

3.3 Power Gains from Removing Factor Effects

Removing factor effects from the test statistics in the manner described can result in increased power when the BH procedure is applied to the adjusted test statistics. This section introduces a characterization of power for multiple testing procedures and then presents a heuristic explanation for why it increases when factor effects are removed. A theorem appears in Section 3.4 which relates the power gains to the loadings of the non-null test statistics upon the factors and to the factor variances.

3.3.1 A Characterization of Power over Multiple Tests

Of the hypotheses $H_1, \ldots, H_N$, let $I_0 \subset \{1, \ldots, N\}$ be the set of indices corresponding to null hypotheses and $I_1 = \{1, \ldots, N\} \setminus I_0$ be the set of indices correponding to non-null hypotheses. When considering the power of a multiple testing procedure for $H_1, \ldots, H_N$, we may characterize it as the expected proportion of non-nulls rejected as a function of some rejection threshold and the overall non-null state.

If a multiple testing procedure rejects H_i when the corresponding test statistic Z_i falls into a critical region $\mathcal{C}_z$ for $i = 1, \ldots, N$, then letting N_1 be the total number of non-null

hypotheses, we can express the power of the procedure as

$$P(\mathcal{C}_z) = E\{\textstyle\sum_{i\in I_1} \mathbb{I}(Z_i \in \mathcal{C}_z)\}/N_1, \tag{3.8}$$

where $\mathbb{I}(\cdot)$ is the indicator function and the expectation is taken with respect to the non-null distributions of the Z_i for $i \in I_1$. The critical region $\mathcal{C}_z$ is typically of one of the forms $\mathcal{C}_z = (-\infty, z]$, $\mathcal{C}_z = [z, \infty)$ and $\mathcal{C}_z = \{(\infty, -|z|] \cup [|z|, \infty)\}$. Power increases as the rejection region $\mathcal{C}_z$ is made larger.

The BH procedure chooses the rejection region as a function of a user-specified q, the level at which it is desired to control the FDR, and the observed $Z_1, \ldots, Z_N$, such that

$$\mathcal{C}_z(q, Z_1, \ldots, Z_N) = \sup\{\mathcal{C}_z : N\Phi(\mathcal{C}_z)/\textstyle\sum_{i=1}^{N} \mathbb{I}(Z_i \in \mathcal{C}_z) \leq q\}, \tag{3.9}$$

where $\Phi(\mathcal{C}_z)$ is the probability mass conferred to $\mathcal{C}_z$ by the null distribution $\Phi(\cdot)$ of the test statistics, which is assumed to be common to all Z_i, $i \in I_0$.

Without loss of generality, assume that $\mathcal{C}_z$ is of the form $\mathcal{C}_z = [z, \infty)$, corresponding to one-sided hypotheses against which there is greater evidence as $Z_1, \ldots, Z_N$ assume greater positive values. When $\mathcal{C}_z$ is of this form, $Z_i \in \mathcal{C}_z \iff Z_i \geq z$. If $z_\alpha = \Phi^{-1}(1-\alpha)$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function for Z_i, $i \in I_0$, then the BH choice of $\mathcal{C}_z$ becomes $\mathcal{C}_z = [z_{\alpha(q,Z_1,\ldots,Z_n)}, \infty)$, where

$$\alpha(q, Z_1, \ldots, Z_N) = \sup\{\alpha : N\alpha/\textstyle\sum_{i=1}^{N} \mathbb{I}(Z_i \geq z_\alpha) \leq q\}. \tag{3.10}$$

The power of the BH procedure may then be expressed (combining (3.8), (3.9), and (3.10)) as

$$P(q) = E\{\textstyle\sum_{i\in I_1} \mathbb{I}(Z_i \geq Z_{\alpha(q,Z_1,\ldots,Z_N)})\}/N_1, \tag{3.11}$$

where the power $P(\cdot)$ is now a function of the choice of FDR bound q.

3.3.2 Power of the BH Procedure Under Factor Model Assumptions

Suppose that the test statistics $Z_1, \ldots, Z_N$ are such that

$$Z_i \;=\; \begin{cases} \boldsymbol{\ell}_i'\boldsymbol{f} + K_i & i \in I_0 \\ \delta_i + \boldsymbol{\ell}_i'\boldsymbol{f} + K_i & i \in I_1, \end{cases}$$

where $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_N$ are known $m \times 1$ vectors and $\boldsymbol{f} = (f_1, \ldots, f_m)' \sim \text{Normal}(0, \boldsymbol{\Delta}_m)$ independently of $(K_1, \ldots, K_N)' \sim \text{Normal}(0, \mathbf{D})$, where $\boldsymbol{\Delta}_m$ and $\mathbf{D}$ are diagonal matrices.

Letting $e_i = \boldsymbol{\ell}_i'\boldsymbol{f} + K_i$ for $i = 1, \ldots, N$, the power of the BH procedure can be expressed as

$$P(q) = E\{\textstyle\sum_{i \in I_1} \mathbb{I}(\delta_i + e_i \geq z_{\alpha(q, Z_1, \ldots, Z_N)})\}/N_1, \tag{3.12}$$

where

$$\alpha(q, Z_1, \ldots, Z_N) = \sup\left[\alpha : N\alpha / \{\textstyle\sum_{i \in I_0} \mathbb{I}(e_i \geq z_\alpha) + \sum_{i \in I_1} \mathbb{I}(\delta_i + e_i \geq z_\alpha)\} \leq q\right]. \tag{3.13}$$

Now suppose new test statistics $\tilde{Z}_1, \ldots, \tilde{Z}_N$ are defined as in (3.3). Then

$$Z_i \;=\; \begin{cases} \tilde{K}_i & i \in I_0 \\ \tilde{\delta}_i + \tilde{K}_i & i \in I_1, \end{cases}$$

where $\tilde{\delta}_i = \delta_i(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2}$ and $\tilde{K}_i = K_i(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2}$ with $\text{Var}(\tilde{K}_i) = 1$. Now the power of the BH procedure on the new test statistics $\tilde{Z}_1, \ldots, \tilde{Z}_N$ can be expressed as

$$P_{FHG}(q) = E\{\textstyle\sum_{i \in I_1} \mathbb{I}(\tilde{\delta}_i + \tilde{K}_i \geq z_{\alpha(q, \tilde{Z}_1, \ldots, \tilde{Z}_N)})\}/N_1, \tag{3.14}$$

where

$$\tilde{\alpha}(q, \tilde{Z}_1, \ldots, \tilde{Z}_N) = \sup\left[\alpha : N\alpha / \{\textstyle\sum_{i\in I_0} \mathbb{I}(\tilde{K}_i \geq z_\alpha) + \sum_{i\in I_1} \mathbb{I}(\tilde{\delta}_i + \tilde{K}_i \geq z_\alpha)\} \leq q\right]. \tag{3.15}$$

3.3.3 Effect of Factor Adjustment on Power

In order to compare $P(q)$ and $P_{FHG}(q)$, we first observe that since $(1 - \boldsymbol{\ell}_i' \boldsymbol{\Delta}_m \boldsymbol{\ell}_i) \leq 1$, the effect size component of $\tilde{Z}_i$ will be $\tilde{\delta}_i = \delta_i(1 - \boldsymbol{\ell}_i' \boldsymbol{\Delta}_m \boldsymbol{\ell}_i)^{-1/2} \geq \delta_i$, so that the signal in $P_{FHG}(q)$ will be boosted by the removal of factor effects and subsequent rescaling. Because of the increased signal size and since $\mathrm{Var}(e_i) = \mathrm{Var}(\tilde{K}_i)$, the adjusted test statistic $\tilde{Z}_i = \tilde{\delta}_i + \tilde{K}_i$ will more often exceed a fixed threshold than its unadjusted counterpart $Z_i = \delta_i + e_i$. Secondly, $\alpha(q, Z_1, \ldots, Z_N)$ will tend to be smaller than $\tilde{\alpha}(q, \tilde{Z}_1, \ldots, \tilde{Z}_N)$, also owing to the rescaling of the signal, as the denominator inside the supremum of (3.15) will tend to be larger than that of (3.13). Thus $z_{\tilde{\alpha}(q,\tilde{Z}_1,\ldots,\tilde{Z}_N)}$ will tend to be smaller than $z_{\alpha(q,Z_1,\ldots,Z_N)}$, producing a more liberal rejection region. This is made rigorous in the next section.

3.4 Main Results

The BH critical region is found by choosing an FDR bound q and then finding the largest critical region of which q will admit according to expression (3.9). To establish power results, it will be more convenient to consider the smallest value of q which will admit of a given critical region: Fix a size α of the critical region and define the corresponding BH false discovery rate bound as

$$q(\alpha, Z_1, \ldots, Z_N) = N\Phi(\mathcal{C}_{z_\alpha}) / \textstyle\sum_{i=1}^{N} \mathbb{I}(Z_i \in \mathcal{C}_{z_\alpha}). \tag{3.16}$$

which is the same as the q value introduced by Storey (2002). This is the lowest FDR bound for which the BH procedure would reject a hypothesis with a test statistic equal to z_α. If for a fixed size α of the critical region, the BH procedure produces a smaller

value of q on the factor-adjusted test statistics $\tilde{Z}_1, \ldots, \tilde{Z}_N$ than on the unadjusted test statistics $Z_1, \ldots, Z_N$, then it follows that at a fixed level of q, the BH critical region defined for $\tilde{Z}_1, \ldots, \tilde{Z}_N$ will be larger than that defined for $Z_1, \ldots, Z_N$. Thus the factor-adjusted test statistics will lead to a more liberal choice of critical region by the BH procedure. Theorem 2 shows that this will occur under the following conditions:

(C.1) Let $H_1, \ldots, H_N$ be two-sided hypotheses and let $I_0 \subset \{1, \ldots, N\}$ be the set of indices for which H_i is null if $i \in I_0$. Let the number of null hypotheses be N_0. Then if $I_1 = \{1, \ldots, N\} \setminus I_0$ has $N_1 = N - N_0$ elements, let $N_0/N \to \pi_0 > 0$ and $N_1/N \to \pi_1 = 1 - \pi_0$ as $N \to \infty$.

(C.2) Let

$$Z_i|\delta_i \;=\; \begin{cases} \boldsymbol{\ell}_i'\boldsymbol{f} + K_i & i \in I_0 \\ \delta_i + \boldsymbol{\ell}_i'\boldsymbol{f} + K_i & i \in I_1, \end{cases}$$

where $\boldsymbol{f} = (f_1, \ldots, f_m)' \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Delta}_m)$ independently of $(K_1, \ldots, K_N)' \sim \text{Normal}(\mathbf{0}, \mathbf{D})$, where $\boldsymbol{\Delta}_m$ and $\mathbf{D} = \text{diag}(d_{11}, \ldots, d_{NN})$ are diagonal matrices such that $\boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i + d_{ii} = 1$ for $i = 1, \ldots, N$.

(C.3) Let $\delta_1, \ldots, \delta_{N_1} \sim \text{Normal}(0, \sigma_\delta^2)$.

Theorem 2 *Under conditions (C.1), (C.2), and (C.3), if new test statistics* $\tilde{Z}_i = (Z_i - \boldsymbol{\ell}_i'\boldsymbol{f})(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2}$ *are defined for* $i = 1, \ldots, N$, *then the BH false discovery rate bound corresponding to the critical region* $C_{z_{\alpha^*}} = \{(-\infty, -|z_{\alpha^*}|] \cup [|z_{\alpha^*}|, \infty)\}$ *defined by* $\tilde{Z}_1, \ldots, \tilde{Z}_N$, *denoted by* $\tilde{q}(\alpha^*, \tilde{Z}_1, \ldots, \tilde{Z}_N)$, *will, as* $N \to \infty$, *be less than or equal to that defined by* $Z_1, \ldots, Z_N$, *denoted by* $q(\alpha^*, Z_1, \ldots, Z_N)$, *such that*

$$\lim_{N\to\infty} \frac{\tilde{q}(\alpha^*, \tilde{Z}_1, \ldots, \tilde{Z}_N)}{q(\alpha^*, Z_1, \ldots, Z_N)} = \frac{\tilde{q}(\alpha^*)}{q(\alpha^*)} \leq Q(\alpha^*, \bar{\Delta}_1^{(\infty)}),$$

where

$$Q(\alpha^*, \bar{\Delta}_1^{(\infty)}) = A(\alpha^*, \sigma_\delta^2)/\{A(\alpha^*, \sigma_\delta^2) + \pi_1 B(\alpha^*, \sigma_\delta^2)\bar{\Delta}_1^{(\infty)}\} \leq 1, \tag{3.17}$$

with

$$\begin{aligned} A(\alpha, \sigma_\delta^2) &= \alpha\pi_0 + \pi_1[1 - \Phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1/2}\}] && (3.18)\\ B(\alpha, \sigma_\delta^2) &= (1/2)(\sigma_\delta^2 + 1)^{-1/2}\phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1/2}\}z_\alpha\{\sigma_\delta^2/(\sigma_\delta^2 + 1)\} && (3.19)\\ \bar{\Delta}_1^{(\infty)} &= \lim_{N\to\infty} N_1^{-1}\textstyle\sum_{i\in I_1} \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i. && (3.20) \end{aligned}$$

Remark 2 *The fact that* $\lim_{N\to\infty} \tilde{q}(\alpha^*, \tilde{Z}_1, \ldots, \tilde{Z}_N)/q(\alpha^*, Z_1, \ldots, Z_N) \leq Q(\alpha^*, \bar{\Delta}_1^{(\infty)}) \leq 1$ *implies that for a fixed false discovery rate bound, the BH procedure on the factor-adjusted test statistics* $\tilde{Z}_1, \ldots, \tilde{Z}_N$ *will choose a critical region of the same size or larger than when carried out on the original test statistics* $Z_1, \ldots, Z_N$*. Furthermore, the difference is a function of the loadings* $\boldsymbol{\ell}_i$*,* $i \in I_1$*, of the non-null test statistics upon the factors* $f_1, \ldots, f_m$ *and of the factor variances contained in* $\boldsymbol{\Delta}_m$*.*

Proof 1 *Under the conditions of the theorem, the BH false discovery rate bound at size* α *of the critical region for the unadjusted test statistics is given by*

$$\begin{aligned} q(\alpha, Z_1, \ldots, Z_N) &= 2\alpha N\{\textstyle\sum_{i=1}^N \mathbb{I}(|Z_i| \geq z_\alpha)\}^{-1} \qquad \textit{(from (3.16))}\\ &= 2\alpha N\{\textstyle\sum_{i\in I_0} \mathbb{I}(|\boldsymbol{\ell}_i'\boldsymbol{f} + K_i| \geq z_\alpha) + \sum_{i\in I_1} \mathbb{I}(|\delta_i + \boldsymbol{\ell}_i'\boldsymbol{f} + K_i| \geq z_\alpha)\}^{-1}\\ \to^p q(\alpha) &\equiv \alpha(\alpha\pi_0 + \pi_1[1 - \Phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1/2}\}])^{-1} \qquad (3.21) \end{aligned}$$

as $N \to \infty$ *since* $\boldsymbol{\ell}_i'\boldsymbol{f} + K_i \sim$ *Normal*$(0, 1)$ *and* $\delta_i + \boldsymbol{\ell}_i'\boldsymbol{f} + K_i \sim$ *Normal*$(0, \sigma_\delta^2 + 1)$ *and by the weak law of large numbers. The adjusted test statistics* $\tilde{Z}_i = (Z_i - \boldsymbol{\ell}_i'\boldsymbol{f})(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2}$

are such that

$$\tilde{Z}_i \;=\; \begin{cases} \tilde{K}_i & i \in I_0 \\ \tilde{\delta}_i + \tilde{K}_i & i \in I_1, \end{cases}$$

where $\tilde{K}_i = K_i(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2} \implies (\tilde{K}_i, \ldots, K_N)' \sim Normal(\mathbf{0}, \boldsymbol{I})$ *and* $\tilde{\delta}_i = \delta_i(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1/2} \implies \delta_i \sim Normal\{0, \sigma_\delta^2(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1}\}$ *for* $i = 1, \ldots, N$.

The BH false discovery rate bound at size α *of the critical region for the factor-adjusted test statistics becomes*

$$\begin{aligned} \tilde{q}(\alpha, \tilde{Z}_1, \ldots, \tilde{Z}_N) &= 2\alpha N\{\textstyle\sum_{i=1}^N \mathbb{I}(|\tilde{Z}_i| \geq z_\alpha)\}^{-1} \\ &= 2\alpha N\{\textstyle\sum_{i\in I_0} \mathbb{I}(|\tilde{K}_i| \geq z_\alpha) + \sum_{i\in I_1} \mathbb{I}(|\tilde{\delta}_i + \tilde{K}_i| \geq z_\alpha)\}^{-1} \\ \to^p \tilde{q}(\alpha) &\equiv \alpha\{\alpha\pi_0 \\ &\quad + \pi_1 \lim_{N\to\infty} N_1^{-1} \textstyle\sum_{i\in I_1} (1 - \Phi[z_\alpha\{\sigma_\delta^2(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1} + 1\}^{-1/2}])\}^{-1}, \end{aligned} \tag{3.22}$$

since $\tilde{\delta}_i + \tilde{K}_i \sim Normal\{0, \sigma_\delta^2(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1} + 1\}$ *for* $i \in I_1$.

By the mean value theorem we can write

$$\begin{aligned} \Phi[z_\alpha\{\sigma_\delta^2(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1} + 1\}^{-1/2}] &= \Phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1/2}\} \\ &\quad + (1/2)(\sigma_\delta^2/c_i + 1)^{-1/2}\phi\{z_\alpha(\sigma_\delta^2/c_i + 1)\}z_\alpha\{\sigma_\delta^2/(\sigma_\delta^2 + c_i)\}(1/c_i)(\boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i) \end{aligned}$$

for some c_i *such that* $1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i \leq c_i \leq 1$. *Since the right hand side of the above equation is increasing in* c_i, *setting* $c_i = 1$ *and subtracting both sides from* 1 *yields the inequality*

$$1 - \Phi[z_\alpha\{\sigma_\delta^2(1 - \boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i)^{-1} + 1\}^{-1/2}] \geq 1 - \Phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1/2}\} + B(\alpha, \sigma_\delta^2)(\boldsymbol{\ell}_i'\boldsymbol{\Delta}_m\boldsymbol{\ell}_i),$$

where $B(\alpha, \sigma_\delta^2)$ is as in (3.19). We may now write that

$$\lim_{N\to\infty} N_1^{-1} \sum_{i\in I_1} 1 - \Phi[z_\alpha \{\sigma_\delta^2 (1 - \boldsymbol{\ell}_i' \boldsymbol{\Delta}_m \boldsymbol{\ell}_i)^{-1} + 1\}^{-1/2}]$$
$$\geq 1 - \Phi\{z_\alpha (\sigma_\delta^2 + 1)^{-1/2}\} + B(\alpha, \sigma_\delta^2) \bar{\Delta}_1^{(\infty)},$$

where $\bar{\Delta}_1^{(\infty)}$ is as in (3.20). Applying this inequality to (3.22), we write

$$\begin{aligned}
\tilde{q}(\alpha) &\leq \alpha\{\alpha\pi_0 + \pi_1[1 - \Phi\{z_\alpha(\sigma_\delta^2 + 1)^{-1}\}] + \pi_1 B(\alpha, \sigma_\delta^2)\bar{\Delta}_m^{(\infty)}\}^{-1} \\
&= \alpha\{A(\alpha, \sigma_\delta^2) + B(\alpha, \sigma_\delta^2)\bar{\Delta}_m^{(\infty)}\}^{-1} \\
&= \alpha A(\alpha, \sigma_\delta^2)^{-1} A(\alpha, \sigma_\delta^2)\{A(\alpha, \sigma_\delta^2) + B(\alpha, \sigma_\delta^2)\bar{\Delta}_m^{(\infty)}\}^{-1} \\
&= q(\alpha) Q(\alpha, \bar{\Delta}_m^{(\infty)}),
\end{aligned}$$

where $A(\alpha, \sigma_\delta^2)$ is as in (3.18), $q(\alpha)$ is as in (3.21) and $Q(\alpha, \bar{\Delta}_m^{(\infty)})$ is as in (3.17). This completes the proof.

3.5 Simulation Studies

3.5.1 Effects of Factor Adjustment on The BH Critical Region

This section describes a simulation study of the effect of factor adjustment on the BH critical region, comparing the effect at $N = 5000$ hypotheses with the limiting effect as $N \to \infty$ given in Theorem 2. Three single-factor models were used to generate sets of test statistics $Z_1, \ldots, Z_N$. The first had compound symmetry dependence among all the test statistics, the second had compound symmetry among only the non-null test statistics, and the third had a single factor upon which the test statistics had sinusoidal loadings.

From each model, 5000 sets of $Z_1, \ldots, Z_N$ were generated with $N = 5000$. The BH two-sided thresholds $z_{\alpha(q, Z_1, \ldots, Z_N)}$ and $z_{\tilde{\alpha}(q, \tilde{Z}_1, \ldots, \tilde{Z}_N)}$ for critical regions of the form $\mathcal{C}_z = \{(-\infty, -|z|] \cup [|z|, \infty)\}$, were found across a range of FDR thresholds q for the original and factor-adjusted test statistics. The average over the 5000 simulated values

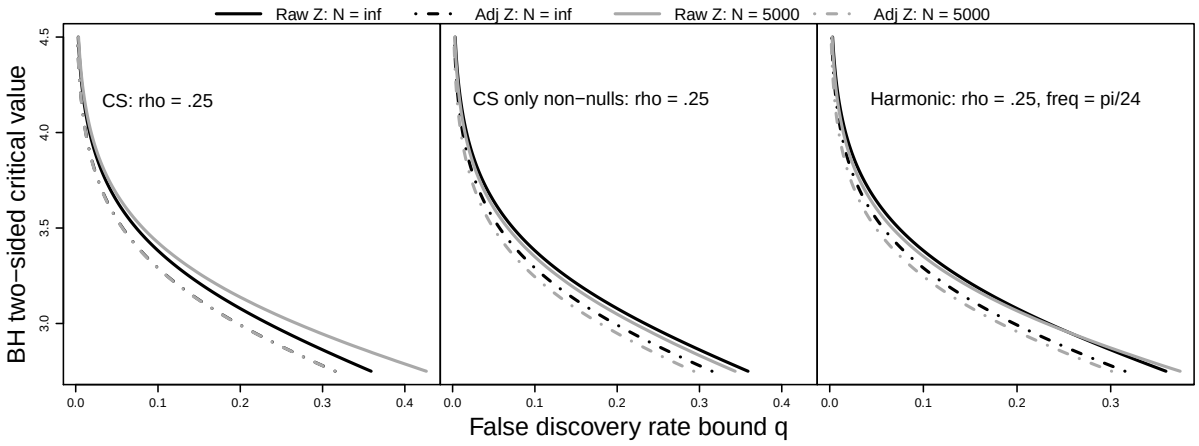

Figure 3.1: Limiting BH-selected two-sided critical values as $N \to \infty$ (in black) as well as when $N = 5000$ (in gray) against the chosen FDR bound q when the BH procedure is carried out on the original and factor-adjusted test statistics.

of the BH two-sided rejection thresholds for the adjusted and unadjusted Z values are plotted as gray lines in Figure 3.1. The black lines are not computed from simulated data but are the limiting values as $N \to \infty$ as derived in Theorem 2. The relationship between the solid (for unadjusted) and dashed (for factor-adjusted) curves for the simulated data mimics that between the solid and dashed curves for the theoretical limit, indicating that the factor adjustment affords gains in power for finite N.

For all three models, the total number of hypotheses was $N = 5000$, the number of non-nulls was $N_1 = 250$, and the signals were generated such that $\delta_1, \ldots, \delta_{N_1} \sim \text{Normal}(0, 4)$. The three models were

(i) $Z_i = \delta_i \mathbb{I}(i \leq 250) + (.25)^{1/2} f + (.75)^{1/2} K_i$

(ii) $Z_i = \delta_i \mathbb{I}(i \leq 250) + (.25)^{1/2} f \mathbb{I}(i \leq 250) + (.75)^{1/2} K_i \mathbb{I}(i \leq 250) + K_i \mathbb{I}(i > 250)$

(iii) $Z_i = \delta_i \mathbb{I}(i \leq 250) + (.25)^{1/2} \{U_1 \cos(\pi i/24) + U_2 \sin(\pi i/24)\} + (.75)^{1/2} K_i,$

where $\delta_i \sim \text{Normal}(0, 4)$, $f, U_1, U_2, K_i \sim \text{Normal}(0, 1)$, all independently of each other for $i = 1, \ldots, 5000$. Note that the variance of K_i is scaled such that the variance of Z_i is equal to 1 for all $i = 1, \ldots, N$.

For model (i), $\bar{\Delta}_1^{(\infty)} = .25$, since the covariance matrix of $(Z_1, \ldots, Z_N)'$ has a single nonzero eigenvalue equal to $(.25)N$ and the eigenvector is $\mathbf{1}_N N^{-1/2}$, so $N_1^{-1} \sum_{i \in I_1} \ell_i' \Delta_m \ell_i = N^{-1/2}\{(.25)N\}N^{-1/2} = .25$. The covariance matrix induced by model (ii) also gives $\bar{\Delta}_1^{(\infty)} = .25$, since it has a single nonzero eigenvalue of $(.25)N_1$ and the single eigenvector has elements equal to $N_1^{-1/2}$ for non-null i and equal to 0 for null i. Thus $N_1^{-1} \sum_{i \in I_1} \ell_i' \Delta_m \ell_i = N_1^{-1/2}\{(.25)N_1\}N_1^{-1/2} = .25$. Model (iii) also has $\bar{\Delta}_1^{(\infty)} = .25$ since $N_1^{-1} \sum_{i \in I_1} \ell_i' \Delta_m \ell_i = N_1^{-1} \sum_{i \in I_1} \{\cos^2(\pi i/24)(.25) + \sin^2(\pi i/24)(.25)\} = .25$.

Models (i) and (ii) differed only in the loadings of the null test statistics on the factors, and in agreement with Theorem 1, the power gains from factor adjustment were very similar across the choices of q for the two models; nor were the power gains from factor adjustment significantly different for model (iii), since it induced the same value of $\bar{\Delta}_1^{(\infty)}$ as the first two models.

3.5.2 Power and FDR Control on Simulated Data Sets

The simulations in this section assess the performance of the factor adjustment for serially dependent test statistics in the two-sample setting where the Z values are the two-sample t-statistics $Z_t \equiv (\bar{X}_{t.} - \bar{Y}_{t.})(s_t^2/n_1 + \vartheta_t^2/n_2)^{-1/2}$, $t = 1, \ldots, N$. We compare the power and the false discovery rate control achieved when the factors are defined by (i) the spectral decomposition of $\hat{\mathbf{\Sigma}}_Z$ when $\hat{\mathbf{\Sigma}}_X = \mathbf{S}_X$ and $\hat{\mathbf{\Sigma}}_Y = \mathbf{S}_Y$, where $\mathbf{S}_X$ and $\mathbf{S}_Y$ are the sample covariance matrices for the two samples, by (ii) the spectral decomposition of $\hat{\mathbf{\Sigma}}_Z^{(T)}$, the Toeplitz estimate of $\mathbf{\Sigma}_Z$ described in Section 3.2.4, and by (iii), a harmonic decomposition as described in Section 3.2.3, where the spectral density of $\{Z_t\}_{t \geq 1}$ is estimated from $\hat{\gamma}_Z(\cdot)$.

Let X_{it} denote measurement t on subject i of the first sample and Y_{jt} denote measurement t on subject j of the second sample. Then the two-sample data were generated according to

$$X_{it} \;=\; \sum_{k=1}^{m} \{U_{1ik} \cos(\omega_k t) + U_{2ik} \sin(\omega_k t)\} + e_{it}$$

$$Y_{jt} \quad = \quad \delta_t + \sum_{k=1}^{m} \{V_{1jk} \cos(\omega_k t) + V_{2jk} \sin(\omega_k t)\} + h_{jt}$$

for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, and $t = 1, \ldots, N$. The signals δ_t were generated such that $\delta_t = u_t \mathbb{I}(|u_t| > c_u)(0.5/c_u)$, where $u_t \sim$ AR(1), with AR parameter $\phi = .8$ and innovations from a t distribution with 10 degrees of freedom, and c_u is the $\lfloor \pi_0 N \rfloor$th largest absolute value of $u_1, \ldots, u_N$. Thus in each simulated data set there is a fixed proportion $\pi_1 = 1 - \pi_0$ of non-nulls, and the magnitude of each non-null signal is at least 0.5. The parameter settings for the two simulations were:

Model 1: $n_1 = 45$, $n_2 = 60$, $N = 1000$; $m = 3$ with $(\omega_1, \omega_2, \omega_3) = (\pi/2, \pi/3, \pi/12)$; $U_{1ik}, U_{2ik}, V_{1jk}, V_{2jk} \sim$ Uniform$(-1, 1)$ for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, $k = 1, 2, 3$; $\{e_{jt}\}_{t=1}^{N}, \{h_{jt}\}_{t=1}^{N} \sim$ ARMA$(1, 1)$, with AR parameter $\phi = 0.4$ and MA parameter $\theta = 0.3$ and Normal$(0, 1)$ innovations; $\pi_1 = 0.10$.

Model 2: $n_1 = 45$, $n_2 = 60$, $N = 1000$; $m = 2$ with $(\omega_1, \omega_2) = (\pi/2, \pi/3)$; $U_{1ik}, U_{2ik}, V_{1jk}, V_{2jk} \sim$ Uniform$(-1, 1)$ for $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, $k = 1, 2$; $\{e_{jt}\}_{t=1}^{N}, \{h_{jt}\}_{t=1}^{N} \sim$ AR(1), with AR parameter $\phi = -0.5$ and Normal$(0, 1)$ innovations; $\pi_1 = 0.10$.

The number of factors to retain was chosen by setting $\xi = 0.80$ in (3.6) for the sample covariance and Toeplitz methods and in (3.7) for the harmonic method. Thus 80% of the variability or spectral mass was retained across the three methods to ensure comparability.

After determining the number of factors and the frequencies or eigenvalues to which they corresponded, the realized values of the factors were fitted with robust regression using the function `rlm()` from the R package `MASS` (Venables & Ripley (2002)) under default settings on the middle 80% of $Z_1, \ldots, Z_N$, and the factor adjustments in (3.3) and (3.5) were carried out. The BH procedure is applied to the three sets of adjusted Z values (corresponding to the sample-covariance estimate of $\mathbf{\Sigma}_Z$, the Toeplitz estimate

of $\mathbf{\Sigma}_Z$, and the harmonic decomposition method using the estimated spectral density of $\{Z_t\}_{t=1}^N$) as well as on the unadjusted Z values.

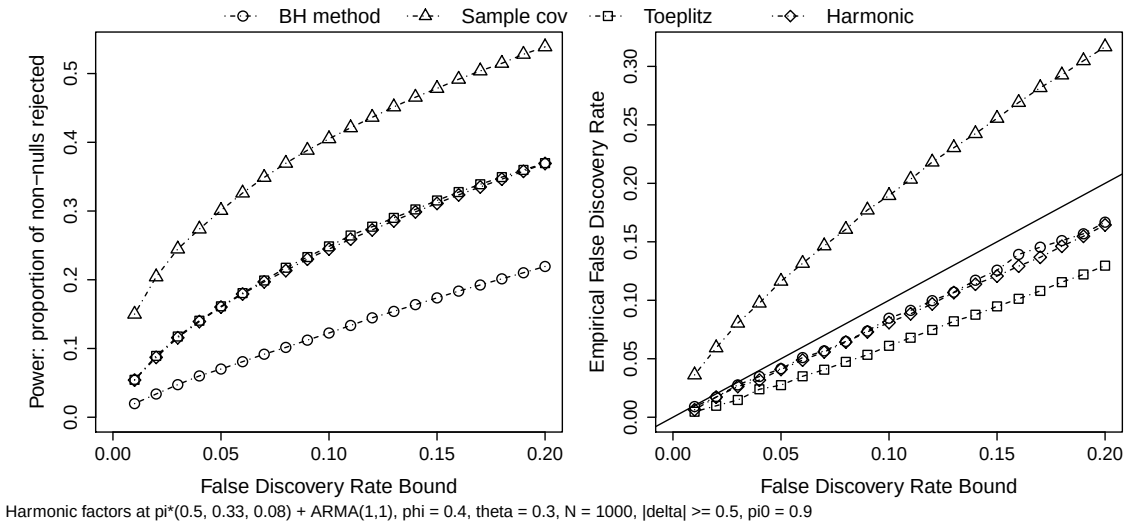

Figure 3.2: Left: Proportion of non-nulls rejected against chosen FDR bound averaged over 500 simulation runs for BH procedure on original Z values and adjusted Z values from the sample covariance, Toeplitz, and harmonic factor adjustments. Right: Simulated FDR against chosen FDR bound.

Power and FDR control results across 500 simulated data sets for Model 1 appear in Figure 3.2. The left hand panel plots the power—the proportion of non-nulls rejected—against the user-specified FDR bound q. The sample covariance, Toeplitz, and harmonic adjustments to the test statistics all result in a substantial increase in power over the BH procedure applied to the unadjusted test statistics—the sample covariance method rejecting by far the most non-nulls as q is increased. In the right-hand panel, the simulated or empirical FDR is plotted against the chosen FDR bound q, and the sample covariance method results in false discovery rates which are far above the chosen bound, its curve lying far above the 45% line. The other three methods keep the FDR below the chosen threshold q, the Toeplitz method doing so most convervatively. Thus the Toeplitz and harmonic methods of test-statistic adjustment increase the power of the BH procedure

substantially without compromising FDR control.

Figure 3.3 exhibits similar behavior for the sample covariance adjustment versus the Toeplitz and harmonic factor adjustments of the test statistics. Much power is gained by the latter two procedures, under which the FDR is still well controlled. In this simulation, the Toeplitz factor adjustment achieved somewhat greater power across the choices of q than the harmonic factor adjustment.

Since the simulated data do not come from a strict factor model—meaning that the harmonic factors do not account for all of the dependence—the harmonic factor adjustment may be disadvantaged by the rigidity of its sinusoidal factor definitions. The Toeplitz factors are more flexible and thus are probably able to capture some of the dependence of the autoregressive errors in the model.

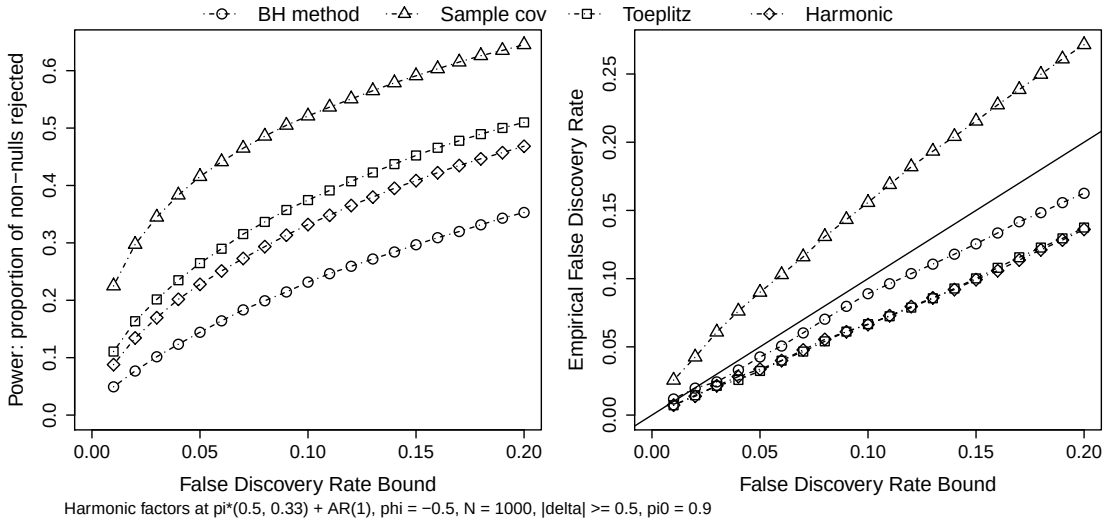

Figure 3.3: Left: Proportion of non-nulls rejected against chosen FDR bound averaged over 500 simulation runs for BH procedure on original Z values and adjusted Z values from the sample covariance, Toeplitz, and harmonic factor adjustments. Right: Simulated FDR against chosen FDR bound.

3.6 Two-Sample Testing for Copy Number Variations

An application in which the test statistics $Z_1, \ldots, Z_N$ can be regarded as a time series is in the analysis of copy number variations. Copy numbers are measured along a chromosome and measure the number of duplications or deletions of DNA sequences in small regions. Deletions or excessive numbers of duplications of DNA sequences at certain chromosomal locations have been linked to disease (Pinkel & Albertson (2005)). Often it is of interest to compare the copy number profiles between two groups of patients in order to identify locations at which the group means differ.

The data we analyze are taken from the Cancer Genome Atlas and consist of copy number measurements taken along each of the 23 chromosomes of 230 patients diagnosed with a type of brain cancer called glioblastoma multiforme. Each chromosome has a p arm and a q arm, and on each arm there are several thousands of measurements. Here we investigate whether the survival times of the patients can be linked to copy number variations at certain chromosomal locations by dividing the 230 patients into a group of 92 long-term survivors (surviving for more than two years after initial diagnosis) and 138 short-term survivors (surviving for less than two years after initial diagnosis), and testing for differences in copy number means at many chromosomal locations. We present an analysis of differences in mean copy numbers for the two patient groups along the p arm of chromosome 3, on which copy numbers are measured at 7531 locations.

Some data pre-processing steps such as double-standardization and the removal of batch effects have been relegated to the Appendix. After these steps were carried out, two-sample t-statistics $t_i = (\bar{X}_{i\cdot} - \bar{Y}_{i\cdot})(s_i^2/n_1 + \vartheta_i^2/n_2)^{-1/2}$, hereafter denoted by Z_i, were computed for each of the 7531 copy number locations. A histogram of $Z_1, \ldots, Z_N$ with the standard normal density overlaid appears in the left hand panel of Figure 3.4. The empirical variance of $Z_1, \ldots, Z_N$ is 0.8591, as shown, which is less than the unit variance we would expect under complete nullity of $H_1, \ldots, H_{7531}$. Moreover, we would expect the empirical variance to exceed 1 if some of the hypotheses were false. More will be said

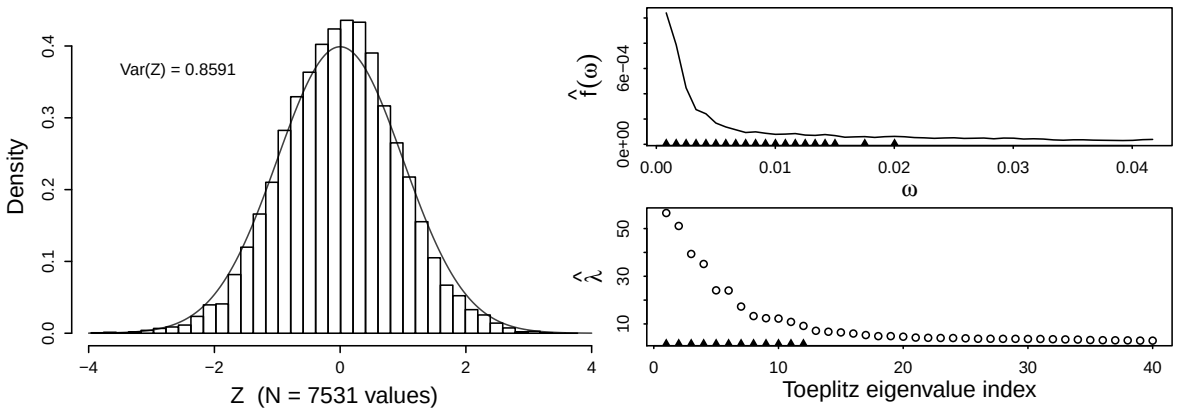

Figure 3.4: Left: A histogram of the 7531 Z values with Normal$(0, 1)$ density overlaid. Right: Estimated spectral density of $\{Z_t\}_{t\geq 1}$ and plot of eigenvalues from Toeplitz estimate of $\mathbf{\Sigma}_Z$. Triangles mark retained frequencies/factors.

about this underdispersion later.

The method of Toeplitz covariance matrix estimation described in Section 3.2.4 was carried out to obtain $\hat{\mathbf{\Sigma}}_Z^{(T)}$ and $\hat{\gamma}_Z(\cdot)$, and from $\hat{\gamma}_Z(\cdot)$ an estimate of the spectral density was obtained. Figure 3.4 displays the estimated spectral density in the upper right hand panel and a plot of the eigenvalues of $\hat{\mathbf{\Sigma}}_Z^{(T)}$ in descending order in the lower right hand panel with triangles marking frequencies and eigenvalues corresponding to retained factors. It was chosen to retain $m = 20$ harmonic factors and $m = 12$ Toeplitz factors. These choices of m satisfied expressions (3.6) and (3.7) at $\xi = 0.58$, so that retained factors accounted for the same proportion of the total variability for both methods. The middle 80% of $Z_1, \ldots, Z_{7531}$ are used for fitting $\boldsymbol{\beta}$ in (3.4) and $\boldsymbol{f}$ in (3.2). The `rlm()` function from the R package `MASS` under default settings was used to obtain the fitted values. Figure 3.5 shows 1000 of the 7531 Z values along a stretch of the p arm of chromosome 3 with the estimated total contribution of the 20 harmonic components overlaid as well as that of the 12 factors defined by spectral decomposition of the Toeplitz covariance matrix.

A normal quantile plot of the adjusted Z values is shown in Figure 3.6 in which the

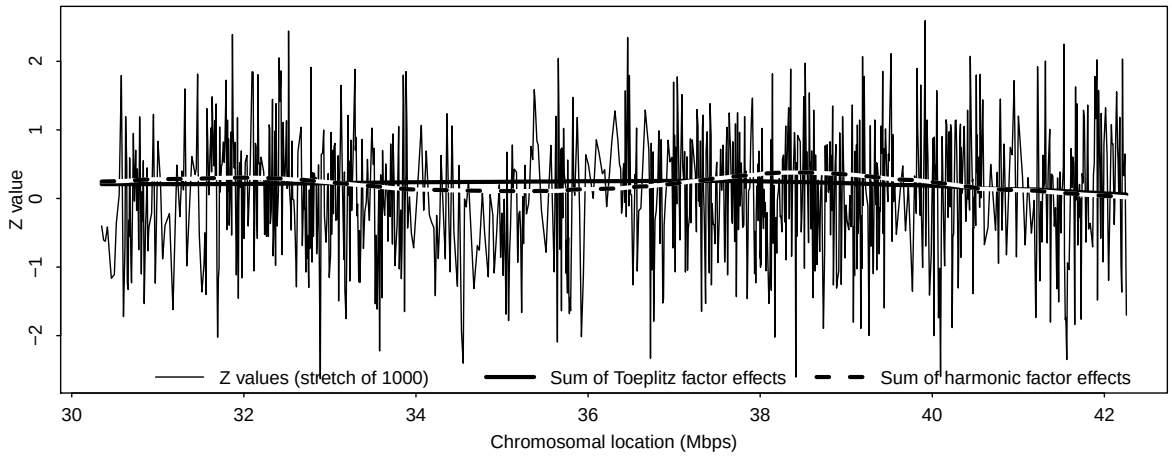

Figure 3.5: Stretch of 1000 Z values along the p arm of chromosome 3 with estimated contribution of harmonic factors and that of factors defined by principal components on the Toeplitz estimate of the covariance matrix.

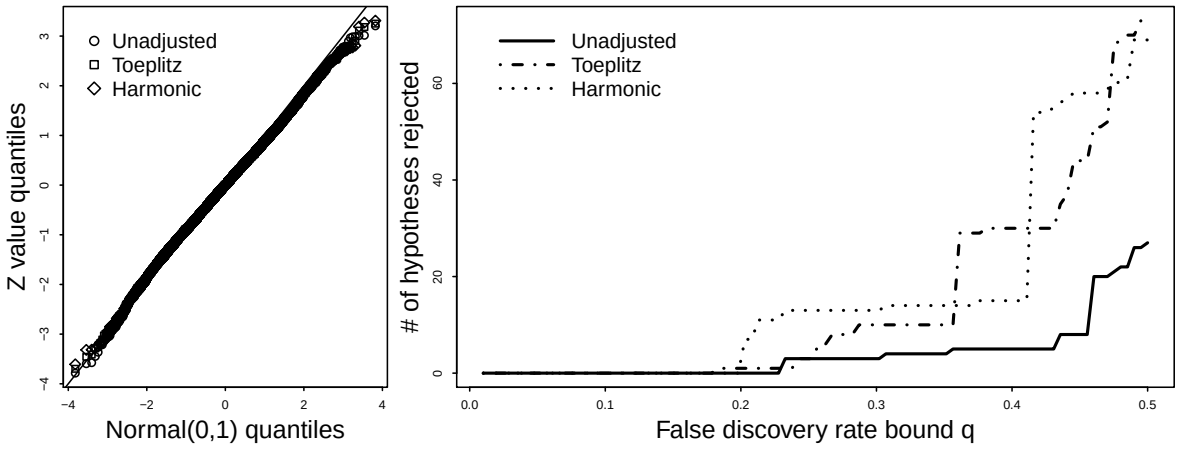

Figure 3.6: Left: Normal quantile plot of raw Z values as well as those FHG-adjusted with harmonic and Toeplitz factors. Right: The numbers of rejections achieved at increasing values of the FDR bound q for the three sets of Z values.

quantiles of the unadjusted Z values appear as well as those adjusted by the Toeplitz and harmonic factors. The right hand panel of Figure 3.6 plots the number of rejections achieved by the three sets of Z values against the FDR threshold q. The curves for the factor-adjusted test statistics rise more quickly than that for the unadjusted test statistics, indicating greater numbers of rejections for smaller values of q. The curve for the harmonic procedure initially climbs more quickly than that for the Toeplitz procedure, yet the curves cross and re-cross each other, leaving it unclear which procedure is preferable for these data.

We also note that the method proposed in Efron (2010a) for rescaling the null distribution of the test statistics with estimates of the null mean and variance was used in the BH step on the three sets of Z values, which all exhibited underdispersion by their less-than-unit slope in the normal quantile plot of Figure 3.6. The `locfdr` function from Efron (2010b) was used to obtain maximum likelihood estimates of the mean and variance of the empirical null distributions.

3.7 Conclusions

An adaptation of the Fan et al. (2012) dependence-adjusted procedure for the case of serially dependent test statistics was developed. Gains in power were demonstrated from removing the effects of harmonic or Toeplitz factors from the test statistics prior to carrying out the Benjamini & Hochberg (1995) procedure. A theoretical result was given showing that the factor-adjusted test statistics lead to a larger choice of critical region by the BH procedure; further, the effect of factor adjustment on the choice of critical region was shown to depend on the loadings of the non-null test statistics upon the factors and on the factor variances. These results were born out in simulation studies. The proposed methodology was shown to be practicable in a real data setting as well as more powerful than the BH procedure on the unadjusted test statistics.

4. A SMOOTH BLOCK BOOTSTRAP FOR STATISTICAL FUNCTIONALS AND TIME SERIES

4.1 Introduction

Many properties of smooth bootstraps have been explored for independent data. To smooth Efron's (1979) iid bootstrap, for example, bootstrap samples are drawn from a kernel density estimate of the population distribution, which is equivalent to resampling observed values $X_1, \ldots, X_n$ after these have each been additively augmented with independent random errors from the underlying kernel density. That is, an iid smooth bootstrap sample $X_1^*, \ldots, X_n^*$ drawn from a kernel density $\widehat{f}_n(x) = (nh)^{-1} \sum_{i=1}^n k((x - X_i)/h)$ estimator (with bandwidth $h > 0$) can be equivalently obtained as $\tilde{X}_i^* + hZ_i^*$ from a sample $\tilde{X}_1^*, \ldots, \tilde{X}_n^*$ drawn with replacement from the observed data values $X_1, \ldots, X_n$ and an (independent) iid sample $Z_1^*, \ldots, Z_n^*$ from a kernel density $k(\cdot)$. Just as kernel density estimators can exhibit advantages over histograms in some inference problems, one might expect a smooth bootstrap to enjoy similar advantages over its unsmooth bootstrap counterpart. This can be particularly true in attempting to approximate the sampling distribution of a statistical functional which depends intricately on unknown "smooth" population quantities. As an illustration, compared to the unsmooth iid bootstrap, Falk & Reiss (1989) and Hall et al. (1989) showed a significant advantage to the smooth iid bootstrap for estimating the distribution of sample quantiles. This is because the asymptotic variance of the pth sample quantile depends crucially on the population density evaluated at the pth population quantile—an unknown quantity which is often difficult to estimate without data smoothing steps. It is in such cases that a smooth bootstrap may be particularly beneficial.

However, unlike the independent data case, smooth bootstrap methods for dependent data have received little attention. Our goal is to extend a smooth bootstrap for time se-

ries based on smoothing modifications to the extended tapered block bootstrap (ETBB). That is, because block bootstraps provide a generally applicable and basic approach for resampling time series (i.e., by resampling blocks of data), it is natural to consider enhancing this time series bootstrap through smoothing steps. While many variants of the block bootstrap have been proposed, including the moving block bootstrap (Künsch (1989); Liu & Singh (1992)), the circular block bootstrap from Politis & Romano (1992) and the stationary bootstrap from Politis & Romano (1994), we focus our development on a smoothed version of the ETBB method. One reason is that the tapered block bootstrap (TBB), introduced by Paparoditis & Politis (2001) and Paparoditis & Politis (2002), offers improvements to the other block bootstraps above by re-weighting observations within data blocks with a taper function (e.g., thereby producing MSE-better variance estimators for approximately linear statistics). In this sense, the TBB represents a state-of-the-art block bootstrap to consider. Furthermore, because of a generalization of the TBB due to Shao (2010), the resulting ETBB can be applied to estimating the distribution of quite general statistical functionals. Examples of such statistics include classes of L-, R-, and M-estimators, which are not necessarily or easily smooth functions of sample averages (as considered originally for the TBB by Paparoditis & Politis (2001) and Paparoditis & Politis (2002)). For such functionals, a smooth ETBB can potentially provide improved inference for time series just as the smoothed iid bootstrap might for independent data.

To frame the results of the paper, suppose that $X_1, \ldots, X_n$ represents an observed stretch from a real-valued stationary time series with marginal distribution F. Denote the target parameter of interest as $\theta = T(F)$ based on some statistical functional $T(\cdot)$, allowing a wide class of parameters to be considered. A natural estimator of θ is then given by

$$\hat{\theta}_n = T(F_n), \tag{4.1}$$

based on the empirical distribution

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

of the data, where δ_x denotes a probability measure with point mass at $x \in \mathbb{R}$. To develop bootstrap versions of F_n and $\hat{\theta}_n = T(F_n)$ for inference, we propose a smooth ETBB empirical distribution F_n^* created as follows: for iid random variables $Z_1^*, \ldots, Z_n^*$ drawn from a kernel density $k(\cdot)$ and using a bandwidth parameter $h > 0$, let F_n^* represent the ETBB empirical distribution (cf. Shao (2010)) constructed by block resampling the *augmented* data $X_1 + hZ_1^*, \ldots, X_n + hZ_n^*$. By this formulation, the smooth ETBB naturally mimics the smoothing mechanics of the iid smooth bootstrap but, in the time series case, the resampling of individual observations is crucially replaced by resampling of data blocks to capture the underling time dependence. A smooth ETBB statistic is then defined as $\hat{\theta}_n^* = T(F_n^*)$ in analogy to $\hat{\theta}_n = T(F_n)$. Under fairly general conditions that allow for a variety of statistical functionals, we show that the smooth ETBB consistently estimates the variance of $\hat{\theta}_n$ and validly approximates the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. Our results expand beyond the smooth function model for time series statistics (i.e., smooth functions of sample averages), representing the formal conditions in previous establishments of both the TBB and the ETBB (Paparoditis & Politis (2001), Paparoditis & Politis (2002), Shao (2010)). In this sense, the smoothed ETBB considered here broadens the scope and applicability of block bootstraps for dependent data.

The rest of the manuscript is organized as follows. Section 4.2 describes the proposed smoothed ETBB procedure. Section 4.3 provides assumptions and examples for statistical functionals, and Section 4.4 gives the main distributional results on the smooth ETBB. Simulation studies of the procedure appear in Section 4.5, where the proposed smooth bootstrap is compared with other block bootstraps. Section 4.6 provides some concluding remarks and proofs of the main results appear in Appendix C.

4.2 The Smooth Extended Tapered Block Bootstrap

Based on some smooth functional $T(\cdot) : \mathbb{P} \to \mathbb{R}$, where $\mathbb{P}$ denotes the space of probability measures on $\mathbb{R}$, recall that the target parameter and its natural estimator are formulated as $\theta = T(F)$ and $\hat{\theta}_n = T(F_n)$, as in (4.1), based on data $X_1, \ldots, X_n$ from a real-valued stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ with the marginal probability distribution F. From $X_1, \ldots, X_n$, we wish to create smooth ETBB versions $\tilde{\theta}_n$ and $\hat{\theta}^*_n$ that adequately mimic both θ and $\hat{\theta}_n$. In which case (and as shown later), the resulting smooth ETBB method (hereafter SETBB) can be applied to estimate the variance $n\text{var}(\hat{\theta})$ of the statistic $\hat{\theta}_n$ or approximate the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ (e.g., for nonparametrically calibrating confidence intervals for θ).

To describe the SETBB method, we first state the ETBB procedure of Shao (2010) for approximating the empirical distribution $F_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ from (4.1) with a ETBB version $\tilde{F}^*_{n,ETBB}$. This ETBB rendition is defined as

$$\tilde{F}^*_{n,ETBB} = \sum_{i=1}^n \pi^*_i \delta_{X_i}, \qquad \sum_{i=1}^n \pi^*_i = 1, \tag{4.2}$$

based on bootstrap empirical weights $\pi^*_1, \ldots, \pi^*_n$ on $X_1, \ldots, X_n$ that are constructed from a process of data block resampling and data tapering as follows. Let $1 \leq \ell < n$ denote an integer block length and $\mathcal{I}_n \equiv \{0, 1, \ldots, n-\ell\}$ denote an index set for overlapping data blocks $(X_{i+1}, \ldots, X_{i+\ell})$, $i \in \mathcal{I}_n$, from $(X_1, \ldots, X_n)$. To resample $b = \lfloor n/\ell \rfloor$ data blocks of length ℓ, let $I^*_1, \ldots, I^*_b$ be iid with a uniform distribution over $\mathcal{I}_n$. Additionally, define a sequence of weights $w_\ell(1), \ldots, w_\ell(\ell)$ in $[0, 1]$ with a tapering window

$$w_\ell(t) \equiv w\left(\frac{t - 0.5}{\ell}\right), \quad \ell = 1, 2, \ldots, \tag{4.3}$$

based on a function $w : \mathbb{R} \to [0, 1]$. Following Künsch (1989), Paparoditis & Politis (2001), Paparoditis & Politis (2002) and Shao (2010), we suppose that $w(t)$ is symmetric

about $t = 1/2$, positive in a neighborhood of $t = 1/2$, nondecreasing for $t \in [0, 1/2]$, and that $w(t) = 0$ if $t \notin [0, 1]$. Then, the empirical weights defining ETBB empirical distribution $\tilde{F}^*_{n,ETBB}$ (4.2) are defined as

$$\pi^*_t = \frac{1}{b\|w_\ell\|_1} \sum_{k=1}^{\ell} w_\ell(k) \sum_{j=1}^{b} \mathbb{I}(t = I^*_j + k), \quad t = 1, \ldots, n,$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and $\|w_\ell\|_1 = \sum_{k=1}^{\ell} w_\ell(k)$. In defining a bootstrap empirical weight π^*_t, note that $b^{-1} \sum_{j=1}^{b} \mathbb{I}(t = I^*_j + k)$ represents the proportion of times that observation X_t falls into a resampled data block in the kth position, $k = 1, \ldots, \ell$, where a taper-based weight $w_\ell(k)/\|w_\ell\|_1$ is further attributed to the kth position of any data block. See Remark 1 below for more details about tapering.

As described in Section 1, the intended SETBB method is then defined by additional data smoothing steps which imitate the smooth bootstrap for iid data. For a choice of kernel density $k(\cdot)$, let $Z^*_1, \ldots, Z^*_n$ be iid random variables from $k(\cdot)$, which are independent of any block resampling, and let $h > 0$ denote a bandwidth parameter (i.e., $h \to 0$ as $n \to \infty$). Then, the SETBB empirical distribution F^*_n results from applying the construction of the ETBB empirical distribution $\tilde{F}^*_{n,ETBB}$ (4.2) to the augmented data $X_1 + hZ^*_1, \ldots, X_n + hZ^*_n$. That is, SETBB empirical distribution can be expressed as

$$F^*_n = \sum_{i=1}^{n} \pi^*_i \delta_{X_i + hZ^*_i}, \qquad \sum_{i=1}^{n} \pi^*_i = 1,$$

with the same bootstrap empirical weights $\pi^*_1, \ldots, \pi^*_n$ as in (4.2). A natural choice of kernel density $k(\cdot)$ is the standard normal density $\phi(\cdot)$ and, for concreteness and simplicity, we assume that $k(\cdot) = \phi(\cdot)$ throughout the remainder.

From the bootstrap empirical distribution F^*_n, SETBB versions of $\hat{\theta}_n = T(F_n)$ and $\theta = T(F) = T(\mathcal{E}(F_n))$ are defined as

$$\hat{\theta}^*_n = T(F^*_n), \qquad \tilde{\theta}_n = T(\mathcal{E}_*(F^*_n)),$$

where $\mathcal{E}_*$ denotes bootstrap expectation (i.e., relative to the distributions of $\{I_j^*\}_{j=1}^b$ and $\{Z_i^*\}_{i=1}^n$) conditional on the data $X_1, \ldots, X_n$. In Section 4.4, we establish that, for a large variety of statistical functionals, the SETBB method validly approximates the variance and sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$. To state the main distributional results, we provide some assumptions on the functional $T(\cdot)$ in the next section, along with some examples.

Remark 1: The tapering of data blocks intends to give reduced weight to observations near the endpoints of a block, which can improve the performance of the block bootstrap (e.g., minimizing bias and MSE in variance estimation; Künsch (1989); Paparoditis & Politis (2001);Paparoditis & Politis (2002). Note that untapered blocks correspond to $w(t) = \mathbb{I}(t \in [0,1])$ as the indicator function of the interval $[0,1]$, in which case the TBB reduces to the original moving block bootstrap (Künsch (1989); Liu & Singh (1992)). In contrast, Paparoditis & Politis (2001) describe advantages of a 'smooth' data taper (4.3), characterized by a self-convolution $(w * w)(t) \equiv \int_{-1}^{1} w(x)w(x + |t|)dx$ being twice continuously differentiable at $t = 0$. One such example is the trapezoidal taper

$$w_c^{\text{trap}}(u) \;=\; \begin{cases} u/c, & \text{if } u \in [0, c] \\ 1, & \text{if } u \in [c, 1-c] \\ (1-u)/c & \text{if } u \in [1-c, 1] \end{cases}$$

where the choice of $c = .43$ has been proposed/used by Paparoditis & Politis (2001), Paparoditis & Politis (2002) and Shao (2010). Conditions on other tuning parameters in the STBB method, such as block length ℓ and bandwidth h, are described in Section 4.4.

4.3 Statistical Functionals: Conditions and Examples

We wish to establish the SETBB method in a general manner for parameters $\theta = T(F)$ and estimators $\hat{\theta}_n = T(F_n)$ as statistical functionals $T(\cdot)$. For illustration, we provide

some brief examples of such functionals in the following. Letting $\mathbb{P}$ denote the space of all probability distributions on $\mathbb{R}$, we denote the distribution function of $F \in \mathbb{P}$ as $F(x) \equiv F((-\infty, x])$, for $x \in \mathbb{R}$.

Example 1 (Smooth Functions of Means): For a function $g : \mathbb{R} \to \mathbb{R}$, consider a functional defined as $\theta = T(F) = g(\int x dF(x))$ based on the mean $\mathcal{E}X_1 = \int x dF(x) < \infty$. Another simple example is the variance functional $\text{var}(X_1) = T(F) = \int (x - \int x dF(x))^2 dF(X)$. See ch. 4 of Lahiri (2003b) for details of other time series statistics falling into this smooth function model (Hall (1992)) of parameters.

Example 2 (L-estimators): For a function $L : [0, 1] \to \mathbb{R}$, an L-functional is defined as $\theta = T(F) = \int xL(F(x))dF(x)$, and $\hat{\theta}_n = T(F_n)$ is an L-estimator. Examples of L-estimators include the sample mean ($L(x) = 1$), a Gini's mean difference ($L(x) = 4t-2$), and trimmed sample means ($L(x) = \mathbb{I}(\alpha < x < \beta)/(\beta - \alpha)$ for some $\alpha < \beta$). See Serfling (1980) and Shao (2003) for further L-estimators.

Example 3 (Rank statistics): Define $\bar{F}(x) = F(x) - \lim_{y \uparrow -x} F(y)$ for $x > 0$ and $\bar{F}(x) = 0$ otherwise, and let $R : [0, 1] \to \mathbb{R}$ with a bounded derivative R'. Then define a functional $T(F) = \int_0^\infty R(\bar{F}(x))dF(x)$ (e.g., $T(F) = 0$ when F is symmetric, $F(x) = 1 - F(x)$) so the corresponding estimator $T(F_n)$ is a signed rank statistic. For example, the case $R(t) = t$ corresponds to the Wilcoxon signed rank statistic (cf. Shao (2003)) as a robust assessment of location. See Tran (1988), Hallin & Puri (1991) and Andrews (2008), and references therein, for other rank-based estimation with time series.

Example 4 (M-estimators): For a function $\Psi(x, t)$, an M-estimator $T(F_n)$ can be defined as the solution to $\int \Psi(x, t)dF_n(x) = 0$, estimating a parameter $T(F)$ for which $\int \Psi(x, T(F))dF(x) = 0$ holds. This class of estimators can contain maximum likelihood estimators and various robust estimators for time series models. See Bustos (1982), Martin & Yohai (1986) and Bustos & Yohai (1986) and the references therein.

Example 5 (Sample Quantiles): For a $p \in (0, 1)$, define $\theta \equiv T(F) = \inf\{x \in \mathbb{R} : F(x) \geq$

$p\}$ as the pth quantile of F and denote $\hat{\theta}_n = T(F_n)$ as the pth sample quantile, where the choice $p = 0.5$ corresponds to the sample median.

In establishing bootstrap methods for statistics as statistical functionals, a compounding factor is formulating a suitable, but general, notion of Taylor expansions of $T(\cdot)$ around F involving an appropriate derivative (or differential) $T_F^{(1)}(\cdot)$. We next state differentiability conditions on the functional $T : \mathbb{P} \to \mathbb{R}$ and assumptions on the marginal distribution function F of $\{X_t\}$.

For this, we require some notation. Let $\mathbb{D}$ denote the space of all real-valued functions on $[-\infty, \infty]$ that are right continuous with left limits, which we equip with the Skorohod metric (cf. Billingsley (1968)), denoted as $d_S(H_1, H_2)$ for $H_1, H_2 \in \mathbb{D}$. Additionally, for $H_1, H_2 \in \mathbb{D}$, define the Kolmogorov norm $\|H_1\|_\infty = \sup_{x\in\mathbb{R}}|H_1(x)|$, the L^1 norm $\|H_1\|_1 = \int_{-\infty}^{\infty}|H_1(x)|dx$ and L^1 distance $d_1(H_1, H_2) = \|H_1 - H_2\|_1$. Let $D_0 \equiv \{a_1(G_1 - G_2) : G_1, G_2 \in \mathbb{P}, a \in \mathbb{R}\} \subset \mathbb{D}$.

Conditions:

(C.1) $F(x) = P(X_1 \leq x)$, $x \in \mathbb{R}$, is continuous and satisfies $\|F(x) - F(x + a)\|_\infty \leq C|a|$, for any $a \in \mathbb{R}$ and some $C > 0$.

(C.2) $T(\cdot)$ is differentiable at F in the sense that

(i) there exists a linear functional $T_F^{(1)} : \mathbb{D}_0 \to \mathbb{R}$ such that

$$T(G) - T(F) = T_F^{(1)}(G - F) + R(G - F)$$

holds for any $G \in \mathbb{P}$ with a remainder term satisfying $|R(G-F)| \leq C[\rho\|G - F\|_\infty^{\lambda+1} + (1 - \rho)\|G - F\|_1\}^{1+\lambda}]$ for some $C > 0$, $\lambda > 0$ and $\rho \in [0, 1]$; when $\rho < 1$, assume $\mathcal{E}|X_1| = \int|x|dF(x) < \infty$.

(ii) $T_F^{(1)}(\cdot)$ is continuous, in $d_{S,1}$-distance, at the zero function $\mathbf{0}$ (i.e., $\mathbf{0}(x) = 0$, $x \in \mathbb{R}$) and $|T_F^{(1)}(H)| \leq A_1 \exp[A_2 d_{S,1}(\mathbf{0}, H)]$ holds for some $A_1, A_2 > 0$, where $d_{S,1}$ is defined as

either d_S or d_1.

Remark 2: The expansion in C.2(i) does not have to hold for any $G \in \mathbb{P}$; it suffices if this condition holds (w.p.1) for G supported on the data $X_1, \ldots, X_n$.

To motivate the differential $T_F^{(1)}(\cdot)$, note that under condition C.2(i) the *influence function* (Hampel (1974)) is given as

$$T_F^{(1)}(\delta_x - F) \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon}[T((1-\epsilon)F + \epsilon\delta_x) - T(F)], \quad x \in \mathbb{R}. \tag{4.4}$$

The conditions above are meant to be compatible with forms of differentiability for statistical functionals, such as Hadamard or Fréchet differentiability, which have been studied for a variety of statistical functionals (cf. Serfling (1980); Huber (1981); Fernholz (1983); Ren & Sen (1991), Ren & Sen (1995); van der Vaart & Wellner (1996); Shao (1993); Shao (2003)). A major complication in formulating differentiability assumptions on statistical functionals is that this aspect can depend intricately on metric used (e.g., $\|\cdot\|_1$ or $\|\cdot\|_\infty$ based) for probability measures (cf. Shao (1993)). For this reason, the conditions above allow for both $\|\cdot\|_1$ or $\|\cdot\|_\infty$-based distances in describing remainder terms. Where allowable, we have also attempted to relax assumptions by using Skorohod distance in place of Kolmogorov distance (i.e., $d_S(H_1, H_2) \leq \|H_1 - H_2\|_\infty$). Additionally, while the differential $T_F^{(1)}$ in (C.2)(i) is assumed to have the typical linearity property (i.e., $T_F^{(1)}(a_1G_1 + a_2G_2) = a_1T_F^{(1)}(G_1) + a_2T_F^{(1)}(G_2)$, $G_1, G_2 \in \mathbb{D}_0$), we need not assume that this functional be generally continuous. Condition (C.2)(i) is also perhaps weaker than strong Fréchet differentiability used in other studies of early block bootstraps (cf. Liu & Singh (1992)).

We next briefly return to the previous examples to illustrate how different statistical functions fit into the assumptions above and, thus, can be validly approximated by the SETBB method.

Example 1 (Smooth Functions of Means): If g has a derivative $g' : \mathbb{R} \to \mathbb{R}$ satisfying a Lipschitz condition $|g'(x) - g'(y)| \leq C|x - y|^{\delta}$ for some $\delta > 0$, then Condition C.2 holds (using d_1 distance) with a remainder bounded by $C\|F - G\|_1^{1+\delta}$ and differential $T_F^{(1)}(\Delta) = g'(\int x dF(x)) \int x d\Delta(x)]$, $\Delta \in \mathbb{D}_0$. Likewise, the variance functional satisfies Condition C.2 with $T_F^{(1)}(\Delta) = \int [x^2 - 2 \int x dF(x)] d\Delta(x)]$ with a remainder bounded by $C\|F - G\|_1^2$.

Example 2 (L-estimators): If the function $L : [0, 1] \to \mathbb{R}$ satisfies $|L(x) - L(y)| \leq C|x - y|^{\delta}$ for some $\delta > 0$, for example, then Condition C.2 holds (with a remainder bounded by $C\|F - G\|_{\infty}^{\delta}\|F - G\|_1$) with $T_F^{(1)}(\Delta) = -\int \Delta(x) J(F(x)) dx$, $\Delta \in \mathbb{D}_0$.

Example 3 (Rank statistics): Considering signed rank statistic $R(t) = t$, for example, Condition C.2 holds (with remainder bounded by $C\|F - G\|_{\infty}^2$) with a differential $T_F^{(1)}(\Delta) = \int_0^{\infty} \bar{\Delta}(x) dF(x) + \int_0^{\infty} \bar{\Delta}(x) dF(x)$.

Example 4 (M-estimators): For simplicity, if one assumes $\Psi(x, t)$ is bounded, Lipschitz of order $\delta > 0$ in t (for any x), and that $\psi(t) \equiv \int \Psi(x, t) dF(x)$ is differentiable in t and with $\psi'(t)$ bounded away from 0, then Condition C.2 holds (with remainder bounded by $C\|F - G\|_{\infty}^{1+\delta}$) for $T_F^{(1)}(\Delta) = -[\psi'(T(F))]^{-1} \int_0^{\infty} \Psi(x, T(F)) d\Delta(x)$, $\Delta \in \mathbb{D}_0$.

Example 5 (Sample Quantiles): Assuming F has a positive derivative/density f around the pth quantile $\theta \equiv T(F) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$, the corresponding differential $T_F^{(1)}(\Delta) = -[f(\theta)]^{-1} \int_{-\infty}^{\theta} \Delta(x) dx$, $\Delta \in \mathbb{D}_0$ depends intricately on the density f at θ. Sample quantiles are difficult to place into the conditions above, but these could also be validated for the SETBB method through alternative techniques, such as Bahadur-Kiefer representations of sample quantiles and order statistics (cf. Serfling (1980)).

While we have reviewed some examples and conditions in this section, it is important to iterate that *implementation* of the SETBB method (with results described next) does not require a differential $T_F^{(1)}$ to be explicitly determined or applied in practice. In this, SETBB differs from other block bootstrap approaches which do require and use a direct

form for $T_F^{(1)}$ in each inference instance (cf. Paparoditis & Politis (2002)). As noted by Shao (2010), observing that the process density $f(\cdot)$ appears in the differential for sample quantiles (e.g., $T_F^{(1)}(\Delta) = -[f(\theta)]^{-1} \int_{-\infty}^{\theta} \Delta(x)dx$ above), such bootstrap approaches directly requiring $T_F^{(1)}$ break down when $T_F^{(1)}$ depends on smooth or infinite dimensional process parameters. It is not hard to find other statistical functionals with this behavior, where for example an M-estimator $T(F_n)$ producing a trimmed sample mean (cf. Huber (1964); Shao (2003)) based on $\Phi(x,t) = (t-x)\mathbb{I}(|t-x| \leq \alpha)$, $\alpha > 0$, has an associated differential $T_F^{(1)}(\Delta) = -\beta_{\theta,\alpha}^{-1} \int \Psi(x,\theta)d\Delta(x)$, with $\theta = T(F)$ and

$$\beta_{\theta,\alpha} = F(\theta+\alpha) - F(\theta-\alpha) - \alpha[f(\theta+\alpha) - f(\theta-\alpha)],$$

that depends intricately on a smooth density $F' = f$. Other examples given above also indicate statistical functionals with complicated differential forms. It is in these cases where the additional smoothing steps associated with the SETBB method are potentially beneficial for improved inference.

4.4 Main Results

To state the main bootstrap approximation results, recall $\hat{\theta}_n^* \equiv T(\tilde{F}_n^*)$ and $\tilde{\theta}_n \equiv T(\mathcal{E}_* \tilde{F}_n^*)$ as defined in Section 4.2 (i.e., based on the SETBB empirical distribution) are the SETBB versions of $\hat{\theta}_n = T(F_n)$ and $\theta = T(F)$. We estimate the distribution of $\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}[T(F_n) - T(F)]$ with the following bootstrap analog

$$m_\ell^{1/2}\sqrt{n}(\hat{\theta}_n^* - \tilde{\theta}_n)$$

where $m_\ell = \|w_\ell\|_1^2/[\ell\|w_\ell\|_2^2]$ represents a scalar depending on norms $\|w_\ell\|_1 = \sum_{k=1}^{\ell} w_\ell(k)$ and $\|w_\ell\|_2 = \{\sum_{k=1}^{\ell} w_\ell^2(k)\}^{1/2}$ of the taper weights from (4.3). The factor m_ℓ adjusts for the effect of the data taper based on length ℓ data blocks, and similar adjustments appear for the TBB/ETBB applied to sample mean inference (or smooth functions of sample

means); see Paparoditis & Politis (2001), Paparoditis & Politis (2002) and Shao (2010). However, unlike in some applications of the block bootstrap (e.g., long-memory series, Lahiri (1993); Kim & Nordman (2011)), this correction should *not* be interpreted as an order adjustment, because $m_\ell \to [\int_0^1 w(t)dt]^2/[\int_0^1 w^2(t)dt] > 0$ converges to a constant as $n \to \infty$. Additionally, we may define a SETBB estimator of the variance $n\text{var}(\hat{\theta}_n)$ as $m_\ell n\text{var}_*(\hat{\theta}_n^*)$, where var_* denotes variance with respect to the SETBB resampling mechanism.

Theorem 1 next shows that the SETBB provides consistent estimators of both variances and sampling distributions over a large class of statistical functions for time series (i.e., as prescribed by the conditions in Section 4.3). Recall that we assume a kernel density for data smoothing as standard normal (cf. Section 4.2). We prescribe weak dependence of the process $\{X_t\}$ in terms of strong mixing coefficients defined as $\alpha(k) = \sup\{|P(A \cap B) - P(A)P(B)|\colon A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty\}$, where $\mathcal{F}_{-\infty}^0, \mathcal{F}_k^\infty$ are the σ-algebras generated by $\{X_t : t \leq 0\}$ and $\{X_t : t \geq k\}$, respectively; see Doukhan (1994). In the following, let $Y_t = T_F^{(1)}(\delta_{X_t} - F)$, $t \in \mathbb{Z}$, denoting the evaluation of observations in the influence function (4.4), and let $\sigma_\infty^2 \equiv \sum_{k=-\infty}^\infty \text{cov}(Y_0, Y_k)$.

Theorem 1 *In addition to Conditions C.1-C.2 (with $\lambda > 0$ from C.2), suppose the data taper satisfies (4.3) and that the SETBB block length ℓ and smoothing bandwidth h satisfy $\ell^{-1} + nh^{2(1+\lambda)} = o(1)$ and $\ell^2/n = O(1)$ as $n \to \infty$. Suppose also that $\sigma_\infty^2 > 0$ and, for some $\gamma > 0$, it holds that $\mathcal{E}|Y_1|^{2+\gamma} < \infty$ and $\sum_{k=1}^\infty k^{c-2}\alpha(k)^{\gamma/(c+\gamma)} < \infty$ for $c = 2\max\{\lceil\lambda\rceil, 4\lceil\gamma/2\rceil\} + 4$. Then, the following hold as $n \to \infty$.*

(i) For the estimator $\hat{\theta}_n$ of θ,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, \sigma_\infty^2) \qquad \textit{and} \qquad n\textit{var}(\hat{\theta}_n) \to \sigma_\infty.$$

(ii) For the SETBB variance estimator,

$$m_\ell nvar_*(\hat{\theta}_n^*) - nvar(\hat{\theta}_n) \xrightarrow{p} 0.$$

(iii) For the SETBB version $m_\ell^{1/2}\sqrt{n}(\hat{\theta}_n^* - \tilde{\theta}_n)$ *of* $\sqrt{n}(\hat{\theta}_n - \theta)$*,*

$$\sup_{x\in\mathbb{R}} \left| P_* \left(m_\ell^{1/2}\sqrt{n}(\hat{\theta}_n^* - \tilde{\theta}_n) \leq x \right) - P\left(\sqrt{n}(\hat{\theta}_n - \theta) \leq x \right) \right| \xrightarrow{p} 0,$$

where P_* *denotes bootstrap probability.*

Theorem 1 shows that, under mild mixing and moment conditions on the time process (i.e., consistent with other mixing assumptions for the block bootstrap, cf. Künsch (1989); Paparoditis & Politis (2001)), the SETBB method is valid for a variety of statistical functionals. The conditions on the block length ℓ are quite general and allow a range of block sizes that include the MSE-optimal block lengths known for block bootstraps in problems of variance and distributional estimation (e.g., $\ell = n^{1/4}$ or $n^{1/5}$); see Hall et al. (1995), Paparoditis & Politis (2001) and Lahiri (2003b) for these details. In data smoothing, the bandwidth h condition is tied to the order of the remainder error in the generalized expansion of the functional $T(\cdot)$ under Condition C.2(i), Section 4.3. Larger bandwidths to induce more data smoothing are helpful in reducing estimation errors when the statistical functional exhibits an adequate degree of smoothness. The next section examines the performance of the SETBB method, and the selection of its tuning parameters (e.g., block length), through numerical studies.

Remark 3: Although the expansions of the bias and variance of SETBB variance estimators in Theorem 1(ii) are beyond the scope of this work, we anticipate that the SETBB method continues to enjoy the same improvements offered by the TBB and ETBB (Paparoditis & Politis (2001); Shao (2010)) over other block boostraps in terms of reduced bias and MSE in variance estimation. For this a smooth data taper is required as described in Remark 1, Section 4.2.

Remark 4: To facilitate the development and proofs for the SETBB, we have assumed the stationary time series process $\{X_t\}$ to be real-valued. Extensions of the SETBB method to time series of $\mathbb{R}^d$-valued random vectors and associated statistical functions are possible for implementations and inference scenarios with time series as described by ch. 4 of Lahiri (2003b) and Shao (2010).

4.5 Simulation Studies

Here we examine the performances of the SETBB and (unsmoothed) ETBB methods as well as the (extended) moving block bootstrap (MBB) and its smooth version (SMBB). The SETBB/ETBB methods use a trapezoidal window as a smooth data taper while the MBB/SMBB approaches use untapered data blocks (i.e., a window $w(t) = \mathbb{I}(t \in [0, 1])$) as described in Remark 1, Section 4.2. In particular, we compare these block bootstrap approaches applied to variance estimation for sample quantiles and trimmed means. In a variety of settings, the smoothing of the ETBB and the MBB significantly reduce MSEs in variance estimation over wide ranges of block sizes. For each bootstrap, we also consider an empirical method for block size selection based on the cross-validation approach of Hall et al. (1995), referred to as the HHJ method in the following. When the block size is chosen by the HHJ method, the MSEs of the smooth block bootstraps are less than those of their unsmooth counterparts in most cases. As the data smoothing steps involve the standard normal kernel, for simplicity, we typically choose a bandwidth h by a selection method of Sheather & Jones (1991), giving $h \propto n^{-1/5}$.

4.5.1 Sample Quantiles

The MSEs of the MBB, SMBB, ETBB, and SETBB estimators of the quantile variances for the $0.2, 0.5$, and 0.8 sample quantiles were compared for time series of lengths $n = 50$, 200, and 1000 for four models crossed with three innovation distributions. The four models were: (i) ARMA$(1, 1)$ with $\phi = .4$ and $\theta = .3$, (ii) AR(1) with $\phi = .9$, (iii) AR(1) with $\phi = -.5$, and (iv) MA(1000) with $\theta_j = (j + 1)^{-2.5}$ for $j = 1, \ldots, 1000$.

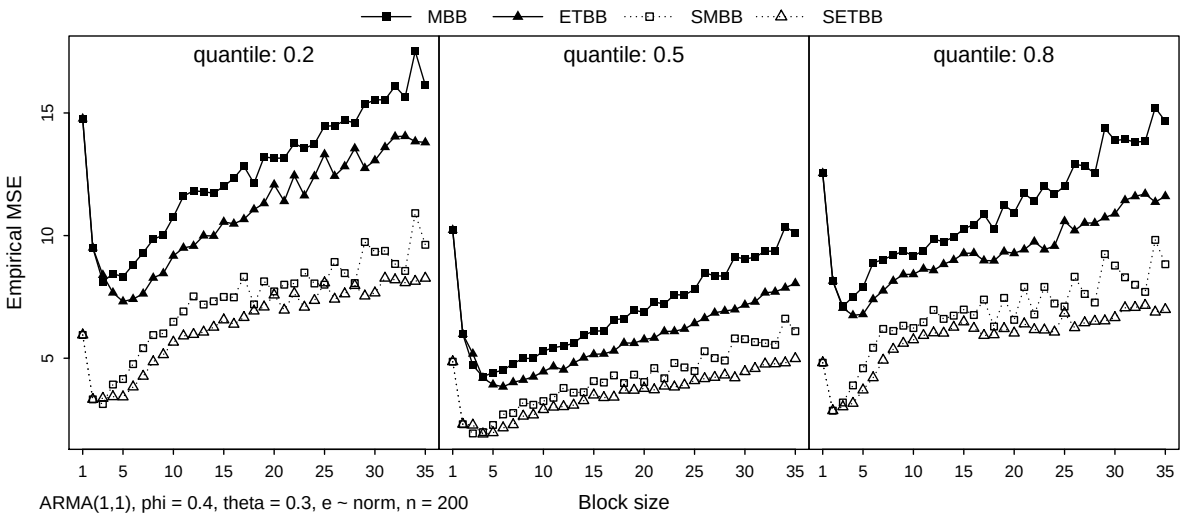

Figure 4.1: Mean squared error achieved for various block sizes by the MBB, SMBB, ETBB, and SETBB estimators of the quantile variance for the 0.2, 0.5, and 0.8 quantiles of a length $n = 200$ realization of an ARMA$(1, 1)$ process with $\phi = 0.4$, $\theta = 0.3$ and Normal$(0, 1)$ innovations. There were 500 simulation runs and the number of bootstrap resamples was set to 500.

The three innovation distributions were: (a) Normal$(0, 1)$, (b) Chi-square$(1) - 1$, and (c), the double exponential with variance equal to 1. Full factorial results for all settings $\{$(i),(ii),(iii),(iv)$\} \times \{$(a),(b),(c)$\}$ are provided in the Supplementary Material, and certain cases are highlighted here.

Figure 4.1 depicts the MSE of the MBB, SMBB, ETBB, and SETBB estimators of the quantile variance of the $0.2, 0.5$, and 0.8 quantiles as a function of the block size for model (i) with Normal$(0, 1)$ innovations for $n = 200$. Smoothing greatly reduced the mean squared error of the MBB and the ETBB estimators across all block sizes. At the optimal block size, the MSEs for the SMBB and SETBB were nearly equal.

To assess the performance of the four bootstrap methods when the block size is chosen using the HHJ empirical method, the block size selection procedure was implemented on each simulated data set and the selected block size was recorded. The resulting MSE in estimating the variance of the sample median with the HHJ-selected block size, as

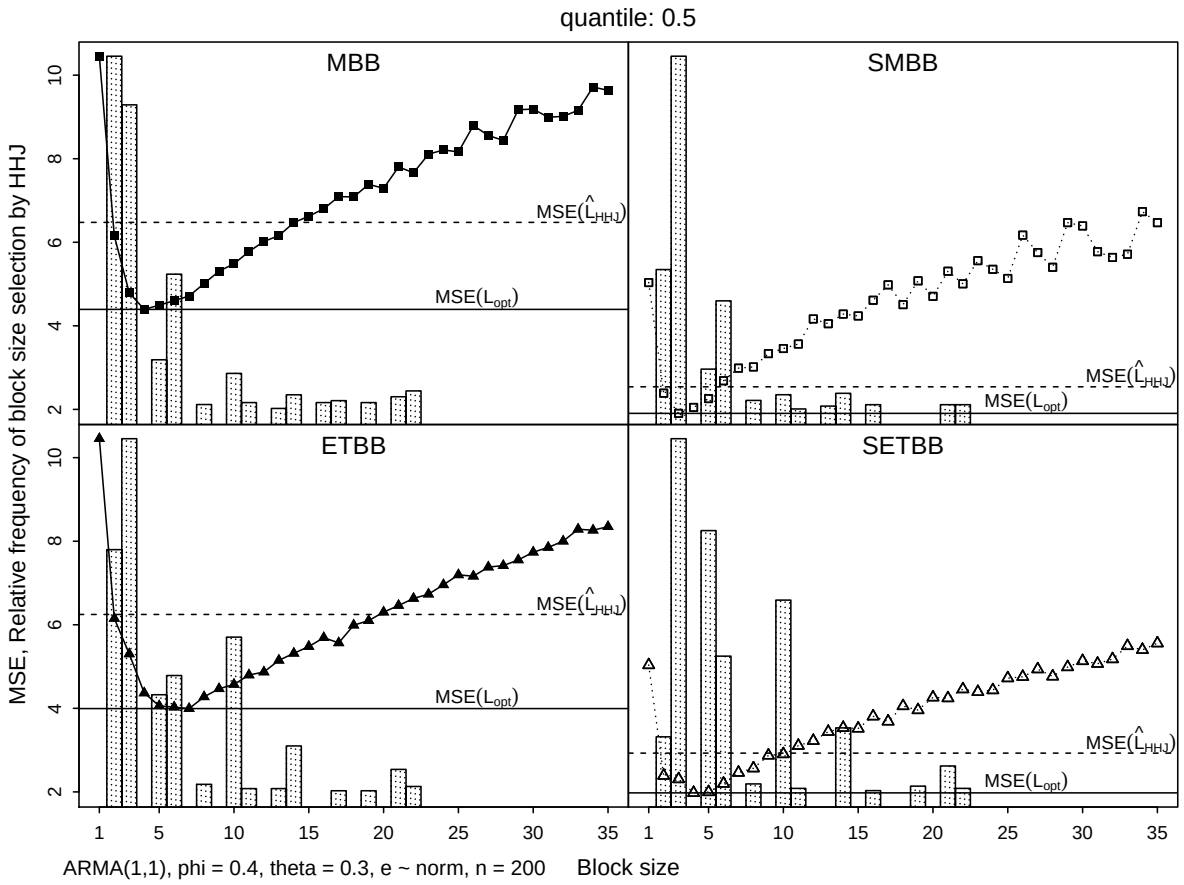

Figure 4.2: Mean squared error achieved by the MBB, SMBB, ETBB, and SETBB estimators of the variance of the median of a length $n = 200$ realization of an ARMA$(1, 1)$ process with parameters $\phi = .4$ and $\theta = .3$ with Normal$(0, 1)$ innovations. The mean squared error at the optimal block size and when the HHJ-selected block size is used are shown as well as the selection frequency of each block size.

well as the frequency with which each block size was selected, is depicted in Figure 4.2. The minimum MSE achieved by each method across all choices of the block size is also indicated. The results shown are for model (i) under Normal$(0, 1)$ innovations for a sample size of $n = 200$. In this case, data smoothing greatly reduced the MSEs of the MBB and ETBB methods, and even appeared to aid the HHJ algorithm in block selection.

Table 4.1 displays the root MSEs of the MBB, SMBB, ETBB, and SETBB estimators of the quantile variance when the block size is chosen by the HHJ empirical method for all combinations of models and innovation distributions {(i),(ii),(iii),(iv)} × {(a),(b),(c)} for the 0.5 and 0.8 quantiles when $n = 200$. For the double exponential innovations, the smoothing bandwidth was set to $h = 2n^{-1/3}$ as this innovation distribution lacks the smoothness of others considered. Except for the case of model (iii), in which the AR(1) parameter was negative, and for the median when double exponential innovations were paired with the MA(1000) model, smoothing again reduced the root MSEs of the MBB and ETBB estimators under the HHJ block selection method.

$n = 200$		0.5 quantile				0.8 quantile			
Model	Innov	MBB	SMBB	ETBB	SETBB	MBB	SMBB	ETBB	SETBB
ARMA(1,1)	norm	2.54	1.59	2.50	1.71	3.25	2.19	3.19	2.15
$\phi = .4$, $\theta = .3$	chisq	3.47	2.75	3.50	2.83	13.34	11.95	12.89	11.67
	dblexp	1.93	1.23	1.91	1.30	3.25	2.29	3.18	2.31
AR(1)	norm	83.85	79.29	82.44	77.92	99.62	92.75	95.24	88.57
$\phi = .9$	chisq	148.86	144.55	147.90	142.49	259.23	251.18	254.41	247.99
	dblexp	78.53	77.54	78.43	76.92	97.76	95.90	95.05	92.88
AR(1)	norm	0.81	1.18	0.72	0.94	1.17	1.54	1.10	1.28
$\phi = -.5$	chisq	0.51	0.66	0.45	0.53	4.32	3.49	3.95	3.35
	dblexp	0.47	0.84	0.44	0.71	1.20	1.11	1.09	0.97
MA(∞)	norm	0.92	0.63	0.87	0.65	1.30	0.86	1.21	0.82
$\theta_j = (j+1)^{-2.5}$	chisq	0.92	0.69	0.89	0.66	5.87	5.50	5.62	5.15
	dblexp	0.52	0.56	0.50	0.54	1.37	0.89	1.30	0.84

Table 4.1: Root mean squared error of the MBB, SMBB, ETBB, and SETBB estimators for the quantile variance when the block size is chosen by the HHJ empirical method for models (i)–(iv) under innovation distributions (a), (b), and (c) for $n = 200$.

4.5.2 The Trimmed Mean

The α-trimmed mean, which is the mean of the middle $(1-2\alpha)100\%$ of the data values, corresponds to an L-functional (cf. Example 2, Section 4.3) given by

$$T(F) = (1-2\alpha)^{-1} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF(x) = (1-2\alpha)^{-1} \int_{\alpha}^{1-\alpha} F^{-1}(u) du$$

with a corresponding L-estimator $T(F_n) = (n-2[\alpha n])^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} X_{(i)}$, where $[x]$ denotes the integer part of x and $X_{(i)}$ is the ith order statistic of the observed data. Intuitively, since the α-trimmed mean approaches the median as α approaches 0.5, the bootstrap estimator of its variance should benefit from smoothing as in the quantile case. A simulation study by Künsch (1989), under settings from Carlstein (1986), of the performance of the jackknife for estimating the variance of the 20%-trimmed mean is replicated and expanded upon here, and the SETBB again demonstrates a marked improvement over the ETBB.

The observations were generated from an AR(1) model with $\phi = 0.8$, with innovations from a mixture of normal distributions such that $e_t \sim (.7)\text{Normal}(0,1) + (.3)\text{Normal}(0,10)$. The MSEs of the variance estimators for the α-trimmed means for $\alpha = 0.1$, 0.2, and 0.3 were computed for the MBB, SMBB, ETBB, and SETBB methods. The three panels of Figure 4.3 display the MSEs achieved by the four bootstrap methods across the block sizes $\ell = 1, \ldots, 23$ for the three choices of α. Each panel shows a reduction in MSE across all block sizes due to smoothing.

Figure 4.4 depicts the performance of the four block bootstrap methods for the $\alpha = 0.2$ case when the block size was chosen according to the HHJ empirical method. As in the case of the sample quantiles, the MSE achieved when using the HHJ-selected block size is seen to be much lower for the smoothed block bootstrap methods than for their unsmoothed counterparts.

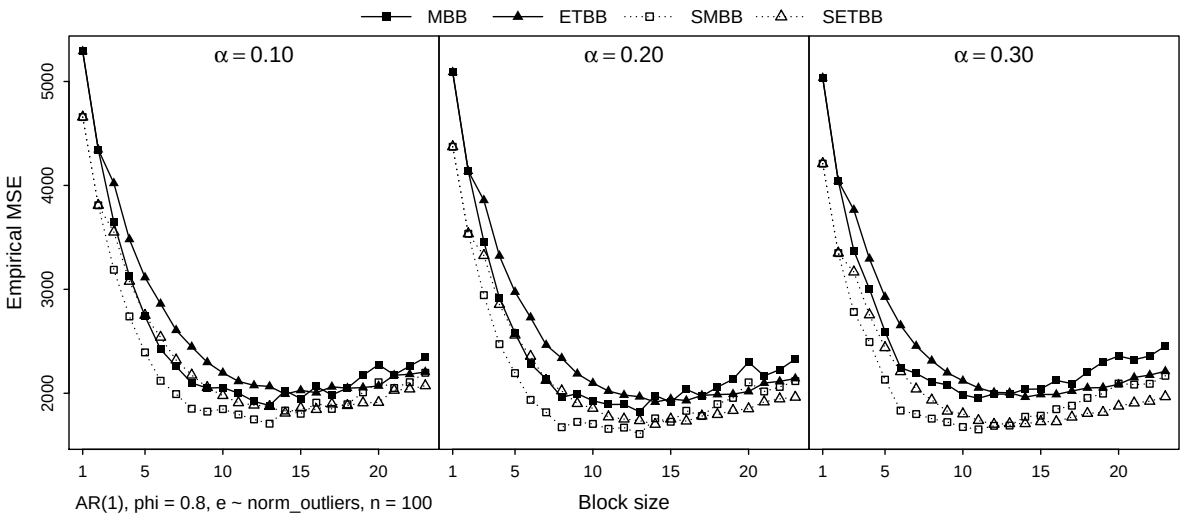

Figure 4.3: MSE achieved by the MBB, SMBB, ETBB, and SETBB estimators of the variance of the α-trimmed mean for $\alpha = 0.1, 0.2, 0.3$ of a length $n = 100$ realization of an AR(1) process with $\phi = 0.8$ and $e_t \sim (.7)\text{Normal}(0, 1) + (.3)\text{Normal}(0, 10)$.

4.6 Conclusions

We have attempted to address a methodological gap where smoothing to improve bootstraps for time series has received little consideration, which contrasts largely to the independent data case. To this end, we proposed a smooth extended tapered block bootstrap (SETBB) based on data smoothing modifications to the (extended) tapered block bootstrap (a general bootstrap for time series that has advantages over other first generation block bootstrap variants). The SETBB method mimics the iid smooth bootstrap by smoothing/augmenting a time series data set with independent random variables drawn from a kernel density (e.g., standard normal) with a bandwidth parameter, prior to applying (block) resampling steps. The purpose of such smoothing within resampling mechanics is to provide improvements to bootstrap distributional approximations, particularly for statistics (e.g., sample quantiles) with distributions depending on unknown, smooth process quantities such as marginal densities.

The SETBB was shown to provide valid inference in estimating the sampling distri-

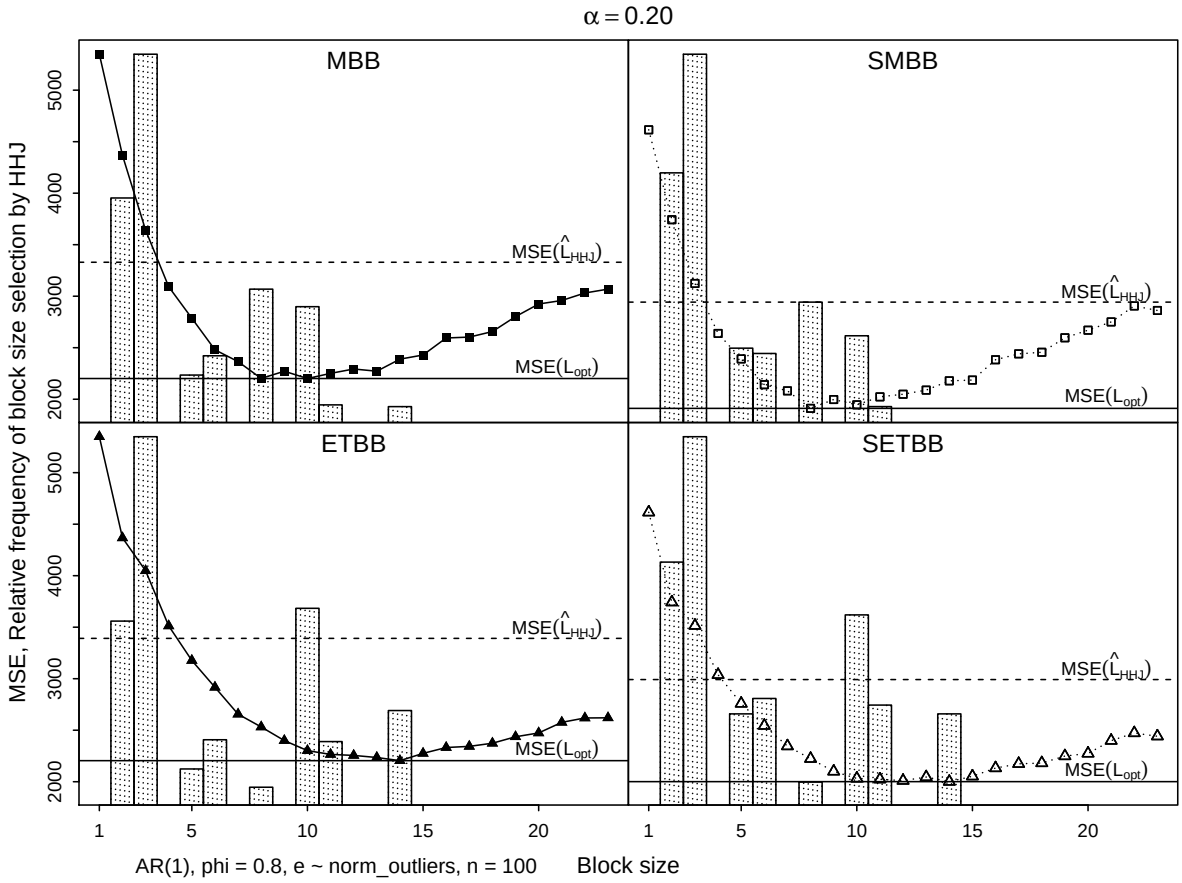

Figure 4.4: MSE achieved by the MBB, SMBB, ETBB, and SETBB estimators of the variance of the 20%-trimmed mean of a length $n = 100$ realization of an AR(1) process with $\phi = 0.8$ and $e_t \sim (.7)\text{Normal}(0, 1) + (.3)\text{Normal}(0, 10)$. The MSE at the optimal block size and when the HHJ-selected block size is used are shown as well as the selection frequency of each block size.

bution of a large class of time series statistics framed in terms of statistical functions. Hence, the formal validity of the SETBB method has been established in a context beyond previous treatments of the tapered block bootstrap, expanding the applicability of the bootstrap for time series inference. The improved performance of the SETBB over unsmooth bootstrap counterparts was also supported by several numerical studies.

Open questions remain concerning the best selection of block lengths, bandwidths and kernel densities for the SETBB approach to achieve optimal convergence rates and coverage accuracy. For concreteness, we have focused on real-valued time series in our development. We anticipate that the SETBB method applies equally to multivariate time series with similar improvements, but the vector-valued case requires further technical work and investigation.

5. SUMMARY

In this work, a novel method for testing equality of mean vectors from two populations in the large-p-small-n setting was introduced. It performed well under the assumption of a serial dependence structure, of which two examples—copy number data from two patient groups and a time series of mitochondrial concentration of Ca^{2+} gathered through the course of an hour from cardiac tissue in mice—were given.

A power-increasing multiple testing procedure for a large number of two-sample univariate equal-means hypotheses was also presented. This was an adaptation of the procedure introduced by Fan et al. (2012), which suggested estimating and removing the effects of latent factors from the test statistics. The serial structure of the dependence in our setting allowed reliable estimation of the covariance matrix of the test statistics, given that a Toeplitz structure could be assumed. The effects of harmonic factors or factors defined via eigendecomposition of the Toeplitz covariance matrix could then be estimated and removed from the test statistics. Gains in power from the procedure were established theoretically as well as demonstrated in simulation.

Lastly, a smooth version of the extended tapered blocks bootstrap from Shao (2010) was introduced and its consistency was proven for a broader class of statistics than originally considered. Simulation studies showed that our smoothing step substantially improves estimation of the sampling variance of quantiles and the trimmed mean.

REFERENCES

Anderson, O. D. (1977). *Time Series Analysis and Forecasting: The Box-Jenkins Approach*. 19 Cummings Park, Woburn, MA, 01801: Butterworths.

Andrews, B. (2008). Rank-based estimation for autoregressive moving average time series models. *Journal of Time Series Analysis* **29**, 51–73.

Athreya, K. B. & Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. New York: Springer.

Bai, Z. & Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica* **6**, 311–329.

Baladandayuthapani, V., Ji, Y., Talluri, R., Nieto-Barajas, L. E. & Morris, J. S. (2010). Bayesian random segmentation models to identify shared copy number aberrations for array cgh data. *Journal of the American Statistical Association* **105**, 1358–1375.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300.

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188.

Bickel, P. J. & Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196–1217.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory*. Holden-Day series in time series analysis. Holden-Day.

Brockwell, P. & Davis, R. (2009). *Time Series: Theory and Methods*. Springer Series in Statistics. New York: Springer.

Bustos, O. H. (1982). General m-estimates for sontaminated pth-order autoregres-

sive processes: consistency and asymptotic normality. *Wahrscheinlichkeitstheor. Verw. Geb.* , 491–504.

Bustos, O. H. & Yohai, V. J. (1986). Robust estimates for arma models. *Journal of the American Statistical Association* **81**, 155–168.

Cai, T., Liu, W. & Luo, X. (2011). A constrained l-1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106**, 594–607.

Cai, T., Liu, W. & Xia, Y. (2013a). Two-sample covariance matrix testing and support recovery. *Journal of the American Statistical Association* **108**, 265–277.

Cai, T. T., Liu, W. & Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Methodological)* **76**, 349–372.

Cai, T. T., Ren, Z. & Zhou, H. H. (2013b). Optimal rates of convergence for estimating toeplitz covariance matrices. *Probability Theory and Related Fields* **156**, 101–143.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The Annals of Statistics* **14**, 1171–1179.

Chen, S. X. & Qin, Y. L. (2010). A two sample test for high dimensional data with applications to gene-set testing. *The Annals of Statistics* **38**, 808–835.

Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics*, vol. 85. New York: Springer-Verlag.

Efron, B. (2009). Are a set of microarrays independent of each other? *The Annals of Applied Statistics* **3**, 922–942.

Efron, B. (2010a). Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association* **105**, 1042–1055.

Efron, B. (2010b). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics monographs. Cambridge, UK: Cambridge University Press.

Falk, M. & Reiss, R. D. (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *The Annals of Probability* **17**, 362–372.

Fan, J., Han, X. & Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association* **107**, 1019–1035.

Febrero-Bande, M. & Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software* **51**, 1–28.

Fernholz, L. T. (1983). *Von Mises Calculus for Statistical Functionals, Vol 19 of Lecture Notes in Statistics*. New York: Springer-Verlag.

Giordan, M. (2014). A two-stage procedure for the removal of batch effects in microarray studies. *Statistics in Biosciences* **6**, 73–84.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hall, P., DiCiccio, T. J. & Romano, J. P. (1989). On smoothing and the bootstrap. *The Annals of Statistics* **17**, 692–704.

Hall, P., Horowitz, J. L. & Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574.

Hall, P., Jing, B.-Y. & Lahiri, S. N. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica* **8**, 1189–1204.

Hallin, M. & Puri, M. L. (1991). Time series analysis via rank order theory:signed rank tests for arma models. *Journal of Multivariate Analysis* **39**, 1–29.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69**.

Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35**, 73–101.

Huber, P. (1981). *Robust Statistics*. New York: Wiley.

Kim, Y. M. & Nordman, D. J. (2011). Large sample properties of a block bootstrap

method under long-range dependence. *Sankhya: Series A* **73**, 79–109.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics* **17**, 1217–1241.

Lahiri, S. N. (1993). On the moving block bootstrap under long range dependence. *Statistics & Probability Letters* **18**, 405–413.

Lahiri, S. N. (2003a). Central limit theorems for weighted sums of spatial processes under a class of stochastic and fixed designs. *Sankhya: Series A* **65**, 356–388.

Lahiri, S. N. (2003b). *Resampling Methods for Dependent Data*. New York: Springer.

Leonov, V. P. & Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory of Probability and Its Applications* **4**, 319–329.

Liu, R. Y. & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap*, R. LePage & L. Billard, eds. New York: John Wiley & Sons, pp. 225–248.

Martin, R. D. & Yohai, V. J. (1986). Influence functionals for time series. *The Annals of Statistics* **14**, 781–855.

Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics* **5**, 557–572.

Pang, H., Liu, H. & Vanderbei, R. (2013). *fastclime: A fast solver for parameterized lp problems and constrained l1 minimization approach to sparse precision matrix estimation*. R package version 1.2.3.

Paparoditis, E. & Politis, D. N. (2001). Tapered block bootstrap. *Biometrika* **88**, 1105–1119.

Paparoditis, E. & Politis, D. N. (2002). The tapered block bootstrap for general statistics from stationary sequences. *Econometrics Journal* **5**, 131–148.

Pinkel, D. & Albertson, D. G. (2005). Array comparative genomic hybridization and its applications in cancer. *Nature Genetics Supplement* **37**, S11–S17.

Politis, D. & Romano, J. P. (1992). A circular block resampling procedure for stationary data. In *Exploring the limits of bootstrap*, R. LePage & L. Billard, eds. New York: Wiley, pp. 263–270.

Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89**, 1303–1313.

Politis, D. N. & Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* **16**, 67–104.

Ren, J. J. & Sen, P. K. (1991). On Hadamard differentiability of extended statistical functionals. *Journal of Multivariate Analysis* **39**, 30–43.

Ren, J. J. & Sen, P. K. (1995). Hadamard differentiability on $d[0,1]^p$. *Journal of Multivariate Analysis* **55**, 14–28.

Ruiz-Meana, M., Garcia-Dorado, D., Pina, P., Inserte, J., Agulló, L. & Soler-Soler, J. (2003). Cariporide preserves mitochondrial proton gradient and delays atp depletion in cardiomyocytes during ischemic conditions. *Am J Physiol Heart Circ Physiol* **285**, H999–H1006.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Seshan, V. E. & Olshen, A. (2013). *DNAcopy: DNA copy number data analysis*. R package version 1.36.0.

Shao, J. (1993). Differentiability of statistical functionals and consistency of the jackknife. *The Annals of Statistics* **21**, 61–75.

Shao, J. (2003). *Mathematical Statistics*. New York: Springer, 2nd ed.

Shao, X. (2010). Extended tapered block bootstrap. *Statistica Sinica* **20**, 000–000.

Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)* , 683–690.

Srivastava, M. (2007). Multivariate theory for analyzing high dimensional data. *J.*

Japan Statist. Soc. **37**, 53–86.

Srivastava, M. S. & Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis* **115**, 204–216.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479–498.

Tran, L. T. (1988). Rank order statistics for time series models. *Ann. Inst. Statist. Math.* **40**, 247–260.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer, 4th ed. ISBN 0-387-95457-0.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89**, 82–93.

Wu, Y., Genton, M. C. & Stefanski, L. A. (2006). A multivariate two-sample test for small sample size and missing data. *Biometrics* **62**, 877–885.

APPENDIX A

PROOFS FOR THE TWO-SAMPLE TEST FOR EQUALITY OF MEANS IN HIGH DIMENSION

A.1 Proofs of Main Results

Proof of Theorem 1: By an adaptation of the big-block-little-block argument to the triangular array it can be shown that $p^{-1/2}\sum_{j=1}^{p}[t_{nj}^2 - Et_{nj}^2] \to \text{Normal}(0, \tau_\infty^2)$, where

$$\begin{aligned}\tau_\infty^2 &= \lim_{n\to\infty}\text{Var}(p^{-1/2}\textstyle\sum_{j=1}^{p}t_{nj}^2) = \lim_{n\to\infty}p^{-1}\sum_{k=0}^{p-1}\sum_{|j_1-j_2|=k}\text{Cov}(t_{nj_1}^2, t_{nj_2}^2)\\ &= \gamma(0) + 2\textstyle\sum_{k=1}^{\infty}\gamma(k), \end{aligned} \tag{A.1}$$

where $\gamma(k) = \lim_{n\to\infty}(p-k)^{-1}\sum_{j=1}^{p-k}\text{Cov}(t_{nj}^2, t_{n(j+k)}^2)$, $k \geq 0$. To prove (A.1), use the moment and α-mixing conditions to show that for any $M \geq 1$,

$$\begin{aligned}p^{-1}\sum_{k=M+1}^{p-1}\sum_{|j_1-j_2|=k}|\text{Cov}(t_{nj_1}, t_{nj_2})| &\leq 2\sum_{k>M}p^{-1}(p-k)\{\alpha(k)^{\delta/(2+\delta)}\bigvee_{j=1}^{p}(E|t_{nj}|^{2+\delta})^{\frac{2}{2+\delta}}\}\\ &\leq C\sum_{k=M+1}^{\infty}\alpha(k)^{\delta/(2+\delta)} \to 0\end{aligned}$$

as $M \to \infty$. Thus,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|P(\sqrt{p}[T_n - p^{-1}\textstyle\sum_{j=1}^{p}E(t_{nj}^2)] \leq x) - \Phi(x/\tau_\infty)| = o(1)\\ \Longrightarrow\quad &\sup_{x\in\mathbb{R}}|P(T_n - p^{-1}\textstyle\sum_{j=1}^{p}E(t_{nj}^2) \leq x) - \Phi(\sqrt{p}x/\tau_\infty)| = o(1)\\ \Longrightarrow\quad &\sup_{x\in\mathbb{R}}|P(T_n - 1 \leq x) - \Phi(\sqrt{p}[x - n^{-1}a_n - n^{-2}b_n]/\tau_\infty)| = o(1),\end{aligned}$$

where a_n and b_n are bounded sequences such that

$$p^{-1}\textstyle\sum_{j=1}^{p}E(t_{nj}^2) = 1 + n^{-1}a_n + n^{-2}b_n + O(n^{-3}). \tag{A.2}$$

Lemma 1 provides c_{nj} and d_{nj} for $j = 1, \ldots, p$ such that $a_n = (c_{n1} + \cdots + c_{np})/p$ and $b_n = (d_{n1} + \cdots + d_{np})/p$ satisfy (A.2).

Lemma 1 *Let $X_{1j}, \ldots, X_{nj}$ and $Y_{1j}, \ldots, Y_{mj}$ be independent identically distributed random samples with $Var(X_{1j}) = \sigma_{1j}^2$ and $Var(Y_{1j}) = \sigma_{2j}^2$ and $EX_{1j} = EY_{1j}$ for all $j = 1, \ldots, p$. Assume that $\max\{E|X_{1j}|^{16}, E|Y_{1j}|^{16}, j = 1, \ldots, p\} = O(1)$ and that $\min\{\sigma_{1j}^2, \sigma_{2j}^2\} > c > 0$ (The first moment condition may be reduced further by means of truncation, but this would considerably lengthen the proof. The discussion of heteroscedasitiy in Section 2.4.4 illustrates the importance of bounding the component variances away from zero). Let $t_{nj}^2 = n(\overline{X}_{nj} - \overline{Y}_{mj})^2\{s_{nj}^2 + (n/m)\vartheta_{mj}^2\}^{-1}$, where s_n^2 and ϑ_m^2 are the two sample variances and let $m \sim n$ as $n \to \infty$. Then $E(t_{nj}^2) = 1 + n^{-1}c_{nj} + n^{-2}d_{nj} + O(n^{-3})$ for*

$$c_{nj} = \tau_{nj}^{-2}\{\sigma_{1j}^2 + (n/m)^2\sigma_{2j}^2\} + 2\tau_{nj}^{-6}\{\mu'_{3j} + (n/m)^2\eta'_{3j}\}^2 \tag{A.3}$$

and

$$\begin{aligned} d_{nj} = {} & \tau_{nj}^{-4}[\{\sigma_{1j}^2 + (n/m)^2\sigma_{2j}^2\} - \{(\mu'_{4j} - 3\sigma_{1j}^4) + (n/m)^4(\eta'_{4j} - 3\sigma_{2j}^4)\}] \\ & + \tau_{nj}^{-6}\{\sigma_{1j}^2 + (n/m)^2\sigma_{2j}^2\}\{(\mu'_{4j} - \sigma_{1j}^4) + (n/m)^3(\eta'_{4j} - \sigma_{2j}^4)\} \\ & - 4\tau_{nj}^{-6}\{\mu'_{3j} + (n/m)^2\eta'_{3j}\}\{\mu'_{3j} + (n/m)^3\eta'_{3j}\} \\ & - 2\tau_{nj}^{-6}\{(\mu'_{3j})^2 + (n/m)^5(\eta'_{3j})^2\} \\ & - 6\tau_{nj}^{-8}\{\mu'_{3j} + (n/m)^2\eta'_{3j}\}\{\mu'_{5j} - 2\mu'_{3j}\sigma_{1j}^2 + (n/m)^4(\eta'_{5j} - 2\eta'_{3j}\sigma_{2j}^2)\} \\ & - 3\tau_{nj}^{-8}\{(\mu'_{4j} - \sigma_{1j}^4) + (n/m)^3(\eta'_{4j} - \sigma_{2j}^4)\}^2 \\ & + 6\tau_{nj}^{-8}\{\sigma_{1j}^2 + (n/m)^2\sigma_{2j}^2\}\{\mu'_{3j} + (n/m)^2\eta'_{3j}\}^2 \\ & + 3\tau_{nj}^{-10}\{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}\{(\mu'_{4j} - \sigma_{1j}^4) + (n/m)^3(\eta'_{4j} - \sigma_{2j}^4)\}^2 \\ & + 12\tau_{nj}^{-10}\{\mu'_{3j} + (n/m)^2\eta'_{3j}\}^2\{(\mu'_{4j} - \sigma_{1j}^4) + (n/m)^3(\eta'_{4j} - \sigma_{2j}^4)\}, \end{aligned} \tag{A.4}$$

where $\tau_{nj}^2 = \{\sigma_{1j}^2 + (n/m)\sigma_{2j}^2\}$ and μ'_{kj} and η'_{kj} are the kth central moments of X_{1j} and

Y_{1j}, *respectively.*

Proof of Lemma 1: For ease of syntax, ignore the subscript j, and, without loss of generality, assume that $EX_{1j} = EY_{1j} = 0$. Let $\Delta_n = s_n^2 - \sigma_1^2 + (n/m)(\vartheta_m^2 - \sigma_2^2)$ and let t_n^2 be approximated by the expansion

$$\tilde{t}_n^2 = n(\overline{X}_n - \overline{Y}_m)^2(\tau_n^{-2} - \tau_n^{-4}\Delta_n + \tau_n^{-6}\Delta_n^2 - \tau_n^{-8}\Delta_n^3 + \tau_n^{-10}\Delta_n^4), \tag{A.5}$$

so that $t_n^2 = \tilde{t}_n^2 + O_p(n^{-3})$. An expression for the expected value $E(\tilde{t}_{nj}^2)$ would thus involve the quantities $n\tau_n^{-2k}E(\overline{X}_n - \overline{Y}_m)^2\Delta_n^{k-1}$ for $k = 1, \ldots, 5$. These expectations must be computed such that they retain terms out to the order of $O(n^{-3})$.

Let $\chi_{|B|}(\{X_j : j \in B\})$ represent the joint cumulant of the random variables in the set $\{X_j : j \in B\}$, where $|B|$ is the cardinality of B. Then the formula

$$E(X_1 \ldots X_k) = \Sigma_\pi \Pi_{B \in \pi} \chi_{|B|}(\{X_j : j \in B\}) \tag{A.6}$$

from Leonov & Shiryaev (1959) gives the expected value of a product of random variables in terms of joint cumulants, where Σ_π denotes summation over all possible partitions of $\{X_1, \ldots, X_k\}$, and $\Pi_{B \in \pi}$ denotes the product over all cells of the partition π. Using (A.6) to compute $E(\overline{X}_n - \overline{Y}_m)^2\Delta_n^{k-1}$ to within $O(n^{-4})$of their true values for $k = 1, \ldots, 5$ involves the joint cumulants tabulated below, where $\Delta \equiv \Delta_n$, $\overline{X} \equiv \overline{X}_n$, and $\overline{Y} \equiv \overline{Y}_m$.

	0	1	2
0		$\chi_1(\overline{X} - \overline{Y})$	$\chi_2(\overline{X} - \overline{Y}, \overline{X} - \overline{Y})$
1	$\chi_1(\Delta)$	$\chi_2(\Delta, \overline{X} - \overline{Y})$	$\chi_3(\Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y})$
2	$\chi_2(\Delta, \Delta)$	$\chi_3(\Delta, \Delta, \overline{X} - \overline{Y})$	$\chi_4(\Delta, \Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y})$
3	$\chi_3(\Delta, \Delta, \Delta)$	$\chi_4(\Delta, \Delta, \Delta, \overline{X} - \overline{Y})$	$\chi_5(\Delta, \Delta, \Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y})$

If $\kappa(i, j)$ denotes the ijth member of the table of joint cumulants, then (A.6) gives

$$E(\overline{X} - \overline{Y})^2 \;=\; \kappa(0, 2) + O(n^{-4}) \tag{A.7}$$

$$
\begin{aligned}
E(\overline{X}-\overline{Y})^2\Delta &= \kappa(1,2)+\kappa(0,2)\kappa(1,0)+O(n^{-4}) && \text{(A.8)}\\
E(\overline{X}-\overline{Y})^2\Delta^2 &= \kappa(2,2)+2\kappa(1,0)\kappa(1,2)+\kappa(0,2)\kappa(2,0)\\
&\quad +2\kappa^2(1,1)+\kappa(0,2)\kappa^2(1,0)+O(n^{-4}) && \text{(A.9)}\\
E(\overline{X}-\overline{Y})^2\Delta^3 &= \kappa(0,2)\kappa(3,0)+6\kappa(1,1)\kappa(2,1)\\
&\quad +3\kappa(2,0)\kappa(1,2)+3\kappa(1,0)\kappa(2,0)\kappa(0,2)\\
&\quad +6\kappa(1,0)\kappa^2(1,1)+O(n^{-4}) && \text{(A.10)}\\
E(\overline{X}-\overline{Y})^2\Delta^4 &= 3\kappa(0,2)\kappa^2(2,0)+12\kappa^2(1,1)\kappa(2,0)+O(n^{-4}),
\end{aligned}
$$

after removing cumulant products of order smaller than $O(n^{-4})$ and noting that $\kappa(0,1)=0$.

Each cumulant is simplified using rules found in Brillinger (1981), and the formula

$$
\chi_k(X_1,\ldots,X_k)=\Sigma_\pi(-1)^{(|\pi|-1)}(|\pi|-1)!\,\Pi_{B\in\pi}E(\Pi_{i\in B}X_i) \qquad \text{(A.11)}
$$

from Leonov & Shiryaev (1959) provides expressions for the simplified cumulants in terms of moments. The cumulants are computed below, where each cumulant is either given exactly, or is approximated to the order necessary for the cumulant products in (A.7)–(A.11) to lie within $O(n^{-4})$ of their true values.

$$
\begin{aligned}
\kappa(0,1) &= \chi_1(\overline{X}-\overline{Y})=E(\overline{X}-\overline{Y})=0\\
\kappa(0,2) &= \chi_2(\overline{X}-\overline{Y},\overline{X}-\overline{Y})=E(\overline{X}-\overline{Y})^2-\{E(\overline{X}-\overline{Y})\}^2\\
&= n^{-1}\{\sigma_1^2+(n/m)\sigma_2^2\}\\
\kappa(1,0) &= \chi_1(\Delta)=E\{(s^2-\sigma_1^2)+(n/m)(\vartheta^2-\sigma_2^2)\}=-\{\sigma_1^2/n-(n/m)\sigma_2^2/m\}\\
&= -n^{-1}\{\sigma_1^2+(n/m)^2\sigma_2^2\}\\
\kappa(1,1) &= \chi_2(\Delta,\overline{X}-\overline{Y})=\chi_2(\overline{X^2}-\overline{X}^2,\overline{X})+(n/m)\chi_2(\overline{Y^2}-\overline{Y}^2,\overline{Y})\\
&= \chi_2(\overline{X^2},\overline{X})-\chi_2(\overline{X}^2,\overline{X})+(n/m)\chi_2(\overline{Y^2}-\overline{Y}^2,\overline{Y})
\end{aligned}
$$

$$
\begin{aligned}
&= n^{-1}\chi_2(X_1^2, X_1) - n^{-3}\chi_2(\Sigma_i X_i^2 + \Sigma_{i\neq j} X_i X_j, \Sigma_i X_i) + (n/m)\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y}) \\
&= n^{-1}\mu_3' - n^{-2}\chi_2(X_1^2 + X_1\Sigma_{j=2}^n X_j, X_1) + (n/m)\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y}) \\
&= (n^{-1} - n^{-2})\mu_3' + n^{-2}(n-1)\chi_2(X_1X_2, X_1) + (n/m)\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y}) \\
&= (n^{-1} - n^{-2})\mu_3' + (n/m)(m^{-1} - m^{-2})\eta_3' \\
\kappa(1,2) &= \chi_3(\Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y}) = \chi_3(\overline{X^2} - \overline{X}^2, \overline{X}, \overline{X}) + (n/m)\chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y}, \overline{Y}) \\
&= n^{-2}\chi_3(X_1^2, X_1, X_1) - n^{-3}\chi_3(X_1^2, X_1, X_1) - n^{-4}\chi_3(\Sigma_i X_i X_j, \Sigma_i X_i, \Sigma_i X_i) \\
&\quad + (n/m)\chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y}, \overline{Y}) \\
&= (n^{-2} - n^{-3})(\mu_4' - \sigma_1^4) + (n/m)(m^{-2} - m^{-3})(\eta_4' - \sigma_2^4) \\
\kappa(2,0) &= \chi_2(\Delta, \Delta) = \chi_2(\overline{X^2} - \overline{X}^2, \overline{X^2} - \overline{X}^2) + (n/m)^2\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= \chi_2(\overline{X^2}, \overline{X^2}) - 2\chi_2(\overline{X^2}, \overline{X}^2) + (\overline{X}^2, \overline{X}^2) + (n/m)^2\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= n^{-1}\chi_2(X_1^2, X_1^2) - 2n^{-3}\{\chi_2(\Sigma_i X_i^2, \Sigma_i X_i^2) + \chi_2(\Sigma_{i\neq j} X_i X_j, \Sigma_i X_i^2)\} \\
&\quad + n^{-4}\{\chi_2(\Sigma_i X_i^2, \Sigma_i X_i^2) - 2\chi_2(\Sigma_{i\neq j} X_i X_j, \Sigma_i X_i^2) \\
&\quad + \chi_2(\Sigma_{i\neq j} X_i X_j, \Sigma_{i\neq j} X_i X_j)\} + (n/m)^2\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= (n^{-1} - 2n^{-2} + n^{-3})(\mu_4' - \sigma_1^4) + n^{-4}\chi_2(\Sigma_{i\neq j} X_i X_j, \Sigma_{i\neq j} X_i X_j) \\
&\quad + (n/m)^2\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= (n^{-1} - 2n^{-2} + n^{-3})(\mu_4' - \sigma_1^4) + \frac{2(n-1)}{n^3}\sigma_1^4 + (n/m)^2\chi_2(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= n^{-3}(n-1)^2\mu_4' - n^{-3}(n-1)(n-3)\sigma_1^4 \\
&\quad + (n/m)^2\{m^{-3}(m-1)^2\eta_4' - m^{-3}(m-1)(m-3)\sigma_2^4\} \\
&= (n^{-1} - 2n^{-2})\mu_4' - (n^{-1} - 4n^{-2})\sigma_1^4 \\
&\quad + (n/m)^2\{(m^{-1} - 2m^{-2})\eta_4' - (m^{-1} - 4m^{-2})\sigma_2^4\} + O(n^{-3}) \\
\kappa(2,1) &= \chi_3(\Delta, \Delta, \overline{X} - \overline{Y}) \\
&= \chi_3(\overline{X^2} - \overline{X}^2, \overline{X^2} - \overline{X}^2, \overline{X}) + (n/m)^2\chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2, \overline{Y}) \\
&= \chi_3(\overline{X^2}, \overline{X^2}, \overline{X}) + (n/m)^2\chi_3(\overline{Y^2}, \overline{Y^2}, \overline{Y}) + O(n^{-3}) \\
&= n^{-2}\chi_3(X_1^2, X_1^2, X_1) + (n/m)^2 m^{-2}\chi_3(Y_1^2, Y_1^2, Y_1) + O(n^{-3})
\end{aligned}
$$

$$
\begin{aligned}
&= n^{-2}(\mu_5' - 2\mu_3'\sigma_1^2) + (n/m)^2 m^{-2}(\eta_5' - 2\eta_3'\sigma_2^2) + O(n^{-3}) \\
&= n^{-2}\{(\mu_5' - 2\mu_3'\sigma_1^2) + (n/m)^4(\eta_5' - 2\eta_3'\sigma_2^2)\} + O(n^{-3}) \\
\kappa(2,2) &= \chi_4(\Delta, \Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y}) \\
&= \chi_4(\overline{X^2} - \overline{X}^2, \overline{X^2} - \overline{X}^2, \overline{X}, \overline{X}) + (n/m)^2 \chi_4(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2, \overline{Y}, \overline{Y}) \\
&= n^{-3}\chi_4(X_1^2, X_1^2, X_1, X_1) + (n/m)^2 m^{-3}\chi_4(Y_1^2, Y_1^2, Y_1, Y_1) + O(n^{-4}) \\
&= n^{-3}[\mu_6' - 3\sigma_1^2\mu_4' - 2(\mu_3')^2 + 2\sigma_1^6 + (n/m)^5\{\eta_6' - 3\sigma_2^2\eta_4' - 2(\eta_3')^2 + 2\sigma_2^6\}] \\
&\quad + O(n^{-3}) \\
\kappa(3,0) &= \chi_3(\Delta, \Delta, \Delta) = \chi_3(\overline{X^2} - \overline{X}^2, \overline{X^2} - \overline{X}^2, \overline{X^2} - \overline{X}^2) \\
&\quad + (n/m)^3 \chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= \chi_3(\overline{X^2}, \overline{X^2}, \overline{X^2}) - 3\chi_3(\overline{X^2}, \overline{X^2}, \overline{X}^2) \\
&\quad + 3\chi_3(\overline{X^2}, \overline{X}^2, \overline{X}^2) + \chi_3(\overline{X}^2, \overline{X}^2, \overline{X}^2) \\
&\quad + (n/m)^3 \chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) \\
&= n^{-2}\chi_3(X_1^2, X_1^2, X_1^2) \\
&\quad + (n/m)^3 \chi_3(\overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2, \overline{Y^2} - \overline{Y}^2) + O(n^{-3}) \\
&= n^{-2}\{(\mu_6' - 3\sigma_1^2\mu_4' + 2\sigma_1^6) + (n/m)^5(\eta_6' - 3\sigma_2^2\eta_4' + 2\sigma_2^6)\} + O(n^{-3}) \\
\kappa(3,2) &= \chi_5(\Delta, \Delta, \Delta, \overline{X} - \overline{Y}, \overline{X} - \overline{Y}) = O(n^{-4})
\end{aligned}
$$

Plugging the above expressions into (A.7)–(A.11) and dropping terms of smaller order than $O(n^{-3})$ yields

$$
\begin{aligned}
n\tau_n^{-2} E(\overline{X}_n - \overline{Y}_m)^2 &= 1 \\
n\tau_n^{-4} E(\overline{X}_n - \overline{Y}_m)^2 \Delta_n &= n^{-1}\tau_n^{-4}[(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)] \\
&\quad - n^{-1}\tau_n^{-2}[\sigma_1^2 + (n/m)^2\sigma_2^2] \\
&\quad - n^{-2}\tau_n^{-4}[(\mu_4' - \sigma_1^4) + (n/m)^4(\eta_4' - \sigma_2^4)] \\
n\tau_{m,n}^{-6} E(\overline{X}_n - \overline{Y}_m)^2 \Delta_n^2 &= n^{-1}\tau_n^{-4}[(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)]
\end{aligned}
$$

$$
\begin{aligned}
&+ n^{-2}\tau_n^{-4}[\sigma_1^2 + (n/m)^2\sigma_2^2]^2 \\
&- n^{-2}\tau_n^{-4}[(2\mu_4' - 4\sigma_1^4) + (n/m)^4(2\eta_4' - 4\sigma_1^4)] \\
&+ 2n^{-1}\tau_n^{-6}[\mu_3' + (n/m)^2\eta_3']^2 \\
&- 4n^{-2}\tau_n^{-6}[\mu_3' + (n/m)^2\eta_3'][\mu_3' + (n/m)^3\eta_3'] \\
&- 2n^{-2}\tau^{-6}[\sigma_1^2 + (n/m)^2\sigma_2^2][(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)] \\
&+ n^{-2}\tau_n^{-6}[\mu_6' - 3\sigma_1^2\mu_4' - 2(\mu_3')^2 + 2\sigma_1^6] \\
&+ n^{-2}\tau_n^{-6}(n/m)^5[\eta_6' - 3\sigma_2^2\eta_4' - 2(\eta_3')^2 + 2\sigma_2^6] \\
&+ O(n^{-3})
\end{aligned}
$$

$$
\begin{aligned}
n\tau_n^{-8}E(\overline{X}_n - \overline{Y}_m)^2\Delta_n^3 &= n^{-2}\tau_n^{-6}[\mu_6' - 3\sigma_1^2\mu_4' + 2\sigma_1^6 + (n/m)^5(\eta_6' - 3\sigma_2^2\eta_4' + 2\sigma_2^6)] \\
&\quad - 3n^{-2}\tau_n^{-6}[\sigma_1^2 + (n/m)^2\sigma_2^2][(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)] \\
&\quad + 6n^{-2}\tau_n^{-8}[\mu_3' + (n/m)^2\eta_3'] \\
&\qquad \times [(\mu_5' - 2\mu_3'\sigma_1^2) + (n/m)^4(\eta_5' - 2\eta_3'\sigma_2^2)] \\
&\quad + 3n^{-2}\tau_n^{-8}[(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)]^2 \\
&\quad - 6n^{-2}\tau_n^{-8}[\sigma_1^2 + (n/m)^2\sigma_2^2][\mu_3' + (n/m)^2\eta_3']^2 \\
&\quad + O(n^{-3})
\end{aligned}
$$

$$
\begin{aligned}
n\tau_n^{-10}E(\overline{X}_n - \overline{Y}_m)^2\Delta_n^4 &= 3n^{-2}\tau_n^{-10}[\sigma_1^2 + (n/m)\sigma_2^2][(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)]^2 \\
&\quad + 12n^{-2}\tau_n^{-10}[\mu_3' + (n/m)^2\eta_3']^2[(\mu_4' - \sigma_1^4) + (n/m)^3(\eta_4' - \sigma_2^4)] \\
&\quad + O(n^{-3}).
\end{aligned}
$$

Adding and subtracting these quantities according to the expansion in (A.5) and gathering terms out of which n^{-1} and n^{-2} can be factored yields c_n from (A.3) and d_n from (A.4), respectively, thus completing the proof.

A.2 A Central Limit Theorem for Strongly Mixing Bounded Random Variables

We here establish a central limit theorem for a triangular array of strongly mixing random variables which are bounded, which is simpler than in the unbounded case. This

illustrates the steps of the proof for the unbounded case. We shall need the following corollary as found in Athreya & Lahiri (2006), as well as Lemma 2 which follows.

Corollary 1 *Let X and Y be two random variables with $\alpha(\sigma\langle X\rangle, \sigma\langle Y\rangle) = \alpha \in [0,1]$.*

(i) (Davydov's inequality). Suppose that $E|X|^p < \infty$, $E|Y| < \infty$ for some $p, q \in (0,\infty)$ with $\frac{1}{p} + \frac{1}{q} < 1$. Then $E|XY| < \infty$ and

$$|Cov(X,Y)| \leq 2r(2\alpha)^{1/r}(E|X|^p)^{1/p}(E|Y|^q)^{1/q}, \tag{A.12}$$

where $\frac{1}{r} = 1 - (\frac{1}{p} + \frac{1}{q})$.

(ii) If $P(|X| \leq c_1) = 1 = P(|Y| \leq c_2)$ for some constants $c_1, c_2 \in (0,\infty)$, then

$$|Cov(X,Y)| \leq 4c_1c_2\alpha. \tag{A.13}$$

Lemma 2 *Suppose that the conditions of Theorem 3 hold. Then*

$$\sup\left\{E\left(\sum_{i=m}^{m+p_n-1} X_{ni}\right)^4 : 1 \leq m \leq r_n - p_n + 1\right\} = o(p_n^3) \tag{A.14}$$

for $p_n \in [\sqrt{r_n}, r_n]$ as $n \to \infty$ (which means $r_n \to \infty$ and thus also $p_n \to \infty$).

Proof of Lemma 2:

$$\begin{aligned}
E\left[\sum_{i=m}^{m+p_n-1} X_{ni}\right]^4 &= \sum_{i,j,k,l} EX_{ni}X_{nj}X_{nk}X_{nl} \\
&= \binom{4}{4}\sum_i EX_{ni}^4 + \binom{4}{3}\sum_{i\neq j} EX_{ni}^3X_{nj} + \binom{4}{2}\frac{1}{2}\sum_{i\neq j} EX_{ni}^2X_{nj}^2 \\
&\quad + \binom{4}{2}\sum_{i\neq j\neq k} EX_{ni}^2X_{nj}X_{nk} + \binom{4}{0}\sum_{i\neq j\neq k\neq l} EX_{ni}X_{nj}X_{nk}X_{nl} \\
&\equiv I_{1p_n} + I_{2p_n} + I_{3p_n} + I_{4p_n} + I_{5p_n},
\end{aligned}$$

where $1 \leq i, j, k, l \leq m - p_n + 1$. *Note that since* $P(X_{ni} < c) = 1$ *for all* $1 \leq i \leq r_n$, $n \geq 1$,

$$|I_{1p_n}|+|I_{2p_n}|+|I_{3p_n}| \leq p_n c^4 + 4p_n(p_n-1)c^4 + 3p_n(p_n-1)c^4 = 7p_n^2 c^4. \tag{A.15}$$

By Corollary 1 (ii),

$$\begin{aligned}
|I_{4p_n}| &= 2(6)\sum_{i<j<k}\left[EX_{ni}^2X_{nj}X_{nk} + EX_{ni}X_{nj}^2X_{nk} + EX_{ni}X_{nj}X_{nk}^2\right] \\
&= 12\sum_{i<j<k}\left[\left|Cov\left(X_{ni}^2X_{nj}, X_{nk}\right)\right| + \left|Cov\left(X_{ni}X_{nj}^2, X_{nk}\right)\right| + \left|Cov\left(X_{ni}, X_{nj}X_{nk}^2\right)\right|\right] \\
&\leq 12\sum_{i<j<k}\left[4c^4\alpha(k-j) + 4c^4\alpha(k-j) + 4c^4\alpha(j-i)\right] \\
&= 48c^4\sum_{i=1}^{p_n-2}\sum_{j=i+1}^{p_n-1}\sum_{k=j+1}^{p_n}\left[2\alpha(k-j) + \alpha(j-i)\right] \\
&= 48c^4\sum_{i=1}^{p_n-2}\sum_{s=1}^{p_n-1-i}\sum_{r=1}^{p_n-s-i}\left[2\alpha(r) + \alpha(s)\right] \\
&\leq 48c^4p_n\left(p_n\sum_{r=1}^{p_n-1}2\alpha(r) + p_n\sum_{s=1}^{p_n-1}\alpha(s)\right) \\
&= 144c^4p_n^2\sum_{r=1}^{p_n-1}\alpha(r).
\end{aligned} \tag{A.16}$$

Similarly, and by the monoticity of $\alpha(\cdot)$,

$$\begin{aligned}
|I_{5p_n}| &\leq \sum_{i\neq j\neq k\neq l}|EX_{ni}X_{nj}X_{nk}X_{nl}| \\
&= 4!\sum_{i<j<k<l}|EX_{ni}X_{nj}X_{nk}X_{nl}| \\
&= 24\sum_{i<j<k<l}|Cov(X_{ni}, X_{nj}X_{nk}X_{nl}) \wedge Cov(X_{ni}X_{nj}X_{nk}, X_{nl})| \\
&\leq 24(4)c^4\sum_{i<j<k<l}\alpha(j-i)\wedge\alpha(l-k) \\
&= 96c^4\sum_{i=1}^{p_n-3}\sum_{j=i+1}^{p_n-2}\sum_{k=j+1}^{p_n-1}\sum_{l=k+1}^{p_n}\alpha(j-i)\wedge\alpha(l-k)
\end{aligned}$$

$$
\begin{aligned}
&= 96c^4 \sum_{i=1}^{p_n-3} \sum_{s=1}^{p_n-2-i} \sum_{k=s+i+1}^{p_n-1} \sum_{r=1}^{p_n-k} \alpha(s) \wedge \alpha(r) \\
&\leq 96c^4 p_n^2 \sum_{s=1}^{p_n-1} \sum_{r=1}^{p_n-1} \alpha(s) \wedge \alpha(r) \\
&\leq 192c^4 p_n^2 \sum_{r=1}^{p_n-1} r\alpha(r) \\
&= 192c^4 p_n^2 \left[\sum_{r=1}^{\lfloor \sqrt{p_n} \rfloor} r\alpha(r) + \sum_{r=\lfloor \sqrt{p_n} \rfloor + 1}^{p_n-1} r\alpha(r) \right] \\
&= 192c^4 p_n^2 \left[p_n^{1/2} \sum_{r=1}^{\infty} \alpha(r) + p_n \sum_{r \geq \lfloor p_n^{1/2} \rfloor + 1} \alpha(r) \right] \\
&= o(p_n^3), \qquad \text{(A.17)}
\end{aligned}
$$

since $\sum_{r=1}^{\infty} \alpha(r) = O(1)$ *and* $\sum_{r \geq \lfloor p_n^{1/2} \rfloor + 1} \alpha(r) = o(1)$. *Thus by (A.15)-(A.17), (A.14) holds.*

Theorem 3 *Let* $\{X_{n1}, \ldots, X_{nr_n}\}_{n \geq 1}$ *be a triangular array of random variables on* $(\Omega_n, \mathcal{F}_n, P_n)$ *such that* $EX_{ni} = 0$ *and* $0 < EX_{ni}^2 < \infty$ *for* $1 \leq i \leq r_n, n \geq 1$. *Let* $S_n = X_{n1} + \cdots + X_{nr_n}$ *and* $s_n^2 = Var(S_n)$. *Let*

$$
\alpha_n(k) = sup \left\{ |P(A \cap B) - P(A)P(B)| \colon A \in \mathcal{F}_{1,m}^{(n)}, B \in \mathcal{F}_{m+k,r_n}^{(n)}, 1 \leq m \leq r_n - k \right\},
$$

where $\mathcal{F}_{i,j}^{(n)} = \sigma\langle \{X_{ni}, \ldots, X_{nj}\} \rangle$. *Suppose that there exists some* $c \in (0, \infty)$ *such that* $P(|X_{ni}| \leq c) = 1$ *for all* $1 \leq i \leq r_n, n \geq 1$ *and some* $\sigma_\infty^2 \in (0, \infty)$ *such that*

$$
\gamma_{p_n} \equiv \sup \left\{ \left| p_n^{-1} Var \left(\sum_{i=j}^{j+p_n-1} X_{ni} \right) - \sigma_\infty^2 \right| : 1 \leq j \leq r_n - p_n + 1 \right\} \to 0 \qquad \text{(A.18)}
$$

for any $p_n \in [\sqrt{r_n}, r_n]$ *as* $n \to \infty$. *Suppose also that there exists a function* $\alpha(\cdot) : \mathbb{N} \to$

$[0,1]$ *such that* $|\alpha_n(k)| < \alpha(k)$ *for all* $n \geq 1, k \geq 1$, *and that*

$$\sum_{k=1}^{\infty} \alpha(k) < \infty. \tag{A.19}$$

Then

$$\frac{S_n}{\sqrt{r_n}} \to^d N(0, \sigma_\infty^2). \tag{A.20}$$

Proof of Theorem 3: Let $p \equiv p_n$, $q \equiv q_n = \lfloor r_n^{1/2} \rfloor$, $n \geq 1$ be integers such that

$$q/p + p/r_n = o(1) \tag{A.21}$$

as $n \to \infty$. Let $m_n \equiv m = p + q$ and $K \equiv K_n = \lfloor r_n/m \rfloor$. Then, for $j = 1, \ldots, K$, let

$$\begin{aligned} B_{nj} &= \sum_{i=(j-1)m+1}^{(j-1)m+p} X_{ni} \\ L_{nj} &= \sum_{i=(j-1)m+p+1}^{jm} X_{ni} \\ R_{nr_n} &= \sum_{i=mK+1}^{r_n} X_{ni}. \end{aligned}$$

Since $q/p = o(1)$, the above provides a decomposition of the row sums S_n of the triangular array into sums of big blocks B_{nj}, little blocks L_{nj}, and a remainder term R_{nr_n} such that

$$\frac{1}{\sqrt{r_n}} S_n = \frac{1}{\sqrt{r_n}} \sum_{j=1}^{K} B_{nj} + \frac{1}{\sqrt{r_n}} \sum_{j=1}^{K} L_{nj} + \frac{1}{\sqrt{r_n}} R_{nr_n}. \tag{A.22}$$

It is first shown that the last two terms converge in probability to zero as $n \to \infty$. By Corollary 1 (ii),

$$E\left(R_{nr_n}/\sqrt{r_n}\right)^2 \leq r_n^{-1} \left(\sum_{i=mK+1}^{r_n} EX_{ni}^2 + \sum_{i \neq j} |EX_{ni}X_{nj}| \right)$$

$$
\begin{aligned}
&= r_n^{-1}\left(\sum_{i=mK+1}^{r_n}|\mathrm{Cov}(X_{ni},X_{ni})|+\sum_{i\neq j}|\mathrm{Cov}(X_{ni},X_{nj})|\right)\\
&\leq r_n^{-1}\left((r_n-mK)4c^2\alpha(0)+2\sum_{i=mK+1}^{r_n-1}\sum_{j=i+1}^{r_n}|\mathrm{Cov}(X_{ni},X_{nj})|\right)\\
&\leq r_n^{-1}\left((r_n-mK)4c^2+2\sum_{i=mK+1}^{r_n-1}\sum_{l=1}^{r_n-i}|\mathrm{Cov}(X_{ni},X_{n(i+l)})|\right)\\
&\leq r_n^{-1}\left((r_n-mK)4c^2+2(r_n-mK)4c^2\sum_{l=1}^{\infty}\alpha(l)\right)\\
&\leq r_n^{-1}(r_n-mK)8c^2\left(1+\sum_{l=1}^{\infty}\alpha(l)\right)\\
&= O\left(\frac{m}{r_n}\right)\to 0. \qquad (A.23)
\end{aligned}
$$

Note that between any random variables X_{ni} and X_{nl} involved in L_{nj} and $L_{n(j+k)}$, respectively, there are at least $(k-1)m+p\geq kp$ intermediate random variables, so that the maximum strong mixing coefficient between X_{ni} and X_{nl} cannot exceed $\alpha_n(kp)$, by the monotonicity of $\alpha_n(\cdot)$. Hence

$$
\begin{aligned}
E\left(\sum_{j=1}^{K}L_{nj}/\sqrt{r_n}\right)^2 &\leq r_n^{-1}\left(\sum_{j=1}^{K}EL_{nj}^2+\sum_{i\neq j}|EL_{ni}L_{nj}|\right)\\
&= r_n^{-1}\left(\sum_{j=1}^{K}EL_{nj}^2+2\sum_{l=1}^{K-1}\sum_{j=1}^{K-l}|\mathrm{Cov}(L_{nj},L_{n(j+l)})|\right)\\
&\leq r_n^{-1}\left(q\sum_{j=1}^{K}\frac{1}{q}\mathrm{Var}\left(\sum_{i=(j-1)m+p+1}^{jm}X_{ni}\right)+2\sum_{l=1}^{K-1}(K-l)4c^2q^2\alpha(lp)\right)\\
&\leq r_n^{-1}\left(Kq(\sigma_\infty^2+\gamma_q)+8Kc^2q^2\sum_{l=1}^{K-1}\sum_{j=1}^{p}\alpha(lp-j)/p\right)\\
&= r_n^{-1}\left(Kq(\sigma_\infty^2+\gamma_q)+Kq^2p^{-1}8c^2\sum_{l=1}^{\infty}\alpha(l)\right)\\
&= O\left(\frac{q}{p}\right)+O\left(\frac{q^2}{p^2}\right)\to 0 \qquad (A.24)
\end{aligned}
$$

as $n\to\infty$, since $pK/r_n\to 1$ as $n\to\infty$.

Now, as q increases, the big blocks grow further apart, and eventually the triangular array $\left\{B_{n1}/\sqrt{r_n}, \ldots, B_{nK}/\sqrt{r_n}\right\}_{n\geq 1}$ can be replaced with a triangular array of independent random variables $\{\tilde{B}_{n1}/\sqrt{r_n}, \ldots, \tilde{B}_{nK}/\sqrt{r_n}\}_{n\geq 1}$, such that $\tilde{B}_{nj} =^d B_{nj}$ for all $1 \leq j \leq K, n \geq 1$. Note that $\alpha(\sigma\langle B_{nj}\rangle, \sigma\langle\{B_{nl} : l \geq j+1\}\rangle) \leq \alpha(q)$, since the big blocks are separated from one another by no fewer than q random variables. Letting $Y_{nj} = \exp\left(\iota t B_{nj}/\sqrt{r_n}\right)$ for any $t \in \mathbb{R}$ and applying Corollary 1 (ii), it is seen that

$$\begin{aligned}
\left|E\prod_{j=1}^{K} Y_{nj} - \prod_{j=1}^{K} EY_{nj}\right| &\leq \left|E\prod_{j=1}^{K} Y_{nj} - EY_{n1}E\prod_{j=2}^{K} Y_{nj}\right| \\
&+ \left|EY_{n1}E\prod_{j=2}^{K} Y_{nj} - EY_{n1}EY_{n2}E\prod_{j=3}^{K} Y_{nj}\right| \\
&+ \left|EY_{n1}EY_{n2}E\prod_{j=3}^{K} Y_{nj} - \prod_{j=1}^{3} EY_{nj}E\prod_{j=4}^{K} Y_{nj}\right| \\
&+ \cdots + \left|\prod_{j=1}^{K-2} EY_{nj}E\prod_{j=K-1}^{K} Y_{nj} - \prod_{j=1}^{K} EY_{nj}\right| \\
&= \left|\mathrm{Cov}\left(Y_{n1}, \prod_{j=2}^{K} Y_{nj}\right)\right| + \sum_{j=2}^{K}\left|\prod_{i=1}^{j-1} EY_{ni}\right|\left|\mathrm{Cov}\left(Y_{nj}, \prod_{i=j+1}^{K} Y_{ni}\right)\right| \\
&\leq 4\sum_{j=1}^{K}\left|\mathrm{Cov}\left(Y_{nj}, \prod_{i=j+1}^{K} Y_{ni}\right)\right| \\
&\leq 16K\alpha(q) \\
&= O\left(q\alpha(q)\frac{r_n}{pq}\right) \to 0,
\end{aligned}$$

where the last step follows from noting that $\sum_{q=1}^{\infty}\alpha(q) < \infty \implies q\alpha(q) \to 0$ as $q \to \infty$. Thus

$$\phi_{\left(\frac{1}{\sqrt{r_n}}\sum_{j=1}^{K} B_{nj}\right)}(t) \to \prod_{j=1}^{K}\phi_{\left(\tilde{B}_{nj}/\sqrt{r_n}\right)}(t) = \phi_{\left(\frac{1}{\sqrt{r_n}}\sum_{j=1}^{K}\tilde{B}_{nj}\right)}(t)$$

for all $t \in \mathbb{R}$, where $\phi_X(t)$ is the characteristic function of X evaluated at t. Hence

$$\frac{1}{\sqrt{r_n}} \sum_{j=1}^{K} B_{nj} \to^d \tilde{S}_n \equiv \frac{1}{\sqrt{r_n}} \sum_{j=1}^{K} \tilde{B}_{nj}. \tag{A.25}$$

It is now shown that $\tilde{s}_n^2 \equiv \text{Var}\left(\sum_{j=1}^{K} \tilde{B}_{nj}/\sqrt{r_n}\right) \to \sigma_\infty^2$.

$$\begin{aligned}
\left| \frac{1}{r_n} \text{Var}\left(\sum_{j=1}^{K} \tilde{B}_{nj} \right) - \sigma_\infty^2 \right| &= \left| \frac{1}{r_n} \sum_{j=1}^{K} E\tilde{B}_{nj}^2 - \sigma_\infty^2 \right| \\
&= \left| \frac{1}{r_n} \sum_{j=1}^{K} EB_{nj}^2 - \sigma_\infty^2 \right| \\
&\leq \frac{1}{r_n} \sum_{j=1}^{K} \left| EB_{nj}^2 - \frac{r_n}{K}\sigma_\infty^2 \right| \\
&\leq \frac{1}{r_n} \sum_{j=1}^{K} \left| EB_{nj}^2 - p\sigma_\infty^2 \right| + \sigma_\infty^2 \left| \frac{Kp}{r_n} - 1 \right| \\
&= \frac{p}{r_n} \sum_{j=1}^{K} \left| \frac{1}{p} \text{Var}\left(\sum_{i=(j-1)m+1}^{(j-1)m+p} X_{ni} \right) - \sigma_\infty^2 \right| + \sigma_\infty^2 \left| \frac{Kp}{r_n} - 1 \right| \\
&\leq \frac{Kp}{r_n}\gamma_p + \sigma_\infty^2 \left| \frac{Kp}{r_n} - 1 \right| \to 0
\end{aligned}$$

as $n \to \infty$. It is now shown that the triangular array of independent random variables $\{\tilde{B}_{n1}/\sqrt{r_n}, \ldots, \tilde{B}_{nK_n}/\sqrt{r_n}\}_{n \geq 1}$ satisfies the Lyapounov condition

$$\lim_{n \to \infty} \tilde{s}_n^{-(2+\delta)} \sum_{j=1}^{K_n} E \left| \frac{\tilde{B}_{nj}}{\sqrt{r_n}} \right|^{(2+\delta)} = 0.$$

for $\delta = 2$. Let

$$\Gamma(k) = \sup \left\{ E \left(\sum_{i=l}^{l+k-1} X_{ni} \right)^4 k^{-3} : 1 \leq l \leq r_n - k + 1 \right\} \tag{A.26}$$

for $k = 1, \ldots, r_n$ and $\Gamma^*(k) = \sup\{\Gamma(j) : k \leq j \leq r_n\}$ for $k = 1, \ldots, r_n$. Then

$$\begin{aligned}
\sum_{j=1}^{K} E\left(\tilde{B}_{nj}/\sqrt{r_n}\right)^4 &= r_n^{-2} \sum_{j=1}^{K} E B_{nj}^4 \\
&= r_n^{-2} \sum_{i=1}^{K} E\left(\sum_{i=(j-1)m+1}^{(j-1)m+p_n} X_{ni}\right)^4 \\
&\leq r_n^{-2} K p_n^3 \Gamma^*(p_n) \\
&\leq r_n^{-2} \frac{r_n}{p_n + q_n} p_n^3 \Gamma^*(p_n) \\
&\leq r_n^{-1} p_n^2 \Gamma^*(p_n).
\end{aligned}$$

Now choose $p_n = \lfloor \sqrt{r_n}\{\Gamma^*(q_n)^{-1/3} \wedge \log r_n\} \rfloor$. Then by Lemma 2,

$$r_n^{-1} p_n^2 \Gamma^*(p_n) \leq r_n^{-1} (r_n^{1/2} \Gamma^*(q_n)^{-1/3})^2 \Gamma^*(q_n) = \Gamma^*(q_n)^{1/3} \to 0 \tag{A.27}$$

as $n \to \infty$. It is easily verified that (A.21) holds for this choice of p. By (A.27) the Lyapounov condition holds for $\delta = 2$ and Lyapounov's CLT gives that

$$\tilde{S}_n \to^d N(0, \sigma_\infty^2), \tag{A.28}$$

Which implies that $\frac{1}{\sqrt{r_n}} \sum_{j=1}^{K} B_{nj} \to^d N(0, \sigma_\infty^2)$. Thus by (A.23) and (A.24),

$$S_n/\sqrt{r_n} \to^d N(0, \sigma_\infty^2).$$

APPENDIX B

PRE-PROCESSING STEPS FOR COPY NUMBER DATA AS ANALYZED IN SECTION 3

Prior to the analysis of the copy number data, the 230×7531 data matrix $\boldsymbol{Y}$, of which the first $n_1 = 92$ rows correspond to long-term survivors and the last $n_2 = 138$ rows correspond to short-term survivors, was doubly standardized so that each row and column had zero mean and unit variance. This has become a common practice in the analysis of microarray data, (Efron (2010b)).

Each patient was labeled with an identifier of the form

`TCGA-11-2222-01A-01D-3333-01`,

where the set of digits in the position of `3333` in this string correspond to the plate or batch in which the subject's DNA was analyzed. Batch effects in copy number data are a common occurence, and we find that they are markedly present here. The 230×230 subject covariance matrix $\boldsymbol{\Delta} = \boldsymbol{Y}\boldsymbol{Y}'/N$ is depicted in the upper left hand panel of Figure B.1, in which there appears strong evidence of block correlations. Directly beneath is a depiction of the subject covariance matrix after removing the batch effects with the function `ber()` from the R package `ber` from Giordan (2014). The right hand panels of Figure B.1 display histograms of 5000 permutation test statistics, where each test statistic is a measurement of block correlation for the subject covariance matrix after a random permutation of its rows. A dark vertical line marks the position of the test statistic under the original subject ordering—by plate/batch. For the unadjusted data, there is extreme evidence of block correlations, whereas after adjusting for the plate effect using the `ber()` function, the block correlation test statistic for the original subject ordering

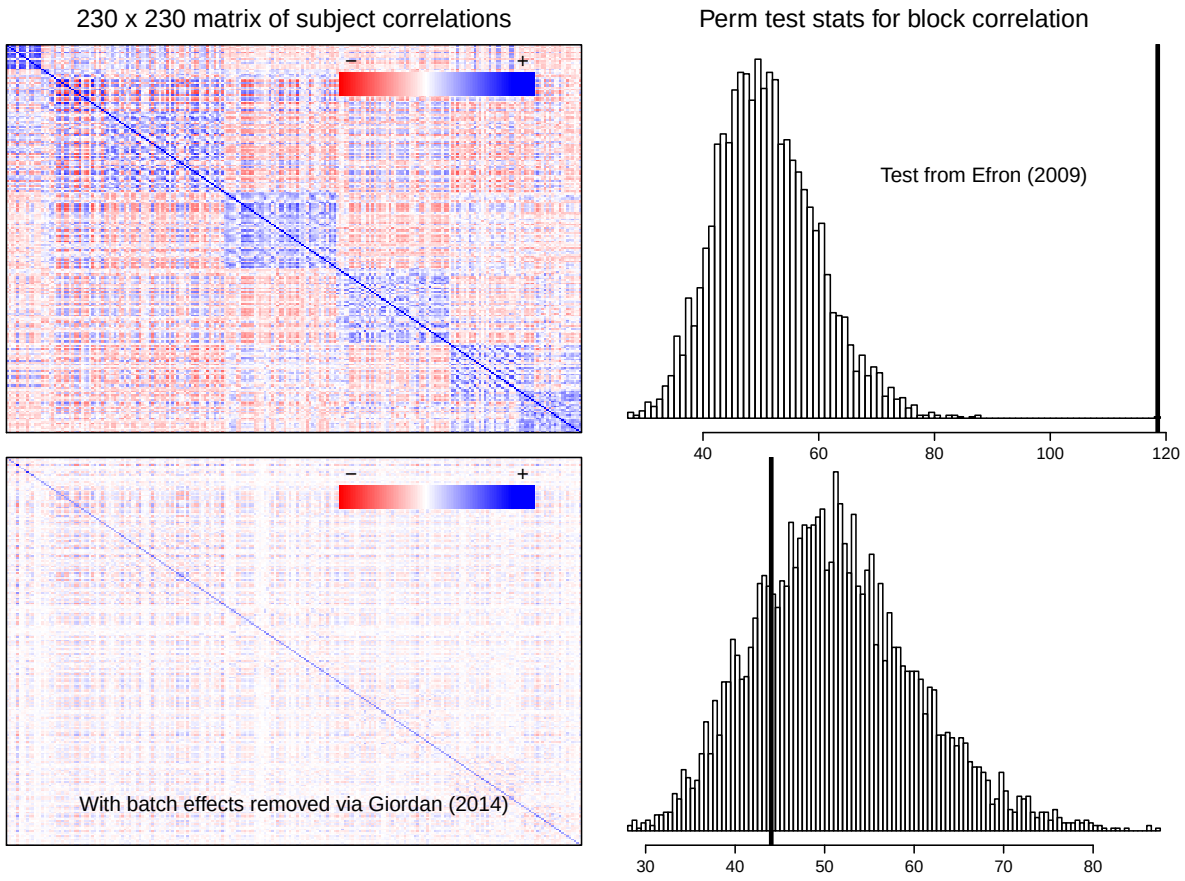

Figure B.1: Left column: Top and bottom panels depict the subject covariance matrix before and after removing the plate (batch) effect in the copy number data. Right column: Top and bottom panels display histograms of the permutation test statistics for block correlation with vertical lines positioned at the observed value of the test statistic.

falls near the center of the histogram of permutation values. For a detailed explanation of the permutation test for block correlations/batch effects, see Efron (2009).

APPENDIX C

PROOFS OF MAIN RESULTS FOR THE SMOOTH BLOCK BOOTSTRAP FOR TIME SERIES

C.1 An Auxiliary Result for The TBB/ETBB

To prove Theorem 1, we require a preliminary result, given in Lemma 1 next. Recall $Y_t = T_F^{(1)}(\delta_{X_t} - F)$, $t \in \mathbb{Z}$, denotes the influence function (4.4) evaluated at X_t, which satisfies $\mathcal{E}Y_t = T_F^{(1)}(\mathcal{E}\delta_{X_t} - F) = T_F^{(1)}(F - F) = 0$ by linearity of $T_F^{(1)}$. Let $\bar{Y}^*_{n,ETBB} = \sum_{i=1}^n \pi_i^* Y_i = \sum_{j=1}^b \sum_{k=1}^\ell w_\ell(k) Y_{I_j^*+k}/[b\|w_\ell\|]$ denote the ETBB/TBB version of the sample mean $\bar{Y}_n = \sum_{i=1}^n Y_i/n$, based on length ℓ blocks and where $\{I_j^*\}_{j=1}^b$ are iid uniform over $\{0, \ldots, n-\ell\}$; for the sample mean, the ETBB and TBB methods are known to match (Shao (2010)).

The next result establishes the validity of the TBB/ETBB approximation of the distribution of $\bar{Y}_n$, under slightly weaker mixing/moment conditions than those considered originally by Paparoditis & Politis (2001) and Paparoditis & Politis (2002) or Shao (2010).

Lemma 1 *Suppose (4.3), the block length ℓ satisfies $\ell^{-1} + \ell/n = o(1)$ as $n \to \infty$, and that $\sigma_\infty^2 \equiv \sum_{k=-\infty}^\infty cov(Y_0, Y_k) > 0$. Suppose also that, for some $\gamma > 0$, $\mathcal{E}|Y_1|^{2+\gamma} < \infty$ and $\sum_{k=1}^\infty \alpha(k)^{\gamma/(2+\gamma)} < \infty$. Then, as $n \to \infty$,*
(i) $\sqrt{n}\bar{Y}_n \xrightarrow{d} Normal(0, \sigma_\infty^2)$ and $nvar(\bar{Y}_n) \longrightarrow \sigma_\infty^2$;
(ii) $m_n n var_*(\bar{Y}^*_{n,ETBB}) \xrightarrow{p} \sigma_\infty^2$;
(iii) $\sup_{x\in\mathbb{R}} |P_*[m_\ell^{1/2}\sqrt{n}(\bar{Y}^*_{n,ETBB} - \mathcal{E}_*\bar{Y}^*_{n,ETBB}) \le x] - P(\sqrt{n}\bar{Y}_n \le x)| \xrightarrow{p} 0$.

Proof of Lemma 1. Part(i) follows by the central limit theorem for mixing sequences (cf. Athreya & Lahiri (2006), Ch. 16.3). For part (ii), write $m_\ell b\ell \mathrm{var}_*(\bar{Y}^*_{n,ETBB}) = \sum_{i=0}^{n-\ell}(U_i - \hat{\mu}_n)^2/(n-\ell+1)$, where $U_i = \sum_{j=i+1}^{i+\ell} Y_i/\|w_\ell\|_2$, $i \ge 0$, and $\hat{\mu}_n = \sum_{i=0}^{n-\ell} U_i/(n-\ell+1)$; this

is the expression of the TBB variance as a block sum sample variance (cf. Paparoditis & Politis (2001)). By writing $V_i = U_i^2 \mathbb{I}(|U_i| < (n/\ell)^{1/8})$, one can show $m_\ell b \ell \mathrm{var}_*(\bar{Y}^*_{n,ETBB}) - \sum_{i=0}^{n-\ell} V_i/(n-\ell+1) \xrightarrow{p} 0$ and $\sum_{i=0}^{n-\ell}(V_i - \mathcal{E}V_0)/(n-\ell+1) \xrightarrow{p} 0$ as in p. 51-53 of Lahiri (2003b). Then, $U_0/\|w_\ell\|_2 \xrightarrow{d} \mathrm{Normal}(0, \sigma_\infty^2)$ holds by a weighted central limit theorem, Theorem 4.3 from Lahiri (2003a), and $\mathcal{E}V_0 \to \sigma_\infty^2$ follows by the dominated convergence theorem. Part(ii) now follows since $\ell b/n \to 1$. To show part(iii), one may use that $m_\ell^{1/2}\sqrt{b\ell}(\bar{Y}^*_{n,ETBB} - \mathcal{E}_*\bar{Y}^*_{n,ETBB}) = b^{-1/2}\sum_{j=1}^b (U_{I_j^*} - \hat{\mu}_n)$ is a sum of conditionally iid variables which, in probability, have a convergent variance by part(ii) and satisfy Lindeberg's condition (i.e., $b^{-1}\sum_{j=1}^b \mathcal{E}_*(U_{I_j^*} - \hat{\mu}_n)^2 \mathbb{I}(|U_{I_j^*} - \hat{\mu}_n| > 2[n/\ell]^{1/4}) \xrightarrow{p} 0$) as in p. 56-57 of Lahiri (2003b). □

C.2 Proof of Theorem 1

Theorem 1(i). We first show $n\mathrm{var}(\hat{\theta}_n) \to \sigma_\infty$. By Condition C.2(i), write

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}[T(F_n) - T(F)] &= \sqrt{n}T_F^{(1)}(F_n - F) + \sqrt{n}R(F_n - F) \\ &= \sqrt{n}\bar{Y}_n + \sqrt{n}R(F_n - F) \qquad \text{(C.1)}\end{aligned}$$

using, by linearity, $T_F^{(1)}(F_n - F) = \bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$ for $Y_i = T_F^{(1)}(\delta_{X_i} - F)$, $i \geq 1$. We assume that the remainder $|R(F_n - F)| \leq C\|F_n - F\|_\infty^{1+\lambda}$ and later describe the treatment of the (simpler) case $|R(F_n - F)| \leq C\|F_n - F\|_1^{1+\lambda}$.

By (C.1), Lemma 1(i) and the Cauchy-Schwarz inequality, $n\mathrm{var}(\hat{\theta}_n) \to \sigma_\infty$ will follow by showing

$$n\mathcal{E}\|F_n - F\|_\infty^{2(1+\lambda)} = o(1), \qquad \text{(C.2)}$$

so that $n\mathcal{E}[R(F_n - F)]^2 = o(1)$. Because F is continuous, we may assume without loss of generality that each X_i is uniformly distributed on $[0,1]$ and $F(t) = t$, $t \in [0,1]$, is the corresponding distribution function. (That is, if $\tilde{F}_n(x)$, $x \in \mathbb{R}$, denotes the empirical distribution function of $F(X_1), \ldots, F(X_n)$, which are uniformly distributed,

then $\tilde{F}_n(F(x)) - F(x) = F_n(x) - F(x)$ for all $x \in \mathbb{R}$ with probability 1.). Define integers $\tau = \max\{\lceil\lambda\rceil, 4\lceil\gamma/2\rceil\} + 2$ and $\kappa = 1 + \lceil\gamma/2\rceil$, and set $D_n(t) = \sqrt{n}[F_n(t) - F(t)]$, for $t \in [0,1]$ and $n \geq 1$. Under the mixing moment assumptions, by Theorem 1.4.1 of Doukhan (1994) and Jensen's inequality, it follows that

$$\mathcal{E}\left|D_n(t) - D_n(s)\right|^{2\tau} \leq C\{\mathcal{E}\left|D_1(t) - D_1(s)\right|^{2+\gamma}\}^{\tau/(2+\gamma)} \leq C\{\mathcal{E}\left|D_1(t) - D_1(s)\right|^{2\kappa}\}^{\tau/(2\kappa)} \tag{C.3}$$

for any $s, t \in [0,1]$ with a constant $C > 0$ (not depending on n or $s, t \in [0,1]$). Using moments of the uniform$[0,1]$ distribution, $\mathcal{E}\left|D_1(t) - D_1(s)\right|^{2\kappa} \leq C_1|t-s|$ holds for some $C_1 > 0$ not depending on $s, t \in [0,1]$ and, by construction, $\tau/(2\kappa) > 1$. By this and (C.3), Theorem 12.2 of Billingsley (1968) yields that, for any integer $m \geq 1$ and $y > 0$, there then exists a constant $K > 0$ such that

$$P\left(\max_{1\leq i\leq m} |D_n(i/m)| \geq y\right) \leq \frac{K}{y^{2\tau}}.$$

Letting $m \to \infty$, this implies that

$$P(\sqrt{n}\|F_n - F\|_\infty \geq y) \leq \frac{K}{y^{2\tau}}$$

for any $y > 0$. Then, (as $\mathcal{E}|V|^r = r\int_0^\infty t^{r-1}P(|V| \geq t)dt$ for a generic variable V and $r > 0$), it holds that

$$\begin{aligned} n^{1+\lambda}\mathcal{E}\|F_n - F\|_\infty^{2(1+\lambda)} &= \int_0^\infty 2(1+\lambda)y^{2\lambda+1}P(\sqrt{n}\|F_n - F\|_\infty \geq y)dy \\ &\leq C + C\int_1^\infty y^{2\lambda+1-2\tau}dy < \infty \end{aligned} \tag{C.4}$$

using that $2\lambda + 1 - 2\tau < -1$. Now (C.2) follows from $n\mathcal{E}\|F_n - F\|_\infty^{2(1+\lambda)} = O(n^{-\lambda}) = o(1)$.

Next (C.2) implies that $\sqrt{n}R(F_n - F) = o_p(1)$, so that the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta)$ follows from (C.3), Lemma 1(i) and Slutsky's theorem.

Finally, a version of (C.2) can also be shown in the L^1 norm when the remainder $|R(F_n - F)| \leq \|F_n - F\|_1^{1+\lambda}$. In this case, define an even integer $m = 2(\lceil \lambda \rceil + 1)$, so that by Jensen's inequality, the Cauchy-Schwarz inequality and Fubini's theorem,

$$\mathcal{E}\|F_n - F\|_1^{2+2\lambda} \leq \left[\int \{\mathcal{E}[F_n(x) - F(x)]^m\}^{1/m} dx\right]^{2+2\lambda}.$$

By the mixing/moment assumptions and using that $|\mathbb{I}(X_i \leq x) - F(x)| \leq \max\{F(x), 1 - F(x)\}$ for each $x \in \mathbb{R}$, it holds that $\mathcal{E}[F_n(x) - F(x)]^m \leq Cn^{-m/2}[\max\{F(x), 1 - F(x)\}]^m$ (for $C > 0$ not depending on n or x) using a standard covariance bound based on α-mixing and bounded random variables (cf. p. 10 of Doukhan (1994);p. 510 of Athreya & Lahiri (2006)) and arguments as in Theorem 1.4.1 of Doukhan (1994) . Hence, $n\mathcal{E}\|F_n - F\|_1^{2+2\lambda} \leq Cn^{-\lambda} = o(1)$ follows from $\mathcal{E}|X_1| < \infty$. $\square$

Theorem 1(ii). Under Condition C.2 and recalling $Y_i = T_F^{(1)}(\delta_{X_i} - F)$, write

$$\begin{aligned} \hat{\theta}_n^* - \tilde{\theta}_n &= T[F_n^*] - T[F] - \{T[\mathcal{E}_* F_n^*] - T[F]\} \\ &= \sum_{i=1}^n \pi_i^* Y_i + L_n^* + R(F_n^* - F) + R(\mathcal{E}_* F_n^* - F) \end{aligned} \tag{C.5}$$

where

$$L_n^* \equiv \sum_{i=1}^n \pi_i^* T_F^{(1)}(\delta_{X_i + Z_{i,n}^*} - \delta_{X_i}) - \mathcal{E}_* \sum_{i=1}^n \pi_i^* T_F^{(1)}(\delta_{X_i + Z_{i,n}^*} - F)$$

by linearity of $T_F^{(1)}$ and using $\sum_{i=1}^n \pi_i^* = 1$. By Lemma 1(ii), $m_\ell n \mathrm{var}_*(\bar{Y}_{n,ETBB}^*) \xrightarrow{p} \sigma_\infty^2$ where $\bar{Y}_{n,ETBB}^* = \sum_{i=1}^n \pi_i^* Y_i$. Hence, by the Cauchy-Schwarz inequality and Condition C.2 (note $m_\ell = O(1)$), Theorem 1(ii) will follow by showing

$$n\mathrm{var}_*(L_n^*) + n\mathcal{E}_*\|F_n^* - \mathcal{E}_* F_n^*\|_\infty^{2(1+\lambda)} + n\|\mathcal{E}_* F_n^* - F\|_\infty^{2(1+\lambda)} \xrightarrow{p} 0; \tag{C.6}$$

above we are assuming a remainder $|R(\cdot)| \leq C\|\cdot\|_\infty^{1+\lambda}$ bounded in the Kolmogorov metric

and later describe the L^1-metric case $|R(\cdot)| \leq C\|\cdot\|_1^{1+\lambda}$.

To handle $n\|\mathcal{E}_* F_n^* - F\|_\infty^{2(1+\lambda)}$ in (C.6), we require some notation. Let $\Phi_{\mu,\sigma}$ denote the distribution for a normal with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$, and let ϕ denote the standard normal density function. By independence of $\{I_j\}_{j=1}^b$ and iid standard normal $\{Z_i^*\}_{i=1}^n$, it holds for a given $i = 1, \ldots, n$ and $k = 1, \ldots, \ell$, that $\mathcal{E}_* \mathbb{I}(I_j^* = i-k)\delta_{X_i + hZ_i^*} = \mathcal{E}_* \mathbb{I}(I_j^* = i-k)\mathcal{E}_* \delta_{X_i + hZ_i^*}$ and $\mathcal{E}_* \delta_{X_i + hZ_i^*} = \Phi_{X_i, h}$. Hence,

$$\begin{aligned}
\mathcal{E}_* F_n^* = \frac{b}{\|w_\ell\|_1} \sum_{j=1}^b \sum_{k=1}^\ell w_\ell(k) \mathcal{E}_* \sum_{i=1}^n \mathbb{I}(I_j^* = i-k) \Phi_{X_i, h} &= \frac{b}{\|w_\ell\|_1} \sum_{j=1}^b \sum_{k=1}^\ell w_\ell(k) \mathcal{E}_* \Phi_{X_{k+I_j^*}, h} \\
&= \sum_{k=1}^\ell \frac{w_\ell(k)}{\|w_\ell\|_1} \sum_{m=0}^{n-\ell} \frac{1}{n-\ell+1} \Phi_{X_{k+m}, h} \\
&= \Phi_{0,h} * F_{1n},
\end{aligned}$$

where the last line denotes the distributional convolution (cf. Sec. 5.4 of Athreya & Lahiri (2006)) between $\Phi_{0,h}$ and the distribution $F_{1n} = \sum_{k=1}^\ell \frac{w_\ell(k)}{\|w_\ell\|_1} \sum_{m=0}^{n-\ell} \frac{1}{n-\ell+1} \delta_{X_{k+m}}$ (i.e., for a Borel set A, $\Phi_{0,h} * F_{1n}(A) = \int_{-\infty}^\infty \int_{-\infty}^\infty \mathbb{I}(x+y \in A) d\Phi_{0,h}(x) dF_{1n}(y)$). Write

$$A_{1n} = \|\Phi_{0,h} * F_{1n} - \Phi_{0,h} * F_n\|_\infty, \; A_{2n} = \|\Phi_{0,h} * F_n - \Phi_{0,h} * F\|_\infty, \; A_{3n} = \|\Phi_{0,h} * F - F\|_\infty.$$

Then, $A_{1n} \leq \|n^{-1}(n-\ell+1)F_{1n} - F_n\|_\infty + \|n^{-1}(n-\ell+1)F_{1n} - F_{1n}\|_\infty \leq 3\ell/n$ holds; $A_{2n} \leq \|F_n - F\|_\infty = O_p(n^{-1/2})$ by (C.4); and $A_{3n} \leq \sup_{x \in \mathbb{R}} \int_{-\infty}^\infty |F(x+hz) - F(x)| \phi(z) dz \leq Ch$ by Condition C.1. It now follows in (C.6) that

$$n\|\mathcal{E}_* F_n^* - F\|_\infty^{2(1+\lambda)} \leq n \sum_{j=1}^3 A_{jn}^{2(1+\lambda)} \leq O(\ell^{2(1+\lambda)}/n^{1+2\lambda}) + O_p(n^{-\lambda}) + O(nh^{2(1+\lambda)}) = o_p(1)$$

by the growth assumptions $\ell^2/n = O(1)$ and $nh^{2(1+\lambda)} = o(1)$.

Consider next $n\mathcal{E}_* \|F_n^* - \mathcal{E}_* F_n^*\|_\infty^{2(1+\lambda)}$ in (C.6). For $k = 1, \ldots, \ell$, define the distribution $F_{k,\ell,n}^* = b^{-1} \sum_{j=1}^b \delta_{X_{I_j^*+k} + hZ_{I_j^*+k}^*}$. Then, $F_n^* - \mathcal{E}_* F_n^* = \|w_\ell\|_1^{-1} \sum_{k=1}^\ell w_\ell(k)[F_{k,\ell,n}^* - \mathcal{E}_* F_{k,\ell,n}^*]$,

so that it follows that (because $\|\cdot\|_\infty$ is a norm), for any $t > 0$,

$$\begin{aligned} P_*\left(\|F_n^* - \mathcal{E}_* F_n^*\|_\infty > t\right) &\leq P_*\left(\bigcup_{k=1}^{\ell}\left\{\|F_{k,\ell,n}^* - \mathcal{E}_* F_{k,\ell,n}\|_\infty > \frac{t\|w_\ell\|_1}{w_\ell(k)}\right\}\right) \\ &\leq \sum_{k=1}^{\ell} P_*\left(\|F_{k,\ell,n}^* - \mathcal{E}_* F_{k,\ell,n}\|_\infty \geq \frac{t\|w_\ell\|_1}{w_\ell(k)}\right). \end{aligned}$$

Because each $F_{k,\ell,n}^* - \mathcal{E}_* F_{k,\ell,n}$ $(k = 1, \ldots, \ell)$ is the centered empirical distribution of b iid random variables $\{X_{I_j^*+k} + hZ_{I_j^*+k}^*\}_{j=1}^b$ (under P_*), by the Dvoretzky-Kiefer-Wolfowitz inequality theorem we have that

$$P_*\left(\|F_{k,\ell,n}^* - \mathcal{E}_* F_{k,\ell,n}\|_\infty \geq \frac{t\|w_\ell\|_1}{w_\ell(k)}\right) \leq 2\exp\{-2b[\|w_\ell\|_1/w_\ell(k)]^2 t^2\}$$

for any $t > 0$. Hence,

$$\begin{aligned} n\mathcal{E}_*\|F_n^* - \mathcal{E}_* F_n^*\|_\infty^{2(1+\lambda)} &\leq n\int_0^\infty t^{2\lambda+1} P_*\left(\|F_n^* - \mathcal{E}_* F_n^*\|_\infty > t\right) dt \\ &\leq 2n\sum_{k=1}^{\ell}\int_0^\infty t^{2\lambda+1}\exp\{-2b[\|w_\ell\|_1/w_\ell(k)]^2 t^2\} dt \\ &\leq Cn\sum_{k=1}^{\ell}\left[\frac{[w_\ell(k)]^2}{\|w_\ell\|_1^2}\frac{1}{b}\right]^{1+\lambda} = O((\ell n)^{-\lambda}) = o(1) \end{aligned}$$

using that $\sum_{k=1}^{\ell}[w_\ell(k)]^{2(1+\lambda)} = O(\ell)$, $b\ell/n \to 1$ and $\|w_\ell\|_1 \propto \ell$ as $n \to \infty$.

Finally, consider $n\mathrm{var}_*(L_n^*)$ in (C.6). Letting $\mathcal{E}_{|I^*}$ and $\mathrm{var}_{|I^*}$ denote bootstrap expectation and variance conditional on variables $\{I_j^*\}_{j=1}^b$ (recall $\{I_j^*\}_{j=1}^b$ and $\{Z_i^*\}_{i=1}^n$ are independent), we have

$$n\mathrm{var}_*(L_n^*) = n\mathcal{E}_*[\mathrm{var}_{|I^*}(L_n^*)] + n\mathrm{var}_*[\mathcal{E}_{|I^*}(L_n^*)]$$

and we next show

$$n\mathcal{E}_*[\mathrm{var}_{|I^*}(L_n^*)] \xrightarrow{p} 0, \qquad n\mathrm{var}_*[\mathcal{E}_{|I^*}(L_n^*)] \xrightarrow{p} 0. \tag{C.7}$$

to establish $n\mathrm{var}_*(L_n^*) \xrightarrow{p} 0$ in (C.6). We may write

$$n\mathcal{E}_*[\mathrm{var}_{|I^*}(L_n^*)] = n\mathcal{E}_*\left[\sum_{i=1}^n [\pi_i^*]^2 \mathrm{var}_{|I^*} T_F^{(1)}(\delta_{X_i+hZ_i^*} - \delta_{X_i})\right] = n\mathcal{E}_*\left[\sum_{i=1}^n [\pi_i^*]^2\right] \cdot V_n$$

for $V_n = \mathrm{var}_*[T_F^{(1)}(\delta_{hZ_1^*} - \delta_0)]$; the above follows using that $\{\pi_i^*\}_{i=1}^n$ are determined by $\{I_j^*\}_{j=1}^b$ and that $T_F^{(1)}(\delta_{X_i+hZ_i^*} - \delta_{X_i}) = T_F^{(1)}(\delta_{hZ_i^*} - \delta_0)]$ (as the distributions are location shifts) along with $\{Z_i^*\}_{i=1}^n$ are iid [standard normal] and independent of $\{I_j^*\}_{j=1}^b$ under P_*. We will show $V_n = o_p(1)$ and $n\mathcal{E}_*\left[\sum_{i=1}^n [\pi_i^*]^2\right] = O_p(1)$ to obtain $n\mathcal{E}_*[\mathrm{var}_{|I^*}(L_n^*)] \xrightarrow{p} 0$ in (C.7). As $\pi_i^* = [b\|w_\ell\|_1]^{-1} \sum_{j=1}^b \sum_{k=1}^\ell w_\ell(k)\mathbb{I}(I_j^* = i+k)$ and $\{I_j^*\}_{j=1}^b$ are iid uniform on $\{0, \ldots, n-\ell\}$,

$$\begin{aligned} n\mathcal{E}_*\left[\sum_{i=1}^n [\pi_i^*]^2\right] &\leq n[b\|w_\ell\|_1]^{-2} \sum_{i=1}^n \Bigg[\sum_{j=1}^b \sum_{k=1}^\ell w_\ell(k) P_*(I_j^* = i+k) \\ &\qquad + 2 \sum_{1\leq j<m\leq b} \sum_{k=1}^\ell \sum_{z=1}^\ell w_\ell(k) w_\ell(z) P_*(I_j^* = i+k) P_*(I_m^* = i+z)\Bigg] \\ &\leq \frac{n^2}{(n-\ell+1)b\|w_\ell\|_1]} + \frac{n^2}{(n-\ell+1)^2} = O(1). \end{aligned}$$

Next note that w.p.1 (P), $hZ_1^* \xrightarrow{P_*} 0$ (converges in distribution to zero in P_*) because $h \to 0$ and that (for $\mathbf{0}(x) = x$, $x \in \mathbb{R}$, in Condition C.2(ii)) in terms of the Skorohod metric $d_S(\mathbf{0}, \delta_{Z_{1,n}^*} - \delta_0) = d_S(\delta_{Z_{1,n}^*}, \delta_0) \leq h|Z_1^*|$ while in terms of the L^1 metric $d_1(\mathbf{0}, \delta_{hZ_1^*} - \delta_0) \leq h\mathcal{E}_*|Z_1^*| \leq h$. Hence, w.p.1 ($P$), $d_{S,1}(\mathbf{0}, \delta_{hZ_1^*} - \delta_0) \xrightarrow{P_*} 0$ so that

$$T_F^{(1)}\delta_{hZ_1^*} - \delta_0) \xrightarrow{P_*} 0 \quad \text{(w.p.1 } (P)\text{)} \tag{C.8}$$

by continuity at $\mathbf{0}$ under Condition C.2(ii). Also, by Condition C.2(ii), it holds w.p.1

(P) that

$$\begin{aligned}\sup_{n\geq 1}\mathcal{E}_*[T_F^{(1)}(\delta_{hZ_1^*}-\delta_0)]^3 &\leq A\sup_{n\geq 1}\mathcal{E}_*\exp[3ad_{S,1}(\mathbf{0},\delta_{hZ_1^*}-\delta_0)]\\ &\leq A\sup_{n\geq 1}\mathcal{E}_*\left(\exp[3ah|Z_1^*|]+\exp[3ah]\right)<\infty,\end{aligned}$$

(as Z_1^* is normal), implying $\{[T_F^{(1)}(\delta_{hZ_1^*}-\delta_0)]^2\}_{n=1}^{\infty}$ is uniformly integrable in P_*. This and (C.8) yield $V_n\to 0$ w.p.1 (P) so that $V_n\xrightarrow{p}0$. Lastly, we consider showing $n\text{var}_*[\mathcal{E}_{|I^*}(L_n^*)]\xrightarrow{p}0$ in (C.7). Again because $T_F^{(1)}(\delta_{X_i+hZ_i^*}-\delta_{X_i})=T_F^{(1)}(\delta_{hZ_i^*}-\delta_0)]$ and $\{Z_i^*\}_{i=1}^n$ are iid normal,

$$\begin{aligned}\mathcal{E}_{|I^*}(L_n^*) &= \sum_{i=1}^n\pi_i^*T_F^{(1)}(\delta_{\Phi_{0,h}}-\delta_0)-\mathcal{E}_*\sum_{i=1}^n\pi_i^*T_F^{(1)}(\delta_{X_i+hZ_i^*}-F)\\ &= T_F^{(1)}(\delta_{\Phi_{0,h}}-\delta_0)-\mathcal{E}_*\sum_{i=1}^n\pi_i^*T_F^{(1)}(\delta_{X_i+hZ_i^*}-F)\end{aligned}$$

using $\sum_{i=1}^n\pi_i^*=1$. Because $\mathcal{E}_{|I^*}(L_n^*)$ is non-stochastic under P_*, $n\text{var}_*[\mathcal{E}_{|I^*}(L_n^*)]=0$ for all $n\geq 1$ w.p.1 (P) and consequently $n\text{var}_*[\mathcal{E}_{|I^*}(L_n^*)]\xrightarrow{p}0$.

This concludes the proof of Theorem 1(ii). In the case of remainders $|R(\cdot)|\leq C\|\cdot\|_1^{1+\lambda}$ bounded in the L^1 metric, one needs to show an analog of (C.6):

$$n\mathcal{E}_*\|F_n^*-\mathcal{E}_*F_n^*\|_1^{2(1+\lambda)}+n\|\mathcal{E}_*F_n^*-F\|_1^{2(1+\lambda)}\xrightarrow{p}0.$$

Using the Mallow/Wasserstein representation of the L^1 metric (cf. Bickel & Freedman (1981)), it is straightforward to show that $n\|\mathcal{E}_*F_n^*-F\|_1^{2(1+\lambda)}=o_p(1)$ (with the same order bounds as $n\|\mathcal{E}_*F_n^*-F\|_\infty^{2(1+\lambda)}$) and $n\mathcal{E}_*\|F_n^*-\mathcal{E}_*F_n^*\|_1^{2(1+\lambda)}\leq n(2h)^{2(1+\lambda)})=o_p(1)$. □

Theorem 1(iii) Re-writing the expansion in (C.5) gives

$$m_\ell^{1/2}\sqrt{n}(\hat{\theta}_n^*-\tilde{\theta}_n)=m_\ell^{1/2}\sqrt{n}\left(\sum_{i=1}^n\pi_i^*Y_i+\mathcal{E}_*L_n^*\right)+R_n^*$$

with a remainder $R_n^* \equiv m_\ell^{1/2}\sqrt{n}[L_n^* - \mathcal{E}_* L_n^* + R(F_n^* - F) + R(\mathcal{E}_* F_n^* - F)]$ and where $\sum_{i=1}^n \pi_i^* Y_i + \mathcal{E}_* L_n^* = \bar{Y}_{n,ETBB}^* - \mathcal{E}_* \bar{Y}_{n,ETBB}^*$. By Lemma 1(i),(iii) and (C.6), for any subsequence $\{n_j\}$ of $\{n\}_{n=1}^\infty$, one may extract a further subsequence $\{n_k\} \subset \{n_j\}$ such that, w.p.1 (P), $m_{\ell_k}\sqrt{n_k}(\bar{Y}_{n_k,ETBB}^* - \mathcal{E}_* \bar{Y}_{n_k,ETBB}^*) \xrightarrow{d}$Normal$(0, \sigma_\infty^2)$ and $R_{n_k} \xrightarrow{p} 0$ in P_*-probability, implying $m_{\ell_k}^{1/2}\sqrt{n_k}(\hat{\theta}^* - \tilde{\theta}_n) \xrightarrow{d}$Normal$(0, \sigma_\infty^2)$ in P_*-probability (w.p.1 (P)). Hence, letting $\Phi(\cdot)$ denote the standard normal distribution function, this last fact implies that

$$\sup_{x \in \mathbb{R}} \left| P_* \left(m_\ell^{1/2}\sqrt{n}(\hat{\theta}_n^* - \tilde{\theta}_n) \right) - \Phi(x/\sigma_\infty) \right| \xrightarrow{p} 0$$

and Theorem 1(iii) now follows from Theorem 1(i). □