Winter 2018

Approximation of Quantiles of Rank Test Statistics Using Almost Sure Limit Theorems

Mark Ledbetter
Old Dominion University

APPROXIMATION OF QUANTILES OF RANK TEST STATISTICS USING ALMOST SURE LIMIT THEOREMS

by

Mark Ledbetter
B.S. July 1998, Charter Oak State College
M.S. December 2014, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
December 2018

Approved by:

Lucia Tabacu (Director)

N. Rao Chaganty (Member)

Norou Diawara (Member)

Juan Du (Member)

Manfred Denker (Member)

ABSTRACT

APPROXIMATION OF QUANTILES OF RANK TEST STATISTICS USING ALMOST SURE LIMIT THEOREMS

Mark Ledbetter
Old Dominion University, 2018
Director: Dr. Lucia Tabacu

There are many problems in statistics where the analysis is based on asymptotic distributions. In some cases, the asymptotic distribution is in an open form or is intractable. One possible solution is the logarithmic quantile estimation (LQE) method introduced by Thangavelu (2005) for rank tests and Fridline (2010) for the correlation coefficient. LQE is derived from an almost sure version of the central limit theorem using the results of Berkes and Csáki (2001), and it estimates the quantiles of a test statistic using only the data. To date, LQE has been used in only a few applications. We extend the use of LQE to three widely analyzed problems.

We investigate the LQE approach using fully nonparametric rank statistics to test for known trend and umbrella patterns in the main effects of three widely used factorial designs: a two-factor fixed effect model, a partial hierarchical repeated measures mixed effect model, and a mixed effect cross-classification repeated measures model. We also test for patterned alternatives in the interaction between the main effect and time in the partial hierarchical repeated measures model. We derive the almost sure central limit theorems for all of these problems and determine the level and power.

The Pettitt (1979) test is a nonparametric test based on the Mann-Whitney statistic used to detect a change in distribution in a sequence of random variables. The proposed statistic has an asymptotic distribution that is the distribution of the supremum of the absolute value of the Brownian bridge, which has an open form. We propose an approximation of the quantiles for the test statistic based on LQE. We provide simulation results for Type I error and power of the logarithmic quantile estimates for the test statistic, and compare the LQE results with other methods for two real data examples.

Thangavelu (2005) considered LQE for the nonparametric Behrens-Fisher problem with some success by introducing new numerically determined coefficients. We

examine the nonparametric two-sample problem using an empirical process of U-statistic structure (Denker and Puri, 1992). Specifically, we investigate using LQE with a second order U-statistic for paired averages within each sample. We provide simulation results to show almost sure convergence of the new test statistic.

Dedicated to my Lord and Savior Jesus Christ, to my wife *Deborah Ledbetter*, to my mother *Gaye Ledbetter*, to the memory of my father *Larry Ledbetter*, to the memory of my grandmother *Parlee Ledbetter*, and to the memory of my sister *Lisa Ledbetter-Loyd*.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the people who supported, guided, mentored, and inspired me. Without these people, this dissertation would not have been possible.

A special thank you to my advisor, Dr. Lucia Tabacu. Your kindness, patience, and encouragement during the last two and a half years have been extraordinary. I am grateful that you accepted me as your Ph.D. student. Thank you for challenging, guiding, and mentoring me during this incredible journey.

Dr. N. Rao Chaganty, thank you for your excellent instruction, support, advice, and insight. I am extremely appreciative that you accepted me into the Ph.D. program, provided a teaching assistantship, and agreed to serve on my dissertation committee.

Dr. Norou Diawara, you are one of the most courteous and hardworking people I know. Thank you for your passion, for agreeing to serve on my dissertation committee, and for always taking time to encourage, help, and teach. I am especially grateful for your exemplary guidance and mentorship during my Masters project.

Dr. Manfred Denker, I am humbled that you agreed to join my dissertation committee. Thank you for making special arrangements in your travels to meet with me, to attend my dissertation defense, and to provide insight and guidance on the field of research that you established.

Dr. Juan Du, thank you so much for agreeing to be on my dissertation committee. I would like to express my gratitude for your time, your feedback, and your suggestions.

Dr. Steve Patch, thank you for introducing me to the world of statistics and inspiring me to pursue a Ph.D. Your enthusiasm, excellent teaching, and personal interest in your students is remarkable and appreciated.

Thank you to the Department of Mathematics and Statistics for providing a nice working environment and financial support during my graduate studies. Thank you to Miriam Venable and Sheila Hegwood for being kind, competent, and helpful. I would like to thank Dr. Wirawan Purwanto for his suggestions for improving the efficiency of a R program.

To my wife Deborah, who supported and encouraged the decision to pursue a Ph.D., my love and heartfelt gratitude. Thank you for always believing in me, loving me, and praying for me. I am so blessed that God brought us together.

To my mom Gaye Ledbetter, you are one of the most giving, kind, and encouraging people I have ever met. Your prayers and love have been instrumental during the pursuit of this Ph.D. To my in-laws Tom and Gerry Williams, thank you for your love, support, and friendship. To my late father Larry Ledbetter, I am grateful that you taught me to have a strong work ethic, to pursue my dreams, and to enjoy life. To my late grandmother Parlee Ledbetter, thank you for your kindness, love, and generosity. To my late sister Lisa Loyd-Ledbetter, thank you for loving me and encouraging me.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In this dissertation we introduce a new approach called logarithmic quantile estimation (LQE) to investigate three different types of statistical problems in a nonparametric setting: tests for patterned alternatives in factorial designs, detecting change-points, and the two-sample problem using an empirical process of U-statistic structure. To be more precise, we use rank tests for three different factorial models and test for the presence of a pattern across the levels of one of the factors using LQE. Trend and umbrella shaped patterns are tested in the alternative hypothesis. Change-point problems involve detecting a change, usually a shift in an (ordered) sequence of data. We limit our LQE investigation to the Pettitt rank test (Pettitt, 1979). The nonparametric two-sample problem involves testing two independent samples to determine if they are from the same population. We propose an empirical second order U-statistic process to investigate the type I error and power for small samples under the LQE approach.

The LQE approach was introduced by Thangavelu (2005), and it has only been investigated for a few statistical problems, which makes it an interesting topic to explore. A brief review of LQE follows immediately, and an extensive literature review is provided in Chapter 2. Thangavelu (2005) proposed a parametric statistic to test if the mean is equal to zero in a normally distributed sample; the results were competitive with the t-test. Thangavelu also studied a parametric and a nonparametric test for the famous Behrens-Fisher problem, which were competitive with the t-test using the Satterthwaite-Smith approximation and various nonparametric tests, respectively. Parametric confidence interval estimations for the correlation coefficient were investigated by Fridline (2010) for bivariate normal distributions. The results were competitive with the bootstrap (Efron, 1979) and classical methods. Tabacu (2014) developed LQE for multiple sample comparisons and for longitudinal factorial models using rank tests (see also Denker and Tabacu, 2014, 2015). Our contribution in this dissertation extends the LQE approach to new and different areas of statistics, which show the flexibility and potential of LQE to address complex problems that may not currently have a viable solution. To understand LQE in this dissertation, we explore its foundational elements of almost sure limit theorems, rank statistics, and nonparametric models.

1.1 BACKGROUND

In this section we discuss general concepts for nonparametric LQE, including some almost sure limit theorems from which LQE is derived. The search for an almost sure (a.s.) version of the central limit theorem (CLT) eluded statisticians until late last century, when Fisher (1987), Brosamler (1988), and Schatte (1988) independently proved an almost sure weak version of the CLT (ASCLT) under varying moment assumptions. The first attempts to obtain an ASCLT involved Cesàro summation. It has been proved that an ASCLT does not exist for Cesàro summation (Berkes, 1998). Instead, the ASCLT was originally proven using logarithmic summation. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d.) random variables with $EX_1 = 0$, $EX_1^2 = 1$, and partial sums $S_n = \sum_{k=1}^{n} X_k$. The simplest form of the ASCLT was given by Lacey and Philipp (1990)

$$\frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{S_k}{\sqrt{k}} \leq t\right) \longrightarrow \Phi(t),\ a.s,\ \forall t \in \mathbb{R}, \tag{1}$$

where log is the natural logarithm, $\mathbb{I}$ denotes the indicator function, and Φ is the standard normal distribution function.

It is known that the convergence of logarithmic sums is very slow. Authors such as Lacey and Philipp (1990) discuss convergence rates, while Hörmann (2005, 2007) discuss optimal weights and their effects on the rate of convergence, but simulation studies and results were not provided. Additionally, a search for programs that used almost sure convergence yielded only one result, a package in R written by Micheaux and Liquet (2009), and the a.s. convergence graphs provided are limited. To the best of our knowledge, there is not an empirical or graphical analysis for the convergence behavior and rate of the ASCLT in literature. To remedy the omission, we provide the following demonstration of the convergence of the ASCLT.

A single random sample of 10^9 observations was simulated from a standard normal distribution. A sequence of the first $n = \{10^3, 5\times10^3, 10^4, 10^5, 10^6, 10^7, 5\times10^7, 10^8, 5\times 10^8, 10^9\}$ observations were used to calculate

$$\tilde{G}_n(t) := \frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\left(\frac{S_k}{\sqrt{k}} \leq t\right) \tag{2}$$

for quantile values t $=\{-2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0\}$. For ease of viewing, a smoothed line was generated and plotted against the standard normal distribution function for the same quantile values. Figure 1 contains profiles for $n = 10^3, 10^4, 10^5, 10^6$. The convergence is very slow, and an increase in n does not guarantee a move closer to the standard normal distribution in every instance. For example in Figure 1, the curve for $n = 10^6$ is farthest from the standard normal for small t, and is further away than the curve for $n = 10^5$ for most values of t.

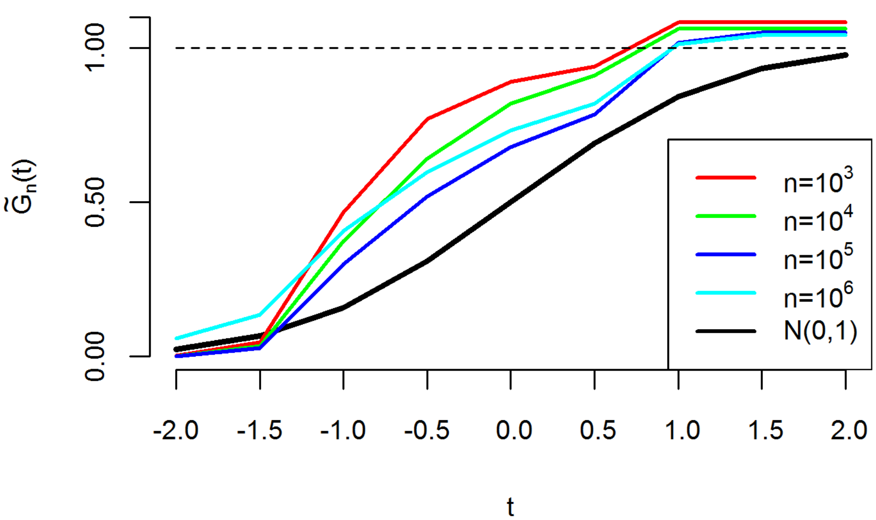

Figure 1: ASCLT convergence using $\log n$ for $n = 10^3, 10^4, 10^5, 10^6$

In Figure 2 with $n \geq 10^8$, each increase in n does result in a shift towards the standard normal, and the convergence rate increases. The slow convergence rate of the ASCLT presents a challenge for applications with small to fairly large sample sizes. Additionally, the values of $\tilde{G}_n(t)$ exceed one and prevent adequate estimation of quantiles.

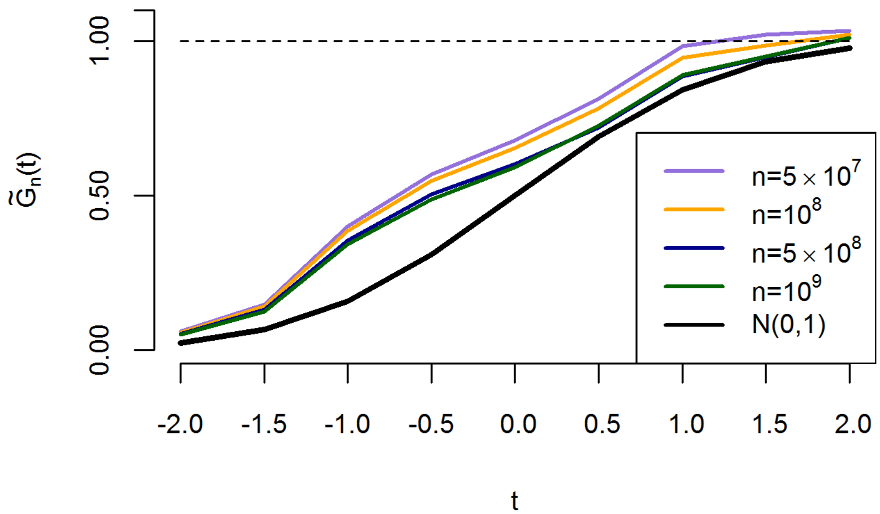

Figure 2: ASCLT convergence using $\log n$ for $n = 10^7, 10^8, 5 \times 10^8, 10^9$

To prevent values greater than one, Thangavelu (2005) proposed replacing $\log n$ with its estimate $C_n = \sum_{k=1}^{n} \frac{1}{k}$ in expression (2), making

$$\hat{G}_n(t) := \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{S_k}{\sqrt{k}} \leq t\right) \tag{3}$$

an empirical distribution function (EDF). Figures 3-4 contain the plots for values of $n = 10^3, 10^4, 10^5, 10^6$ and $n = 10^7, 10^8, 5 \times 10^8, 10^9$, respectively. The behavior is very similar to that in Figures 1-2. The convergence rate is marginally better, but it still very slow.

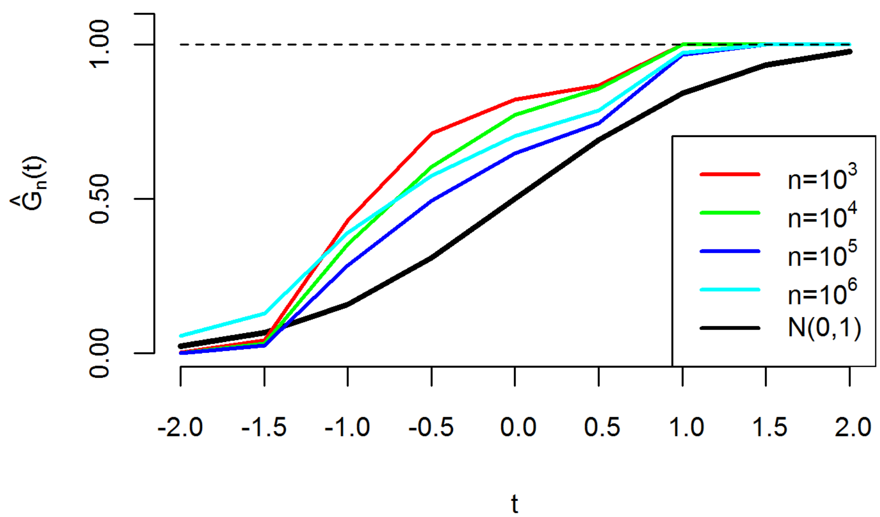

Figure 3: ASCLT convergence using C_n for $n = 10^3, 10^4, 10^5, 10^6$

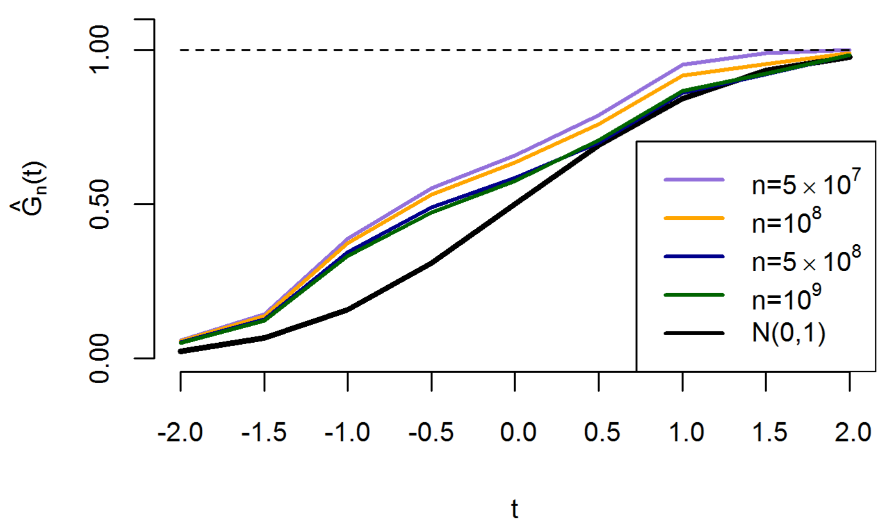

Figure 4: ASCLT convergence using C_n for $n = 10^7, 10^8, 5 \times 10^8, 10^9$

To increase the rate of convergence, we introduce random permutations of the data. We permuted the normally distributed sample of 10^9 observations p=100 times, and computed $\hat{G}_n(t)$. Due to the increase in convergence and computational limitations, $n = 10^7$ was the largest number of observations analyzed. Figure 5 displays a similar convergence pattern to that in Figure 3 for $n \leq 10000$, but the rate of convergence is much improved. Figure 6 shows an increased rate of convergence compared to that in Figure 4, but the pattern of convergence has changed suggesting that the not every increase in n results in a profile closer to the standard normal. The rate of convergence is significantly improved using p=100 permutations, especially in the tails of the distribution, which are of particular interest for hypothesis testing. It becomes apparent that practical applications of the ASCLT for small samples require the use of permutations.

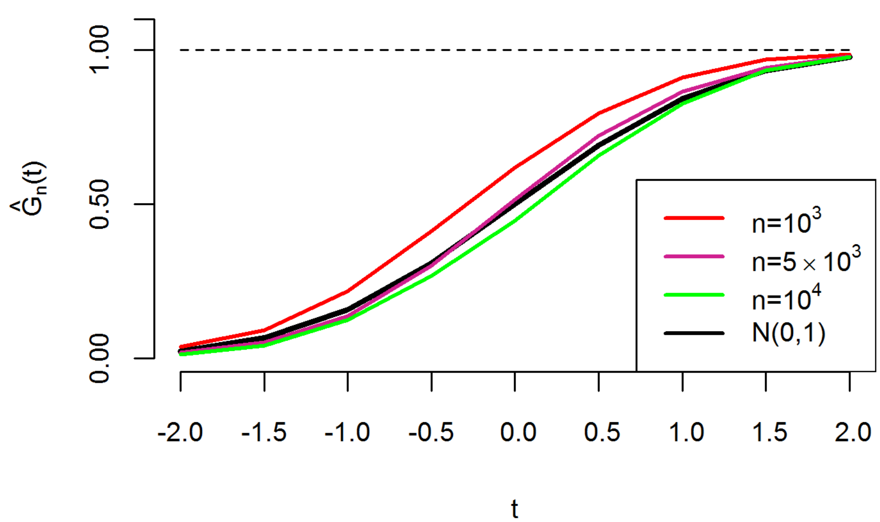

Figure 5: ASCLT convergence using C_n, p=100, and $n = 10^3, 5 \times 10^3, 10^4$

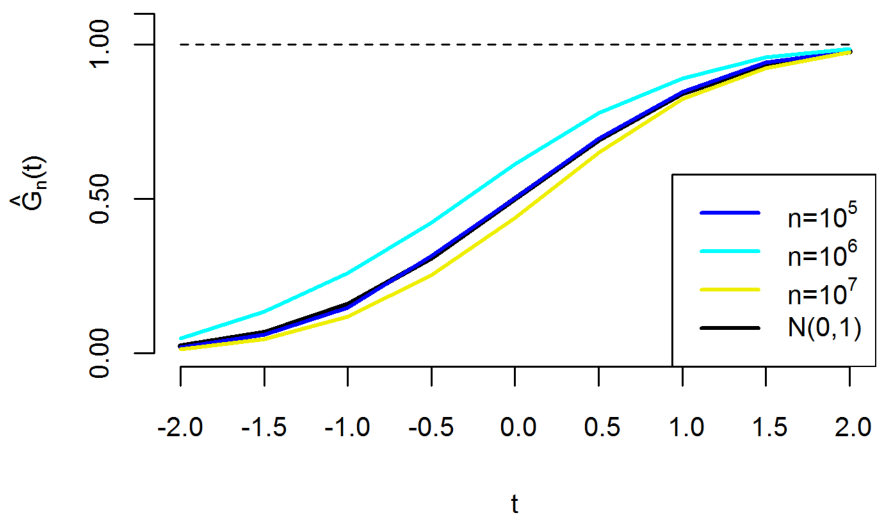

Figure 6: ASCLT convergence using C_n, p=100, and $n = 10^5, 10^6, 10^7$

In an attempt to further improve the convergence, p=500 permutations were applied to the sample. Figure 7 shows an additional increase in the rate of convergence and a similar convergence pattern to that of Figure 5. The comparison of Figures 7 and 5 for $n \geq 10^5$ also reveal a significantly faster convergence, but the convergence pattern has changed. The rate of convergence and the minimum number of permutations for each individual almost sure limit theorem (ASLT) investigated in this dissertation is determined by simulation studies.

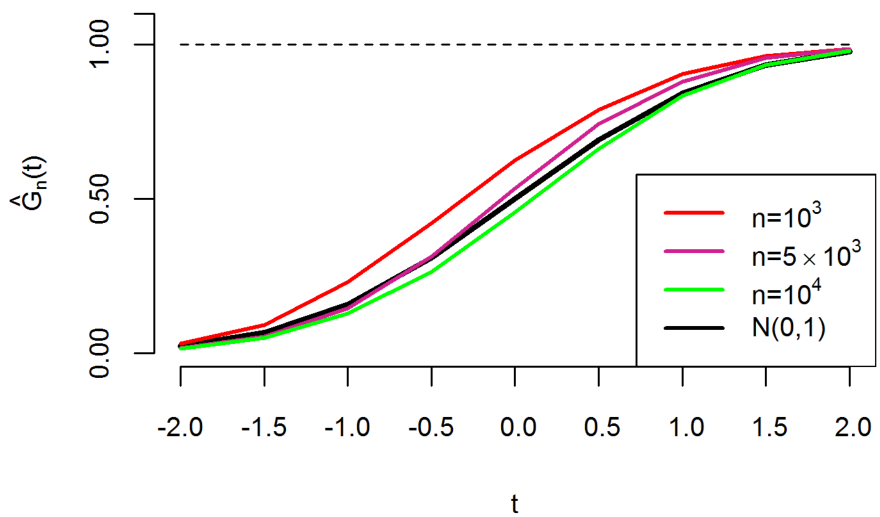

Figure 7: ASCLT convergence using C_n, p=500, and $n = 10^3, 5 \times 10^3, 10^4$

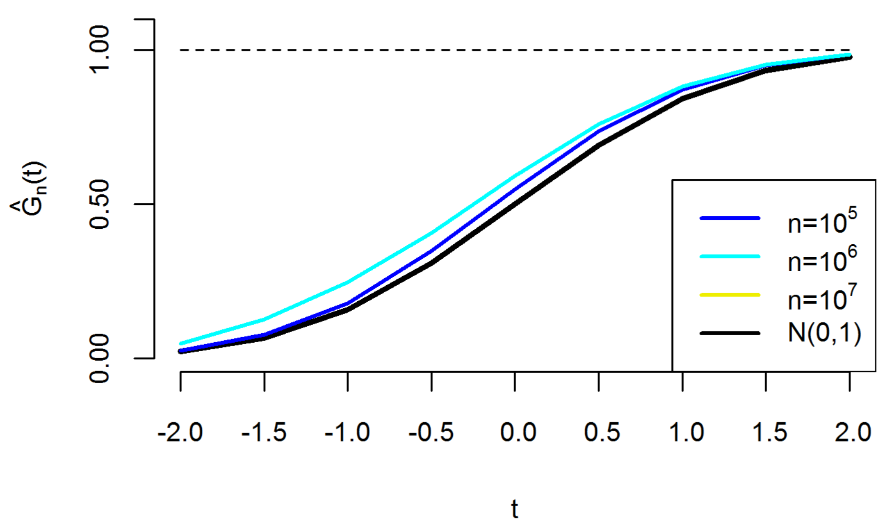

Figure 8: ASCLT convergence using C_n, p=500, and $n = 10^5, 10^6, 10^7$

Now that we have an understanding of how an asymptotic distribution is estimated by the ASCLT, we briefly introduce the contributions that have been made in the field of ASLT. We provide a literature review of ASLT in Section 2.1. For clarity, when the expression convergences almost surely to a normal distribution, it is referred to as an ASCLT. For any other limiting distribution, the theorem it called an almost sure limit theorem (ASLT). After the discovery of the ASCLT, proofs of the ASCLT using alternate methods by authors such as Lacey and Philipp (1990) and Peligrad and Révész (1991) provided additional insight into almost sure limit theory. By 1993, authors were noticing the relationship between convergence in distribution and the corresponding almost sure version (Berkes and Dehling, 1993). Within a few years, ASLT were proven for dependent data (Peligrad and Shao, 1995). A general framework of ASLT for i.i.d. random variables was provided by Berkes and Csáki (2001). The key result of Berkes and Csáki proved that if a convergence in law exists, then an ASLT converging to the same distribution can be found under some mild technical conditions. Many other contributions are listed in 2.1.

The result of Berkes and Csáki (2001) provides the framework for LQE. For a sequence of test statistics, say T_n, if we can prove that its distribution function $P(T_n \leq t)$ convergences in law to some distribution function $G(t)$, then we can derive an ASLT that converges to the same distribution function $G(t)$. Since the distribution function of the test statistic converges weakly to $G(t)$, and $\hat{G}_n(t) = \frac{1}{C_n} \sum_{k=1}^{n} \mathbb{I}(T_n \leq t)$ converges almost surely weakly to $G(t)$, we can use $\hat{G}_n(t)$ to estimate (approximate) the distribution function $P(T_n \leq t)$. Note that we are not estimating the asymptotic distribution function $G(t)$, but the unknown distribution of T_n directly. The quality of the approximation will depend upon the rates of convergence of both $P(T_n \leq t)$ and $\hat{G}_n(t)$ and how close their distributions are to $G(t)$ for small n. More precisely, the distance between $P(T_n \leq t)$ and $\hat{G}_n(t)$ is bounded by the sum of their distances to G(t), but in practice it may be much closer, as known for example in the case of Edgeworth expansion (Hall , 2013). In Section 2.2, we explore the technical aspects of LQE in more detail.

Another important element of the nonparametric LQE investigated in this dissertation is the field of rank statistics. Rank statistics have several desirable properties, including invariance under strictly monotone transformations (Lehmann, 1953), minimal distributional requirements, and robustness. Rank statistics use the overall ranks of the observations in place of their values, which makes the corresponding tests and quantile estimates nonparametric. Many rank tests employ permutations of the data to capitalize on the asymptotic distributional properties. The speed of convergence increases significantly when permutation techniques are applied to classical (weak) limit theorems. Since the purpose of most tests is to determine if two or more conditions (treatments, times, levels of treatments, distributions, etc.) are different, all permutations of the data between, as opposed to within the conditions of interest will result in similar distributions of the ranks under the null hypothesis H_0 : *no difference between conditions*. Rank tests employing permutations have the added benefit of converging to the asymptotic distribution for sufficient sample sizes. As there are $N!$ permutations for any single sample of N values, using the total number of possible permutations is often impractical for computational purposes. It is standard practice to randomly select a significant number of permutations (e.g. $p = 100$, $p = 500$, or $p = 1,000$) when applying rank tests. In some cases, a much smaller number of permutations are required to approximate the asymptotic distribution (e.g. $p = 20$ or $p = 50$). The asymptotic properties of rank tests are critical to the efficacy of the classical nonparametric methods.

Permutation methods for LQE serve a different purpose than those of classical nonparametric tests (Tabacu, 2014). In general, the EDF of a test statistic derived from an ASLT is not symmetric, and the quantiles may depend upon the random order in which the observations were selected Tabacu (2014). More precisely, the calculations for lower values of k are given more weight than subsequent calculations and may dominate the values of $\hat{G}_n(t)$ (Tabacu, 2014). An additional benefit of using permutations in LQE is a faster convergence rate as shown in Figures 1 to 8. The specific number of permutations for each test statistic must be determined by the researcher.

We use rank statistics in the fully nonparametric model developed by Akritas and Arnold (1994), Brunner and Denker (1994), Brunner and Puri (1996), Akritas et al. (1997), Akritas and Brunner (1997), among many others. In this framework, hypothesis are formulated using distribution functions. Let X_{ij}, $1 \leq i \leq a$, $1 \leq j \leq b$ be random variables from a statistical experiment, where the values of i, j denote some set of conditions under which the random variables are generated. The only distributional assumptions we make is that $X_{ij} \backsim F_{ij}$, where F_{ij} is a continuous marginal distribution function. The hypothesis that the random variables are from the same population is given by

$$H_0 : F_{11} = \cdots = F_{1b} = \cdots = F_{a1} = \cdots = F_{ab}. \tag{4}$$

The alternative hypothesis may be expressed specifically to address the desired test. We provide an omnibus alternative hypothesis as an example

$$H_1 : F_{ij} \neq F_{i'j'}, \tag{5}$$

for at least one $(i, j) \neq (i', j')$. Any parametric statement of the hypothesis will be implied by the nonparametric hypothesis. For example, the test that some parameter $\zeta = \zeta_0$ is included in the statement of identical distribution functions in (4). For a detailed discussion, see Akritas and Arnold (1994), Akritas et al. (1997). Hypotheses specific to the proposed tests are provided for each model investigated in Chapters 3-5. An overview of the dissertation is provided in the following section.

1.2 OVERVIEW OF THE DISSERTATION

The organization of this dissertation and an overview of the results are now provided. In Chapter 2, we provide a brief overview of the contributions to almost sure limit theory after the discovery of the ASCLT by Fisher (1987), Brosamler (1988), and Schatte (1988), which lead to the seminal work of Berkes and Csáki (2001). Theorems that are important to the results obtained in Chapters 3 - 5 are also provided. In the second half of the chapter, we present the key concepts and technical requirements for LQE, along with an investigation of some technical computational considerations for LQE.

In Chapter 3 we investigate three different factorial models presented in Akritas and Brunner (1996), and test for the presence of a pattern (trend, umbrella, etc.) across the levels of one of the factors under the LQE approach. Without loss of generality, we test the null hypothesis H_0*: no difference in* for the presence of a pattern across the levels of Factor A in the three models. The models investigated include a two-way fixed effects model, a partial hierarchical repeated measures model, and a cross-classification repeated measures model. We also test for the presence of interaction with a patterned alternative in the partial hierarchical model. Simulation studies are provided for type I error and power for trend and umbrella patterns in the alternative hypotheses for small sample sizes. Real data sets are analyzed for each model, and the results of the LQE tests agree with known corresponding analyses (where available). The type I error results are conservative, but power exceeds 80% for several alternative hypotheses.

Chapter 4 explores LQE for the change-point problem of detecting a change in distribution in a sequence (or stream) of observations. More precisely, we investigate LQE for the test proposed by Pettitt (1979) for small to moderate sample sizes. We analyze several small datasets and compare their results to those of Pettitt (1979), Lombard (1987), and Gombay (1994) where available. Simulation studies are provided for type I error and power. The test is liberal for smaller sample sizes but the type I error approaches the significance level as the sample sizes increase.

A new approach for the two-sample problem is examined in Chapter 5. We consider the ideas in Compagnone and Denker (1996) for increasing the efficiency of the nonparametric Wilcoxon-Mann-Whitney test with respect to the parametric t-test using an empirical process of U-statistic structure. We propose a new test statistic and provide empirical verification that it converges in law and its corresponding AS-CLT. Simulated type I error and power are provided for small independent samples of normally distributed random variables.

We summarize the results of the dissertation in Chapter 6, and we indicate some open problems.

CHAPTER 2

ALMOST SURE QUANTILE ESTIMATION

2.1 ALMOST SURE LIMIT THEOREMS (ASLT)

Almost sure limit theorems (ASLT) form the basis of logarithmic quantile estimation (LQE). The first result was the almost sure central limit theorem (ASCLT), independently proved by Fisher (1987), Schatte (1988), and Brosamler (1988) for different moment conditions. Lacey and Philipp (1990) were able to relax the moment conditions specified by Fisher, Brosamler, and Schatte. The simplest form of the ASCLT given by Lacey and Philipp (1990) follows. Let $(\Omega, \mathcal{B}, \mathcal{P})$ be a probability space. Let $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ be a random sample of independent and identically distributed (i.i.d.) random variables, and let any event $\omega \in \Omega$ be a sequence $\omega = (\omega_1, \omega_2, \ldots)$ of outcomes such that $P(\omega) > 0$ (i.e. $\omega \notin \mathcal{N}$ where $\mathcal{N}$ is the null-set of Ω). Without loss of generality assume $EX_1 = 0$ and $EX_1^2 = 1$. We denote the partial sum of the random variables as $S_n(\omega) = \sum_{k=1}^{n} X_k(\omega)$. $\mathbb{I}(\cdot)$ is the indicator function, and Φ is the standard normal distribution function. Then the simplest version of the ASCLT is (Lacey and Philipp, 1990)

$$\frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{S_k(\omega)}{\sqrt{k}} \leq t \right) \xrightarrow[n\to\infty]{a.s} \Phi(t), \qquad \forall t \in \mathbb{R}, \tag{6}$$

where $\log n$ is the natural logarithm of n. Brosamler (1988) observed that to verify a random number generator using the ASCLT, one only has to use a single (typical) path ω through Ω. In contrast, multiple paths are required for verification using the classical central limit theorem (CLT). In this dissertation, it is understood that ω is fixed, and it is omitted to simplify the notation.

There have been many advancements in the field of almost sure limit theory since the introduction of the ASCLT. Brosamler (1988) proved a functional form of the ASCLT using ergodic theory; Lacey and Philipp (1990) obtained an almost sure invariance principle using a probabilistic method. Lacey and Philipp (1990) also provided a limit for the rate of convergence for the ASCLT as $O\left(\frac{1}{\sqrt{\log n}}\right)$ in probability. Berkes and Dehling (1993) showed under mild technical conditions that partial sums S_k of independent but not necessarily identical random variables for any distribution function G and any Borel set $A \subset \mathbb{R}$ with boundary ∂A such that $G(\partial A) = 0$ and such that the $\mathcal{P}$-null set is independent of A , the expression

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left\{ \frac{S_k - b_k}{a_k} \in A \right\} = G(A), \qquad a.s., \tag{7}$$

is equivalent to the statement

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} P\left\{ \frac{S_k - b_k}{a_k} \in A \right\} = G(A), \tag{8}$$

where $a_k > 0$ and b_k are sequences of real numbers. It is also shown under these same conditions that the ASLT is implied by the convergence in distribution

$$\frac{S_k - a_k}{b_k} \xrightarrow{d} G.$$

Thus, under these mild technical assumptions, the ASLT is a weaker statement than convergence in distribution, unlike ordinary $a.s.$ convergence. Peligrad and Shao (1995) used the almost sure invariance principle from Lacey and Philipp (1990) to develop an ASCLT for stationary sequences with finite covariances, stationary mixing sequences, and strong mixing sequences. Fahrner and Stadtmüller (1998) developed ASLT for the maximum value of a sample. Consider a sequence of functions $T_n(X_1, \ldots, X_n) = \max_{1\le i\le n}\{X_i\}$, where X_i are real-valued i.i.d. random variables and $i, n \in \mathbb{N}$. The ASLT were developed by Fahrner and Stadtmüller (1998) under three categories of max-stable limiting distributions, which include

$$\begin{aligned} &\Lambda(x) = e^{-e^{-x}} && x \in \mathbb{R}, \\ &\Phi_\alpha(x) = e^{-x^{-\alpha}}\mathbb{I}(x > 0), && x \in \mathbb{R},\ \alpha > 0, \text{ and} \\ &\Psi_\alpha(x) = e^{-(-x)^\alpha}\mathbb{I}(x \le 0) + \mathbb{I}(x > 0) && x \in \mathbb{R},\ \alpha > 0. \end{aligned} \tag{9}$$

The specific technical conditions under which each of the three distributions in (9) are max-stable are provided along with conditions under which T_n does not converge almost surely. The same extreme-value distributions in (9) were studied by Cheng et al. (1998); however, a much broader scope of convergence was proven. In fact, whereas Fahrner and Stadtmüller (1998) claimed that only averaging very close to the logarithmic averaging results in convergence, Cheng et al. (1998) proved that *a.s.* convergence occurs for a wide range between logarithmic and Cesàro averaging (e.g. replacing $\log n$ with n or $n+1$). Ibragimov and Lifshits (2000) used characteristic functions to prove several ASLT. An ASLT for independent random vectors, and an ASCLT for i.i.d. random vectors were also confirmed by Ibragimov and Lifshits (2000). They proved an ASLT for weakly dependent random vectors. More precisely, they showed an ASLT for stationary sequences with expectation zero and finite variance under mild technical conditions. Additionally, an ASLT for a sequence of random vectors without the assumptions of independence or identical distribution is provided by Ibragimov and Lifshits (2000). For independent random variables Berkes and Csáki (2001) provided a key result that generalizes several previous observations concerning the relationship between limit theorems for convergence in law and ASLT. The following theorem states that every weak limit theorem (convergence in distribution) has a weighted almost sure version under mild technical conditions.

Theorem 1 (Berkes and Csáki (2001), Theorem 1)**.** *Let $X_1, X_2, \dots$ be independent rvs satisfying the weak limit theorem*

$$f_k(X_1, X_2, \dots, X_n) \xrightarrow{d} G, \tag{10}$$

where $f_k : \mathbb{R}^k \to \mathbb{R}$ $(k = 1, 2, \dots)$ are measurable functions and G is a distribution function. Assume that for each $1 \le k < l$ there exists a measurable function $f_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$ such that $E(|f_l(X_1, \dots, X_l) - f_{k,l}(X_{k+1}, \dots, X_l)| \wedge 1) \le A\,(c_k/c_l)$ with a constant $A > 0$ and a positive, nondecreasing sequence (c_n) satisfying $c_n \to \infty$ and $\frac{c_{n+1}}{c_n} = O(1)$. Put $d_k = \log(c_{k+1}/c_k)$, $D_n = \sum_{k \le n} d_k$. Then

$$\lim_{n\to\infty} \frac{1}{D_n} \sum_{k=1}^{n} d_k \mathbb{I}\{f_k(X_1, \dots, X_k) < x\} = G(x) \ a.s. \ \forall x \in C_G, \tag{11}$$

where C_G is the set of continuity points of G. The result remains valid if we replace the weight sequence (d_k) by any $(d_k^*) \ni\!:\ 0 \le d_k^* \le d_k, \sum d_k^* = \infty$.

The mild technical conditions in Theorem 1 may be summarized as follows (see Berkes and Csáki, 2001):

(i) The convergence in distribution for measurable functions f_k is not affected by removing finitely many random variables.

(ii) The sequence $\log(c_{k+1}/c_k)$ is positive, finite, and its infinite sum is infinity for a nondecreasing sequence $c_k \to \infty$, where the ratio of successive terms is finite (i.e. bounded).

Berkes and Csáki (2001) also noted that if we take $c_k = k^\epsilon$ for some $\epsilon > 0$, then $d_k \sim constant(1/k)$, and (11) becomes

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\{f_k(X_1, \ldots, X_k) < x\} = G(x) \qquad \text{a.s. } \forall x \in C_G. \tag{12}$$

Setting $f_k(X_1, \ldots, X_k) = \frac{S_k}{\sqrt{k}}$ yields the ASCLT (6). Berkes and Csáki (2001) also applied their results to show ASLT for special cases, such as, dependent processes with independent increments, extreme-value distributions similar to those of Cheng et al. (1998) and Fahrner and Stadtmüller (1998), maxima for partial sums, empirical distribution functions, U-statistics, local times, return times, and Darling-Erdös type limit theorems. Lifshits (2001) proved a multivariate ASLT for partial sums of independent random vectors. The ASLT for central order statistics was proven by Peng and Qi (2003). The moment conditions of Berkes and Csáki (2001) for U-statistics were relaxed by Holzmann et al. (2004). Holzmann et al. (2004) proved an ASLT with a stable limiting distribution for nondegenerate U-statistics and its functional version. Hörmann (2005, 2007) proves the convergence of (11) for a broader range of weights

$$d_k = \begin{cases} \frac{1}{k}, & k \in \mathbb{N} \\ \frac{(\log k)^\alpha}{k}, & \alpha > -1, \\ \frac{e^{(\log k)^\alpha}}{k}, & (0 \le \alpha < 1). \end{cases} \tag{13}$$

Increasing d_k makes the convergence in (6) stronger. A relation between D_k and d_k along with optimal criteria for d_k were provided (Hörmann, 2007). Peng et al. (2009) proved the ASLT for the joint distribution function of all order statistics. The vector martingale transform was investigated by Bercu et al. (2009). They showed that the normalized even moments of martingales follow an ASCLT. Denker and Fridline (2010) proved the following almost sure version of Cramér's theorem.

Theorem 2 (Denker and Fridline (2010), Theorem 2.1)**.** *Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a function which is differentiable in a neighborhood of some $\mu \in \mathbb{R}^d$ and its derivative g' is continuous at μ. Let X_n, $n \geq 1$ be a sequence of $\mathbb{R}^d$ valued random vectors satisfying the ASLT*

$$\frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\{\sqrt{k}(X_k-\mu)\leq t\} \xrightarrow[n\to\infty]{a.s} G_X(t), \qquad t \in D(G_X), \tag{14}$$

where G_X is the cumulative distribution function of some random variable X and $D(G_X)$ is the set of continuity points of G_X. If there exists a sequence $\mathbb{N}_0 = \{n_k : k \in \mathbb{N}\}$ of integers such that

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1;k\notin\mathbb{N}_0}^{n}\frac{1}{k} = 0, \qquad \textit{and} \tag{15}$$

$$\lim_{k\to\infty} X_{n_k} = \mu \qquad a.e. \tag{16}$$

then

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}\mathbb{I}\{\sqrt{k}\,(g(X_k)-g(\mu))\leq t\} = G_{g\prime(\mu)X}(t) \qquad t \in D(G_{g\prime(\mu)X}), a.e. \tag{17}$$

An almost sure version of Cramér's theorem has the powerful benefit of increasing the scope of application for almost sure limit theorems. By carefully selecting g so that the derivative $g\prime \in (0,1)$, the resulting asymptotic variance is lower than G_X (Denker and Fridline, 2010). Denker and Tabacu (2015) proved the ASCLT for linear rank statistics. An ASCLT for the ratio of order statistics from an exponential distribution was proved by Miao et al. (2016). There are many contributions to almost sure limit theory not listed, and the body of knowledge continues to progress.

With the large amount of results for almost sure limit theory, an intuitive step was to use ASLT for statistical analyses, such as quantile estimation and by extension, hypothesis tests. The method for such analyses was introduced by Thangavelu (2005) and is called logarithmic quantile estimation (LQE), which is discussed in Section 2.2.

2.2 LOGARITHMIC QUANTILE ESTIMATION (LQE)

LQE is an approach that uses ASLT to estimate the quantiles of a test statistic directly from the data without using the asymptotic distribution. Since tests based on the central limit theorem (CLT) involve estimation of the asymptotic variance, the LQE approach may be especially attractive when those expressions are complex. As we mentioned in the Introduction, the goal of this dissertation is to extend the LQE approach in three directions: analysis of factorial designs for the presence of a patterned alternative across the levels of one factor (Chapter 3), testing for a change-point in a sequence of data (Chapter 4), and investigating a new two-sample problem rank test based on U-statistic structure (Chapter 5). We begin this section with a literature review, which is followed by a discussion of the technical requirements for LQE. We empirically investigate the convergence behavior and rate of the ASCLT in (6), and we explore some technical computational properties of LQE to assist in understanding of the results in Chapters 3-5.

LQE is a relatively new statistical approach, and this fact allows us to provide a complete literature review. LQE was introduced by Thangavelu (2005). Thangavelu proposed a parametric LQE method for testing if the mean of a sample is zero for small sample sizes (10 and 15 observations), and the simulated type I error of the test was comparable to that of the t-test and was closer to the significance level when compared to the bootstrap method. Thangavelu (2005) also investigated a parametric and a nonparametric (rank statistic) LQE test for the Behrens-Fisher problem (BFP). The parametric and nonparametric LQE tests for the Behrens-Fisher problem performed competitively with Welch's t-test and the Wilcoxon-Mann-Whitney (WMW) test, respectively for sample sizes of 10, 15, and 30.

A parametric LQE method for estimating the confidence intervals for the correlation coefficient was developed by Fridline (2010). To arrive at an ASCLT for the Fisher transformation, Fridline proved an almost sure version of Cramér's theorem and applied it to a proposed ASCLT for the correlation coefficient. The estimated quantiles from the ASCLT for the desired significance level were used to calculate the confidence interval. These LQE confidence intervals were similar in width and coverage probabilities to confidence intervals generated using the bootstrap method for large samples with correlation coefficient values between 0.25 and 0.7.

Tabacu (2014) proved an ASCLT for linear rank statistics, which was used for developing LQE for the c-sample problem and a longitudinal factorial model. The c-sample problem compares c samples. Tabacu (2014) did not restrict the analysis to independent samples, resulting in an unknown asymptotic distribution under the null hypothesis that the c samples are all from the same distribution. For independent samples, the LQE quantiles in Tabacu (2014) compared favorably with the asymptotic chi-squared quantiles of the Kruskal-Wallis test statistic (Kruskal and Wallis, 1952). The simulated type I error for independent and dependent samples was very similar, and the results were conservative. Tabacu (2014) derived ASCLT for linear rank statistics to test a three-way longitudinal factorial design, and successfully used the LQE approach to analyze a real dataset with 41 subjects in two treatment groups stratified by gender and having repeated measures. The LQE p-values for the longitudinal study were comparable to those presented in Brunner, Domhof, and Langer (2002). Simulation studies were provided in Denker and Tabacu (2015). The type I error results were conservative.

In this dissertation, we restrict our discussion to nonparametric methods for LQE by replacing observations with their overall rank in the experiment (rank statistics). We define a sequence of rank test statistics, say T_n, $n \in \mathbb{N}$ on a common probability space. The following convergence relations are needed for LQE.

$$\lim_{n\to\infty} P(T_n < t) = G(t), \qquad \forall t \in \mathbb{R}, \tag{18}$$

$$\lim_{n\to\infty} \frac{1}{C_n}\sum_{k=1}^{n} \frac{1}{k}\mathbb{I}(T_k < t) = G(t), \qquad a.s., \forall t \in \mathbb{R}, \tag{19}$$

where G is a distribution function, and $C_n = \sum_{k=1}^{n} \frac{1}{k}$. The limiting distributions G in both (18) and (19) are identical, leading to the concept behind LQE. If the left-hand sides (LHS) of (18) and (19) converge to exactly the same distribution G, then it is intuitive to use the logarithmic summation in the LHS of (19) to estimate the unknown distribution function of the rank test statistic T_n in the LHS of (18). The logarithmic summation in (19) is calculated directly from the data, and it does not include an estimation of the asymptotic variance of T_n. Unlike CLT tests, LQE does not estimate the asymptotic distribution function of T_n, but approximates its actual distribution function.

The use of C_n in the ASCLT (19) instead of $\log n$ was proposed by Thangavelu (2005), because even though

$$\frac{\sum_{k=1}^{n} \frac{1}{k}}{\log n} \longrightarrow 1, \qquad a.s.,$$

this ratio is greater than one for all $n > 1$. Thangavelu noticed that the LHS of (19) is an empirical distribution function (EDF), making it an appropriate estimate for the distribution function of T_n. Fridline (2010) proved that the ASCLT in (19) is equivalent to the ASCLT when $\log n$ is used in place of C_n using an extension of Slutzky's theorem. Fridline (2010, Lemma 2.2, page 22) extended Slutzky's theorem to the almost sure weak version, and used the lemma to prove that exchanging $\log n$ for C_n resulted in the equivalent ASCLT in (19).

To estimate quantiles for the distribution of T_n, one can invert the EDF defined by Thangavelu (2005) and refined by Tabacu (2014)

Definition 2.2.1 (Tabacu, 2014, Definition 3.1.3, p. 27)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables defined on the same probability space. Let $T_n = T_n(X_1, \dots, X_n)$ be a sequence of test statistics where T_n is a function of $X_1, \dots, X_n$. Then the logarithmic type empirical distribution function is defined as

$$\hat{G}_n(t) = \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}(T_k \leq t), \qquad \forall t \in \mathbb{R}. \tag{20}$$

In general, $\hat{G}_n(t)$ converges almost surely to $G(t)$. Logarithmic empirical α-quantiles (i.e. logarithmic α-quantile estimates) are defined as the inverse of the logarithmic EDF (Thangavelu, 2005).

Definition 2.2.2 (Logarithmic α-quantile estimates)**.** Let $\alpha \in [0, 1]$, $n \in \mathbb{N}$, and $\hat{G}_n(t)$ be defined as in Definition 2.2.1 above. Then the (α, n)- logarithmic quantile estimate of a sequence of test statistics T_n is

$$\hat{t}_{\alpha}^{(n)} = \hat{G}_n^{-1}(\alpha) = \begin{cases} \sup\{t|\hat{G}_n(t) = 0\} & \text{, for } \alpha = 0 \\ \sup\{t|\hat{G}_n(t) < \alpha\} & \text{, for } \alpha \in (0, 1) \\ \inf\{t|\hat{G}_n(t) = 1\} & \text{, for } \alpha = 1. \end{cases} \tag{21}$$

We refer to $\hat{t}_\alpha^{(n)}$ throughout the dissertation as a logarithmic α-quantile estimate, where the dependence upon n is understood.

Several properties for logarithmic α-quantile estimates were provided by Thangavelu (2005) and are stated below for convenience. See Tabacu (2014) for detailed proofs.

1. Let t_α denote the true α-quantile of continuous distribution function $G(t)$, then

$$\lim_{n\to\infty} \hat{t}_\alpha^{(n)} = t_\alpha, \qquad a.s.$$

2. When using the LQE approach, the type I error of a test converges a.s. to the significance level under the null hypothesis, which for the nonparametric case is that all random variables are from the same distribution function.

3. Under any specific alternative hypothesis, the power of the test converges a.s. to 1.

We now investigate some of the computational properties of LQE by examining the equations of $\hat{G}_n(t)$ and $\hat{t}_\alpha^{(n)}$. In practice, the logarithmic α-quantile estimate in (21) for all $\alpha < 1$ becomes

$$\hat{t}_\alpha^{(n)} = \max\left\{ t \mid \frac{1}{C_n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}(T_k \le t) \le \alpha \right\}. \tag{22}$$

$\hat{G}_n(t)$ has a discrete number of possible values which are determined by n. This in turn limits the values of α for which $\hat{t}_\alpha^{(n)}$ can be precisely determined. To fully understand the potential effects of this restriction, we use a small value of n. For example, if $n = 5$, then the two largest possible values $\hat{G}_5(t)$ can take (rounded to 4 decimal places) are

$$\frac{\sum_{k=1}^{4}\frac{1}{k}}{\sum_{k=1}^{5}\frac{1}{k}} = 0.9124, \qquad \text{and} \qquad 1.$$

For $\alpha = 1$ in (21), we choose the smallest value of t such that $\hat{G}_n(t) = 1$, say $\hat{t}_1^{(5)}$. Then for any $0 < \delta \le 1 - 0.9124$, $\alpha = 1 - \delta$ in equation (22) results in a $t = \hat{t}_{1-\delta}^{(5)} = \hat{t}_{0.9124}^{(5)} < \hat{t}_1^{(5)}$. For example, if a value for α is arbitrarily chosen as, say $\alpha = 0.95$ when

$n = 5$, then the quantile $\hat{t}^{(5)}_{0.95} = \hat{t}^{(5)}_{0.9124}$, and if the value of the test statistic T_5 exceeds $\hat{t}^{(5)}_{0.95}$ for a one-sided test, without any further information, a p-value of less than 0.05 would be reported, instead of p-value $< 1 - 0.9124 = 0.0876$. Hence, we do not have enough information to determine a p-value < 0.0876, and caution should be used when selecting values of α for small sample sizes. The minimum p-value that can be determined for $5 \leq n \leq 16$ are provided in Table 1 for convenience.

Table 1: Precision of LQE (minimum p-value)

n	p-value	n	p-value	n	p-value
5	0.0876	6	0.0680	7	0.0551
8	0.0460	9	0.0393	10	0.0341
11	0.0301	12	0.0269	13	0.0242.
14	0.0220	15	0.0201	16	0.0185

The p-values listed are the smallest p-value that can be precisely determined by LQE for the corresponding value of n.

Another technical computational property was introduced by Fridline (2010). The random order of the observations may have a significant effect on the calculations, because we are only selecting observations from one sequence of random variates, and the largest weights for $\frac{1}{k}$ $(1, 0.5, \frac{1}{3}, \ldots)$ are applied to the earliest selected observations. Fridline (2010) showed that omitting a small number (relative to n), say k_0, of the initial calculations reduces the impact of the larger weights, which may increase the rate of convergence, and results in an expression that is an ASLT with the same limiting distribution.

2.3 CONCLUSION

In this chapter, we introduced the ASCLT and described important theoretical developments in almost sure limit theory. We then showed how LQE is an extension of almost sure limit theory to hypothesis testing. Some properties and technical aspects of LQE were reviewed, and are summarized here. The convergence of ASLT is very slow; however, permutations of the sequence of random variables increase the symmetry of the logarithmic summation and increases the rate of convergence of the ASLT. Logarithmic quantile estimates approach the true quantiles of the asymptotic distribution as the number of observations increase. The type I error and power estimates approach the significance level and one almost surely, respectively. The distribution to which the ASLT converges is not affected by removing finitely many initial random variables (vectors), and the effects of the random order of the observations on logarithmic quantile estimate convergence may be mitigated if a relatively small number of calculations (compared to the sample size) are omitted. The precision of the p-value is restricted by the number of terms in the logarithmic summation, and consideration of the precision should be considered during decision-making. Now that the theory and practical aspects of LQE have been explored, in the following three chapters, we investigate the three statistical analyses described in the Introduction.

CHAPTER 3

TESTS FOR PATTERNED ALTERNATIVES

3.1 INTRODUCTION

In this chapter we introduce nonparametric tests for patterned alternatives in factorial designs via logarithmic quantile estimation (LQE). It is well known that tests for patterned alternatives result in higher power compared to global tests. Nonparametric tests for patterned alternatives were developed by Terpstra (1952) and Jonckheere (1954) for independent samples. Tests for dependent samples were developed by Page (1954). The method for independent samples proposed by Hettmansperger and Norton (1987) was generalized to mixed effects factorial models by Akritas and Brunner (1996), Brunner and Puri (2001), Brunner, Domhof, and Langer (2002), Callegari and Akritas (2004), among several others.

Results of tests for increasing trend and umbrella type patterned alternatives are also provided for three different factorial models in this chapter. The three models considered are a fixed effects two-way factorial model, a partial hierarchical repeated measures model with two fixed factors, and a cross-classification repeated measures model with two crossed fixed factors as described in Akritas and Brunner (1996). The three models are fully nonparametric factorial designs as proposed and developed by Akritas and Arnold (1994), Brunner and Denker (1994), Brunner and Puri (1996), Akritas and Brunner (1997), Akritas et al. (1997), Brunner and Puri (2001), among others. The hypotheses for fully nonparametric models are formulated using only the distribution functions, and the test statistics are defined using the overall ranks of the observations in place of the values. These rank tests have the advantage of being invariant under monotone transformations of the response variable unlike the corresponding linear parametric models. Additionally, the assumptions of normality and homoscedasticity (homogeneity of variances) are not requirements under these models.

Our contribution in this chapter is the extension of LQE to patterned alternatives in fully nonparametric factorial designs. For the models considered in this chapter, Akritas and Brunner (1996) propose the use of linear rank statistics divided by the square root of their asymptotic variance to test for patterned alternatives. For small sample sizes, these tests use a central t-distribution requiring the estimation of the degrees of freedom using the Satterthwaite-Smith approximation. When using LQE, there is no need for estimating the asymptotic variance or calculating the Satterthwaite-Smith degrees of freedom.

The chapter is organized as follows. In Section 3.2 we introduce the models, the hypotheses, and the test methods. The simulated type one error and power of the tests along with applications to datasets using LQE are provided in Section 3.3. Section 3.4 is the conclusion, and Section 3.5 is the appendix containing a sketch of the proofs.

3.2 MODELS AND TESTS STATISTICS

We investigate three distinct models in this chapter. The two-factor fixed effects model consists of multiple levels in each of the factors under experimentation. The individual observations are exposed to exactly one level of each factor (referred to as cells) and are independent of other observations in the model. The partial hierarchical design consists of randomly chosen experimental units nested under one treatment level (say factor B) and measured at multiple time points or locations (factor A) under the only one treatment level. For simplicity we will refer to the repeated measurements as time points without loss of generality. The final design is a cross-classification repeated measures model where each experimental unit is randomly selected, and measurements are obtained for all combinations of the two fixed-effect treatments (factors A and B). Both factors A and B are repeated measurements where factor A is nested within factor B. We look at these models from a nonparametric point of view in the context given by Akritas and Brunner (1997).

In the fully nonparametric setting, hypotheses are stated in terms of the distribution functions, and the test statistics are expressed in terms of relative treatment effects. Relative treatment effects describe the tendency of a marginal distribution function, say for a specific combination of levels of factor A and factor B, with respect to the overall distribution function of the experiment. To explain the relationship between relative treatment effects and the distribution functions, it is first necessary to define several quantities which will be used throughout the chapter. Let X_{ijk} represent an observation for the i^{th} level of factor A, the j^{th} level of factor B, and the k^{th} subject, where $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n$. Although the models presented are valid for unbalanced designs, we limit our discussion to balanced models ($n_{ij} \equiv n$), which results in less complex expressions and matches the simulations performed by Akritas and Brunner (1996). We denote the total number of observations as $N = abn$. We assume X_{ijk} has the continuous marginal distribution function $F_{ij}(x) = P\left(X_{ijk} \leq x\right)$, for all real x. The vector of distribution functions for the experiment is given as

$$F = (F_{11}, \ldots, F_{1b}, \ldots, F_{a1}, \ldots, F_{ab})'.$$

The corresponding empirical marginal distribution function is

$$\hat{F}_{ij}(x) = n^{-1} \sum_{k=1}^{n} \mathbb{I}(X_{ijk} \leq x),$$

where $\mathbb{I}(A)$ is the indicator function of set A. The overall distribution function for the experiment is defined by

$$H(x) = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} F_{ij}(x). \tag{23}$$

The empirical form of $H(t)$ is expressed by

$$\hat{H}(x) = \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \mathbb{I}(X_{ijk} \leq x). \tag{24}$$

Let the vector of relative treatment effects be

$$p = (p_{11}, \ldots, p_{1b}, \ldots, p_{a1}, \ldots, p_{ab}),$$

where

$$p_{ij} = \int H dF_{ij}.$$

Relative treatment effects describe the probabilistic tendency of the marginal distribution function F_{ij} with respect to the overall mean distribution function H (Brunner, Domhof, and Langer, 2002). If $p_{ij} < \frac{1}{2}$, then F_{ij} tends to lie in the region to the left of H. When $p_{il} > \frac{1}{2}$, then F_{ij} tends to lie to the right of H, and if $p_{ij} = \frac{1}{2}$, then F_{ij} does not tend to lie on either side of H (Brunner, Domhof, and Langer, 2002). It is of interest to test the null hypothesis that the relative treatment effects are the same (i.e. no treatment effect exists)

$$H_0^p : p_{11} = \cdots = p_{1b} = \cdots = p_{a1} = \cdots = p_{ab}. \tag{25}$$

For convenience, we wish to express the hypothesis in terms of distribution functions. Brunner, Domhof, and Langer (2002) notes that the hypothesis in (25) is implied by the hypothesis

$$H_0^\mu : F_{11} = \cdots = F_{1b} = \cdots = F_{a1} = \cdots = F_{ab}. \tag{26}$$

For an overview of the origins of relative treatment effects, their use in testing nonparametric models, and the derivation of asymptotically valid inference procedure having good small sample properties see Brunner et al. (2017).

The use of ranks in place of actual observations for testing nonparametric hypotheses results in the robustness of the test statistics when outliers are present (Akritas and Brunner, 1997) and the invariance of the data under monotone transforms, unlike the corresponding classical linear models where the main effect may disappear or be reversed (Akritas et al., 1997). Let R_{ijk} be the rank of the observation X_{ijk} among all N observations in the experiment. We define the average ranks for cell (i, j) and for treatment i of factor A by

$$\bar{R}_{ij\cdot} = \frac{1}{n}\sum_{k=1}^{n} R_{ijk}, \tag{27}$$

$$\bar{R}_{i\cdot\cdot} = \frac{1}{b}\sum_{j=1}^{b} \bar{R}_{ij\cdot}, \tag{28}$$

respectively. The unbiased and consistent estimate of the relative treatment effect p_{ij} is expressed in terms of ranks as follows

$$\hat{p}_{ij} = \int \hat{H} d\hat{F}_{ij} = N^{-1}(\bar{R}_{ij\cdot} - 0.5).$$

In order to formulate a test statistic for patterned alternatives, Hettmansperger and Norton (1987) assigned weights to the rank means corresponding to the hypothesized pattern in the factor of interest. Without loss of generality, we assume that factor A contains the hypothesized pattern, and we assign integer weights in accordance with the recommendations of Hettmansperger and Norton (1987). The vector of weights is

$$w = (w_1, \ldots, w_a)'.$$

The general form of the test statistic for patterned alternatives (Brunner and Puri, 1996) is given as

$$P_N = \sqrt{N} w' C \hat{p}, \tag{29}$$

which under the specific null hypothesis has an asymptotic normal distribution with mean 0 and variance

$$\sigma^2 = w' C V C' w, \tag{30}$$

where the contrast matrix C is determined by the hypothesis. A general formulation of the contrast matrices for various hypotheses are provided in Akritas and Brunner (1997). The structure of asymptotic covariance matrix V of the ranks of the observations is specific to each factorial model. We refer to Brunner and Puri (2001) for a detailed discussion. In the sequel, we provide the appropriate form of P_N in (29) for each of the investigated models. The form of w is

$$w_i = i, \qquad\qquad i = 1, \ldots, a, \tag{31}$$

for the increasing trend alternative and is

$$w_i = \begin{cases} i & \text{for } i < l, \\ 2l - i & \text{for } l \leq i \leq a. \end{cases} \tag{32}$$

for the umbrella pattern, where l corresponds to the known location of the peak, i.e. the level of factor A demonstrating the greatest relative treatment effect (equivalently largest average rank). The test statistic used by Brunner and Puri (1996) and Akritas and Brunner (1996) is $L_N = P_N / \hat{\sigma}_N$, where $\hat{\sigma}_N^2 = w' C \hat{V}_N C' w$ is a consistent estimator of σ^2 in (30). For brevity, we refer to Akritas and Brunner (1997) for the details of V and $\hat{V}_N$.

As we mentioned in the Introduction of this chapter, the aim is to approximate quantiles of the rank test statistics for patterned alternatives using the almost sure quantile estimation approach (LQE). The LQE approach uses only the data to estimate the distribution of the test statistic without any estimation of the asymptotic variance. Hence, we use the expression of P_N in (29) to derive the LQE. In the following subsections we discuss each model, explicit forms of the corresponding rank test statistic, and the main result that allows us to estimate quantiles almost surely. These results are almost sure central limit theorems (ASCLT) for each rank test statistic. For the proofs we use the ideas of Denker and Tabacu (2014), and define a sequence of independent random vectors Z_k (which have a closed form for each design) for which we need to assume the following.

Assumption 1. Each Z_k, $1 \le k \le n$, has a finite covariance matrix Σ_k such that

$$\frac{\Sigma_1 + \ldots \Sigma_n}{n} \to \Sigma, \quad \text{as } n \to \infty. \tag{33}$$

3.2.1 TWO-FACTOR FIXED EFFECT MODEL

The two-factor fixed effect model consists of factors A (levels $i = 1, \ldots, a$) and B (levels $j = 1, \ldots, b$) with the levels chosen or fixed by the researcher. The independent observations X_{ijk} represent a unique individual that is exposed to exactly one (i, j) combination of treatments. The n randomly selected individuals in each (i, j) cell are identically distributed such that $X_{ijk} \backsim F_{ij}$.

The hypothesis of no main effect of factor A is stated in terms of distribution functions

$$H_0^F(A) : \bar{F}_{1\cdot} = \cdots = \bar{F}_{a\cdot}, \tag{34}$$

where $\bar{F}_{i\cdot} = b^{-1} \sum_{j=1}^{b} F_{ij}$. The alternative hypotheses of an increasing trend across the levels of factor A is given by

$$H_1^F(A) : \bar{F}_{1\cdot} \ge \cdots \ge \bar{F}_{a\cdot}, \tag{35}$$

with at least one strict inequality. Likewise, the hypothesis for an umbrella shaped pattern with a peak at level l of factor A may be expressed by

$$H_1^F(A) : \bar{F}_{1\cdot} \geq \cdots \geq \bar{F}_{l\cdot} \leq \cdots \leq \bar{F}_{a\cdot}, \qquad \text{for } 1 < l < a, \tag{36}$$

with at least one strict inequality on both sides of $\bar{F}_{l\cdot}$. Let $P_a = I_a - \frac{1}{a}1_a 1_a'$ denote the projection matrix, where I_a is the identity matrix of dimension a and 1_a is the vector of ones with length a. The average weight is defined as $\bar{w} = a^{-1}\sum_{i=1}^{a} w_i$. By using the contrast matrix $C_A = P_a \otimes \frac{1}{b}1_b'$ in (29), we can express the test statistic as

$$P_N^{(fix)}(A) = \frac{1}{\sqrt{N}} \sum_{i=1}^{a} (w_i - \bar{w})\, \bar{R}_{i\cdot\cdot}, \tag{37}$$

where the (fix) denotes the fixed effects model, and A identifies the factor tested. The asymptotic variance in (30) is consistently estimated (see Akritas and Brunner, 1996) by

$$\hat{\sigma}^2_{N(fix)}(A) = \frac{1}{Nb^2 n(n-1)} \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} \left[(w_i - \bar{w})\left(R_{ijk} - \bar{R}_{ij\cdot}\right)\right]^2. \tag{38}$$

For small sample sizes, the distribution of $P_N^{(fix)}(A)/\hat{\sigma}_{N(fix)}(A)$ proposed by Akritas and Brunner (1996) is approximated by a central $t_{\nu_{(fix)}}$-distribution where the degrees of freedom $\nu_{(fix)}$ are derived from the Satterthwaite-Smith approximation in the form

$$\nu_{(fix)} = \frac{(n-1)\left[\sum_{i=1}^{a}(w_i - \bar{w})^2 \sum_{j=1}^{b}\sum_{k=1}^{n}\left(R_{ijk} - \bar{R}_{ij\cdot}\right)^2\right]^2}{\sum_{i=1}^{a}(w_i - \bar{w})^4 \sum_{j=1}^{b}\left[\sum_{k=1}^{n}\left(R_{ijk} - \bar{R}_{ij\cdot}\right)^2\right]^2}.$$

Our goal is to compute the empirical logarithmic quantiles of $P_N^{(fix)}(A)$ using following almost sure central limit theorem. In this case there is no need for the calculation of the variance estimate $\hat{\sigma}^2_{N(fix)}(A)$ in (38), or the Satterthwaite-Smith approximation of degrees of freedom, $\nu_{(fix)}$.

Proposition 1. For the two-way fixed effects model under Assumption 1 the statistic $P_N^{(fix)}(A)$ satisfies the almost sure central limit theorem

$$\lim_{N\to\infty} \frac{1}{C_N} \sum_{k=1}^{N} \frac{1}{k} \mathbb{I}\left(P_k^{(fix)}(A) \leq t\right) = G(t) \tag{39}$$

almost surely $\forall t \in \mathbb{R}$, where G is $\mathcal{N}(0, \sigma^2)$ with σ^2 defined in (30), and $C_N = \sum_{m=1}^{N} \frac{1}{m}$ with $N = abn$.

Note that there are $N = abn$ summations corresponding to the number of independent random vectors of length one. A sketch of the proof is provided in the appendix. The simulation results are provided in Section 3.3.2.

3.2.2 PARTIAL HIERARCHICAL MODEL

Akritas and Brunner (1996, 1997) describe the partial hierarchical model as a three-way model with two fixed factors and a random factor. The fixed factors A with levels $i = 1, \ldots, a$ (repeated measures) and B with levels $j = 1, \ldots, b$ (treatment group/level) are crossed. The individual subjects form the different levels of the random factor and are nested within factor B. An example occurs when subjects are randomly divided into several treatment groups and repeated measurements are taken at several subsequent times. The subjects nested in each group form independent random vectors. These independent random vectors are given as

$$X_{jk} = (X_{1jk}, \ldots, X_{ajk})',$$

where $1 \leq j \leq b$ are the treatment levels of factor B, and $1 \leq k \leq n$ are the subjects in the j^{th} treatment level. There are bn subjects and hence bn independent random vectors in the study, and the total number of observations is $N = abn$. Each observation X_{ijk} is distributed as $F_{ij}(x)$, $1 \leq i \leq a$, $1 \leq j \leq b$. Unlike the two-factor fixed effects model in Section 3.2.1, the repeated measurements within each subject may have some level of dependency. We will test for patterned alternatives in factor A and in the interaction between factors A and B. We now provide the details and the main results of this section.

Main effect for factor A

We test the null hypothesis in (34) against the patterned alternatives given in (35) and (36). Using the contrast matrix C_A defined in Section 3.2.1, the expression for P_N in (29) for this model may be stated in terms of the independent random vectors

$$P_{bn}^{(ph)}(A) = \frac{1}{b\sqrt{a}} \frac{1}{\sqrt{bn}} \sum_{i=1}^{a} \sum_{j=1}^{b} (w_i - \bar{w}) \bar{R}_{ij\cdot}, \tag{40}$$

where (ph) identifies the partial hierarchical model, A indicates the factor tested, and the consistent estimate of the asymptotic variance given in (30) is

$$\hat{\sigma}^2_{N(ph)}(A) = \frac{1}{Nb^2 n(n-1)} \sum_{i=1}^{a} (w_i - \bar{w})^2 \sum_{k=1}^{n} \left[\sum_{j=1}^{b} (R_{ijk} - \bar{R}_{ij\cdot}) \right]^2 . \tag{41}$$

The Satterthwaite-Smith estimated degrees of freedom for small sample sizes in the central $t_{\nu_{(ph)}}$-distribution approximation of $P_{bn}^{(ph)}(A)/\hat{\sigma}_{N(ph)}(A)$ is

$$\nu_{(ph)} = \frac{(n-1)\left[\sum_{i=1}^{a} \sum_{k=1}^{n} \left[\sum_{j=1}^{b} (w_i - \bar{w})(R_{ijk} - \bar{R}_{ij\cdot}) \right]^2 \right]^2}{\sum_{i=1}^{a} \left[\sum_{k=1}^{n} \left[\sum_{j=1}^{b} (w_i - \bar{w})(R_{ijk} - \bar{R}_{ij\cdot}) \right]^2 \right]^2}. \tag{42}$$

The almost sure central limit theorem for testing the main effect across the repeated measures is provided in the following Proposition.

Proposition 2. For the partial hierarchical model under Assumption 1 the statistic $P_{bn}^{(ph)}(A)$ satisfies the almost sure central limit theorem

$$\lim_{n\to\infty} \frac{1}{C_{bn}} \sum_{k=1}^{bn} \frac{1}{k} \mathbb{I}\left(P_k^{(ph)}(A) \leq t \right) = G(t) \tag{43}$$

almost surely $\forall t \in \mathbb{R}$, where G is $\mathcal{N}(0, \sigma^2)$ with σ^2 defined in (30), and $C_{bn} = \sum_{m=1}^{bn} \frac{1}{m}$.

It is important to note that the number of summations has been reduced to bn due to the dependence structure of the model. The sketch of the proof provided in the Appendix is similar to that of Proposition 1. Results of simulation studies are provided in Section 3.3.3.

Interaction effect

Often, researchers encounter differences in the behavior of subjects across repeated measurements depending upon the level of treatment (factor B) to which they are exposed. We consider a commonly occurring case where factor B has two levels such as treatment group ($j = 1$) and control group ($j = 2$), although the analysis is valid for any number of levels of factor B. The difference in the profiles of the treatment groups across repeated measurements (e.g. time) may be tested under the hypothesis of no interaction between factors A and B:

$$H_0^F(AB) : F_{i1} - F_{i2} = \bar{F}_{\cdot 1} - \bar{F}_{\cdot 2}, \qquad i = 1, \ldots, a. \tag{44}$$

The alternative hypotheses for increasing trend and umbrella patterns, respectively, are

$$H_1^F(AB) : F_{11} - F_{12} \geq \cdots \geq F_{a1} - F_{a2}, \tag{45}$$

with at least one strict inequality, and

$$H_1^F(AB) : F_{11} - F_{12} \geq \cdots \geq F_{l1} - F_{l2} \leq \ldots, \leq F_{a1} - F_{a2} \tag{46}$$

where $1 < l < a$ and $F_{11} - F_{12} > F_{l1} - F_{l2} < F_{a1} - F_{a2}$. Define the contrast matrix $C_{AB} = P_a \otimes (1, -1)$. The test statistic (29) has the form

$$P_{bn}^{(ph)}(AB) = \frac{1}{\sqrt{a}} \frac{1}{\sqrt{bn}} \sum_{i=1}^{a} (w_i - \bar{w}) \left(\bar{R}_{i1\cdot} - \bar{R}_{i2\cdot} \right), \tag{47}$$

where AB indicates the test for an interaction between factors A and B. The asymptotic variance in (30) is consistently estimated by

$$\hat{\sigma}^2_{N(ph)}(AB) = \frac{1}{Nn(n-1)} \sum_{i=1}^{a} \sum_{k=1}^{n} \left[\sum_{j=1}^{2} (w_i - \bar{w})(R_{ijk} - \bar{R}_{ij\cdot}) \right]^2 . \tag{48}$$

The degrees of freedom for the small sample approximation used by Akritas and Brunner (1996) are given in (42) with $b = 2$. We provide the main result for testing the interaction effect between time and group factors.

Proposition 3. For the partial hierarchical model under Assumption 1 the test statistic $P_{bn}^{(ph)}(AB)$ satisfies the almost sure central limit theorem

$$\lim_{n\to\infty} \frac{1}{C_{bn}} \sum_{k=1}^{bn} \frac{1}{k} \mathbb{I}\left(P_k^{(ph)}(AB) \leq t\right) = G(t) \tag{49}$$

almost surely $\forall t \in \mathbb{R}$, where G is $\mathcal{N}(0,\sigma^2)$ with σ^2 defined in (30), and $C_{bn} = \sum_{m=1}^{bn} \frac{1}{m}$.

The proof is similar to the proof of Proposition 2 and a sketch is provided in the Appendix. Simulation results are provided in Section 3.3.3.

3.2.3 CROSS-CLASSIFICATION REPEATED MEASURES MODEL

The selected cross-classification repeated measures model has n randomly selected subjects (levels of the random effect) and the repeated measurements are on two fixed factors A and B with factor A (levels $i = 1, \ldots, a$) nested within factor B (levels $j = 1, \ldots, b$). Each subject is measured at every (i,j) combination. The independent random vectors correspond to each subject and are given as

$$X_k = (X_{11k}, \ldots, X_{1bk}, \ldots, X_{a1k}, \ldots, X_{abk})', \qquad k = 1, \ldots, n.$$

The observations have distributions $X_{ijk} \backsim F_{ij}$. We test the null hypothesis in (34) against the alternative hypotheses for increasing trend and umbrella patterns in (35) and (36), respectively. Let the contrast matrix C_A be as specified in Section 3.2.1. The resulting form of the statistic P_N given in (29) is

$$P_n^{(cc)}(A) = \frac{1}{\sqrt{abn}} \sum_{i=1}^{a} (w_i - \bar{w}) R_{i\cdot\cdot}, \tag{50}$$

where the (cc) identifies the cross-classification repeated measures model, A indicates the factor tested. The asymptotic variance given in (30) is consistently estimated by

$$\hat{\sigma}^2_{N(cc)}(A) = \frac{a}{b} \frac{1}{N^2(n-1)} \sum_{k=1}^{n} \left[\sum_{i=1}^{a} \sum_{j=1}^{b} (w_i - \bar{w})(R_{ijk} - \bar{R}_{ij\cdot}) \right]^2 . \tag{51}$$

The small sample approximation of $P_n^{(cc)}(A)/\hat{\sigma}_{N(cc)}(A)$ in Akritas and Brunner (1996) is a central t_{n-1} distribution. The following Proposition allows for the almost sure quantile estimation using the algorithm in Section 3.3.1.

Proposition 4. For the cross-classification repeated measures model under Assumption 1 the test statistic $P_n^{(cc)}(A)$ satisfies the almost sure central limit theorem

$$\lim_{n\to\infty} \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(P_k^{(cc)}(A) \leq t\right) = G(t) \tag{52}$$

almost surely $\forall t \in \mathbb{R}$, where G is $\mathcal{N}(0, \sigma^2)$ with σ^2 defined in (30), and $C_n = \sum_{m=1}^{n} \frac{1}{m}$.

It is of interest to note that the number of summations match the n independent random vectors in the model. The proof is similar to the proof of Proposition 2 and is omitted. Results of simulations studies and the analysis of a dataset are provided in Section 3.3.4.

3.3 ANALYSES

In this section we provide the algorithm for computing logarithmic quantiles and the results of simulation studies for power and type I error along with the analysis of a dataset under the cross-classification repeated measures model. Extensive simulation studies for type I error and power were performed for each of the three models. We have provided some of the results in Sections 3.3.2 through 3.3.4.

3.3.1 ALGORITHM

A form of the following algorithm was first proposed by Thangavelu (2005) for testing the nonparametric Behrens-Fisher problem. Denker and Tabacu (2014 and 2015) used the same type of algorithm to test nonparametric hypotheses for quadratic rank statistics. We provide the algorithm here for convenience.

1. For a sample of N independent random vectors, permute the order of selection of the vectors "*nper*" times.

2. Calculate the logarithmic α-quantile estimate for the i^{th} permutation using

$$\hat{t}_{\alpha}^{i,(N)} = max\left\{t \left| C_N^{-1}\sum_{k=1}^{N} k^{-1}\mathbb{I}\left(P_k < t\right) \leq \alpha \right.\right\}, \tag{53}$$

where P_k is the appropriate linear rank test statistic for patterned alternatives of the form given in the corresponding Proposition, and $C_N = \sum_{k=1}^{N}\frac{1}{k}$. Recall from Section 2.2 that the summation may start from some small value k_0 relative to n, which reduces the influence of the initial observations.

3. Calculate the estimated logarithmic α-quantile for each simulation as

$$\bar{t}_{\alpha}^{(N)} = \frac{\hat{t}_{\alpha}^{i,(N)}}{nper}. \tag{54}$$

4. Reject the null hypothesis when $P_N > \bar{t}_{\alpha}^{(N)}$.

3.3.2 TWO-WAY FIXED EFFECT MODEL

Akritas and Brunner (1996) obtain type I error approximations that agree closely to the nominal levels ($\alpha = 0.10$, 0.05, and 0.01) when they use $P_N^{(fix)}(A)/\hat{\sigma}_{N(fix)}(A)$ to test (34) against the increasing trend alternative (35). They use random variables from a discrete uniform $U(1, 3)$ distribution with small sample sizes $n = 6$ in each (i, j) cell, $i = 1, \ldots, a$, $j = 1, \ldots, b$, with $b = 2$ levels of factor B. They employ $5,000$ simulations. They test designs with $a = 3$ and $a = 20$ levels of factor A. We have included their results in Table 3. Likewise, we use 5000 simulations. The results are stable for 50 permutations. For sample sizes 6, 8, and 10, we use LQE with the statistic $P_N^{(fix)}(A)$ in (37) to test for the same increasing trend alternative under the same designs and for $Gamma(4, \frac{1}{2})$, $N(0, 1)$, $Exp(1)$, and $U(1, 3)$ random variables in Table 2. The almost sure results for the exponential, normal, discrete uniform, and the gamma distributions are slightly conservative at the 0.10 level for $a = 3$ and slightly liberal for $a = 20$. The results become increasingly conservative as the significance level (α) decreases. The same analysis performed for a decreasing trend pattern achieved similar results.

Table 2: Type I error for main effect: trend pattern (fixed effects model)

		a=3			a=20		
Distribution	n	.10	.05	.01	.10	.05	.01
Exp(1)	6	.0702	.0162	.0006	.1138	.0290	.0000
	8	.0806	.0226	.0000	.1062	.0296	.0000
	10	.0724	.0210	.0000	.1064	.0296	.0002
N(0, 1)	6	.0672	.0134	.0000	.1130	.0256	.0000
	8	.0756	.0226	.0006	.1070	.0258	.0000
	10	.0716	.0214	.0000	.1052	.0294	.0002
U(1,3)	6	.0754	.0210	.0008	.1072	.0278	.0002
	8	.0770	.0220	.0002	.1172	.0280	.0000
	10	.0888	.0240	.0008	.1070	.0276	.0000
Gamma(4,$\frac{1}{2}$)	6	.0788	.0222	.0008	.1084	.0274	.0000
	8	.0732	.0198	.0006	.1102	.0268	.0000
	10	.0660	.0204	.0002	.1054	.0254	.0000

Results are from S=5,000 simulations with nper=20 permutations, and b=2.

Table 3: Type I error for main effect in fixed effect model (Akritas and Brunner, 1996)

Note: table converted from percentile to type I error

$b = 2$, $n_{ij} \equiv n = 6$		
Level	a=3	a=20
0.10	.094	.106
0.05	.047	.053
0.01	.010	.009

Note: table converted from percentile to type 1 error. Simulation results for the small sample approximation of the null distribution of $P_N^{(fix)}(A)/\hat{\sigma}_{N(fix)}(A)$ for a discrete rectangular distribution (5,000 simulations).

Table 4 contains the type I error simulation results when an umbrella pattern is specified in factor A with the peak located at $i = 2$ and $i = 11$ for the models with $a = 3$ and $a = 20$, respectively. The results are similar to those in Table 2, but slightly less conservative. Simulation results for a corresponding u-shaped alternative are very similar to those in Table 4.

Table 4: Type I error for main effect: umbrella pattern (fixed effects model)

		a=3			a=20		
Distribution	n	.10	.05	.01	.10	.05	.01
Exp(1)	6	.0804	.0214	.0006	.1050	.0232	.0000
	8	.0820	.0246	.0004	.1098	.0280	.0000
	10	.0864	.0238	.0002	.1086	.0302	.0000
N(0, 1)	6	.0780	.0202	.0006	.1124	.0258	.0002
	8	.0800	.0230	.0002	.1032	.0278	.0000
	10	.0844	.0234	.0002	.0964	.0282	.0000
U(1,3)	6	.0866	.0210	.0008	.1104	.0302	.0002
	8	.0926	.0292	.0000	.1164	.0270	.0000
	10	.0916	.0278	.0008	.1166	.0340	.0004
Gamma($4, \frac{1}{2}$)	6	.0866	.0230	.0006	.1086	.0302	.0000
	8	.0774	.0206	.0004	.0964	.0282	.0000
	10	.0894	.0302	.0006	.1084	.0300	.0002

Results are from S=5,000 simulations with nper=20 permutations, and b=2.

The simulated power for increasing trend and umbrella patterns are provided in Tables 5 and 6, respectively. We use a shift parameter δ_i to denote the amount of shift in the mean of the selected distribution at level i of factor A from the distribution at level 1 ($\delta_1 = 0$). The amount of change between consecutive levels i and $i + 1$ are equal (i.e. $|\delta_{i+1} - \delta_i| = |\delta_{i+2} - \delta_{i+1}|$ for $1 \leq i \leq a$) in all models. For the trend alternative with $a = 3$ levels of factor A, the total shift from the first to last levels is two: $\delta_1 = 0$, $\delta_2 = 1$, and $\delta_3 = 2$. The power is close to one at the 10% and 5% levels for all three sample sizes. For the 1% level, the power does not consistently exceed 80% unless $n \geq 10$. For $a = 20$, the total shift from first to last levels is one: $\delta_1 = 0$, $\delta_2 = \frac{1}{19}$, $\ldots$, and $\delta_{20} = 1$ for Table 5. The results are very similar to those with $a = 3$.

Table 5: Power for main effect: trend pattern (fixed effects model)

		a=3 ($\delta_3 = 2$)			a=20 ($\delta_{20} = 1$)		
Distribution	n	.10	.05	.01	.10	.05	.01
Exp(1)+δ_i	6	.9988	.9810	.6732	1.0000	.9996	.7702
	8	1.0000	.9980	.7424	1.0000	1.0000	.9438
	10	1.0000	.9998	.9166	1.0000	1.0000	.9918
N(δ_i, 1)	6	.9964	.9782	.6266	.9980	.9798	.3166
	8	1.0000	.9984	.7136	1.0000	.9966	.5976
	10	1.0000	.9996	.9158	1.0000	.9996	.8150
U(1, 3)+δ_i	6	1.0000	.9998	.9164	.9996	.9938	.4708
	8	1.0000	1.0000	.9670	1.0000	.9998	.7578
	10	1.0000	1.0000	.9986	1.0000	1.0000	.9258
Gamma(4,$\frac{1}{2}$)+δ_i	6	.9980	.9822	.6496	.9998	.9912	.4410
	8	1.0000	.9960	.7174	.9998	.9988	.7226
	10	1.0000	.9998	.9148	1.0000	1.0000	.9014

Results are from S=5,000 simulations with nper=20 permutations, and b=2. Note $\delta_1 = 0$.

For an umbrella alternative, the maximum shift from the first level of factor A to the peak l is one for $a = 3$ and $a = 20$. For an odd number of levels, such as $a = 3$, with a peak located at the median level, $\delta_1 = \delta_a = 0$. For $a = 20$ and $l = 11$, $\delta_{11} = 1$ and $\delta_{20} = 0.1$. In general, the power is lower for the umbrella alternative when compared with the trend alternative. For the 10% level, the power is consistently above 80%. For $a = 3$ and $n \leq 8$, the power exceeds 80% for the 5% level; however, at the 1% level, sample sizes of at least 20 are required to assure power is greater than 80%. As the number of levels of factor A increases, the power typically increases with the same or smaller overall shift in location across those levels. For $a = 20$, a sample size of 10 consistently results in power of more than 80%.

Table 6: Power for main effect: umbrella pattern (fixed effects model)

			a=3, $l=2$			a=20, $l=11$		
Distribution	δ_l	n	.10	.05	.01	.10	.05	.01
Exp(1)+δ_i	1	6	.9782	.9110	.4350	1.0000	.9990	.7244
		8	.9956	.9736	.5408	1.0000	1.0000	.9266
		10	.9992	.9966	.7842	1.0000	1.0000	.9876
N(δ_i, 1)	1	6	.8566	.6662	.1806	.9966	.9646	.2788
		8	.9350	.8180	.2094	.9996	.9962	.5304
		10	.9704	.9016	.3790	1.0000	.9992	.7572
U(1, 3)+δ_i	1	6	.9492	.8324	.3058	.9998	.9940	.4750
		8	.9832	.9244	.3690	.9998	.9986	.6634
		10	.9942	.9716	.5978	1.0000	1.0000	.8656
Gamma(4,$\frac{1}{2}$)+δ_i	1	6	.9268	.7762	.2526	.9992	.9886	.3876
		8	.9756	.9022	.3066	1.0000	1.0000	.6766
		10	.9914	.9608	.5478	1.0000	1.0000	.8644

Results are from S=5,000 simulations with nper=20 permutations, and b=2. Peak location is denoted as l. $\delta_1 = 0$.

In Montgomery (2013), pages 227-228, a study from a manufacturer of men's shirts is provided in which the product quality of fabric in the manufacturer's dyeing process is measured and compared to a standard for fabric dyed at three cycle times (Factor A) and two operating temperatures (Factor B). Let the levels of factor A ($i = 1, 2, 3$) correspond to cycle times 40, 50, and 60, respectively. Let the levels of factor B ($j = 1, 2$) denote operating temperatures 300^{o}C and 350^{o}C, respectively. The data is presented in Table 7. Figure 9 illustrates the potential for an umbrella pattern across the cycle times. Table 8 provides the logarithmic quantile estimates using the vector of weights $w = (1, 2, 1)'$. We obtain a test statistic of $P_N^{(fix)}(A) = 1.5839$ corresponding to a p-value=0.005. There exists evidence to support the claim of an umbrella pattern across the levels of cycle time with peak corresponding to a cycle time of 50.

Table 7: Fabric quality score data (Montgomery, 2013, page 228)

	Temperature					
Cycle Time	300^{o}C			350^{o}C		
40	23	27	31	24	38	34
	24	28	32	23	36	36
	25	26	29	28	35	39
50	36	34	33	37	34	34
	35	38	34	39	38	36
	36	39	35	35	36	31
60	28	35	26	26	36	28
	24	35	27	29	37	26
	27	34	25	25	34	24

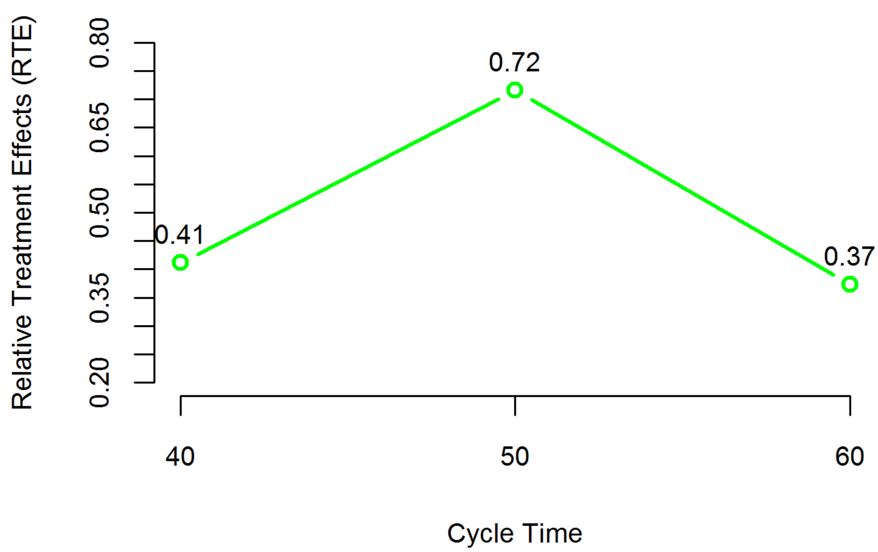

Figure 9: Main effect for fabric quality study

Table 8: Logarithmic quantiles for fabric study cycle times

						LQE Quantiles		
n	N	a	b	*nper*	$P_{bn}^{(ph)}(AB)$	10%	5%	1%
9	54	3	2	20	1.5839	1.2900	1.4150	1.5325

The test statistic value corresponds to a p-value=0.005.

3.3.3 PARTIAL HIERARCHICAL DESIGN

The partial hierarchical model given in Section (3.2.2) allows for dependence within each subject across the repeated measures of factor A. It is customary to use an AR(1) covariance structure to model repeated measures which accounts for the diminishing association as the distance or time between measurements increases. In order to create the AR(1) dependence structure we simulated a multivariate normal distribution with the specified covariance matrix (see Rizzo, 2008 for more details) for each level of factor B. For the exponential and gamma distributions, we used the Gaussian copula to transform the observations from multivariate normal to the multivariate exponential distribution with the same dependence structure using the ideas of Cario and Nelson (1997). The differences in means for the multivariate gamma and exponential distributions were accomplished by adding a shift parameter.

Main effect

Tables 9-12 present the type I error and power results for both increasing trend and umbrella pattern alternatives in factor A under an AR(1) covariance matrix with correlation coefficient values $\rho = (0.3, 0.6, 0.9)$, which represent weak to strong associations within each subject. The type I error results for the test statistic $P_{bn}^{(ph)}(A)$ given in (40) are provided for the trend and umbrella alternatives in Tables 9 and 10, respectively. The simulated type I error is quite conservative for all significance levels. The trend and umbrella patterns result in similar type I error levels. However, the type I error decreases significantly as ρ increases from 0.6 to 0.9. Additionally, as the number of repeated measures increase from $a = 3$ to $a = 10$, the error increases slightly, where the increase is more notable when $\rho = 0.9$. Similar results were obtained when a constant shift exists between the two groups (levels of factor B). Simulations with $n = 10$ subjects per group resulted in very similar values for type I error. The simulated power (Tables 11 and 12) is presented for the trend and umbrella alternatives where the total change in the shift parameter is of magnitude two across the levels of the repeated measurements of factor A. The power at the 10% and 5% levels are above 0.9. Among the many simulations performed, it was observed that if a shift is introduced between the levels of factor B, the power decreases significantly. As expected, an increase in amount of shift δ_i across the repeated measures results in a higher power. An increase in sample size to $n = 10$ results in only a slight increase in power.

Table 9: Type I error for main effect: trend pattern (partial hierarchical model)

		a=3			a=10		
ρ	Distribution	.10	.05	.01	.10	.05	.01
0.3	Exp(1)	.0341	.0105	.0001	.0347	.0082	.0000
	N(0, 1)	.0357	.0090	.0000	.0317	.0079	.0001
	Gamma(4,$\frac{1}{2}$)	.0340	.0109	.0004	.0318	.0076	.0000
0.6	Exp(1)	.0225	.0046	.0001	.0266	.0060	.0000
	N(0, 1)	.0244	.0054	.0003	.0312	.0075	.0000
	Gamma(4,$\frac{1}{2}$)	.0203	.0040	.0001	.0327	.0088	.0000
0.9	Exp(1)	.0030	.0004	.0000	.0132	.0021	.0000
	N(0, 1)	.0024	.0002	.0000	.0105	.0021	.0000
	Gamma(4,$\frac{1}{2}$)	.0036	.0005	.0000	.0106	.0015	.0000

Results are from 10,000 simulations with 20 permutations, b=2, and $n = 6$ with AR(1) covariance structure between levels of factor A.

Table 10: Type I error for main effect: umbrella pattern (partial hierarchical model)

		a=3			a=10		
ρ	Distribution	.10	.05	.01	.10	.05	.01
0.3	Exp(1)	.0343	.0100	.0007	.0292	.0074	.0000
	N(0, 1)	.0352	.0102	.0006	.0339	.0079	.0000
	Gamma(4,$\frac{1}{2}$)	.0369	.0113	.0012	.0342	.0074	.0074
0.6	Exp(1)	.0232	.0055	.0007	.0250	.0049	.0000
	N(0, 1)	.0214	.0054	.0002	.0248	.0061	.0000
	Gamma(4,$\frac{1}{2}$)	.0213	.0045	.0007	.0268	.0064	.0000
0.9	Exp(1)	.0050	.0006	.0006	.0082	.0015	.0000
	N(0, 1)	.0050	.0007	.0001	.0089	.0012	.0000
	Gamma(4,$\frac{1}{2}$)	.0057	.0012	.0001	.0085	.0007	.0000

Results are from 10,000 simulations with 20 permutations, b=2, and $n = 6$ with AR(1) covariance structure between levels of factor A.

Table 11: Power for main effect: trend pattern (partial hierarchical model)

		a=3, $\delta_3 = 2$			a=10, $\delta_{10} = 2$		
ρ	Distribution	.10	.05	.01	.10	.05	.01
0.3	Exp(1)+δ_i	.9935	.9639	.5077	.9997	.9944	.2550
	N(δ_i, 1)	.9931	.9628	.5073	.9976	.9776	.1520
	Gamma(4,$\frac{1}{2}$)+δ_i	.9925	.9615	.5167	.9990	.9868	.1916
0.6	Exp(1)+δ_i	.9956	.9649	.4897	.9971	.9662	.1526
	N(δ_i, 1)	.9978	.9798	.5271	.9860	.9169	.0851
	Gamma(4,$\frac{1}{2}$)+δ_i	.9968	.9743	.5192	.9923	.9362	.1120
0.9	Exp(1)+δ_i	.9979	.9746	.4610	.9866	.9046	.1026
	N(δ_i, 1)	.9999	.9933	.5449	.9874	.9005	.0424
	Gamma(4,$\frac{1}{2}$)+δ_i	.9995	.9863	.5219	.9879	.9007	.0665

Results are from 10,000 simulations with 20 permutations, b=2 treatment groups, and $n = 6$ with AR(1) covariance structure between levels of factor A. Note: $\delta_1 = 0$.

Table 12: Power for main effect: umbrella pattern (partial hierarchical model)

ρ	Distribution	a=3, $l = 2$			a=10, $l = 5$		
		.10	.05	.01	.10	.05	.01
0.3	Exp(1)+δ_i	.9825	.9248	.5122	.9780	.8850	.0787
	N(δ_i, 1)	.8739	.7006	.2433	.8528	.6318	.0208
	Gamma(4,$\frac{1}{2}$)+δ_i	.9462	.8324	.3698	.9105	.7167	.0315
0.6	Exp(1)+δ_i	1.0000	.9995	.8847	1.0000	.9987	.3986
	N(δ_i, 1)	1.0000	.9999	.9079	.9999	.9941	.3150
	Gamma(4,$\frac{1}{2}$)+δ_i	.9869	.9248	.4848	.8802	.6700	.0242
0.9	Exp(1)+δ_i	1.0000	.9989	.7978	1.0000	.9913	.2552
	N(δ_i, 1)	1.0000	.9999	.8785	.9999	.9950	.2259
	Gamma(4,$\frac{1}{2}$)+δ_i	.9913	.9272	.4346	.9370	.7247	.0189

Results are from 10,000 simulations with 20 permutations, b=2, and $n = 6$ with AR(1) covariance structure between levels of factor A. Peak location is denoted by l. For both $a = 3$ and $a = 10$, $\delta_1 = 0$ and $\delta_l = 2$. For $a = 10$, $\delta_{10} = -0.5$.

We now provide analysis of an example, which we will also analyze for interaction in the following section. When the CD4 cell count in a persons blood stream drops below 200, the person is diagnosed with autoimmune deficiency syndrome (AIDS). Table 13 contains data for 22 males subjects with AIDS selected from a study published in Abrams et al. (1994). The subjects within each drug treatment group are indexed by $k = 1, \ldots, 11$. The values in Table 13 are the square root of the CD4 cell counts. The subjects included in the study were either non-responsive or intolerant to the drug AZT. The subjects were randomly assigned to one of two drug treatment groups: ddC or ddl (factor B). The CD4 cell counts measurements were repeated at four times (factor A) for each subject: start of study and every six months for a total of 18 months. A successful treatment would result in a stable level of CD4 counts or even an increase in CD4 counts in the best case. After reviewing the relative treatment effects for each level of factor A (see Figure 10), we analyzed the data with the LQE approach for a decreasing trend across time using a vector of weights $w = (4, 3, 2, 1)'$. The results of the analysis are contained in Table 14. The values of the test statistic for a partial hierarchical repeated measures model is

$P_{bn}^{(ph)}(A) = 2.8516$ and corresponds to a p-value of 0.062.

Table 13: Square root of CD4 cell counts for 22 male subjects with AIDS (see Abrams et al., 1994)

Drug	ddC Observation Time (months)				ddl Observation Time (months)			
k	0	6	12	18	0	6	12	18
1	4.123	2.236	1.414	1.732	6.325	8.124	4.583	5.000
2	2.000	1.414	4.583	4.359	17.176	20.273	17.059	13.601
3	8.062	6.782	2.236	6.083	12.530	7.141	6.856	6.325
4	14.036	11.619	10.488	5.568	9.434	6.557	3.000	2.449
5	6.481	5.477	3.317	3.873	5.657	6.782	5.477	5.831
6	10.954	10.954	11.402	8.944	7.348	4.796	3.742	4.359
7	16.763	12.649	8.426	7.874	4.000	3.162	1.414	2.000
8	3.464	5.831	3.742	3.742	2.828	4.123	3.464	2.828
9	17.321	17.029	16.432	18.439	16.523	8.062	6.164	4.583
10	1.732	2.828	2.449	1.732	7.211	9.165	7.937	4.472
11	3.606	2.449	2.646	2.646	11.747	12.410	10.954	11.225

Analysis was performed on original data using six decimal places.

Table 14: Logarithmic quantiles for main effect in AIDS study (Abrams et al., 1994)

						Level		
n	N	a	b	*nper*	$P_{bn}^{(ph)}(A)$	10%	5%	1%
11	88	4	2	20	2.8516	2.7075	2.9175	3.1075

The test statistic value corresponds to a p-value of 0.062.

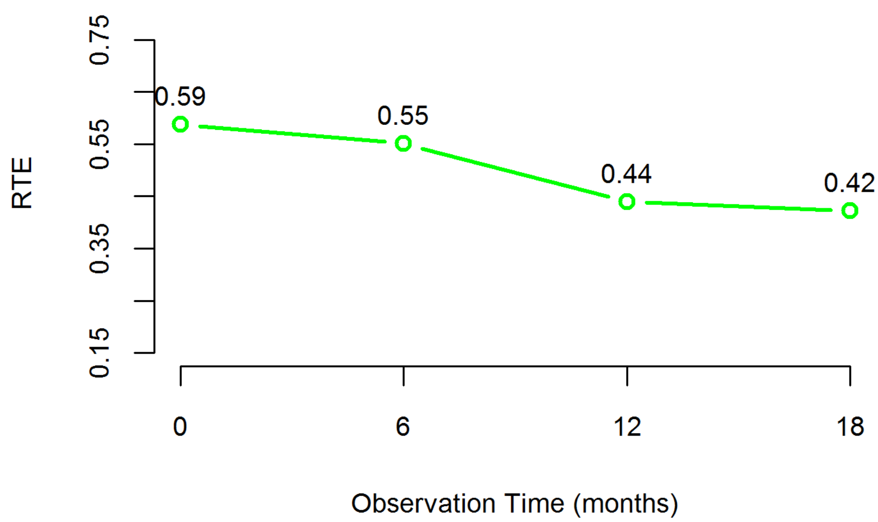

Figure 10: Main effect: AIDS study

Interaction effect

In this section, we explore the simulated significance level and power for the interaction test statistic $P_{bn}^{(ph)}(AB)$ provided in (47) when a pattern in the main effect factor A (repeated measurements) exists. Tables 15 and 16 provide the simulated type I errors when the distributions are the same across the levels of factor B (i.e. group) but experience a trend or umbrella pattern across the levels of factor A. To be more precise, the same pattern exists for both levels of factor B: identical profiles for relative treatment effects. The covariance structure within each treatment group is AR(1), where the value of rho is provided in the table. We denote the shift parameter for the i^{th} time point in the j^{th} group as δ_{ij}, $1 \leq j \leq 2$, where $\delta_{1j} = 0$. The simulated type I error levels are quite conservative. The type I error increases as ρ increases for $a = 3$ levels of factor A under both the trend and umbrella alternatives. As the number of levels of factor A increases to $a = 10$, the simulated type I error levels for $\rho = 0.9$ becomes less conservative. Results for $n = 10$ subjects in each group are very similar to those provided. In additional simulations, we observed that the type

I error decreases when a constant difference in means between the levels of factor B is introduced while maintaining parallel profiles.

Table 15: Type I error for interaction effect: trend pattern (partial hierarchical model)

		a=3				a=10			
ρ	Distribution	δ_{3j}	.10	.05	.01	$\delta_{10,j}$	.10	.05	.01
	Exp(1)+δ_{ij}		.0288	.0073	.0000		.0327	.0081	.0000
0.3	N(δ_{ij}, 1)	2	.0293	.0083	.0001	2	.0314	.0058	.0000
	Gamma(4,$\frac{1}{2}$)+δ_{ij}		.0288	.0073	.0000		.0292	.0068	.0000
	Exp(1)+δ_{ij}		.0385	.0101	.0004		.0301	.0078	.0000
0.6	N(δ_{ij}, 1)	2	.0372	.0099	.0005	2	.0301	.0060	.0000
	Gamma(4,$\frac{1}{2}$)+δ_{ij}		.0173	.0041	.0001		.0287	.0076	.0001
	Exp(1)+δ_{ij}		.0680	.0229	.0014		.0349	.0089	.0000
0.9	N(δ_{ij}, 1)	2	.0623	.0202	.0017	2	.0301	.0073	.0000
	Gamma(4,$\frac{1}{2}$)+δ_{ij}		.0014	.0002	.0000		.0121	.0020	.0000

Results are from 10,000 simulations, 20 permutations, b=2 groups, and $n = 6$.

Table 16: Type I error for interaction effect: umbrella pattern (partial hierarchical model)

		a=3, $l = 2$			a=10, $l = 5$		
ρ	Distribution	.10	.05	.01	.10	.05	.01
	Exp(1)+δ_{ij}	.0271	.0066	.0005	.0308	.0062	.0000
0.3	N(δ_{ij}, 1)	.0278	.0080	.0001	.0304	.0075	.0000
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.0268	.0071	.0001	.0329	.0069	.0000
	Exp(1)+δ_{ij}	.0495	.0160	.0023	.0296	.0069	.0000
0.6	N(δ_{ij}, 1)	.0541	.0207	.0031	.0308	.0069	.0001
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.0161	.0040	.0001	.0259	.0052	.0000
	Exp(1)+δ_{ij}	.0861	.0385	.0095	.0454	.0102	.0000
0.9	N(δ_{ij}, 1)	.0725	.0303	.0099	.0485	.0128	.0001
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.0025	.0002	.0000	.0098	.0013	.0000

Results are from 10,000 simulations, 20 permutations, b=2 groups, and $n = 6$. l is the peak location. Note: $\delta_{1j} = 0$ and $\delta_{lj} = 2$. For $a = 3$, $\delta_{3j} = 0$. For $a = 10$, $\delta_{10,j} = -0.5$.

There are many examples of studies where individuals from the same population are randomly assigned to two treatment groups (e.g. placebo and treatment), and it is therefore reasonable to assume that the distributions of the two groups are identical initially at time point $i = 1$ of factor A. If the treatment is effective we may expect an increasing trend, whereas the subjects in the placebo group are expected to experience a decline in condition. Similarly, the treatment group may experience an umbrella pattern while the placebo group experiences a slightly decreasing trend. As a result of the above reasoning, Tables 17 and 18 contain the simulated power levels when the distributions at the first level of factor A are identical. The second group (placebo, i.e. baseline) follows a slightly decreasing trend while the first group (e.g. treatment) follows the specified pattern (trend and umbrella, respectively). The overall main effect has the specified pattern, albeit slightly different from the treatment group due to the averaging across the groups for each level of factor A. The covariance structure between the levels of factor A within each group is AR(1). The shift parameter for the i^{th} time point in the j^{th} group is denoted as δ_{ij}, $1 \leq j \leq 2$. The results indicate that the power is above 0.8 for $\alpha = 0.10$ under a trend alternative. For $\alpha = 0.05$, the power is marginal under the trend model with $a = 3$ levels in factor A. The corresponding power under an umbrella model is smaller and is marginal for $\alpha = 0.05$. As the number of levels in factor A increases, the power decreases when the change in means across A is maintained. Results for $n = 10$ subjects per group have only slightly higher power than that for $n = 6$. There is a higher level of power for higher values of ρ. Additional simulations showed that the power is significantly diminished when the placebo group has a constant mean across the levels of A (i.e. no change in condition). A potential solution for the situation where an umbrella pattern and interaction coexist is to rearrange the levels of factor A to form a trend pattern and perform the analysis with a higher resulting power as suggested by Akritas and Brunner (1996).

Table 17: Power for interaction effect: trend pattern (partial hierarchical model)

		a=3			a=10		
ρ	Distribution	.10	.05	.01	.10	.05	.01
0.3	Exp(1)+δ_{ij}	.9183	.7481	.1144	.9765	.8653	.0300
	N(δ_i, 1)	.8732	.6460	.0535	.9465	.7744	.0137
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.8842	.6677	.0610	.9564	.7940	.0159
0.6	Exp(1)+δ_{ij}	.9371	.7737	.1285	.9226	.7256	.0149
	N(δ_i, 1)	.9023	.6720	.0535	.8593	.5961	.0051
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.9141	.7012	.0655	.8774	.6222	.0076
0.9	Exp(1)+δ_{ij}	.9752	.8535	.1836	.8845	.6294	.0117
	N(δ_i, 1)	.9524	.7531	.0688	.8042	.4793	.0025
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.9595	.7846	.1032	.8203	.5142	.0032

Results are from 10,000 simulations, 20 permutations, b=2 groups, and $n = 6$. Note: $\delta_{1j} = 0$, $\delta_{a1} = 2$, and $\delta_{a2} = -1$.

Table 18: Power for interaction effect: umbrella pattern (partial hierarchical model)

		a=3, $l = 2$			a=10, $l = 5$		
ρ	Distribution	.10	.05	.01	.10	.05	.01
0.3	$Exp(1) + \delta_{ij}$	.7685	.5065	.0768	.7593	.4488	.0022
	$N(\delta_{ij}, 1)$	.7836	.5422	.0980	.7615	.4662	.0043
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.7728	.5182	.0878	.7532	.4510	.0031
0.6	$Exp(1) + \delta_{ij}$	.8739	.6500	.1221	.7070	.3866	.0019
	$N(\delta_{ij}, 1)$	.8885	.6870	.1596	.7156	.4208	.0032
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.8853	.6675	.1408	.7048	.3985	.0027
0.9	$Exp(1) + \delta_{ij}$	.9540	.8104	.2720	.7622	.4461	.0026
	$N(\delta_{ij}, 1)$	.9620	.8227	.2705	.8062	.5158	.0054
	Gamma(4,$\frac{1}{2}$)+δ_{ij}	.9560	.8075	.2744	.7793	.4700	.0039

Results are from 10,000 simulations, 20 permutations, b=2 groups, and $n = 6$. Peak locations are denoted as l. Note: $\delta_{1j} = 0$, $\delta_{l1} = 2$, and $\delta_{a2} = -1$. For $a = 3$, $\delta_{31} = 0$. For $a = 10$ $\delta_{10,1} = -0.5$.

We now revisit the AIDS data introduced in 3.3.3 (Table 13) and test for interaction between drug and time. The relative treatment effects for each drug plotted across the observation times is provided in Figure 11. The general pattern is a downward trend for both drugs; however, the profiles are not parallel suggesting the potential presence of interaction. Table 19 contains the logarithmic quantile estimates. The test statistic $P_{bn}^{(ph)}(AB) = 1.8171$ corresponds with a p-value < 0.001, providing evidence to support the claim of the presence of interaction between drug and time where a decreasing trend exists across observation time.

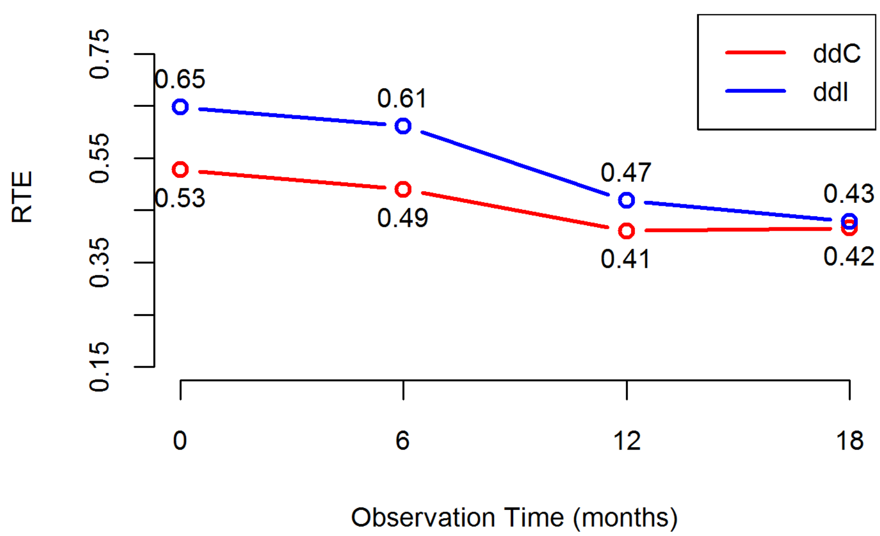

Figure 11: Interaction effect for AIDS study

Table 19: Logarithmic quantiles for interaction effect in AIDS study (Abrams et al., 1994)

						Level		
n	N	a	b	*nper*	$P_{bn}^{(ph)}(AB)$	10%	5%	1%
11	88	4	2	20	1.8171	1.500	1.5200	1.5350

The test statistic value corresponds to a p-value < 0.001.

3.3.4 CROSS-CLASSIFICATION REPEATED MEASURES DESIGN

We begin with the analysis of an example of a cross-classified repeated measures design. The study involved 14 volunteers. Saliva α-amylase levels of each volunteer were measured on two fixed days at four fixed time points throughout the day. The α-Amylase study data (Brunner, Domhof, and Langer, 2002) is provided in Table 20 below. Researchers from a previous study postulated that the α-amylase levels are lower in the morning, increase until late afternoon, and then decrease. α-amylase levels were also surmised to change over the course of the week. Quantile-quantile plots of the data reveal severe departures from normality and support the choice of nonparametric methods for analysis.

Table 20: α-amylase activity data (see Brunner, Domhof, and Langer, 2002, p. 132)

	Monday				Thursday			
	Time of Day (hours)				Time of Day (hours)			
Subject	8	12	17	21	8	12	17	21
1	146.8	167.0	107.2	161.8	90.8	151.6	123.0	142.8
2	818.2	1314.2	1578.8	932.5	378.8	759.5	1881.2	572.6
3	394.4	1157.4	585.2	629.2	171.0	538.4	729.8	412.1
4	100.2	140.4	234.4	244.8	121.6	154.6	221.8	170.6
5	169.8	99.9	184.2	168.8	103.0	170.0	342.0	162.2
6	107.2	262.8	198.4	465.1	178.8	312.6	261.6	450.5
7	272.0	551.2	265.2	453.2	133.4	560.4	977.9	402.0
8	51.8	144.4	125.4	203.8	122.2	71.4	434.9	191.2
9	273.6	351.6	510.0	354.0	403.0	665.4	420.4	566.0
10	367.2	435.6	783.3	523.1	221.8	601.2	1028.5	713.4
11	519.2	264.6	321.4	1433.8	137.2	345.6	884.9	331.8
12	88.6	135.0	88.6	86.2	164.2	190.4	301.0	173.2
13	218.0	109.2	167.6	179.4	162.8	185.6	193.6	183.2
14	117.2	151.0	150.0	218.0	178.2	151.0	165.2	170.0

The values are in U/ml representing α-amylase units per milliliter. An α-amylase unit is defined as the amount that will liberate 1.0 mg of maltose from starch in 3 minutes at pH=6.9 at 20°.

In our design we designate factor A as the time of day and factor B as the day of the week. Let the levels of factor A be denoted as $i = 1, 2, 3, 4$ corresponding to

8 a.m., 12 p.m., 5 p.m., and 9 p.m., respectively. The levels of factor B are Monday ($j = 1$) and Thursday ($j = 2$). The vector of weights applied to the levels of factor A are $w = (1, 2, 4, 3)'$ for both days matching the test in Akritas and Brunner (1996). The plot of relative treatment effects by time of day is shown in Figure 12, and suggests an umbrella pattern across the time of day. We tested the null hypothesis in (34) against the umbrella alternative in (36). We obtained $P_n^{(cc)}(A) = 3.996$ for the test statistic in (50) and an estimated logarithmic $\alpha-$quantile of 3.89 and $p-value < 0.022$, which supports the umbrella patterned alternative. The $p-value$ obtained by Akritas and Brunner (1996) is 0.000044 and agrees with our results.

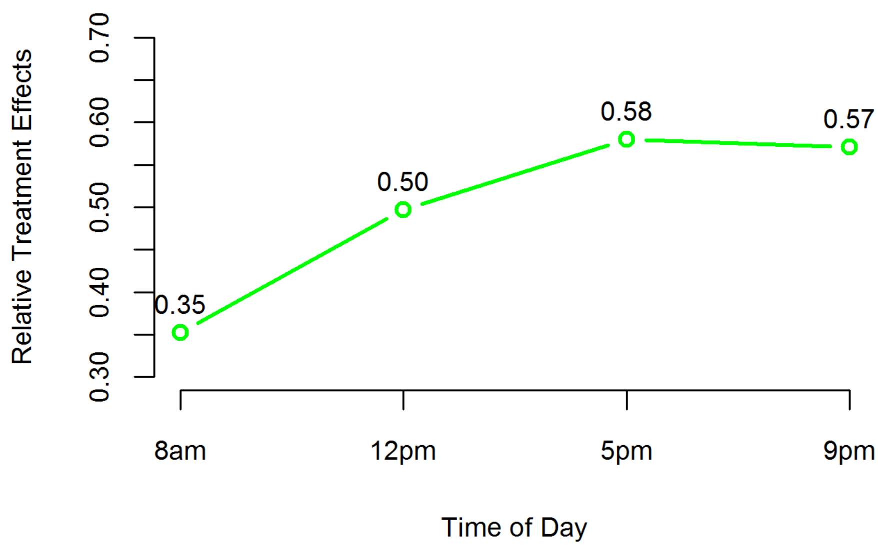

Figure 12: Main effect for α-amylase data

Results from 5000 simulations with 50 permutations are provided for type I error and power in the Tables 21-24 below. The data was simulated with an AR(1) covariance structure within each level of factor B, where ρ denotes the correlation coefficient. Since the measurements between levels of factor B belong to the same subject, a constant correlation $\phi = 0.2$, $\phi = 0.5$, and $\phi = 0.8$ was assigned between the levels of factor B for $\rho = 0.3$, $\rho = 0.6$, and $\rho = 0.9$, respectively. The results for $n = 5$ and $n = 10$ subjects are provided. In Tables 21 and 22, we observe that the type I error becomes more conservative as ρ increases. In general, there is very little difference in type I error for the different levels of factor A and for both the trend and umbrella alternatives. However, as the number of levels for factor A increases from $a = 3$ to $a = 5$ for small sample sizes, the type I error decreases slightly. The umbrella alternative results are slightly less conservative than the trend alternative results.

Table 21: Type I error for main effect: trend pattern (cross-classification repeated measures model)

			a=3		a=4		a=5	
ρ	Distribution	n	.10	.05	.10	.05	.10	.05
0.3	Exp(1)	5	.0556	.0082	.0536	.0072	.0526	.0064
		10	.0434	.0184	.0472	.0154	.0430	.0134
	N(0, 1)	5	.0574	.0118	.0526	.0058	.0554	.0036
		10	.0460	.0164	.0444	.0158	.0484	.0146
	Gamma(4,$\frac{1}{2}$)	5	.0556	.0082	.0536	.0072	.0526	.0064
		10	.0434	.0184	.0472	.0154	.0430	.0134
0.6	Exp(1)	5	.0308	.0034	.0324	.0026	.0318	.0016
		10	.0220	.0056	.0242	.0060	.0230	.0048
	N(0, 1)	5	.0282	.0030	.0330	.0034	.0312	.0010
		10	.0226	.0074	.0246	.0064	.0240	.0090
	Gamma(4,$\frac{1}{2}$)	5	.0308	.0034	.0344	.0026	.0318	.0016
		10	.0220	.0056	.0242	.0060	.0230	.0048
0.9	Exp(1)	5	.0084	.0006	.0058	.0006	.0072	.0004
		10	.0024	.0000	.0040	.0004	.0024	.0002
	N(0, 1)	5	.0060	.0006	.0056	.0002	.0028	.0002
		10	.0014	.0002	.0024	.0000	.0028	.0000
	Gamma(4,$\frac{1}{2}$)	5	.0084	.0006	.0058	.0006	.0072	.0004
		10	.0024	.0000	.0040	.0004	.0024	.0002

Results are from 5,000 simulations with 50 permutations, b=2 levels of factor B having an AR(1) covariance structure between levels of factor A within each group with correlation values 0.3, 0.6, and 0.9 (ρ), and with constant correlation values between each group of $\phi = 0.2$, $\phi = 0.5$, and $\phi = 0.8$, respectively.

Table 22: Type I error for main effect: umbrella pattern (cross-classification repeated measures model)

			a=3		a=4		a=5	
ρ	Distribution	n	.10	.05	.10	.05	.10	.05
0.3	Exp(1)	5	.0630	.0160	.0550	.0084	.0598	.0076
		10	.0478	.0196	.0424	.0134	.0420	.0148
	N(0, 1)	5	.0658	.0172	.0566	.0074	.0658	.0088
		10	.0474	.0180	.0470	.0162	.0472	.0152
	Gamma(4,$\frac{1}{2}$)	5	.0630	.0160	.0550	.0084	.0598	.0076
		10	.0478	.0196	.0424	.0134	.0420	.0148
0.6	Exp(1)	5	.0386	.0106	.0338	.0106	.0386	.0040
		10	.0242	.0070	.0232	.0042	.0206	.0054
	N(0, 1)	5	.0372	.0084	.0334	.0056	.0306	.0040
		10	.0226	.0066	.0258	.0066	.0242	.0060
	Gamma(4,$\frac{1}{2}$)	5	.0386	.0106	.0338	.0084	.0386	.0040
		10	.0242	.0070	.0232	.0042	.0206	.0054
0.9	Exp(1)	5	.0144	.0048	.0082	.0010	.0100	.0014
		10	.0030	.0010	.0026	.0006	.0018	.0000
	N(0, 1)	5	.0130	.0048	.0086	.0006	.0076	.0006
		10	.0024	.0006	.0038	.0008	.0034	.0008
	Gamma(4,$\frac{1}{2}$)	5	.0144	.0048	.0082	.0010	.0100	.0014
		10	.0030	.0010	.0026	.0006	.0018	.0000

Results are from 5,000 simulations with 50 permutations, b=2 levels of factor B having an AR(1) covariance structure between levels of factor A within each group with correlation values 0.3, 0.6, and 0.9 (ρ), and with constant correlation values between each group of $\phi = 0.2$, $\phi = 0.5$, and $\phi = 0.8$, respectively.

Table 23 contains the power under a trend alternative with a total shift of two across the levels of factor A. For the 10% level, the power is well above 0.8 for $n = 5$. A sample size of at least $n = 8$ is necessary to assure the power is above 80%. As the number of levels of factor A increases the power generally decreases slightly. Table 24 contains the simulated power under an umbrella alternative with a shift of one from the first level of factor A to the peak l. A sample size of $n = 10$ is necessary to assure the power is above 0.8 for the 5% level.

Table 23: Power for main effect: trend pattern (cross-classification repeated measures model)

ρ	Distribution	n	a=3 .10	a=3 .05	a=4 .10	a=4 .05	a=5 .10	a=5 .05
0.3	Exp(1)+δ_i	5	.9806	.6894	.9782	.6264	.9844	.6132
		10	1.0000	.9980	.9998	.9986	1.0000	.9996
	N(δ_i, 1)	5	.9848	.7388	.9792	.6376	.9800	.5922
		10	1.0000	.9996	1.0000	.9996	1.0000	.9994
	G(4,$\frac{1}{2}$)+δ_i	5	.9826	.7176	.9784	.6492	.9838	.6070
		10	1.0000	.9994	1.0000	.9998	1.0000	.9998
0.6	Exp(1)+δ_i	5	.9260	.6326	.9438	.5352	.9404	.4944
		10	1.0000	.9970	.9994	.9950	1.0000	.9962
	N(δ_i, 1)	5	.9774	.6706	.9634	.5420	.9558	.4710
		10	1.0000	1.0000	1.0000	.9988	1.0000	.9978
	G(4,$\frac{1}{2}$)+δ_i	5	.9692	.6568	.9586	.5558	.9438	.4994
		10	1.0000	.9982	.9996	.9962	1.0000	.9968
0.9	Exp(1)+δ_i	5	.9144	.5788	.8760	.4700	.8638	.4100
		10	1.0000	.9942	.9998	.9842	.9994	.9766
	N(δ_i, 1)	5	.9450	.5356	.8878	.3734	.8656	.3280
		10	1.0000	.9994	1.0000	.9972	1.0000	.9920
	G(4,$\frac{1}{2}$)+δ_i	5	.9266	.5556	.8932	.4204	.8618	.3696
		10	1.0000	.9978	.9996	.9914	.9998	.9946

Note: G(4,$\frac{1}{2}$) represents the Gamma(4,$\frac{1}{2}$) distribution.
Results are from 5,000 simulations with 50 permutations, b=2 levels of factor B having an AR(1) covariance structure between levels of factor A within each group with correlation values 0.3, 0.6, and 0.9 (ρ), and with constant correlation values between each group of $\phi = 0.2$, $\phi = 0.5$, and $\phi = 0.8$, respectively. Note: $\delta_1 = 0$ and $\delta_a = 2$.

Table 24: Power for main effect: umbrella pattern (cross-classification repeated measures model)

			a=3, $l = 2$		a=4, $l = 2$		a=5, $l = 3$	
ρ	Distribution	n	.10	.05	.10	.05	.10	.05
0.3	Exp(1)+δ_i	5	.9506	.6692	.9838	.6788	.9072	.3994
		10	1.0000	.9956	1.0000	.9998	.9978	.9806
	N(δ_i, 1)	5	.8384	.4824	.9858	.6882	.7482	.2372
		10	.9904	.9620	1.0000	.9990	1.0000	.9976
	G(4,$\frac{1}{2}$)+δ_i	5	1.0000	.9518	1.0000	.9796	.9976	.8000
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.6	Exp(1)+δ_i	5	.9256	.5812	.9558	.5986	.8348	.3104
		10	.9996	.9952	.9998	.9978	.9962	.9640
	N(δ_i, 1)	5	.8312	.4652	.9706	.6206	.6998	.1996
		10	.9982	.9850	1.0000	.9996	.9714	.8900
	G(4,$\frac{1}{2}$)+δ_i	5	.9976	.9140	.9998	.9490	.9904	.7018
		10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.9	Exp(1)+δ_i	5	.8026	.4264	.8962	.4880	.6354	.2016
		10	.9984	.9678	1.0000	.9902	.9804	.8536
	N(δ_i, 1)	5	.6800	.2820	.9012	.4024	.4590	.1250
		10	.9936	.9564	1.0000	.9976	.9512	.8004
	G(4,$\frac{1}{2}$)+δ_i	5	.9860	.8060	.9980	.8824	.9262	.4784
		10	1.0000	1.0000	1.0000	1.0000	1.0000	.9978

Note: G(4,$\frac{1}{2}$) represents the Gamma(4,$\frac{1}{2}$) distribution.
Results are from 5,000 simulations with 50 permutations, b=2 levels of factor B having an AR(1) covariance structure between levels of factor A within each group with correlation values 0.3, 0.6, and 0.9 (ρ), and with constant correlation values between each group of $\phi = 0.2$, $\phi = 0.5$, and $\phi = 0.8$, respectively. $\delta_a = 0$ for $a = 3, 5$ and $\delta_a = -1$ for $a = 4$. $\delta_1 = 0$ and $\delta_l = 1$ for $a = 3, 4, 5$.

3.4 SUMMARY

In this chapter, the LQE approach has been considered for testing three different complex fully nonparametric factorial designs for patterned alternatives. Unlike the proposed methods of Akritas and Brunner (1996), the LQE approach presented does not require the calculation of the asymptotic variance estimate or the Satterthwaite degrees of freedom for small samples. The type I error simulation results are generally conservative to very conservative depending upon the design. The only exception is for the two way fixed effects factorial model at the $\alpha = 0.10$ significance level for $a = 20$ levels of factor A, which has slightly liberal results. The simulated power is acceptable for small sample sizes under reasonable alternatives in most cases. The analysis of a clinical data set with $n = 14$ subjects supports an umbrella alternative hypothesis and agrees with the findings of Akritas and Brunner (1996). LQE for testing patterned alternatives is a viable approach for complex fixed effects and repeated measures factorial designs.

3.5 PROOFS OF PROPOSITIONS

We sketch the proofs of Propositions 1 through 4 following the ideas from Denker and Tabacu (2014, 2015). For convenience, we present here the formulas from Denker and Tabacu (2014) that we use in our derivations. Denker and Tabacu (2014) showed that a simple linear rank statistic based on n independent random vectors of possibly unequal lengths satisfies the almost sure central limit theorem. In our case we use n independent random vectors of equal length $X_1, \ldots, X_n$, having components $X_u = (X_{u1}, \ldots, X_{um})$ $(u = 1, \ldots, n)$ and $N(n) = nm$ total number of observations. We now define the simple linear rank statistic

$$T_n = \frac{1}{N(n)} \sum_{u=1}^{n} \sum_{v=1}^{m} \lambda_{uv}^{(n)} \frac{R_{uv}(n)}{N(n)+1} - \int_{-\infty}^{\infty} H dF_n, \tag{55}$$

which has the following term that appears in its Taylor decomposition

$$B_n = \int_{-\infty}^{\infty} H d(\hat{F}_n - F_n) + \int_{-\infty}^{\infty} (\hat{H}_n - H) dF_n, \text{ and} \tag{56}$$

$$\sigma_n^2 = N^2(n) Var(B_n). \tag{57}$$

$R_{uv}(n)$ denotes the rank of observation $X_{uv} \backsim F_{uv}$ among all $N(n)$ random variables, and $\lambda_{uv}^{(n)}$ are known regression constants satisfying the constraint

$$\max_{1 \le u \le n, 1 \le v \le m} |\lambda_{uv}^{(n)}| = 1.$$

We define the distribution function weighted by the regression constants for the n independent random vectors

$$F_n(x) = \frac{1}{mn} \sum_{u=1}^{n} \sum_{v=1}^{m} \lambda_{uv}^{(n)} F_{uv}(x), \tag{58}$$

and its empirical form

$$\hat{F}_n(x) = \frac{1}{mn} \sum_{u=1}^{n} \sum_{v=1}^{m} \lambda_{uv}^{(n)} \mathbb{I}(X_{uv} \leq x). \tag{59}$$

In the proofs equations (55) through (59) are specified for each model and used to prove the almost sure central limit theorems.

Proof of Proposition 1:

In the two way fixed-effects model introduced in Section 3.2.1, $X_{ijk} \backsim F_{ij}$ are independent random variables, and $N = abn$ is the total number of subjects. We define independent random vectors $X_{ik} = (X_{i1k}, \ldots, X_{ibk})'$, $1 \leq i \leq a$, $1 \leq k \leq n$. Let $1 \leq l \leq a$ denote the l^{th} level of factor A and define

$$\lambda_{ij}^{(n)} = \begin{cases} 1, & i = l,\ j = 1, \ldots, b \\ 0, & otherwise. \end{cases} \tag{60}$$

Equations (58) and (59) become

$$F_n^{(l)}(x) = \frac{1}{ab} \sum_{j=1}^{b} F_{lj}(x), \text{ and}$$

$$\hat{F}_n^{(l)}(x) = \frac{1}{abn} \sum_{j=1}^{b} \sum_{k=1}^{n} \mathbb{I}(X_{ljk} \leq x),$$

which are the average distribution function across the levels of factor B (groups) for a fixed time ($i = l$) of factor A, and its empirical analog, respectively. The linear rank statistic is

$$T_n^{(l)} = \frac{1}{abn(abn+1)} \sum_{j=1}^{b} \sum_{k=1}^{n} R_{ljk} - \frac{1}{ab} \sum_{j=1}^{b} \int_{-\infty}^{\infty} H(x) dF_{lj}(x),$$

where $H(x)$ is defined in (23), and

$$B_n^{(l)} = \frac{1}{N}\sum_{k=1}^{n}\sum_{j=1}^{b} H(X_{ljk}) - \frac{2}{N}\sum_{k=1}^{n}\sum_{j=1}^{b}\int_{-\infty}^{\infty} H(x)dF_{lj}(x) + \frac{1}{N}\sum_{k=1}^{n}\sum_{s=1}^{a}\sum_{j=1}^{b}\int_{-\infty}^{\infty}\mathbb{I}(X_{sjk}\le x)d\left(\frac{1}{ab}\sum_{t=1}^{b}F_{lt}(x)\right).$$

We note that under $H_0^F(A)$ in (34) the test statistic $P_N^{(fix)}(A)$ in (37) can be expressed as

$$P_N^{(fix)}(A) = \frac{a(N+1)}{\sqrt{N}}\sum_{i=1}^{a}(w_i - \bar{w})T_n^{(i)}. \tag{61}$$

The almost sure weak convergence of the vector

$$\tilde{T}_n = \left(\frac{a(N+1)}{\sqrt{N}}(w_i - \bar{w})T_n^{(i)}\right)_{1\le i\le a} \tag{62}$$

can be obtained by showing the almost sure weak convergence of the vector

$$\tilde{B}_n = \left(\frac{a(N+1)}{\sqrt{N}}(w_i - \bar{w})B_n^{(i)}\right)_{1\le i\le a}. \tag{63}$$

As in Denker and Tabacu (2014), we can express $\tilde{B}_n$ as a sum of a dimensional independent random vectors Z_k, which are bounded with $EZ_k = 0$

$$\tilde{B}_n = \sqrt{\frac{a}{b}}\left(\frac{abn+1}{abn}\right)\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_k, \tag{64}$$

where Z_k has components

$$Z_{ki} = (w_i - \bar{w})\sum_{j=1}^{b}\left[H(X_{ijk}) - 2\int H dF_{ij} + \frac{1}{ab}\sum_{s=1}^{a}\int \mathbb{I}\left(X_{sjk}\le x\right)d\left(\sum_{t=1}^{b}F_{it}\right)\right].$$

By the multivariate central limit theorem (MVCLT) and under Assumption 1, the vector

$$\tilde{B}_n \xrightarrow{d} \mathcal{N}\left(0, \frac{a}{b}\Sigma\right), \qquad \text{as } n\to\infty. \tag{65}$$

It can be verified using Theorem 3.1 of Lifshits (2001) that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\frac{1}{\sqrt{k}} \sum_{l=1}^{k} Z_l \leq x \right) = G_X(x), \;\; a.s., \;\; \forall x \in \mathbb{R}, \tag{66}$$

where G_X is the distribution function of $X \backsim \mathcal{N}(0, \Sigma)$. Lemma 2.2 of Fridline (2010) implies

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(\tilde{B}_k \leq x \right) = G_X\left(x\sqrt{\frac{a}{b}} \right), \;\; a.s., \;\; \forall x \in \mathbb{R}. \tag{67}$$

For the continuous function $f : \mathbb{R}^a \longrightarrow \mathbb{R}$, $f(x_1, \ldots, x_a) = \sum_{i=1}^{a} x_i$, and by Lifshits (2001) and the techniques in Tabacu (2014) equation (67) is equivalent to

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(f(\tilde{T}_k) \leq x \right) = G_{f(X)}\left(x\sqrt{\frac{a}{b}} \right), \;\; a.s., \;\; \forall x \in \mathbb{R}. \tag{68}$$

Hence, the almost sure weak convergence of the test statistic (61) follows. $\square$

Proof of Proposition 2:

Recall that each of the bn subjects in the partial hierarchical model of Section 3.2.2 can be expressed as independent random vectors $X_{jk} = (X_{1jk}, \ldots, X_{ajk})'$, $1 \leq j \leq 2, 1 \leq k \leq n$, where $X_{ijk} \backsim F_{ij}$. For fixed l,v such that $1 \leq l \leq a$, $1 \leq v \leq b$ let

$$\lambda_{ij}^{(n)} = \begin{cases} 1, & i = l, \; j = v \\ 0, & otherwise. \end{cases} \tag{69}$$

Then the linear rank statistic has the form

$$T_n^{(l,v)} = \frac{1}{N} \sum_{k=1}^{n} \frac{R_{lvk}}{N+1} - \frac{n}{N} \int_{-\infty}^{\infty} H(x) dF_{lv}(x), \tag{70}$$

and

$$B_n^{(l,v)} = \frac{1}{N} \sum_{k=1}^{n} \left[H(X_{lvk}) - 2\int H dF_{lv} + \frac{1}{N} \sum_{s=1}^{a} \sum_{t=1}^{b} \sum_{u=1}^{n} \int \mathbb{I}\left(X_{stu} \leq x \right) dF_{lv} \right]. \tag{71}$$

The test statistic in (40) can be written

$$P_{bn}^{(ph)}(A) = \sqrt{\frac{a}{b}}\ \frac{(abn+1)}{\sqrt{n}} \sum_{i=1}^{a} \sum_{j=1}^{b} (w_i - \bar{w}) T_n^{(i,j)}. \tag{72}$$

Consider the vectors

$$\tilde{T}_n = \left(\frac{\sqrt{a}(N+1)(w_i - \bar{w})}{\sqrt{bn}} T_n^{(i,j)} \right)_{1\le i\le a, 1\le j\le b}, \tag{73}$$

and

$$\tilde{B}_n = \frac{\sqrt{a}(abn+1)}{\sqrt{bn}} \left(w_i - \bar{w}) B_n^{(i,j)} \right)_{1\le i\le a, 1\le j\le b}. \tag{74}$$

$\tilde{B}_n$ can be expressed as a sum of a dimensional independent random vectors

$$\tilde{B}_n = \sqrt{\frac{a}{b}}\ \frac{(abn+1)}{abn} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k, \tag{75}$$

where components of the vector Z_k are

$$Z_{ki} = \left((w_i - \bar{w}) \left[H(X_{ijk}) + \frac{1}{N} \sum_{s=1}^{a} \sum_{t=1}^{b} \sum_{u=1}^{n} \int \mathbb{I}(X_{stu} \le x) dF_{ij} \right] \right)_{1\le j\le b}, \tag{76}$$

for $1 \le i \le a$. By the multivariate central limit theorem (MVCLT) and under Assumption 1, the vector

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k \to \mathcal{N}(0, \Sigma), \qquad\qquad \text{as } n \to \infty, \tag{77}$$

and hence,

$$\tilde{B}_n \to \mathcal{N}(0, \frac{a}{b}\Sigma), \qquad\qquad \text{as } n \to \infty. \tag{78}$$

The proof follows as in Proposition 1 by taking the function $f : \mathbb{R}^{ab} \to \mathbb{R}$, $f(x_1, \ldots, x_{ab}) = \sum_{i=1}^{ab} x_i$. □

Proof of Proposition 3: Partial Hierarchical Model (Interaction).

Due to the similarity of the proof to that of Proposition 2, we provide only those formulas that differ from those in the preceding proof. The test statistic in (47) when expressed in terms of (70) using i and $j = 1, 2$ in place of l and v, respectively is

$$P_{nb}^{(ph)}(AB) = \frac{\sqrt{ab(N+1)}}{\sqrt{n}} \sum_{i=1}^{a} (w_i - \bar{w}) \left[T_n^{(i,1)} - T_n^{(i,2)} \right]. \tag{79}$$

Define the vectors

$$\tilde{T}_n^{(AB)} = \left(\frac{(-1)^{j-1}\sqrt{2a}(N+1)}{\sqrt{n}} (w_i - \bar{w}) T_n^{(i,j)} \right)_{1 \le i \le a, 1 \le j \le 2}, \tag{80}$$

$$\tilde{B}_n^{(AB)} = \left(\frac{(-1)^{j-1}\sqrt{2a}(N+1)}{\sqrt{n}} (w_i - \bar{w}) B_n^{(i,j)} \right)_{1 \le i \le a, 1 \le j \le 2}. \tag{81}$$

We can express (81) as a sum of independent random vectors $Z_k^{(AB)}$

$$\tilde{B}_n = \sqrt{ab}\, \frac{(abn+1)}{abn} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k^{(AB)}. \tag{82}$$

The vector elements of the vector $Z_k^{(AB)}$ are

$$Z_{ki}^{(AB)} = \frac{(w_i - \bar{w})}{(-1)^{j-1}} \left[H(X_{ijk}) + \frac{1}{N} \sum_{s=1}^{a} \sum_{t=1}^{2} \sum_{u=1}^{n} \int \mathbb{I}(X_{stu} \le x) dF_{ij} \right], \tag{83}$$

for $1 \le j \le 2$. Following the ideas in the proof of Proposition 1 above, taking a function $f : \mathbb{R}^{ab} \to \mathbb{R}$, the result follows. $\square$

Due to the similarity of the proof of Proposition 4 to that of Proposition 2, it has been omitted.

CHAPTER 4

CHANGE POINT ANALYSIS

4.1 INTRODUCTION

One of the basic questions for many applications is whether a process has changed. In particular, it is desired to know if a process has changed across either time or space. Without loss of generality, we limit our discussion to the concept of time. The general category of analysis for these questions is referred to as change-point problems. The change-point problem originated from the need to control the quality of manufactured goods (Ferger, 1994). Our motivation for applying the logarithmic quantile estimation method to change-point problems involves the ability to estimate the quantiles of test statistics with intractable asymptotic distributions in a relatively simple manner. In this chapter we investigate the logarithmic quantile estimation approach for a change-point problem. More precisely, we focus on estimating quantiles for Pettitt's rank test (Pettitt, 1979). In Pettitt (1979) the rank statistic for testing for a change in distribution converges weakly to the supremum of the absolute value of the Brownian bridge. In this chapter we propose the almost sure limit theorem for Pettitt's rank test, provide simulation results for the significance level and power, and compute approximations of quantiles of this test for several data sets.

There is an extensive literature on parametric and nonparametric statistics for hypothesis testing and estimation in change-point analysis. Initial literature focused upon parametric or semi-parametric assumptions for continuous distributions, such as Page (1954, 1955, 1957). Hinkley (1970) introduced maximum likelihood estimation of the change point. Later McGilchrist and Woodyer (1975) explored using nonparametric cumulative sums to detect a single change in the median. Sen and Srivastava (1975) introduced the supremum of normalized partial sums of indicator functions of the observations relative to the sample median as a nonparametric test statistic whose Monte Carlo power evaluations outperformed Bayesian methods of the time. Pettitt (1979) modified the famous two-sample test of Mann and Whitney (1947) to develop a nonparametric test based upon the maxima of partial sums. Nearly a decade later methods for abrupt, smooth single and multiple changes in both scale and location were developed by Lombard (1987). From a hypothesis testing perspective, permutation and bootstrap methods were developed to obtain approximations of the critical values for different test statistics. Hušková (2004) gives a survey of permutation tests and resampling techniques for detecting one or multiple changes in location models with i.i.d. data. Hušková and Slabý (2001), Antoch and Hušková (2001) provide approximations to the critical values for kernel generated tests for multiple changes in location or scale models. Permutation methods applied to cumulative sum and moving sum were studied by Berkes et al. (2004) and permutation methods for U-statistics type tests were investigated by Horváth and Hušková (2005). To detect changes in monthly precipitation, Gombay and Horváth (1999) proposed tests based on empirical distributions and Kendall's tau (Kendall, 1938) for random vectors and approximated their distributions using weighted bootstrap. Permutation methods for abrupt and gradual changes for dependent data were studied by Kirch and Steinebach (2006). To the existing methods we add the logarithmic quantile estimation as another way of approximating quantiles of change-point test statistics.

The chapter is organized as follows. In Section 4.2 we introduce the rank test from Pettitt (1979) and propose its almost sure limit theorem. Section 4.3 contains the numerical results for a several data sets along with simulations for significance level and power.

4.2 ALMOST SURE LIMIT THEOREM FOR PETTITT'S RANK TEST

Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables. Testing for a change-point is equivalent to testing the null hypothesis

$$H_0 : X_1, ..., X_n \text{ are identically distributed}$$

against the alternative

$$H_1 : X_1, ..., X_\tau \sim F(x) \text{ and } X_{\tau+1}, ..., X_n \sim G(x).$$

If the change point τ is unknown, Pettitt (1979) introduced the test statistic

$$K_n = \max_{1 \le j < n} |U_{j,n}|, \tag{84}$$

where

$$U_{j,n} = \sum_{i=1}^{j} \sum_{l=j+1}^{n} sgn(X_i - X_l)$$

and $sgn(x) = 1$ if $x > 0, 0$ if $x = 0, -1$ if $x < 0$. The statistic can be expressed as $K_n = \max(K_n^+, K_n^-)$, where $K_n^+ = \max_{1 \le j < n} U_{j,n}$ and $K_n^- = -\min_{1 \le j < n} U_{j,n}$ are tests for change in one direction.
For continuous observations, Pettitt (1979) re-expressed $U_{j,n}$ as a rank statistic

$$U_{j,n} = 2W_j - j(n+1), \text{ where } W_j = \sum_{i=1}^{j} R_i$$

and $R_1, ..., R_j$ are the ranks of the first j observations $X_1, ..., X_j$ in the overall sample of n observations. Thus, the nonparametric test

$$\frac{1}{n}\sqrt{\frac{3}{n+1}} K_n = \frac{1}{n}\sqrt{\frac{3}{n+1}} \max_{1 \le j < n} |2 \sum_{i=1}^{j} R_i - j(n+1)|, \tag{85}$$

has the limiting distribution given by the supremum of the absolute value of the Brownian bridge and the significance probabilities associated with the value k of K_n are approximated by

$$p_{OA} = 2 \sum_{r=1}^{\infty} (-1)^{r+1} \exp\{-6kr^2/(T^3 + T^2)\} \approx 2\exp\{-6k^2/(T^3 + T^2)\}. \tag{86}$$

The approximation is valid only for small p-values, and often results in values greater than 1 for samples with no change-point. The test introduced by Pettitt (1979) performs two roles: test for the presence of a change-point τ and then provides an estimate of τ when H_0 is rejected. The estimated value of τ is the value of j for which the maximum value of $|U_{j,n}|$ is obtained. It is important to note that LQE only estimates quantiles for the test statistic at specified levels of significance, and hence does not influence the estimation of τ.

We now introduce some notation and terminology to be more precise. For a score function ϕ satisfying $0 < \int_0^1 \phi^2(t)dt < \infty$, let

$$\bar{\phi} = n^{-1} \sum_{i=1}^{n} \phi(i/(n+1)) \text{ and } A^2 = (n-1)^{-1} \sum_{i=1}^{n} [\phi\{i/(n+1)\} - \bar{\phi}]^2,$$

and define the rank score of X_i as

$$s(R_i) = \frac{\phi\{R_i/(n+1)\} - \bar{\phi}}{A}, \text{ for } i = 1, ..., n. \tag{87}$$

Lombard (1987) and Koziol (1987) showed that under the null hypothesis of no change in distribution, the process

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} s(R_i)$$

converges in distribution to the Brownian bridge $(B(t))_{0 \le t \le 1}$, as $n \to \infty$. This convergence holds in space of cadlag functions $D[0,1]$ (continuous on the right with left limits), and it is obtained using Theorem 24.1 in Billingsley (1968). If the Wilcoxon's score function $\phi(t) = t$ is considered, then the process $Y_n(t)$ becomes

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \sqrt{\frac{12(n+1)}{n}} \left(\frac{R_i}{n+1} - \frac{1}{2} \right) \tag{88}$$

which is equivalently written as

$$Y_n(t) = \begin{cases} \frac{1}{n}\sqrt{\frac{3}{n+1}} U_{j,n}, \text{ if } \frac{j}{n} \le t < \frac{j+1}{n}, 1 \le j < n \\ 0, \text{ if } t = 1 \text{ or } 0 \le t < \frac{1}{n} \end{cases} \tag{89}$$

and the limiting distribution of $\frac{1}{n}\sqrt{\frac{3}{n+1}} \max_{1 \le j < n} |U_{j,n}|$ is that of $\sup_{0 \le t \le 1} |B(t)|$, where

$$P(\sup_{0 \le t \le 1} |B(t)| \le a) = 1 + 2 \sum_{i=1}^{\infty} (-1)^i \exp(-2i^2 a^2).$$

As we mentioned in the introduction, the aim is to obtain approximations for the quantiles of Pettitt's rank test and propose an almost sure limit theorem for the statistic $\frac{1}{n}\sqrt{\frac{3}{n+1}}\max_{1\le j<n}|U_{j,n}|$. We propose the following almost sure limit theorem without proof, and we verify it empirically using simulation studies for type I error and power.

Proposition 5. Let $X_1, ..., X_n$ be a sequence of independent random variables from a continuous distribution F. The statistic $\frac{1}{n}\sqrt{\frac{3}{n+1}}\max_{1\le j<n}|U_{j,n}|$ satisfies the almost sure limit theorem:

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\left(\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\le j<k}|U_{j,k}|\le x\right)=G(x) \qquad \text{a.s. } \forall x, \tag{90}$$

where $G(x)$ is the distribution function of $\sup_{0\le x\le 1}|B(x)|$ and $B(x)$ is the Brownian bridge.

4.3 NUMERICAL RESULTS

We present numerical results for the Pettitt test using simulated data, some data sets from Pettitt (1979) and Lombard (1987), and the logarithmic quantile estimation. When an almost sure limit theorem and a weak convergence for a test statistic hold with the same limiting distribution, Thangavelu (2005), Denker and Tabacu (2014, 2015) showed that the logarithmic quantile estimation method approximates the true quantiles almost surely. The weak convergence for the test in (85) was given in Pettitt (1979) and in Propositon 5 we propose its almost sure version with the same limiting distribution:

$$\lim_{n\to\infty}P\left(\frac{1}{n}\sqrt{\frac{3}{n+1}}\max_{1\le j<n}|U_{j,n}|\le t\right)=P(\sup_{0\le u\le 1}|B_n(u)|\le t)$$

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}\frac{1}{k}\mathbb{I}\left(\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\le j<k}|U_{j,k}|\le t\right)=G(t) \text{ a.s. for any } t,$$

where $G(t)$ is the distribution of $\sup_{0\le t\le 1}|B(t)|$.

We compare our results with the approximation given in equation (86) using Page's data set (see Page, 1955) and the industrial data as they appear in Pettitt

(1979). We also compute quantiles for the data set from Lombard (1987) and compare the results with the conclusions from Gombay (1994). For our final data analysis, we estimate the quantiles for the Dow Jones Industrial Average (DJIA) immediately around the stock market crash of 2008-2009 with a known change-point at March 9, 2009. Various simulation results for type I error and power are presented at the end of the section.

Recall from Chapter 2 that both weak convergence and the almost sure version of the weak convergence are independent of a finite number of initial random variables. More precisely, the convergence of the ASLT in Propositions 1-5 is unaffected by excluding values of $\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\leq j<k}|U_{j,k}|$ for $k=1,\ldots,k_0$, where $k_0 << n$. Fridline (2010) observed that smaller values of k correspond to larger weights $\frac{1}{k}$, which may have a significant influence on the values of the estimated quantiles, especially for smaller sample sizes. Fridline (2010) proposed a modification to the logarithmic quantile estimation algorithm to reduce the influence of the initial weights on the estimates without changing the asymptotic properties of the ASLT. Using the same theoretical principles for exchanging $\log n$ with $\sum_{k=1}^{n}\frac{1}{k}$, Fridline (2010) proposed using $C_n=\sum_{k=k_0}^{n}\frac{1}{k}$ and starting the algorithm using $k=k_0$ for k_0 suitably smaller than n. The modified version of the algorithm for computing the logarithmic quantiles introduced in Thangavelu (2005) and employed by Fridline (2010), Denker and Tabacu (2014, 2015) follows.

The upper and lower $\alpha/2$ empirical logarithmic quantiles of $\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\leq j<k}|U_{j,k}|$ in Proposition 5 for the i^{th} permutation of the data are estimated by

$$\hat{t}^{i,(n)}_{\alpha/2}=\max\left\{t\mid\frac{1}{C_n}\sum_{k=k_0}^{n}\frac{1}{k}\mathbb{I}\left(\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\leq j<k}\mid U_{j,k}\mid< t\right)\leq\alpha/2\right\},\ \text{and}\tag{91}$$

$$\hat{t}^{i,(n)}_{1-\alpha/2}=\max\left\{t\mid\frac{1}{C_n}\sum_{k=k_0}^{n}\frac{1}{k}\mathbb{I}\left(\frac{1}{k}\sqrt{\frac{3}{k+1}}\max_{1\leq j<k}\mid U_{j,k}\mid< t\right)\leq 1-\alpha/2\right\},\tag{92}$$

respectively, where $C_n=\sum_{i=1}^{n}\frac{1}{i}$, and $2\leq k_0<<n$ (note the test statistic is 0 when $k=1$). The upper and lower $\alpha/2$ logarithmic quantiles are estimated by

$$\bar{t}^{(n)}_{\alpha/2}=\frac{\sum_{i=1}^{\mathrm{p}}\hat{t}^{i,(n)}_{\alpha/2}}{\mathrm{p}},\ \text{and}\tag{93}$$

$$\bar{t}_{1-\alpha/2}^{(n)} = \frac{\sum_{i=1}^{\mathrm{p}} \hat{t}_{1-\alpha/2}^{i,(n)}}{\mathrm{p}}, \tag{94}$$

respectively, where p is the number of permutations chosen by the user. Unlike the approximation given in (86), the significance probabilities computed from the LQE cannot exceed one.

4.3.1 EXAMPLES

Page's Data

Table 25 contains a modified version of Page's data (Page, 1955) from Table 1 in Pettitt(1979, page 130). The first 20 values were generated from a $\mathcal{N}(5, 1)$ distribution. The remaining 20 values were generated from a $\mathcal{N}(6, 1)$ distribution. Pettitt (1979) subtracted five from all values.

Table 25: Page's data (see Pettitt, 1979, Table 1, p. 130)

i	1	2	3	4	5	6	7	8	9	10
x_i	-1.05	0.96	1.22	0.58	-0.98	-0.03	-1.54	-0.71	-0.35	0.66
i	11	12	13	14	15	16	17	18	19	20
x_i	0.44	0.91	-0.02	-1.42	1.26	-1.02	-0.81	1.66	1.05	0.97
i	21	22	23	24	25	26	27	28	29	30
x_i	2.14	1.22	-0.24	1.6	0.72	-0.12	0.44	0.03	0.66	0.56
i	31	32	33	34	35	36	37	38	39	40
x_i	1.37	1.66	0.1	0.8	1.29	0.49	-0.07	1.18	3.29	1.84

Note: Values displayed have been reduced by a value of 5 to match Table 1 in Pettitt (1979).

Figure 13 provides a plot of the data along with the sample means for the first and last 20 values. Table 26 provides the estimated logarithmic quantiles for Page's data ($n = 40$). Pettitt (1979) obtained an estimated p-value $p_{OA} = 0.014$ and a corresponding estimated change-point location $\tau = 17$. Pettitt noted that Page's test also estimated $\tau = 17$ with a p-value of approximately 0.01. Using logarithmic quantile estimation we obtain a p-value < 0.0152 which agrees with Pettitt (1979).

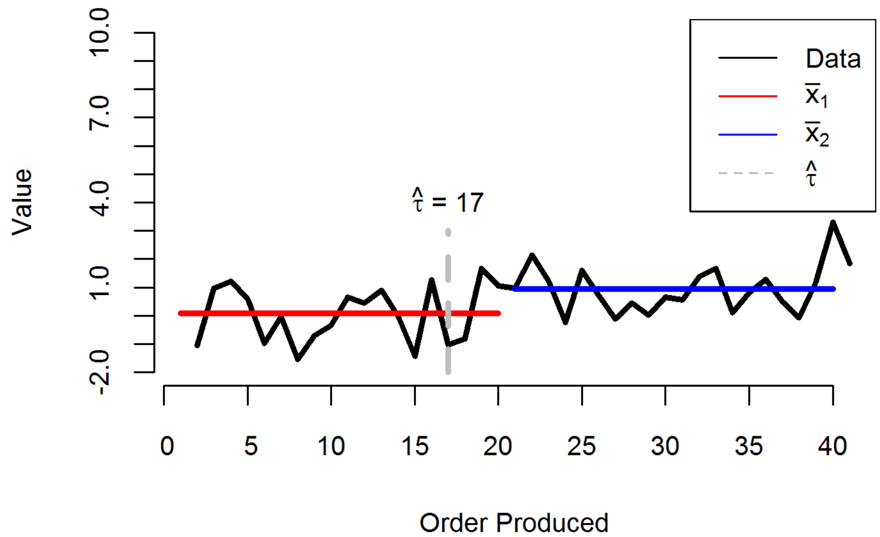

Figure 13: Page's data

Table 26: Logarithmic quantiles for Page's data

			$1-\alpha/2$			$\alpha/2$		
n	k_0	$\frac{1}{n}\sqrt{\frac{3}{n+1}}K_n$	0.9924	0.975	0.950	0.050	0.025	0.0076
40	2	1.5689	.4169	.4257	.4396	.9935	1.0442	1.0922

Values were obtained for a two-sided test with p=500 permutations. k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92).

Pettitt's Industrial Data Pettitt (1979) provides industrial data of the percentage of a certain material from 27 batches taken in order from a production source, and no other information is provided. Table 27 contains the Industrial Data provided by Pettitt (1979).

Table 27: Pettitt's industrial data (see Pettitt, 1979, Table 3, p. 133)

i	1	2	3	4	5	6	7	8	9
x_i	7.1	8.1	8.2	11.1	6.6	4.9	4	17.7	6.5
i	10	11	12	13	14	15	16	17	18
x_i	4.6	8.8	11.6	6.8	7.5	6.9	8.1	9.3	7.5
i	19	20	21	22	23	24	25	26	27
x_i	10	8.7	9.1	8.9	9.1	9.6	8.1	9.8	8.2

A value of $K_n = 90$ with a change-point location $\tau = 16$ corresponding to a two-sided test significance probability $p_{OA} = 0.185$ was obtained by Pettitt (1979). In order to support a change in the data to match indications from graphical analyses, Pettitt (1979) used a one-sided test with a corresponding significance probability $p_{OA} = 0.092$. The logarithmic quantile estimates are provided in Table 28. Using the LQE approach, the value of the test statistic corresponds to a p-value < 0.0256 providing evidence to support the claim that a change in distribution exists.

Table 28: Logarithmic quantiles for Pettitt's industrial data

			$1-\alpha/2$			$\alpha/2$		
n	k_0	$\frac{1}{n}\sqrt{\frac{3}{n+1}}K_n$	0.9872	0.975	0.950	0.050	0.025	0.0128
27	2	1.0911	.4106	.4166	.4299	.9431	.9969	1.0284

Values were obtained for a two-sided test with p=500 permutations. k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92). Observed p-value < 0.0256.

A plot of the data is provided in Figure 14. The lines for the average of the first 16 values, the average of the last 11 values, and the estimated change-point $\tau = 16$ are provided in the plot.

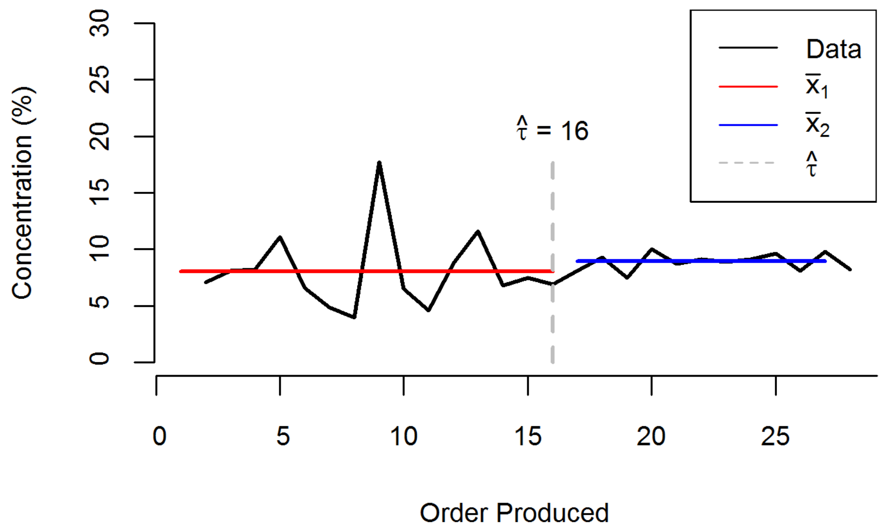

Figure 14: Pettitt's industrial data

Lombard's Data

Lombard (1987) provided a dataset from a manufacturing process. A circular indentation was cut into each part using a milling machine. The radii in centimeters were measured for $n = 100$ consecutive parts. A constant value of 3.9 was subtracted by Lombard from the measurements provided in Table 29. It was known that two servicing and resetting routines were performed during the production. The times (or parts) at which the routines were performed were either unknown or were not provided. A plot of the data is provided in Figure 15. A review of the plotted data does not readily reveal any possible change-point locations.

Lombard (1987) introduced several tests for change-points including smooth and abrupt changes for known and unknown change-point locations. For abrupt changes, he introduced a general test for multiple changes change-points. Lombard analyzed the dataset for one, two and thre change-points. For a single change-point, he obtained a p-value < 0.20. The test for two changes in distribution provided a p-value approximately 0.1, while his test for three possible change-points resulted in a p-value approximately 0.05. Gombay (1994) proposed a test which confirmed the presence of one change in Lombard's data with p-value = 0.0131. Gombay also performed a two-sided Pettitt test and obtained $p_{OA} = 0.1324$. Table 30 lists the estimated logarithmic quantiles, and the corresponding p-value = 0.0538 for a two-sided test of change in distribution.

Table 29: Lombard's data (Lombard, 1987, Table 3, p. 620)

1.010	1.066	0.975	0.921	1.165	1.027	1.100	0.981	0.977	1.106
0.932	0.990	0.940	0.877	0.987	0.958	1.112	0.878	1.029	0.971
1.004	1.087	1.038	1.119	0.768	1.096	1.114	1.007	0.978	0.957
0.884	1.004	1.032	1.130	0.961	1.066	1.029	1.107	1.150	1.190
1.152	1.049	1.183	0.933	1.161	0.988	1.087	1.034	0.889	1.109
1.196	1.098	0.954	0.986	0.943	1.058	0.960	1.073	0.904	1.171
1.060	1.189	1.019	1.213	1.204	1.148	1.033	1.023	1.145	0.994
1.147	1.054	1.059	0.972	1.141	1.082	0.931	0.848	1.039	1.043
1.016	1.027	0.932	0.879	0.754	0.911	0.971	1.180	0.849	0.870
1.003	0.834	1.018	1.145	0.995	0.895	1.085	1.055	0.992	1.141

Table 30: Logarithmic quantiles for Lombard's data

			$1-\alpha/2$			$\alpha/2$		
n	k_0	$\frac{1}{n}\sqrt{\frac{3}{n+1}}K_n$	0.995	0.975	0.950	0.050	0.025	0.005
100	2	1.1116	.4143	.4297	.4517	1.0581	1.1183	1.1893

Values were obtained for a two-sided test for one change-point using Pettitt's method with p=500 permutations. k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92). Observed p-value < 0.0538.

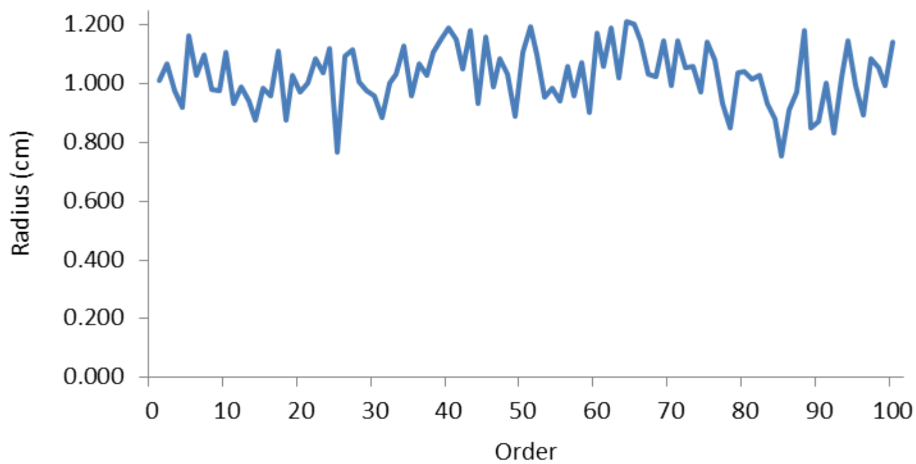

Figure 15: Lombard's data

Dow Jones Industrial Average (DJIA) Data

The daily closing values for the DJIA from August 8, 2008 through April 23, 2010 were obtained from The Wall Street Journal website (quotes.wsj.com). It is well-known that the market reached its lowest closing value during that crisis on March 10, 2009 (change-point). The logarithmic quantile estimates are provided in Table 31 for the entire data and for two shorter periods of time around the bottom of the market. Analysis of all $n = 430$ values resulted in a test statistic of 6.0485 corresponding to an LQE p-value = .00082 (Pettitt's $p_{OA} = 1.67 \times 10^{-32}$). The change-point was identified as October 7, 2008 (see Figure 16). A shorter interval of time from January 05, 2009 to April 1, 2009 ($n = 61$) was chosen, and a test statistic of 3.3320 was obtained corresponding to an LQE p-value = .0152 ($p_{OA} = 2.27 \times 10^{-10}$). The change-point was identified as Feb 13, 2009 (see Figure 17). Finally, a shorter interval of time was chosen between February 23, 2009 and March 11, 2009 ($n = 13$). The test statistic value 1.4243 corresponds to a p-value $< .0704$ ($p_{OA} = .0346$).The indicated change-point is on February 27, 2009. This analysis demonstrates the ability of the Pettitt test to detect a change-point when the change occurs as a trend instead of abruptly. The plot of the DJIA data clearly shows a decreasing trend followed by an increasing trend. Two open problems are to extend the LQE approach to change-point analysis of multiple changes and to time series data.

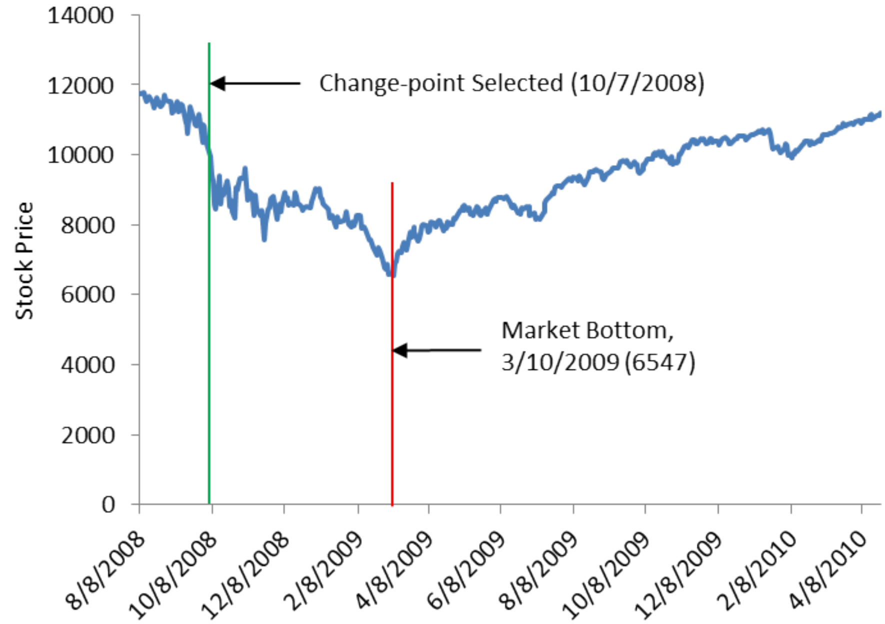

Figure 16: Dow Jones Industrial Average data (08/08/2008 - 04/23/2010)

Table 31: Logarithmic quantiles for Dow Jones Industrial Average data

			$1-\alpha/2$			$\alpha/2$		
n	k_0	$\frac{1}{n}\sqrt{\frac{3}{n+1}}K_n$	0.995	0.975	0.950	0.050	0.025	0.005
430	2	6.0485	.4016	.4222	.4512	1.1416	1.2114	1.3079
61	2	3.3320	.4156	.4289	.4459	1.0288	1.0841	1.1555
13	2	1.4243	.4416*	.4416*	.4502	.8738	.9104**	.9104**

* The smallest quantile occurred for $\alpha/2 = .0352$.

** The largest quantile occurred for $1-\alpha/2 = .9648$

Values were obtained for a two-sided test with p=500 permutations. k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92).

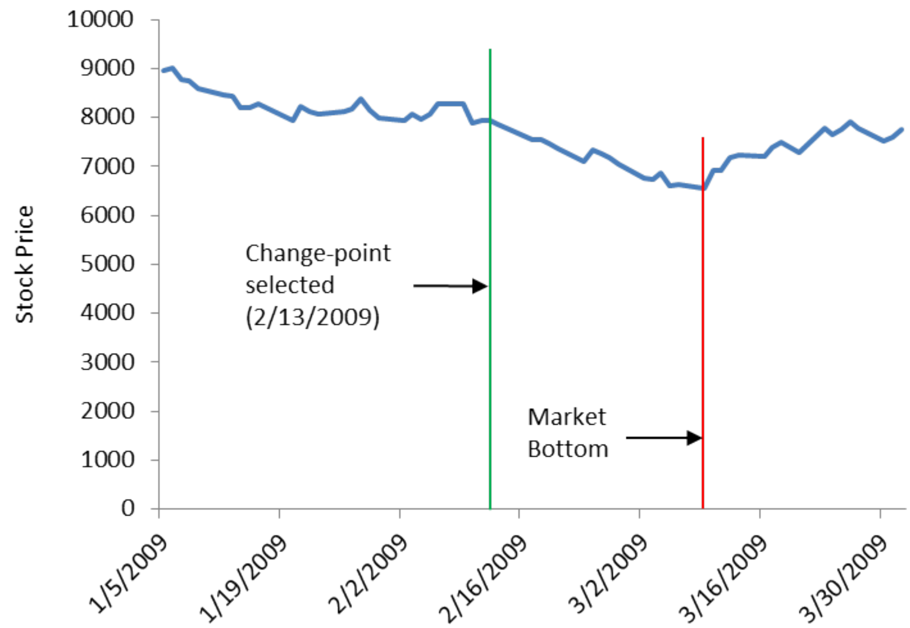

Figure 17: Dow Jones Industrial Average data (01/05/2009 - 04-04/2009)

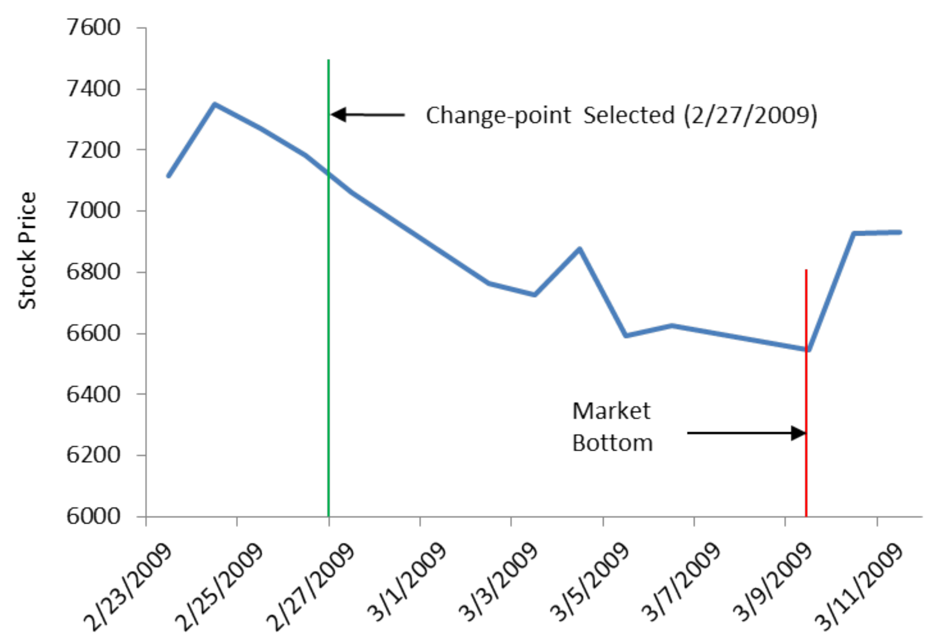

Figure 18: Dow Jones Industrial Average data (02/23/2009 - 03/11/2009)

4.3.2 SIMULATION RESULTS

Tables 32 and 33 provide the type I error simulation results for the $Exp(1)$, $\mathcal{N}(0,1)$, and $Gamma(4, \frac{1}{2})$ distributions from 1000 simulations and 500 permutations each. The simulated type I error results are liberal. The test is more liberal for small n, but slowly approaches the true level as n increases. One possible solution to address the liberalness of the test was to start with larger values of k_0 to reduce the influence of the first few terms. Simulation studies were performed for each selected distribution for $1 \leq k_0 \leq 10$. Tables 32 and 33 include values of $k_0 \leq 9$ for brevity.

A review of the effects of k_0 on the type I error in Tables 32 and 33 reveal that the type I error gets lower as k_0 increases to some value and then increases again. In both tables, the lowest (i.e. closest to target) values for each combination of level, n, and distribution have been placed in boldface. There is not a consistent value for k_0 that delivers the best type I error. More importantly, the amount of reduction in type I error provided by fine tuning k_0 is not significantly closer to the true level than it is for $k_0 = 2$. The most significant improvement is provided at $n = 500$, yet the true values remain at best 40% above the target value. Addressing the liberalness of the results is an area for future investigation.

Table 32: Simulated type I error for Pettitt's test ($n = 30$, $n = 50$)

		$n=30$			$n=50$		
		Level			Level		
Distribution	k_0	10%	5%	1%	10%	5%	1%
Exp(1)	2	0.266	0.210	0.178	0.251	0.199	**0.141**
	3	0.255	0.206	0.178	0.236	0.192	0.146
	4	0.247	**0.201**	**0.174**	0.223	0.185	0.150
	5	**0.243**	0.204	0.183	**0.219**	**0.180**	0.147
	6	0.253	0.208	0.189	0.222	0.183	0.154
	7	0.267	0.220	0.206	0.224	0.185	0.157
	8	0.274	0.234	0.223	0.232	0.189	0.167
	9	0.291	0.242	0.242	0.235	0.195	0.170
	10	0.302	0.260	0.26	0.246	0.204	0.175
N(0,1)	2	0.255	0.207	0.182	0.240	0.184	**0.138**
	3	0.241	**0.198**	**0.184**	0.230	0.187	0.143
	4	0.228	**0.198**	0.187	0.222	0.182	0.140
	5	**0.226**	0.201	0.190	0.219	0.180	0.141
	6	0.244	0.211	0.197	**0.217**	**0.177**	0.145
	7	0.259	0.229	0.217	0.224	0.186	0.152
	8	0.269	0.248	0.240	0.230	0.185	0.163
	9	0.283	0.248	0.248	0.240	0.199	0.168
	10	0.294	0.268	0.268	0.242	0.200	0.176
Gamma($4,\frac{1}{2}$)	2	0.312	0.261	**0.210**	0.251	0.201	**0.143**
	3	0.296	**0.244**	0.214	0.241	0.193	0.149
	4	**0.290**	0.252	0.217	**0.230**	0.188	0.146
	5	0.294	0.253	0.225	**0.230**	**0.180**	0.149
	6	0.298	0.265	0.236	0.236	0.187	0.159
	7	0.305	0.273	0.250	0.236	0.195	0.162
	8	0.311	0.267	0.257	0.245	0.199	0.172
	9	0.324	0.269	0.269	0.254	0.211	0.181
	10	0.345	0.296	0.296	0.265	0.214	0.185

Note: k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92). The change-point is denoted by τ.Values were obtained from 1000 simulations and 500 permutations each.

Table 33: Simulated type I error for Pettitt's test ($n = 100$, $n = 500$)

		$n=100$ Level			$n=500$ Level		
Distribution	k_0	10%	5%	1%	10%	5%	1%
Exp(1)	2	0.192	0.148	0.099	0.164	0.115	0.070
	3	0.185	0.143	**0.093**	0.160	0.111	**0.067**
	4	0.185	0.139	0.094	0.152	0.109	0.070
	5	0.182	**0.136**	0.096	0.149	**0.107**	0.070
	6	0.186	0.140	0.101	0.149	0.108	0.070
	7	**0.181**	0.139	0.101	0.150	0.111	0.070
	8	0.197	0.145	0.103	**0.148**	0.112	0.072
	9	0.199	0.149	0.105	0.149	0.111	0.073
	10	0.202	0.156	0.108	0.153	0.112	0.072
N(0,1)	2	0.201	0.151	0.099	0.160	0.096	**0.052**
	3	0.183	0.140	**0.095**	0.153	0.104	0.060
	4	0.182	0.138	**0.095**	0.146	0.099	0.060
	5	**0.179**	**0.134**	0.097	0.145	0.097	0.060
	6	0.180	0.141	0.099	**0.143**	0.098	0.059
	7	0.186	0.139	0.100	0.145	0.092	0.058
	8	0.190	0.144	0.106	0.150	0.092	0.058
	9	0.198	0.145	0.107	0.150	**0.091**	0.059
	10	0.200	0.151	0.112	0.155	0.093	0.057
Gamma($4,\frac{1}{2}$)	2	0.196	0.142	0.100	0.176	0.122	0.078
	3	0.194	0.141	**0.098**	0.157	0.111	0.063
	4	0.189	0.139	0.100	0.151	0.110	**0.062**
	5	**0.179**	**0.137**	0.103	0.152	0.107	0.063
	6	0.185	0.143	0.099	0.154	0.109	0.064
	7	0.186	0.148	0.108	**0.141**	**0.096**	0.069
	8	0.197	0.146	0.109	0.144	0.099	0.069
	9	0.201	0.150	0.112	0.144	0.104	0.070
	10	0.200	0.163	0.115	0.146	0.107	0.071

Note: k_0 is the minimum number of observations used to calculate $\hat{t}^{i,(n)}_{\alpha/2}$ in (91) and $\hat{t}^{i,(n)}_{1-\alpha/2}$ in (92). The change-point is denoted by τ.Values were obtained from 1000 simulations and 500 permutations each.

The liberal results for type I error led to an investigation of the type I error for the Pettitt test (not LQE). A simulation study for the type I error is not found in the literature to the best of our knowledge. Hence, we performed a simulation study in order to compare the type I error of the Pettitt test to the type I error of the LQE approach. The simulation study was implemented for the Pettitt test at the 10%, 5%, and 1% levels using sample sizes $n = 10, 20, \ldots, 100, 200, \ldots, 1000, 2000, \ldots, 10000$, with 2000, 5000, and 10,000 simulations for same three distributions $Exp(1)$, $\mathcal{N}(0,1)$, and $Gamma(4, \frac{1}{2})$.

Figures 19-21 contain the results for the $\mathcal{N}(0,1)$ distribution. Results for the Exp(1) and Gamma(4,$\frac{1}{2}$) were very similar. All graphs have the same scale to prevent misleading interpretations, but for the 1% level graphs, the differences due to the number of simulations is difficult to visually detect. The type I error approaches the actual level of the test around $n = 2000$.

For larger sample sizes ($n \geq 2000$) the values for 2000 and 5000 simulations fluctuate between liberal and conservative results (see Figures 19-21) . The line for 10000 simulations generally remains near the actual level or slightly conservative.

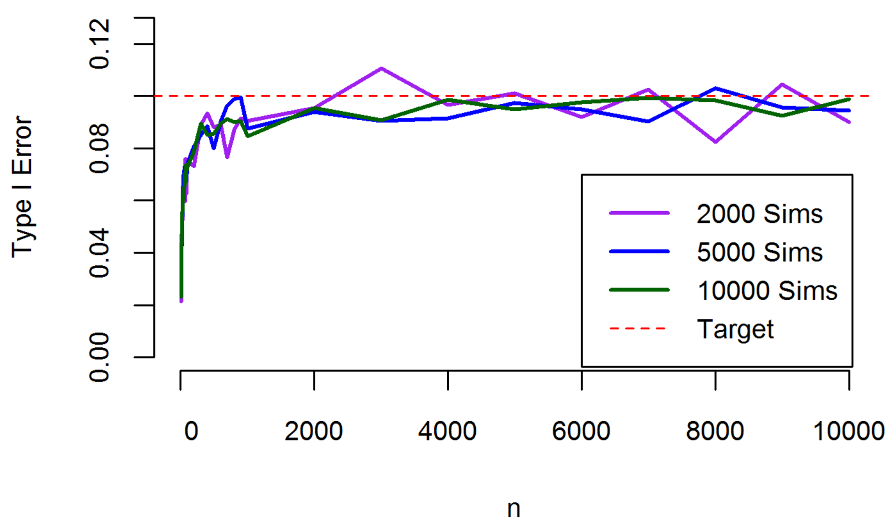

Figure 19: Type I error for Pettitt's test, 10% level, $\mathcal{N}(0,1)$

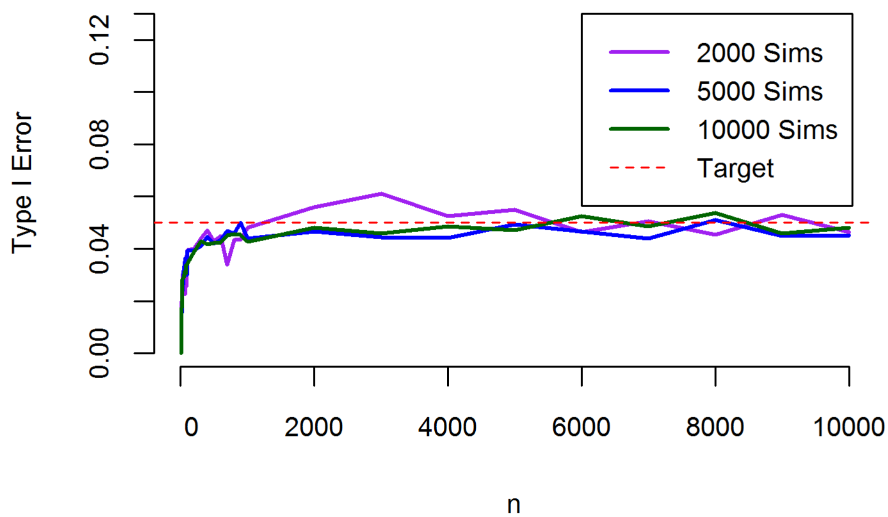

Figure 20: Type I error for Pettitt's test, 5% level, $\mathcal{N}(0, 1)$

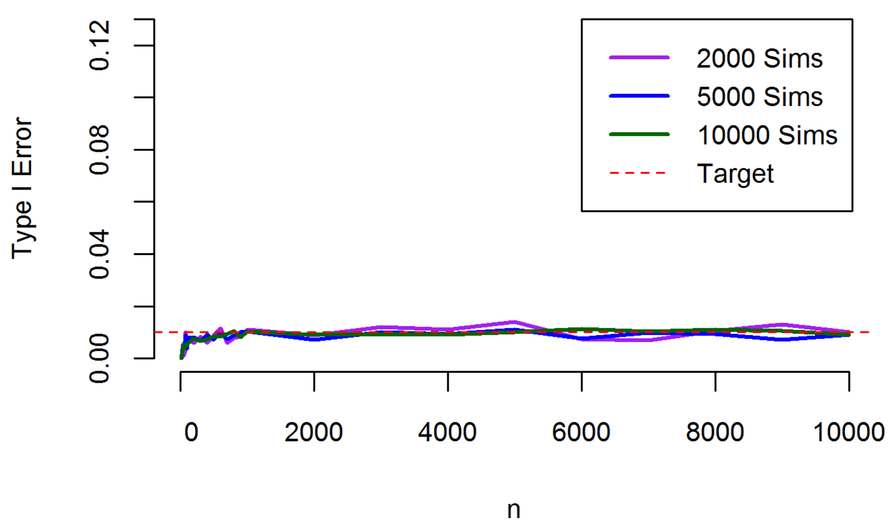

Figure 21: Type I error for Pettitt's test, 1% level, $\mathcal{N}(0, 1)$

The results for sample sizes in the range investigated for LQE ($n \leq 500$) appear to be very conservative, opposite to the liberal LQE results. To better understand the type I error of the Pettitt test for $n \leq 500$, Figures 22-24 are provided for the $\mathcal{N}(0,1)$ distribution. The results are significantly conservative for $n \leq 100$ at the 10% and 5% levels. As samples sizes approach 500, the test diverges from the level, unlike the LQE type I error results which converge to the level as n approaches 500. For the 1% level, the simulated type I error converges to the level when n is approximately 200. The results for Gamma$(4, \frac{1}{2})$ and Exp(1) were similar. These simulation studies show that the Pettitt test is not liberal, and hence is not the cause for the liberal LQE results.

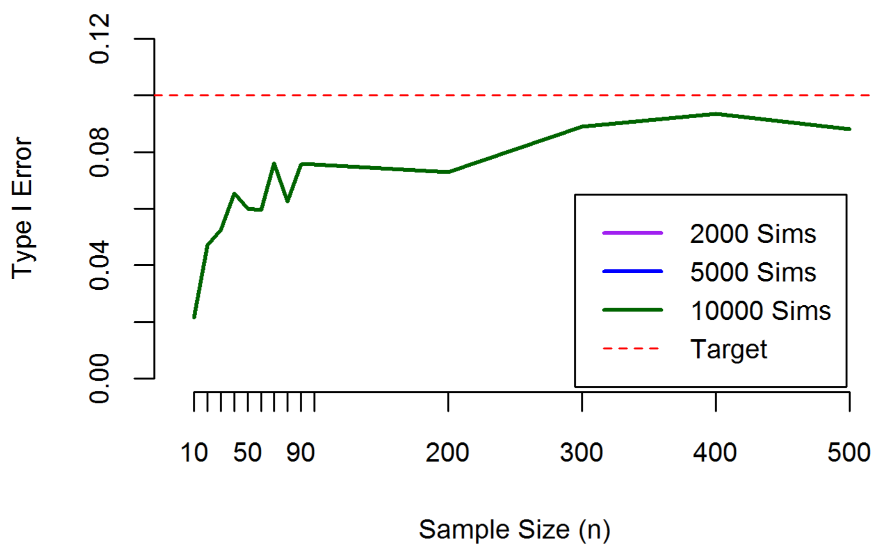

Figure 22: Type I error for Pettitt's test, 10% level, $n \leq 500$, $\mathcal{N}(0,1)$

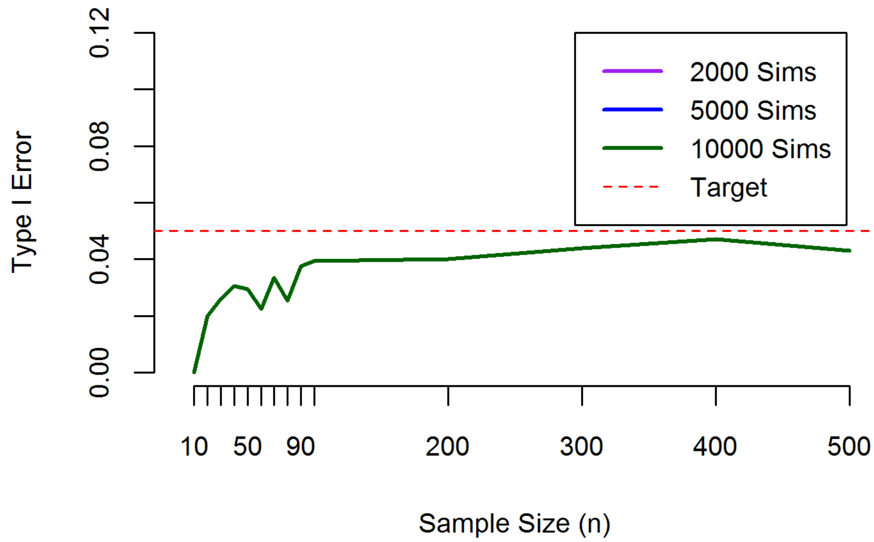

Figure 23: Type I error for Pettitt's test, 5% level, $n \leq 500$, $\mathcal{N}(0, 1)$

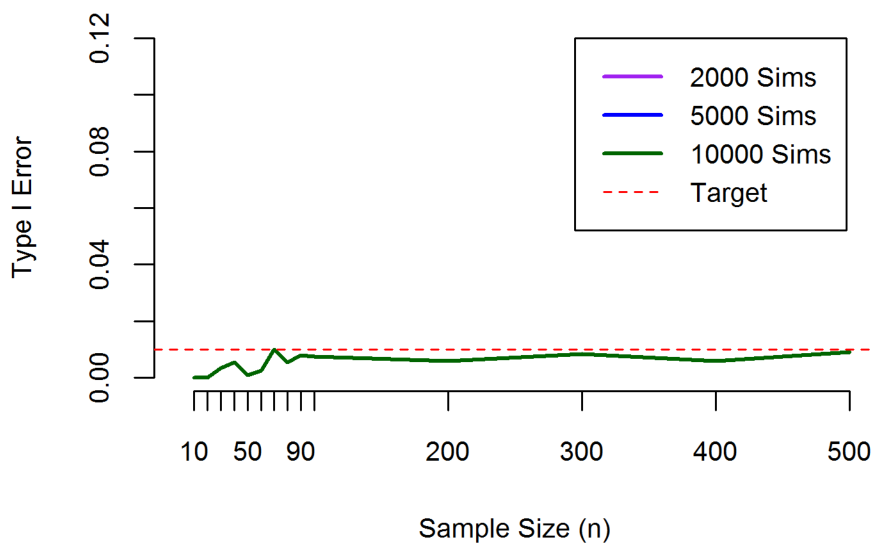

Figure 24: Type I error for Pettitt's test, 1% level, $n \leq 500$, $\mathcal{N}(0, 1)$

The simulated power for the 10%, 5%, and 1% levels for the Exp(1), Gamma($4, \frac{1}{2}$), and $\mathcal{N}$(0,1) distributions are presented in Tables 34-36, respectively. The values of τ denote the change-point position at 10%, 25%, 50%, 75%, and 90% of n. The amount of shift in the distribution at the change-point is given by δ. Two amounts of shift are provided: $\delta = 0.5$ and $\delta = 0.75$. The power is highest when the change-point occurs near the middle of the sample, and it decreases symmetrically as the change-point location approaches either the beginning or end of the data stream. The difference in power when the change-point occurs between $0.25n$ and $0.75n$ is fairly small. The power is significantly lower when τ occurs outside of this range.

Table 34: Power for various τ (exponential distribution)

		Exp(1), Exp(1)+δ					
		$\delta = 0.5$			$\delta = 0.75$		
n	τ	10%	5%	1%	10%	5%	1%
30	3	0.298	0.229	0.192	0.339	0.269	0.223
30	7	0.546	0.498	0.422	0.697	0.637	0.570
30	15	0.659	0.618	0.579	0.850	0.822	0.816
30	22	0.559	0.484	0.476	0.746	0.686	0.692
30	27	0.286	0.231	0.190	0.323	0.260	0.207
50	5	0.328	0.272	0.201	0.428	0.360	0.269
50	12	0.629	0.583	0.531	0.846	0.803	0.760
50	25	0.838	0.797	0.737	0.977	0.965	0.923
50	37	0.697	0.635	0.560	0.916	0.879	0.841
50	45	0.312	0.252	0.194	0.384	0.310	0.236
100	10	0.431	0.357	0.247	0.616	0.526	0.394
100	25	0.862	0.818	0.764	0.976	0.969	0.945
100	50	0.974	0.961	0.926	1.000	0.999	1.000
100	75	0.897	0.870	0.816	0.997	0.991	0.984
100	90	0.375	0.308	0.242	0.565	0.480	0.400
500	50	0.908	0.874	0.815	0.996	0.992	0.983
500	125	0.999	0.999	1.000	1.000	1.000	1.000
500	250	1.000	1.000	1.000	1.000	1.000	1.000
500	375	1.000	1.000	1.000	1.000	1.000	1.000
500	450	0.972	0.946	0.892	1.000	1.000	0.999

Note: $k_0 = 2$ was used in (91) and (92). τ is the change-point.
Values were obtained from 1000 simulations, p=500.

For the following statements, we restrict our considerations to change-points near the middle of the data stream. When the amount of shift is $\delta = 0.5$, the power achieves 0.8 for a sample size somewhat larger than 100. For $\delta = 0.75$, the power exceeds 0.8 for the 10% and 5% level tests for $n \geq 50$. To achieve a power of at least 0.8 at the 1% level, a sample size of $n \geq 100$ is required. An additional simulation study using $\delta = 1$ results in power very close to one for $n \geq 30$.

Table 35: Power for various τ (gamma distribution)

		Gamma(4,$\frac{1}{2}$), Gamma(4,$\frac{1}{2}$)+δ					
		$\delta = 0.5$			$\delta = 0.75$		
n	τ	10%	5%	1%	10%	5%	1%
30	3	0.286	0.231	0.211	0.308	0.241	0.218
30	7	0.427	0.366	0.331	0.566	0.515	0.476
30	15	0.552	0.489	0.449	0.750	0.699	0.697
30	22	0.451	0.382	0.324	0.634	0.572	0.507
30	27	0.268	0.214	0.175	0.286	0.223	0.177
50	5	0.299	0.236	0.162	0.336	0.282	0.207
50	12	0.505	0.439	0.374	0.719	0.665	0.598
50	25	0.653	0.603	0.541	0.890	0.862	0.832
50	37	0.525	0.454	0.405	0.777	0.727	0.671
50	45	0.280	0.218	0.177	0.320	0.254	0.195
100	10	0.297	0.236	0.181	0.434	0.354	0.301
100	25	0.696	0.645	0.559	0.916	0.899	0.862
100	50	0.829	0.796	0.771	0.987	0.977	0.953
100	75	0.720	0.651	0.533	0.940	0.919	0.887
100	90	0.334	0.257	0.186	0.463	0.378	0.304
500	50	0.757	0.676	0.543	0.973	0.958	0.920
500	125	0.997	0.995	0.990	1.000	1.000	1.000
500	250	1.000	1.000	1.000	1.000	1.000	1.000
500	375	0.998	0.997	0.997	1.000	1.000	1.000
500	450	0.773	0.684	0.554	0.992	0.985	0.967

Note: $k_0 = 2$ was used in (91) and (92). τ is the change-point.
Values were obtained from 1000 simulations, p=500.

Table 36: Power for various τ (normal distribution)

		$\mathcal{N}(0,1)$, $\mathcal{N}(\delta,1)$					
		$\delta = 0.5$			$\delta = 0.75$		
n	τ	10%	5%	1%	10%	5%	1%
30	3	0.291	0.234	0.198	0.314	0.248	0.229
30	7	0.402	0.333	0.289	0.534	0.470	0.420
30	15	0.506	0.442	0.405	0.701	0.653	0.621
30	22	0.393	0.339	0.311	0.583	0.520	0.476
30	27	0.286	0.234	0.203	0.310	0.252	0.217
50	5	0.272	0.220	0.163	0.329	0.273	0.201
50	12	0.428	0.369	0.318	0.628	0.571	0.526
50	25	0.590	0.537	0.445	0.834	0.805	0.743
50	37	0.471	0.408	0.351	0.678	0.621	0.572
50	45	0.264	0.205	0.164	0.303	0.248	0.212
100	10	0.302	0.253	0.166	0.423	0.350	0.254
100	25	0.617	0.559	0.486	0.887	0.852	0.807
100	50	0.798	0.758	0.707	0.972	0.962	0.954
100	75	0.629	0.562	0.464	0.883	0.853	0.822
100	90	0.305	0.247	0.174	0.413	0.343	0.278
500	50	0.670	0.571	0.420	0.959	0.931	0.879
500	125	0.995	0.992	0.993	1.000	1.000	1.000
500	250	0.999	0.999	0.999	1.000	1.000	1.000
500	375	0.995	0.993	0.984	1.000	1.000	1.000
500	450	0.690	0.610	0.472	0.965	0.937	0.892

Note: $k_0 = 2$ was used in (91) and (92). τ is the change-point.
Values were obtained from 1000 simulations, p=500.

4.4 SUMMARY

LQE estimates the quantiles of the test statistic directly using only the data, and without using the asymptotic distribution. In this chapter, we have investigated the change-point problem from the logarithmic quantile estimation approach. To be more precise, we have proposed an almost sure limit theorem using the statistic proposed by Pettitt (1979) to test for a change in distribution in a sequence of data. We then provided an algorithm for estimating quantiles of the statistic for a two-sided test. Four datasets were tested for the presence of a change-point. Results agreed with previous analyses where available. The DJIA dataset clearly displayed a decreasing trend followed by an increasing trend instead of an abrupt change, and the Pettitt test detected a change-point. Simulation results for type I error are liberal for small sample sizes and approach the level of the test as the sample sizes increase. The power was simulated with change-points at several locations within the data. For change-points near the center of the data, power is higher, and power decreases symmetrically as the change-point location nears the edges of the sequence of values. For change-points at the center of the data, the power of the test requires large sample sizes when the shift in distribution is only 0.5. For a shift of 0.75 at the change-point, sample sizes of 50 or more had power exceeding 0.8. LQE is a competitive approach for detecting change-points, and warrants investigation for multiple change-points in stable processes and in time-series.

CHAPTER 5

TWO-SAMPLE PROBLEM USING PAIRED AVERAGES

5.1 INTRODUCTION

The two-sample problem has been studied extensively in literature. In general, let $X_1, \ldots, X_n \backsim F$ and $Y_1, \ldots, Y_m \backsim G$ be i.i.d. random variables, and F, G are distribution functions. The original parametric two-sample problem tested for the equality of means. The nonparametric two-sample problem studied in this dissertation tests for the equality of the two distributions. If the variances of the two samples are not assumed equal, the problem is referred to as the famous Behrens-Fisher problem. In this dissertation, we limit our investigation to the nonparametric analysis of the two-sample problem. Our motivation for revisiting the nonparametric two-sample problem follows from the fact that the Wilcoxon-Mann-Whitney (WMW) test (Mann and Whitney, 1947) is not as efficient as the t-test, and hence requires larger sample sizes to obtain a similar power. In many medical applications, it is either impractical or cost-prohibitive to obtain larger sample sizes, hence an alternative approach is required.

In order to provide insight into the logic behind our proposed solution, a brief literature review follows. Three well-known nonparametric tests for the equality of distribution functions are the WMW, the Kolmogorov-Smirnov (KS) (Smirnov, 1939), and the Cramer-von Mises (Mises, 1947) tests. The latter two require continuous distribution functions, whereas an adjustment for ties exists for the WMW test. Brunner and Neumann (1986) investigated the asymptotic properties of the WMW test when the variances are unequal assuming continuous distribution functions. Baumgartner et al. (1998) proposed a modification of the Cramer-von Mises test weighted by its variance. This test was shown to have power comparable to the WMW, Cramer-von Mises and KS tests for continuous distributions.

Brunner and Munzel (2000) relaxed the requirement for continuity, and developed a rank test using a consistent estimate of the asymptotic variance. An extensive simulation study on the WMW test was performed for different scenarios: variance ratios (ratio of the variance of the first sample to the variance of the second sample), modality, and skewness. The study revealed that the MWM test was conservative if the larger sample size had higher variance and was liberal for the reverse situation, where the results were determined by the ratio of the variances and the ratios of the sample sizes not the combined sample size (preserved asymptotically). The performance of their proposed test statistic was accurate for large sample sizes, but was liberal for sample sizes less than 50. Their small sample approximation used the central t-distribution and the Satterthwaite-Smith-Welch (SSW) approximation for the degrees of freedom. This small sample approximation provided type I error results comparable to the parametric t-test using the SSW approximation for samples sizes greater than 10.

Denker and Puri (1992) proved the asymptotic normality of the two-sample linear rank test statistic under an empirical process of U-statistic structure and provided an upper bound for the asymptotic variance along with an estimate for the Pittman efficiency. Compagnone and Denker (1996) extended the proof to the generalized linear rank statistic and provided explicit forms of the expectation and variance. They also show that an increase of one in the smallest sample-size results in a significant increase in the efficiency of the test.

We propose using kernel functions of order two to generate paired averages within each sample thus effectively increasing the sample size to $\binom{n}{2}$. With this test, we hope to remedy the results of the study in Brunner and Munzel (2000) where they state that accurate tests for sample sizes less then 10 cannot be expected in the general nonparametric model. In Section 5.2, we introduce the model and notation under the U-statistic structure for rank tests , and we propose a test statistic using paired averages. In Section 5.3 we provide the algorithms for verifying both its convergence in distribution and the convergence of the proposed ASCLT. We then show empirically the convergence along with simulation results of type I error and power. Section 5.4 is a summary of our findings and the open problems.

5.2 NOTATION AND MODEL

The empirical process of U-statistic structure (Denker and Puri, 1992) is given by

$$\binom{N}{m}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_m \le N} \left(\mathbb{I}\left(\Psi(X_{i_1}, X_{i_2}, \cdots, X_{i_m}) \le t\right) - P(\{\Psi(X_1, \ldots, X_m) \le t\}) \right),$$

where $n = n + m$ is the total number of observations in the experiment, and Ψ is a symmetric measurable function in m variables. We limit our discussion to a two-sample problem, and provide the model below.

We define independent random variables $X_1, X_2, \ldots, X_n \sim F$, and $Y_1, Y_2, \ldots, Y_m \sim G$ from stationary processes. Let

$$\Psi_q^X : \mathbb{R}^q \to \mathbb{R},$$
$$\Psi_p^Y : \mathbb{R}^p \to \mathbb{R}$$

be two symmetric measurable functions. We consider the case where $n = m$ and $p = q = 2$, and define new random variables

$$X_{t_1,t_2} := \Psi_2^X(X_{t_1}, X_{t_2}) = \frac{X_{t_1} + X_{t_2}}{2}, \qquad 1 \le t_1 < t_2 \le n,$$
$$Y_{s_1,s_2} := \Psi_2^Y(Y_{s_1}, Y_{s_2}) = \frac{Y_{s_1} + Y_{s_2}}{2}, \qquad 1 \le s_1 < s_2 \le n.$$

The corresponding U-statistic structure empirical distribution functions are

$$\hat{F}_n(t) = \binom{n}{2}^{-1} \sum_{1 \le t_1 < t_2 \le n} \mathbb{I}\left(X_{t_1,t_2} \le t\right), \qquad E\hat{F}_n(t) = F_n(t),$$

$$\hat{G}_n(t) = \binom{n}{2}^{-1} \sum_{1 \le s_1 < s_2 \le n} \mathbb{I}\left(Y_{s_1,s_2} \le t\right), \qquad E\hat{G}_n(t) = G_n(t).$$

We now define the overall (average) empirical distribution function

$$\hat{H}_n(t) = \frac{\binom{n}{2}}{1 + 2n(n-1)} \left(\hat{F}_n(t) + \hat{G}_n(t) \right),$$

where its expectation is $H_n(t)$. The general form of the two-sample linear rank statistic may be expressed by

$$T_n(J) = \binom{n}{2}^{-1} \sum_{1 \le t_1 < t_2 \le n} J\left(\frac{R_{t_1,t_2}}{1+n(n-1)}\right) - \int_{-\infty}^{\infty} J(H_n(t))dF_n(t),$$

where J is a continuous score function, R_{t_1,t_2} is the rank of X_{t_1,t_2} among all $n(n-1)$ random variables from the combined samples. Throughout this chapter, we use the Wilcoxon score function $J(u) = u$. The asymptotic variance of $T_n(J)$ is

$$\sigma_n^2 = \frac{4\tau_1^2}{n} + \frac{4\tau_2^2}{n}.$$

The components are defined by

$$\tau_1^2 = \int_{-\infty}^{\infty} h_1^2(x)dF(x), \qquad \tau_2^2 = \int_{-\infty}^{\infty} g_1^2(y)dG(y),$$

where

$$\begin{aligned} h_1(x) = E\{H_n(\Psi(x, X_2)) - \int_{-\infty}^{\infty} H_n(t)dF_n(t)\} \\ + \frac{1}{2}\int_{-\infty}^{\infty}(P(\Psi(x, X_2) \le t) - F_n(t))dF_n(t), \end{aligned}$$

and
$$g_1(y) = \frac{1}{2}\int_{-\infty}^{\infty}(P(\Psi(y, Y_2) \le t) - G_n(t))dG_n(t).$$

Denker and Puri (1992) proved that the ratio

$$\frac{T_n(J)}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1).$$

Due to its open form, an upper bound for the asymptotic variance was provided. Specifically for $p, q = 2$, $m = n$, and with some constant C independent of n, we have

$$\sigma_n^2 \le C \left\|J\right\|^2 (0.5) \left(\frac{4}{n}\right) = O\left(\frac{1}{n}\right).$$

In practice, it is very difficult to use an open form for the variance, so we propose a test statistic formulated with the upper bound of σ_n^2 for use in an almost sure central limit theorem. This new test statistic should converge to a $\mathcal{N}(0, \sigma^2)$ for some unknown σ^2 instead of the standard normal distribution. The proposed test statistic is given by

$$L_n = \frac{T_n(J)}{1/\sqrt{n}} = \sqrt{n}\left(\frac{1}{\binom{n}{2}} \sum_{1 \le t_1 < t_2 \le n} \frac{R_{t_1,t_2}}{1 + n(n-1)} - \frac{1}{2}\right). \tag{95}$$

The hypotheses of interest are

$$H_0 : F(t) = G(t) \qquad \text{vs.} \qquad H_1 : F(t) \neq G(t) \qquad \forall t. \tag{96}$$

LQE requires the existence of a weak law and an almost sure limit theorem converging to the same distributions.

L_n in (95) may be used to test the hypotheses in (96), the convergence in distribution and the corresponding convergence of the ASCLT must be investigated.

5.3 CONVERGENCE OF THE TEST STATISTIC

For the new test statistic we proposed in (95), we need to show the following required convergences in order to use the LQE approach, which has not previously been done.

$$L_n = \frac{T_n(J)}{1/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{97}$$

and

$$\lim_{n\to\infty} \frac{1}{C_n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{I}\left(L_n < t\right) = \mathcal{N}(0, \sigma^2), \qquad \text{a.s. } \forall t \in \mathbb{R}, \tag{98}$$

where $\sigma^2 < \infty$ is unknown, and $C_n = \sum_{m=1}^{n} \frac{1}{m}$. Convergence in both distribution and almost surely were verified empirically using the two following algorithm. The algorithm for convergence in distribution follows standard methodology found in many texts (see Rizzo, 2008). The algorithm to verify the convergence of the ASCLT is modified from Thangavelu (2005), Fridline (2010), and Denker and Tabacu (2014,2015).

Algorithm 1: Convergence in Distribution for the Test Statistic

1. Simulate S pairs of independent random samples each of size n.

2. For each simulation

   (a) Create all unique paired averages within each sample,

   (b) Combine the samples and rank,

   (c) Calculate the test statistic for the i^{th} simulation

   $$L_n^{(i)}(t) = \sqrt{n}\left(\frac{1}{\binom{n}{2}}\sum_{1\le t_1<t_2\le n}\frac{R_{t_1,t_2}^{(i)}}{1+n(n-1)} - \frac{1}{2}\right),$$

   where R_{t_1,t_2} is the rank of X_{t_1,t_2} among all $n(n-1)$ random variables (paired averages) in both samples.

3. Calculate empirical quantiles of the test statistic.

4. Calculate the Monte Carlo estimate of variance, σ^2_{mc}.

5. Repeat for larger sample sizes until the results converge.

The results are graphically compared to the quantiles for the distribution function of $\mathcal{N}(0, \hat{\sigma^2_{mc}})$ (see Section 5.3.1). The convergence of the almost sure central limit theorem was confirmed using the algorithm below.

Algorithm 2: Almost Sure Weak Convergence of Test Statistic

1. Simulate S pairs of two independent samples of size n.

2. For the i^{th} simulation, permute each sample independently p times, where p is chosen by the user.

3. For the j^{th} permutation of the i^{th} simulation,

   (a) Calculate

   $$L_k^{(i,j)} = \sqrt{k}\left(\frac{1}{\binom{k}{2}}\sum_{1\le t_1<t_2\le k}\frac{R_{t_1,t_2}^{(i,j)}}{1+k(k-1)} - \frac{1}{2}\right), \qquad k = k_0, \ldots, n, \tag{99}$$

where $R^{(i,j)}_{t_1,t_2}$ is the rank of X_{t_1,t_2} among all $k(k-1)$ random variables in both samples in the j^{th} permutation of the i^{th} simulation.

(b) Select a range of values of t covering the values obtained for $L^{(i,j)}_k$. Calculate the test statistic empirical distribution function (EDF) values for the j^{th} permutation of the i^{th} simulation:

$$\hat{G}^{(i,j)}(t) = \frac{1}{C_n} \sum_{k=k_0}^{n} \frac{1}{k} \mathbb{I}(L^{(i,j)}_k < t), \tag{100}$$

where $C_n = \sum_{k=1}^{n} \frac{1}{k}$.

4. For each value of t, calculate the test statistic EDF values for the i^{th} simulation

$$\hat{G}^{(i)}(t) = \frac{1}{p} \sum_{j=1}^{p} \hat{G}^{(i,j)}(t). \tag{101}$$

5. Estimate the distribution function for each value of t

$$\hat{G}(t) = \frac{1}{S} \sum_{i=1}^{S} \hat{G}^{(i)}(t). \tag{102}$$

6. Compare the graphs of $\hat{G}(t)$ and the distribution function

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\hat{\sigma}^2_{mc}}} e^{(-u^2/2\hat{\sigma}^2_{mc})} du.$$

7. Repeat for larger values of n until the results converge.

5.3.1 EMPIRICAL RESULTS FOR CONVERGENCE

We present the empirical results of the convergence investigation. The verification of a convergence of the ASCLT first requires a convergence in distribution. A sequence of sample sizes $n = (5, 10, 15, 20, 40, 60)$ was used for the Algorithm 1. The number of simulations employed were 1,000, 2,000, 5,000, and 10,000. The Monte Carlo variance estimate $\hat{\sigma}^2_{mc}$ was calculated by taking the sample variance of the values of $L^{(i)}_n$ in Algorithm 1 above. The values of $\hat{\sigma}^2_{mc}$ converge to 0.080 as the number of simulations is increased for $n = 5, 10, 15$ in each sample. In Figure 25 we plot the empirical distribution function for the estimated quantiles from Algorithm 1 and the distribution function at the same quantiles for $\mathcal{N}(0, \hat{\sigma}^2_{mc} = 0.08)$ with 1,000. The plots for 10,000 simulations are provided in Figure 26. The plot for $n = 5$ (10 paired averages per sample) for 1,000 simulations agrees with the plot for $\mathcal{N}(0, 0.080)$, and the results empirically confirm the convergence in distribution of the test statistic.

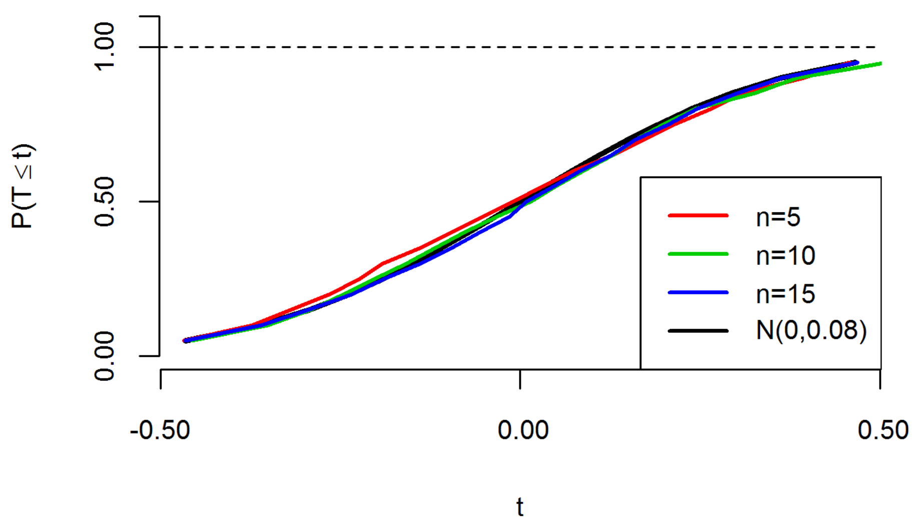

Figure 25: Convergence in distribution $L_n^{(i)}(x)$, 1000 simulations

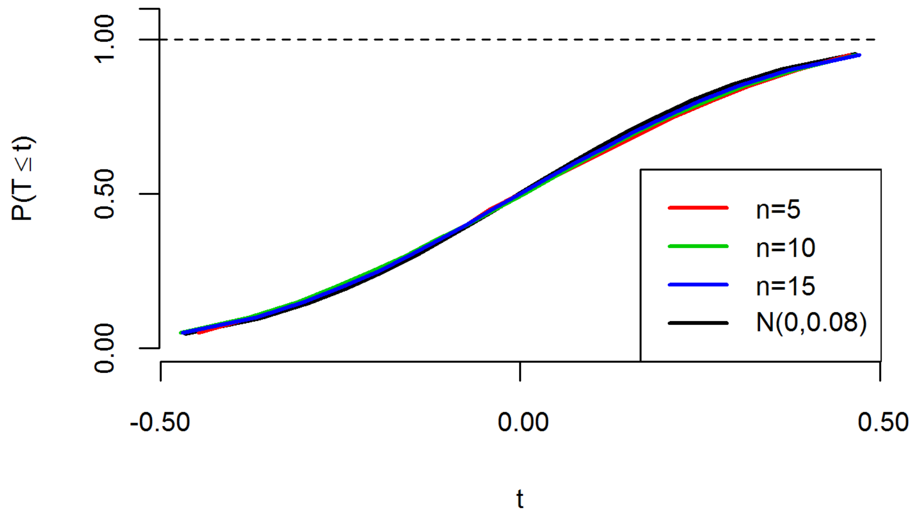

Figure 26: Convergence in distribution $L_n^{(i)}(x)$, 10,000 simulations

For the almost sure convergence, we plotted the values of $\hat{G}(t)$ calculated in Algorithm 2 above for a sequence of quantiles between -0.50 and 0.50 using increments of 0.01 on the same graph containing a plot of the distribution function for $\mathcal{N}(0, \hat{\sigma}^2_{mc} = 0.08)$. Initial results for the convergence of the ASCLT were not as close as desired. We explored improving convergence by starting the calculations with more observations from each sample denoted by k_0 (see Section 2.2 for a detailed explanation). For $n = 5$ (10 paired averages per sample), simulation studies were performed for $k_0 = 2, 3, 4, 5$ and $p = 20, 100, 500$ permutations. Due to the poor results for $k_0 = 6, ..., 10$, the graphs provided only contain profiles for $k_0 \leq 5$. The graphs in Figures 27-29 show the plots of distribution curves of $\hat{G}(t)$ for $n = 5$ observations per sample under $p = 20, 100, 500$ permutations for $k_0 = 2, \ldots, 5$. The profile for $k_0 = 2$ is closest to the curve for the $\mathcal{N}(0, 0.08)$ distribution in all three graphs. Results for $n = 10$ to $n = 25$ for $k_0 = 2, \ldots, 5$ were very similar to those for $n = 5$ and are omitted. In Figures 27-29 it is difficult to determine if larger numbers of permutations improve the convergence, leading to additional investigations.

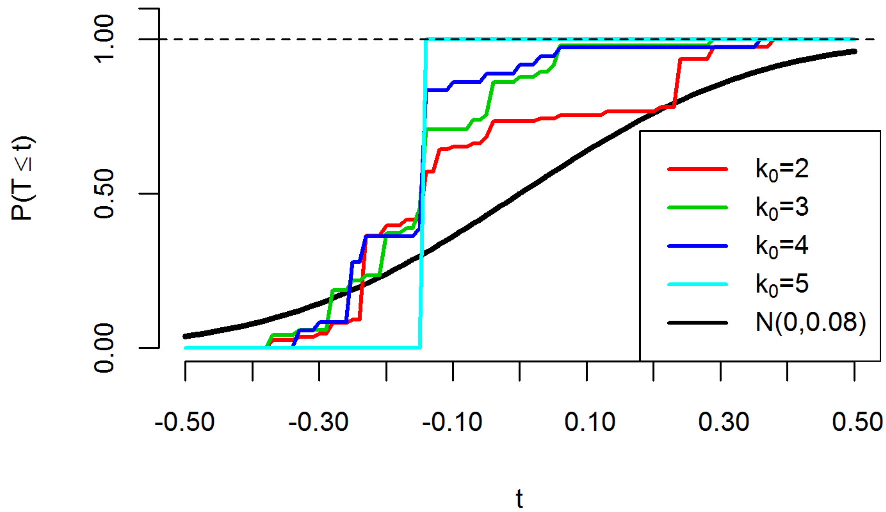

Figure 27: ASCLT convergence of $L_n^{(i)}(x)$, p=20, n=5, $k_0 = 2, \ldots, 5$

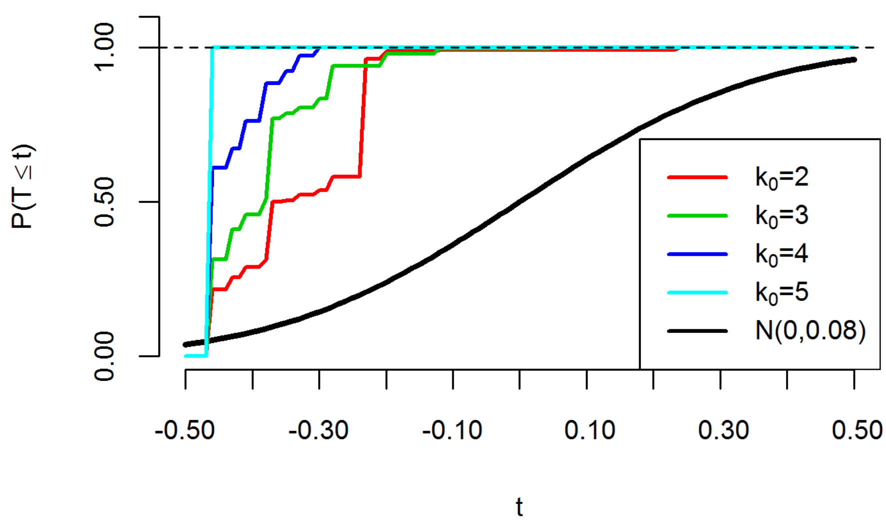

Figure 28: ASCLT convergence of $L_n^{(i)}(x)$, p=100, n=5, $k_0 = 2, \ldots, 5$

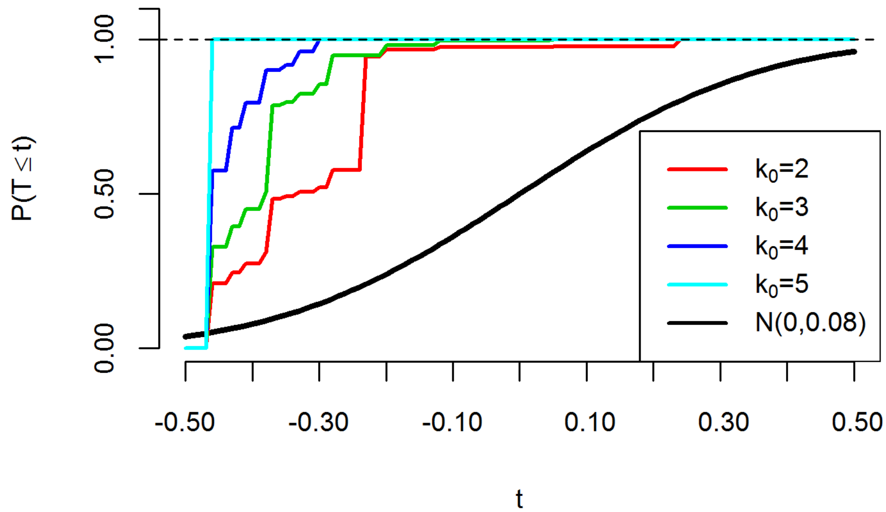

Figure 29: ASCLT convergence of $L_n^{(i)}(x)$,p=500, n=5, $k_0 = 2, \ldots, 5$

The plots in Figure 30 shows the curves of $\hat{G}(t)$ for $k_0 = 2$ and for p=500, 1000, and 10,000 permutations. The curves for 1000 and 10,000 permutations are very similar and are significantly closer to the distribution curve for $\mathcal{N}(0, 0.08)$ than the curve for 500 permutations. This result suggests that 1000 permutations may be needed to obtain satisfactory results for sample sizes of $n = 5$.

Figure 31 contains the results for $n = 10$, $k_0 = 2$ for permutations up to 10,000. The curves for 1,000 and 10,000 are nearly identical, indicating that 1,000 permutations may be needed for adequate analysis. However, the distance of the curves from the target distribution suggests that a larger sample size may be required for valid analyses.

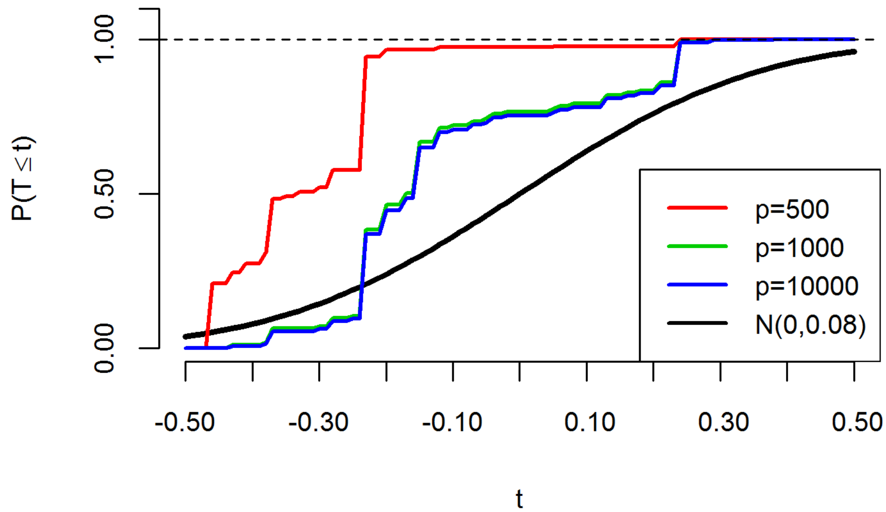

Figure 30: ASCLT convergence of $L_n^{(i)}(x)$, n=5, $k_0 = 2$, p=$500, 1000, 10000$

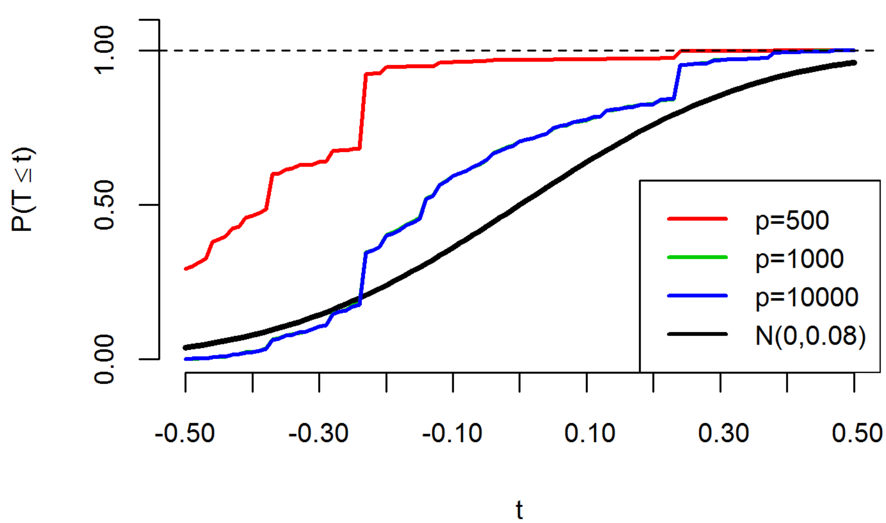

Figure 31: ASCLT convergence of $L_n^{(i)}(x)$, n=10, $k_0 = 2$, p=$500, 1000, 10000$

The results for $n = 15$ (Figure 32) are slightly closer to the target distribution, and the conclusion is the same as that for $n = 10$. The curves for 1,000 and 10,000 are nearly coincident, indicating that 1,000 permutations may be ideal for analyses. However, profile for the 1,000 and 10,000 curves do not appear to be closer to the target distribution than those for $n = 5$, suggesting that $n = 15$ is not a large enough sample size for valid analyses.

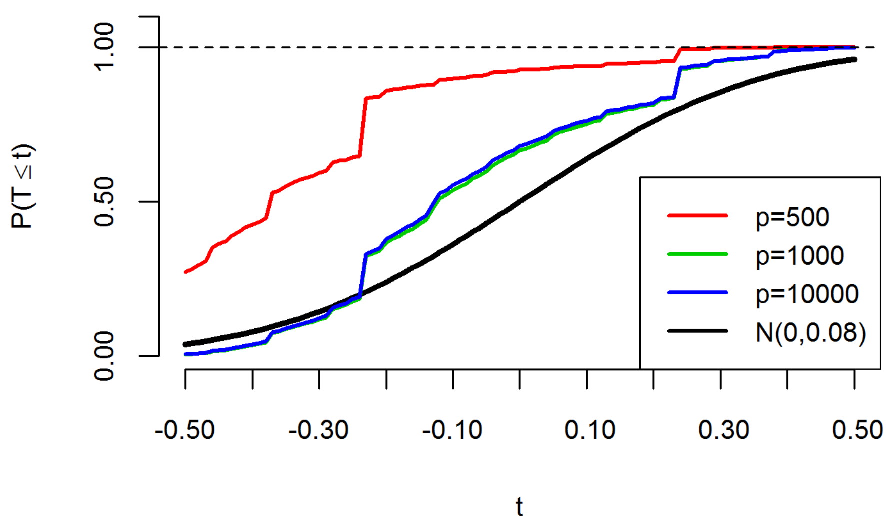

Figure 32: ASCLT convergence of $L_n^{(i)}(x)$, n=15, $k_0 = 2$, p=$500, 1000, 10000$

For $n = 25$, the convergence is also questionable (Figure 33). In fact, the profiles curves are further from the curve for the $\mathcal{N}(0, 0.08)$ distribution than the corresponding profiles for $n = 10$ and $n = 15$, which indicates that $n = 25$ is not a large enough sample size for analysis. An examination of Figure 33 shows that convergence does not improve for permutations over 1,000.

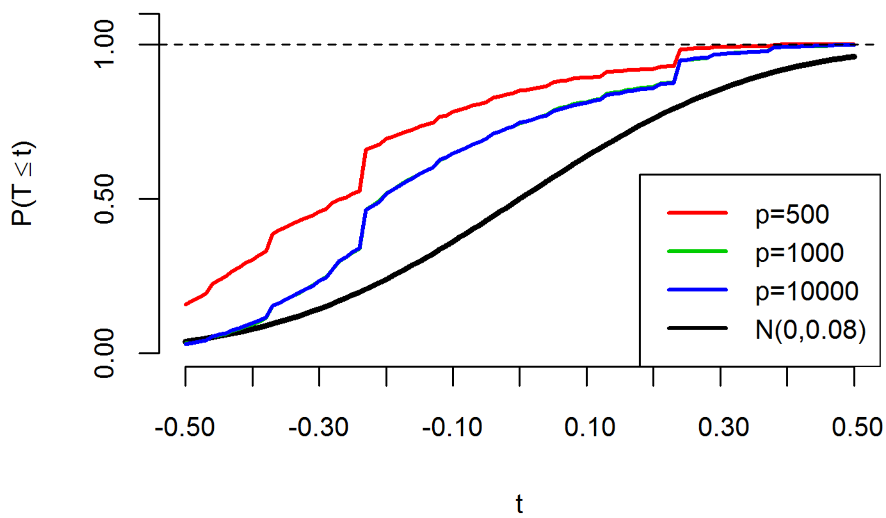

Figure 33: ASCLT convergence of $L_n^{(i)}(x)$, n=25, $k_0 = 2$, p=$500, 1000, 10000$

Figures 34-35 show that the ASCLT is converging to the $\mathcal{N}(0, 0.08)$ distribution function. For $n = 50$, the profiles for 1,000 and 10,000 permutations are similar and indicate that 1000 permutations is sufficient for valid analysis. In Figure 35, the profiles for 500, 1,000, and 10,000 permutations are approximately coincident, which suggests that 500 permutations may provide satisfactory analyses.

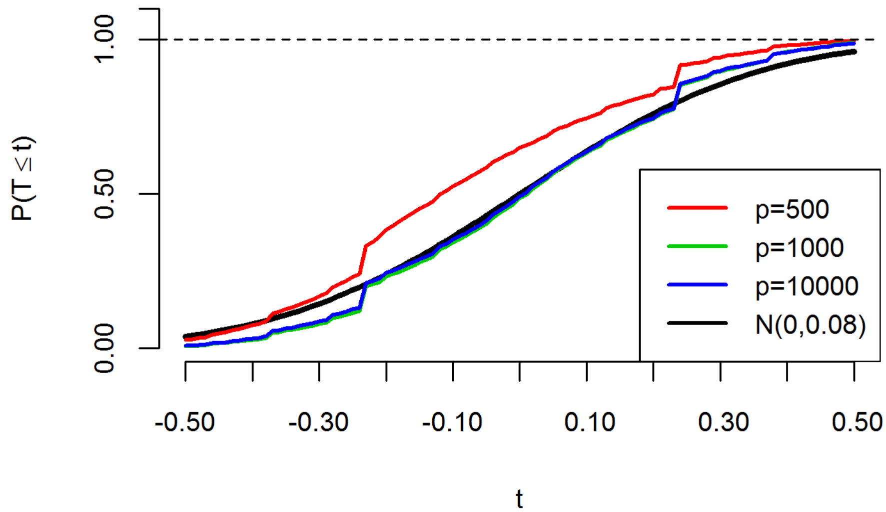

Figure 34: ASCLT convergence of $L_n^{(i)}(x)$, n=50, $k_0 = 2$, p=$500, 1000, 10000$

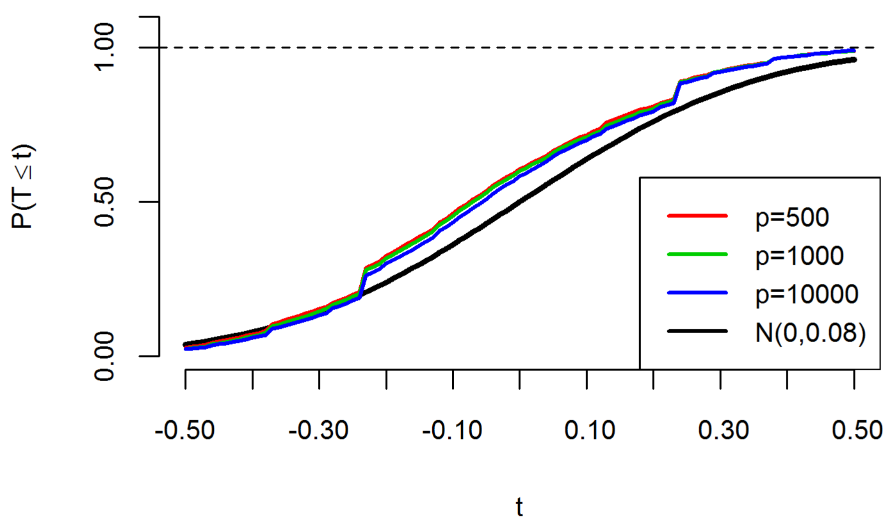

Figure 35: ASCLT convergence of $L_n^{(i)}(x)$, n=100, $k_0 = 2$, p=$500, 1000, 10000$

We have confirmed convergence for sample sizes $n \geq 50$. We now explore the rate of convergence for larger samples ($n = 1000, 2000, 3000$) in Figures 36-41 in order to determine the minimum number of permutations needed. The idea is that larger sample sizes should require lower numbers of permutations to achieve similar results to those for smaller sample sizes. Various values for the initial number of observations k_0 used for calculating the test statistic in its ASCLT. The results for $k_0 = 2$ and $k_0 = 20$ are provided.

Without permutations, the convergence is very slow. Viewing Figures 36-37 alone would not provide confidence in the convergence. Recall (Section 1.1) that because we are using only one sequence of data (or sample) not multiple sequences (samples), the random order of the observations can have a significant effect on the rate of convergence in the absence of permutations. The profiles for $k_0 = 2$ in Figure 36 are closer to the target distribution than those for $k_0 = 20$ in Figure 37.

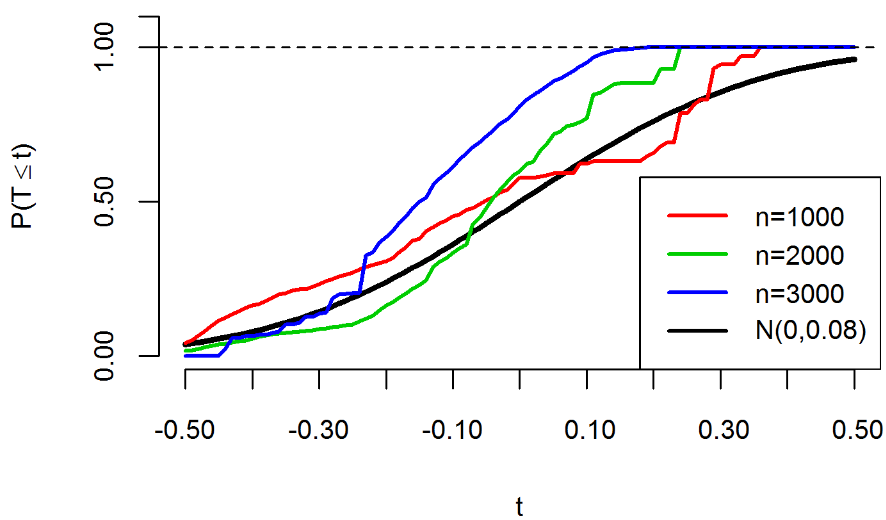

Figure 36: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 2$ (no permutations)

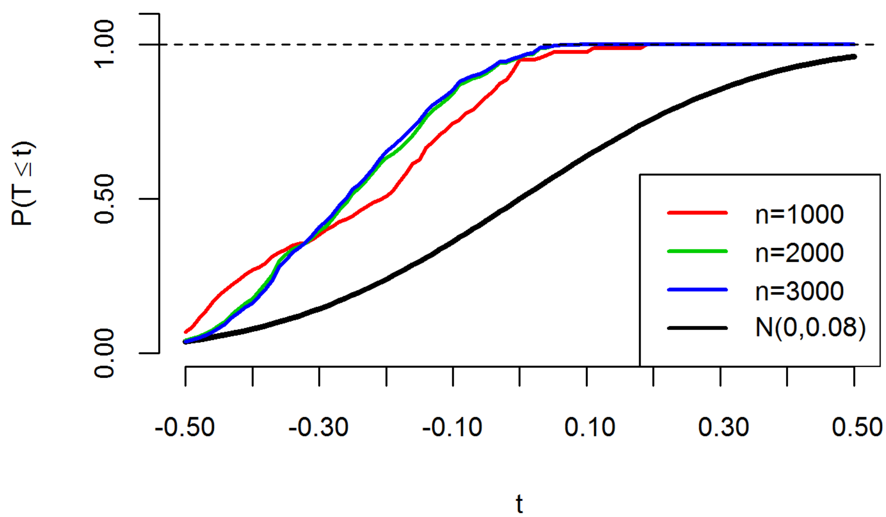

Figure 37: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 20$ (no permutations)

Adding 20 permutations significantly improves the rate of convergence (Figures 38-39). The curves for $k_0 = 2$ are closer to the $\mathcal{N}(0, 0.08)$ distribution curve than the curves for $k_0 = 20$. The profile for $n = 1000$ is not close to the target curve, suggesting that more permutations are required for valid analyses.

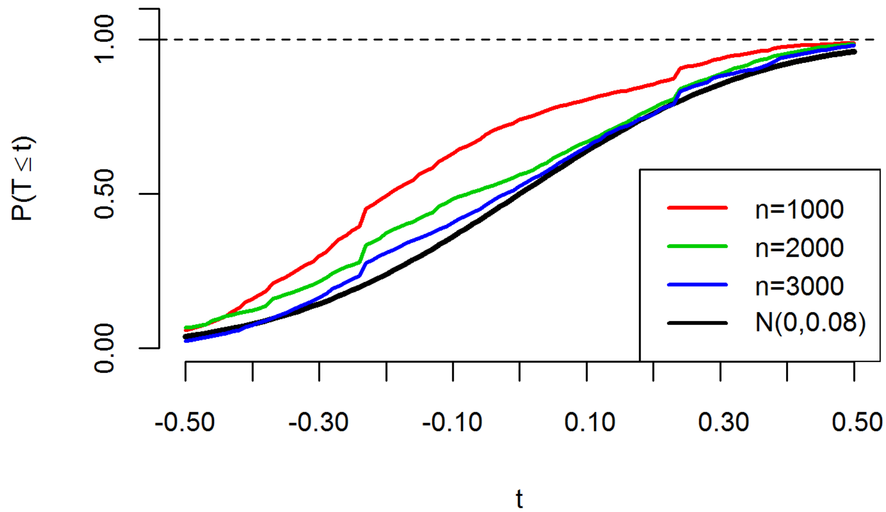

Figure 38: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 2$ (20 permutations)

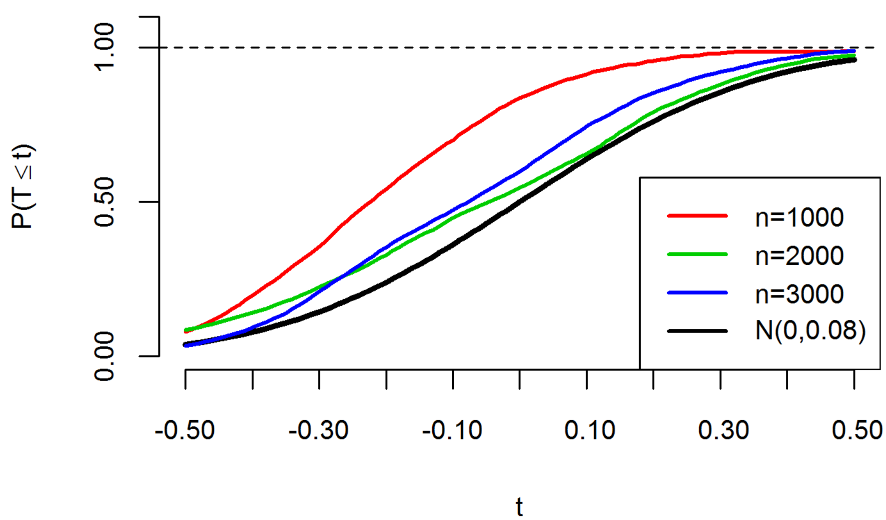

Figure 39: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 20$ (20 permutations)

Increasing the number of permutations to 50 (Figures 40-41) improves the convergence rate as anticipated. The profile curve for all three sample sizes are close to the target normal distribution curve in both tails for $k_0 = 2$. The minimum number of permutations that may be required for adequate test results is 50. The profiles for all three sample sizes using $k_0 = 2$ are closest to the $\mathcal{N}(0, 0.08)$ distribution function profile curve for all the permutation levels all the values of k_0 investigated.

Simulation studies for larger permutation quantities are an on-going endeavor as the computational time for such large quantities $\binom{n}{2}$ is very slow. Simulation studies for unequal variances (Behrens-Fisher Problem) remain an open problem and a next step.

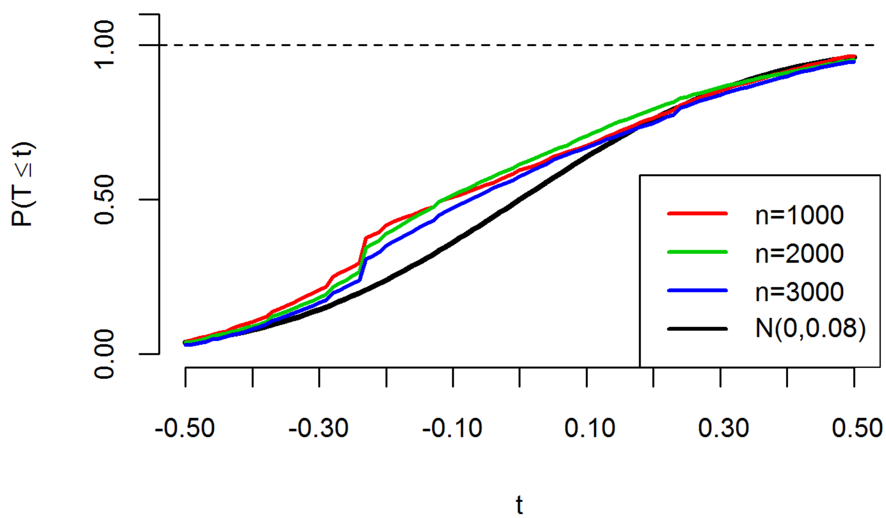

Figure 40: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 2$ (50 permutations)

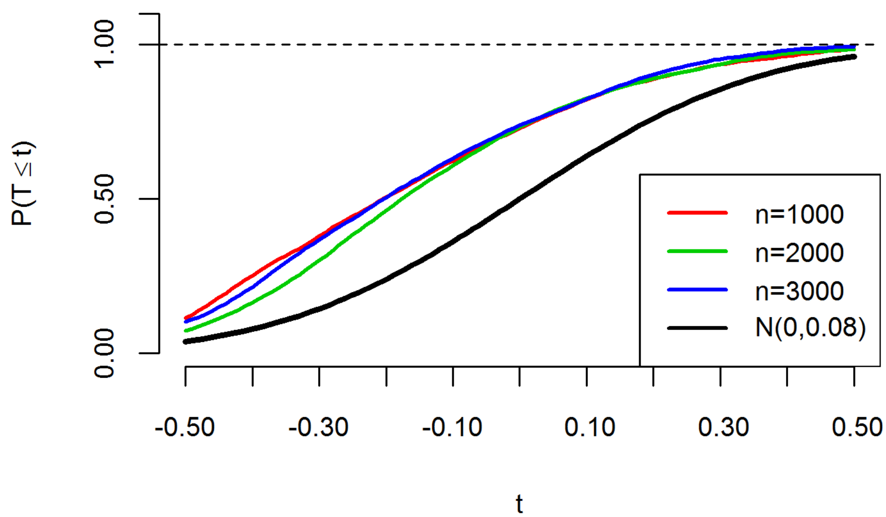

Figure 41: ASCLT convergence of $L_n^{(i)}(x)$, $k_0 = 20$ (50 permutations)

5.3.2 SIGNIFICANCE LEVEL AND POWER

Using the convergence results from Section 5.3.1, we explore the type I error for sample sizes $n = 5, 10, 15, 25, 50, 100$. We apply 1000 permutations for $n \leq 50$ and 500 permutations for $n = 100$. The type I error results will determine which analyses are performed for power. Table 37 contains the simulation results for variables generated from the $\mathcal{N}(0, 1)$. Recall that for small sample sizes, the curves for the EDF contain sharp jumps and obtain the value of one for much lower quantile values compared to those of the distribution of $\mathcal{N}(0, 0.08)$; the results for $n = 5, 10, 15$ reflect this behavior. The type I error values for $n = 50, 100$ are conservative at the 10% level and strongly conservative at the 5% and 1% levels, indicating that larger sample sizes may be needed to obtain satisfactory results.

Table 37: Simulated type I error for the two-sample problem $\mathcal{N}(0,1)$

n	p	k_0	$\mathcal{N}(0,1)$ Level 10%	5%	1%
5	500	2	0.117	0.117	0.117
	1000		0.117	0.117	0.117
10	500	2	0.024	0.024	0.024
	1000		0.024	0.024	0.024
15	500	2	0.047	0.012	0.012
	1000		0.043	0.012	0.012
25	500	2	0.045	0.006	0.000
	1000		0.045	0.006	0.000
50	500	2	0.042	0.012	0.000
	1000		0.040	0.013	0.000
100	500	2	0.044	0.007	0.000
	1000		0.045	0.006	0.000

Note: k_0 is the minimum number of observations used to calculate $L_k^{(i,j)}$ in (99). Values were obtained from 1000 simulations, where p denotes the number of permutations of each simulation.

Table 38 contains the power for normally distributed data where the difference in means between the samples is $\delta = 1$. The power was obtained from 1000 simulations with p=500 permutations each for sample sizes $n = 5, 10, 15, 25, 50, 100$ and calculations were started at $k_0 = 2$ observations per sample. The power for $n \leq 25$ is lower than the type I error, indicating that satisfactory results may not be obtainable with the proposed test statistic for these small values of n. The power for $n = 50, 100$ is close to one for the 10% level is above 0.85 for the 5% level. It appears that larger sample sizes are needed to obtain an acceptable power at the 1% level.

Table 38: Simulated power for the two-sample problem

		$\mathcal{N}(0,1)$, $\mathcal{N}(1,1)$ (Level)		
n	N	0.10	0.05	0.01
5	10	0.003	0.003	0.003
10	20	0.000	0.000	0.000
15	30	0.215	0.000	0.000
25	50	0.767	0.022	0.000
50	100	0.992	0.872	0.000
100	200	1.000	1.000	0.001

Note: $k_0 = 2$ is the minimum number of observations used to calculate $L_k^{(i,j)}$. Values were obtained from p=1000 permutations of 1000 simulations.

5.4 SUMMARY

In this chapter, we investigated a new approach to increase the effective sample size in the two-sample problem with LQE. We introduced a modification to the test statistic in Denker and Puri (1992). Due to the open form for the estimate of the asymptotic variance of the linear rank statistic, we proposed using the upper bound of the variance given in Denker and Puri (1992) in its place. We hypothesized that the new test statistic converges to a normal distribution with some unknown variance. Empirically we verified the convergence in distribution for this new statistic to a $\mathcal{N}(0, 0.08)$ distribution, and we verified the convergence of its corresponding ASCLT to the same distribution. Simulation results indicate that p=1000 permutations are needed for stable results. Type I error for sample sizes $n = 50, 100$ from standard normal distributions was conservative at the 10% level and very conservative at the 5% and 1% levels. For power, we simulated the data from normally distributed populations with a difference in means of one. The power for $n = 50, 100$ was well above 0.8 for the 10% and 5% levels. It appears that a larger sample size is needed for adequate power at the 1% level. Analysis for this problem is on-going. Some future next steps ar provided in Section 6.1.

CHAPTER 6

CONCLUSION

In this dissertation, we introduced a new approach called logarithmic quantiles estimation for analyzing three different types of statistical problems in a nonparametric setting.

In Chapter 3 we investigated three factorial for the presence of a pattern (trend, umbrella, etc.) across the levels of one of the factors under the LQE approach. The models investigated were a two-way fixed effects model, a partial hierarchical repeated measures model, and a cross-classification repeated measures model. The presence of an umbrella pattern in the factor for cycle time was confirmed in the two- factor fixed effect study of the quality of shirts in a manufacturing process. A study for the effectiveness of two drugs in patients diagnosed with AIDS had a partial hierarchical model and showed strong evidence of a decreasing trend across time when interaction between drug treatment and time was considered. The α-Amylase study introduced by Akritas and Brunner (1996) was analyzed with a cross-classification repeated measures model and the test for an umbrella pattern in the α-Amylase levels across the time of day agreed with the results of Akritas and Brunner (1996). The type I error for the three models were conservative; however, high levels of power were achieved for reasonable alternatives.

LQE for the change-point problem was studied in Chapter 4 for the test proposed by Pettitt (1979) for small to moderate sample sizes. We analyzed several small datasets and compared their results to existing results where other analyses were available. The LQE approach to the Pettitt test was liberal for smaller sample sizes but the type I error approaches the significance level as the sample sizes increase. High levels of power were obtained for relatively small sample sizes when a shift in location of 0.75 occurred near the middle of the sequence of values.

A new approach for the two-sample problem was proposed in Chapter 5. We pursued the ideas in Compagnone and Denker (1996) for increasing the efficiency of the nonparametric Wilcoxon-Mann-Whitney test (Wilcoxon, 1945 and Mann and Whitney, 1947) with respect to the parametric t-test using an empirical process of U-statistic structure. We proposed a new test statistic based on a linear rank statistic using a second order empirical U-statistic process to generate paired averages within each independent sample divided by the upper bound for the asymptotic variance. The process effectively increased the sample size from n to $\binom{n}{2}$. We empirically confirmed the convergence in distribution of the new statistic, and the convergence of the corresponding ASCLT for samples sizes $n \geq 50$. Simulated type I error and power were provided for small to moderate independent samples ($n \leq 100$) of normally distributed random variables. The type I error results were conservative. The power for $n = 50, 100$ at the 5% and 10% levels exceed 0.8 for a difference in means of one.

The use of the LQE approach in the three different statistical problems above significantly expands the body of knowledge for LQE. The results obtained in this dissertation confirm the potential viability of LQE for many additional analyses. We include some open problems and areas for future exploration in the following section.

6.1 FUTURE WORK AND OPEN PROBLEMS

We discuss several open problems and topics for future investigations under the LQE approach. The conservativeness of LQE in nonparametric models has been an open problem since its introduction by Thangavelu (2005). Another open issue involves addressing the liberal results of the model for the change-point problem. Change-point and time series are two areas that are quickly advancing to keep abreast of the need to analyze large data streams generated in this era of rapid technological advancement. The LQE approach may be suited to address both change-point problems and time series analyses for large datasets.

We continue to explore the two-sample problem introduced in Chapter 5. Our next steps include investigating more efficient computational approaches, and exploring modifications to the test statistic to improve type I error and power in smaller sample sizes. We will also consider testing for equal variances of the two-samples using the test proposed in Compagnone and Denker (1996) with the LQE approach.

The field of LQE is relatively new, and the results of this dissertation begin to reveal the potential for its extension to many statistical areas.

REFERENCES

Abrams, D. I., Goldman, A. I., Launer, C., Korvick, J. A., Neaton, J. D., Crane, L. R., et al. (1994). A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 330(10), 657-662.

Akritas, M. G., and Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measures designs. *Journal of the American Statistical Association*, 89(425), 336-343.

Akritas, M. G., Arnold, S. F., and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. *Journal of the American Statistical Association*, 92(437), 258-265.

Akritas, M. G., and Brunner, E. (1997). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61(2), 249-277.

Akritas, M. G., and Brunner, E. (1996). Rank tests for patterned alternatives in factorial designs with interactions. *Festschrift on the Occasion of the 65th Birthday of Madam L. Puri*, VSP-International Science Publishers, Utrecht, The Netherlands, 277-288.

Antoch, J., and Hušková, M. (2001). Permutation tests in change point analysis. *Statistics and Probability Letters*, 53(1), 37-46.

Baumgartner, W., Weiß, P., and Schindler, H. (1998). A nonparametric test for the general two-sample problem. *Biometrics*, 1129-1135.

Bercu, B., Cnac, P., and Fayolle, G. (2009). On the almost sure central limit theorem for vector martingales: convergence of moments and statistical applications. *Journal of Applied Probability*, 46(1), 151-169.

Berkes, I. (1998). Results and problems related to the pointwise central limit theorem. *Asymptotic Methods in Probability and Statistics* ,pp. 59-96.

Berkes, I., Horváth, L., Hušková, M., Steinebach, J. (2004). Applications of permutations to the simulations of critical values, *Nonparametric Statistics*, Vol. 16 (1-2), 197-216.

Berkes, I., and Csáki, E. (2001). A universal result in almost sure central limit theory. *Stochastic Processes and Their Applications*, 94(1), 105-134.

Berkes, I., and Dehling, H. (1993). Some limit theorems in log density. *The Annals of Probability*, 1640-1670.

Billingsley, P. (1968). *Convergence of probability measures*, Wiley.

Brosamler, G. A. (1988). An almost everywhere central limit theorem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(3), 561-574.

Brunner, E. and Denker, M. (1994). Rank statistics under dependent observations and applications to factorial designs. *Journal of Statistical Planning and Inference*, 42, 353-378.

Brunner, E., Domhof, S., and Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial designs*, Wiley Series in Probability and Statistics.

Brunner, E., Konietschke, F., Pauly, M., and Puri, M. L. (2017). Rank–based procedures in factorial designs: hypotheses about non-parametric treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5), 1463-1485.

Brunner, E. and Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and small-sample approximation. *Biometrical Journal*, 42, 17-23.

Brunner, E. and Neumann, N. (1986). Two-sample rank tests in general models. *Biometrical Journal*, 28(4), 395-402.

Brunner, E., and Puri, M. L. (1996). 19 Nonparametric methods in design and analysis of experiments. *Handbook of statistics, 13*, 631-703.

Brunner, E., and Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical papers*, 42(1), 1-52.

Callegari, F., and Akritas, M. G. (2004). Rank tests for patterned alternatives in two-way non-parametric analysis of variance. *Journal of Statistical Planning and Inference*, 126(1), 1-23.

Cario, M. C., and Nelson, B. L. (1997). *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix.* Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Cheng, S., Peng, L., and Qi, Y. (1998). Almost sure convergence in extreme value theory. *Mathematische Nachrichten*, 190(1), 43-50.

Compagnone, D., and Denker, M. (1996). Nonparametric tests for scale and location. *Journal of Nonparametric Statistics*, 7(2), 123-154.

Denker, M., and Fridline, M. (2010). The almost sure version of Cramér's theorem. *Dependence in probability, analysis and number theory*, Kendrick Press, Heber City, UT, 195-201.

Denker, M., and Puri, M. L. (1992). Asymptotic normality of two sample linear rank statistics under U-statistic structure. *Journal of Statistical Planning and Inference*, 32(1), 89-110.

Denker, M., and Tabacu, L. (2014). Testing longitudinal data by logarithmic quantiles. *Electronic Journal of Statistics*, 8(2), 2937-2952.

Denker, M., and Tabacu, L. (2015). Logarithmic quantile estimation for rank statistics. *Journal of Statistical Theory and Practice*, 9(1), 146-170.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1-26.

Fahrner, I., and Stadtmüller, U. (1998). On almost sure max-limit theorems. *Statistics and Probability Letters*, 37(3), 229-236.

Ferger, D.(1994). On the power of nonparametric changepoint-tests. *Metrika*, 41, 277-292.

Fisher, A. (1987). Convex-invariant means and a pathwise central limit theorem. *Advances in Mathematics*, 63(3), 213-246.

Fridline, M. M. (2010). Almost sure confidence intervals for the correlation coefficient, *Doctoral dissertation*, Case Western Reserve University.

Gombay, E. (1994). Testing for change-points with rank and sign statistics. *Statistics and Probability Letters*, 20(1), 49-55.

Gombay, E., Horváth, L. (1999). Change-points and bootstrap, *Environmetrics*, 10, 725-736.

Hall, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science and Business Media.

Hettmansperger, T. P., and Norton, R. M. (1987). Tests for patterned alternatives in k-sample problems. *Journal of the American Statistical Association*, 82(397), 292-299.

Hinkley, D.V. (1970). About the change-point in a sequence of random variables. *Biometrika*, 57(1), 1-17.

Holzmann, H., Koch, S., and Min, A. (2004). Almost sure limit theorems for U-statistics. *Statistics and probability letters*, 69(3), 261-269.

Hörmann, S. (2005). Optimal averaging procedures in almost sure central limit theory. *Metodoloski Zvezki*, 2(2), 271.

Hörmann, S. (2007). Critical behavior in almost sure central limit theory. *Journal of Theoretical Probability*, 20(3), 613-636.

Horváth, L., Hušková, M. (2005). Testing for changes using permutations of U-statistics, *Journal of Statistical Planning and Inference*, 128, 351-371.

Hušková, M. (2004). Permutation principle and bootstrap in change point analysis, *Fields Institute Communications*, Vol. 44, 273-291.

Hušková, M., Slabý, A. (2001). Permutation tests for multiple changes, *Kybernetika*, Vol. 37, No. 5, 605-622.

Ibragimov, A., and Lifshits, M. A. (2000). On almost sure limit theorems. *Theory of Probability and Its Applications*, 44(2), 254-272.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2), 133-145.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81-93.

Kirch, C., Steinebach, J. (2006). Permutation principles for the change analysis of stochastic processes under strong invariance, *Journal of Computational and Applied Mathematics*, 186, 64-88.

Koziol, J. A. (1987). A note on non-parametric tests for the change-point problem, *Biometrical Journal 29*, 791-797.

Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. emphJournal of the American Statistical Association, 47(260), 583-621.

Lacey, M.T., and Philipp, W. (1990). A note on the almost sure central limit theorem. *Statistics and Probability Letters*, 9(3), 201-205.

Lehmann, E. L. (1953). The power of rank tests. *The Annals of Mathematical Statistics*, 23-43.

Lifshits, M. A. (2001). Lecture notes on almost sure limit theorems. *Publications IRMA, Lille*, 54, No. 8, 1-23.

Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74, 615-624.

Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50-60.

McGilchrist, C. A., and Woodyer, K. D. (1975). Note on a distribution-free CUSUM technique. *Technometrics*, 17(3), 321-325.

Miao, Y., Wang, R., and Adler, A. (2016). Limit theorems for order statistics from exponentials. *Statistics and Probability Letters*, 110, 51-57.

Lafaye de Micheaux, P., and Liquet, B. (2009). Understanding convergence concepts: A visual-minded and graphical simulation-based approach. *The American Statistician*, 63(2), 173-178.

Mises, R. v. (1947). On the asymptotic distribution of differentiable statistical functions. *The annals of mathematical statistics*, 18(3), 309-348.

Montgomery, D. C. (2013). *Design and analysis of experiments*, New York: J Wiley and Sons.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100-115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4), 523-527.

Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1-2), 248-252.

Peligrad, M. and Révész, P. (1991). On the almost sure central limit theorem.*Almost Everywhere Convergence II*, A. Bellow and R. Jones, eds., Academc Press, New York, 209-225.

Peligrad, M., and Shao, Q. M. (1995). A note on the almost sure central limit theorem for weakly dependent random variables. *Statistics and Probability Letters*, 22(2), 131-136.

Peng, Z., Li, J., and Nadarajah, S. (2009). Almost sure convergence of extreme order statistics. *Electronic Journal of Statistics*, 3, 546-556.

Peng, L., and Qi, Y. (2003). Almost sure convergence of the distributional limit theorem for order statistics. *Probability and Mathematical Statistics-Wroclaw University*, 23(2), 217-228.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics*, 126-135.

Rizzo, Maria L. (2008) *Statistical Computing with R*, Chapman and Hall/CRC.

Schatte, P. (1988). On strong versions of the central limit theorem. *Mathematische Nachrichten*, 137(1), 249-256.

Sen, A., and Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, 98-108.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. emphBulletin Mathématique de L'Université de Moscow, 2(2), 3-14.

Tabacu, L. (2014). Logarithmic Quantile Estimation and Its Applications to Nonparametric Factorial Designs, *Doctoral dissertation*, The Pennsylvania State University.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagations Math.*, 14, 327-333.

Thangavelu, K. (2005). Quantile estimation based on the almost sure central limit theorem, *Doctoral dissertation*, Dissertation, University of Göttingen, Germany.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6), 80-83.

VITA

Mark Ledbetter
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D.	Old Dominion University, Norfolk, VA (December 2018). Major: Computational and Applied Mathematics (Statistics).
MS	Old Dominion University, Norfolk, VA (December 2014). Major: Computational and Applied Mathematics (Statistics).
BS	Charter Oak State College, New Britain, CT. (July 1998) Major: Liberal Studies (Physics and Mathematics)

Relevant Experience

Adjunct Instructor, Department of Mathematics, Thomas Nelson Community College, Hampton, Virginia (09/2017 - present)

Graduate Teaching Assistant Instructor, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA, (01/2015 - 5/2018).

Graduate Assistant, Peer Educator Program, Old Dominion University, Norfolk, VA, (08/2013 - 12/2013).

Typeset using LaTeX.