Applications of Random Rotations in $\mathbb{R}^n$ to test for serial correlation in time series data

by

Amitakshar Biswas

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences
University of Alberta

© Amitakshar Biswas, 2024

Abstract

Motivated by Elizabeth Meckes' work on concentration inequalities using the group $\mathbb{SO}(n)$ (see [9]), this thesis explores the use of random rotations for detecting autocorrelation in time series data. Traditional tests like the Durbin-Watson test assess autocorrelation by analyzing quadratic forms of residual vectors. Our new approach uses concentration inequalities associated with uniform measures on $\mathbb{SO}(n)$ to construct a test without strong distributional assumptions.

We propose a new test statistic for autocorrelation in AR(1) processes, utilizing random rotations of sample autocorrelation function. By establishing a subgaussian concentration inequality, we derive a one-sided test with an upper bound for the p-value. Further refinement through beta adjustment enhances our p-value accuracy. We also generalize it to a two-sided test.

This thesis provides comprehensive background material on group invariance, random rotations, AR(1) processes, and the Durbin-Watson test. Key results, including the concentration inequality proof and beta transformation application, are presented in detail. A simulation study confirms the effectiveness of our rotation test, highlighting its potential in practical statistical applications.

Preface

I intend to publish the results of my thesis with my supervisor Prof. Adam Kashlak. My contributions were primarily behind the formulation of the lemmas and theorems, while the key ideas belonged to my supervisor.

To my Maa, Baba and Didibhai,

For everything.

Acknowledgements

I extend my heartfelt appreciation to Prof. Adam Kashlak, my supervisor, whose unwavering support and guidance have been pivotal throughout my thesis journey. His expertise and encouragement have significantly shaped the trajectory of my research, and I am truly grateful for his mentorship.

I also want to express my gratitude to Prof. Rohana Karunamuni and Prof. Yaozhong Hu for their exceptional teaching in Statistics and Probability. Their clarity and dedication have played a crucial role in enhancing my understanding of these subjects and have greatly contributed to the quality of my research.

A special thank you goes out to my family, my parents and sister, for their continuous love, encouragement, and understanding.

I am indebted to the University of Alberta for providing an enriching academic environment. The university's commitment to research excellence, coupled with the availability of ample resources and funding, has created an ideal atmosphere for my research pursuits.

This thesis is a collective achievement, and I am deeply grateful to each individual mentioned for their unique contributions. Together, they have played a crucial role in my academic and personal development.

Table of Contents

List of Figures

Chapter 1

Introduction

Statistical hypothesis testing has always been a hard problem. Typically, parametric approaches that assume normality, are used to simplify the problem. On the other hand, nonparametric tests, while requiring fewer assumptions, are often challenging, both theoretically and practically. Within the realm of nonparametric tests, a class known as randomization tests offers an unique approach by applying random group actions to a dataset and assessing hypotheses based on the structure of the group, such as symmetry.

One particularly notable group in this context is the permutation group, $\mathbb{S}_n$. The utilization of this group leads to a well-known test called Fisher's Permutation test. Extensive prior research has been conducted on permutation tests (see [2]), as well as on concentration of measure (see [1], [4] and [5]). However, the applications of random rotations to statistical hypothesis testing remains an understudied area, mainly due to computational limitations.

The main motivation behind using random rotations came from the book, 'The Random Matrix Theory of the Classical Compact Groups' by Elizabeth

Meckes ([9]). In Chapter 5 of that book, there's a handy concentration inequality using the group $\mathbb{SO}(n)$ (see Theorem 2.1.5), which helps us get a bound for our p-value.

After the fitting of a linear model to a dataset, it is often desirable to check for autoregressive correlations amidst the residual vector. Several tests exist for assessing serial correlation, such as the well-known Durbin-Watson test. These tests utilize test statistics that involve ratios of quadratic forms of the residual vector. The objective is to determine whether the residuals exhibit statistical characteristics resembling white noise.

We've developed a new test for detecting serial correlation, utilizing the concentration inequalities associated with uniform measures on $\mathbb{SO}(n)$. By leveraging these concentration inequalities, we constructed a robust test that does not rely on strong assumptions about the underlying data distribution.

We considered an AR(1) process, and introduced a new test statistic which randomly rotates the sample autocorrelation function. Then, using established results, we derived a robust subgaussian concentration inequality (see Theorem 3.1.5) related to this test statistic. This enabled us to devise a one-sided test and obtain an upper bound for the associated p-value. Subsequently, we harnessed mathematical techniques, precisely, the beta adjustment, to refine our p-value estimation (see Section 3.2). Upon comparative analysis with the well-known Durbin-Watson test, we observed that our test demonstrates comparable performance, closely resembling that of the DW p-value.

Chapter 2 delves into the essential background material and literature review crucial for understanding the context of the thesis. It initiates with the exploration of Group Invariance, a foundational concept central to the the-

sis. Additionally, the chapter delves into the discussion of invariance in the asymptotic sense when certain condition is not satisfied. A concise overview is provided on topics such as random rotations, the AR(1) process, and the Durbin-Watson test.

Moving on to Chapter 3, it unfolds the core findings of the thesis. The chapter encompasses the proof of key results necessary for deriving the concentration inequality. A notable application of the proposed beta transformation is presented, contributing to the enhancement of p-value accuracy. The chapter culminates with a simulation study, showcasing the performance of our rotation test against the Durbin-Watson test.

Chapter 2

Background Material

In this chapter, we'll cover some important background information that will be handy in Chapter 3. We'll begin by looking into Group Invariance, Random Rotations and then dive into the well-known Durbin-Watson test.

2.1 Group Invariance

Mathematically, achieving certain symmetries asymptotically means that the object maintains its invariance under a set of group actions. For example, the standard Gaussian measure in $\mathbb{R}^n$ is invariant under the orthogonal group $O(n)$. Based on [4], we first setup the general framework. Let

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- $(H, \mathcal{H})$ be a Hilbert space space equipped with the Borel σ-field generated by the norm topology, $\mathcal{H}$.
- $(\mathbb{R}, \mathcal{B})$ denote the real line with the standard Borel σ-field.

- $X : \Omega \to H$ be an H-valued random variable, and $T : H \to \mathbb{R}$ be a measurable function.

- Let G be a compact topological group, i.e., mappings, $(g, h) \to gh$, and $g \to g^{-1}$ are continuous.

We consider G to be equipped with the normalized Haar measure ρ, which is the unique left-translation-invariant measure, i.e.,

$$\rho(gS) = \rho(S) \tag{2.1}$$

for every $g \in G$ and all Borel sets, $S \subseteq G$. We consider the map, $\pi : G \to \mathcal{L}(H)$, where $\mathcal{L}(H)$ is the space of H-endomorphisms with the strong operator topology, defined by, $\pi(g) = \pi_g$ where $\pi_g \in \mathcal{L}(H)$. The Hilbert space H is called an Hilbert G-Module. A set $S \in \mathcal{H}$ is said to be G-invariant if $S = \pi_g S$ for all $g \in G$. Let $\mathcal{S}$ be the collection of all G-invariant sets and $\mathcal{T} = \sigma(T)$.

In the context of hypothesis testing, we impose the following condition: *The measure on* $\mathbb{R}$ *induced by the mapping* $T(X)$ *is G-invariant, i.e., for all* $g \in G$ *and* $B \in \mathcal{B}$,

$$\mathbb{P}(T(X) \in B) = \mathbb{P}(T(\pi_g X) \in B) \tag{2.2}$$

We also define a randomization threshold $t_\alpha(x)$ for a fixed $x \in H$ and $\alpha \in (0, 1)$ by,

$$t_\alpha(x) = \inf\{t \in \mathbb{R} : \rho(\{g \in G : T(\pi_g x) > t\}) \leq \alpha\} \tag{2.3}$$

Theorem 2.1.1. *Under the above condition,* $\mathbb{P}(T(X) \in B|\mathcal{S} \cap \mathcal{T}) = \rho(\{g \in G : T(\pi_g X) \in B\})$*, P-almost surely for any* $B \in \mathcal{B}$*.*

Corollary 2.1.2. *Under the above condition,* $\mathbb{P}(T(X) > t_\alpha(X)) \leq \alpha$*.*

Proof. Define $R = \{x \in H : T(x) > t_\alpha(x)\}$. Then, $R \in \mathcal{H}$ and from Theorem 2.1.1,

$$\begin{aligned}\mathbb{P}(T(X) > t_\alpha(X)) &= \mathbb{E}[\mathbb{P}(T(X) > t_\alpha(X)|\mathcal{S} \cap \mathcal{T})] \\ &= \mathbb{E}[\rho(\{g \in G : T(\pi_g X) > t_\alpha(X)\})] \leq \alpha\end{aligned}$$

as almost sure equality implies equality in mean. □

The corollary essentially implies that by selecting a significance threshold t_α based on the group G and its Haar measure ρ, we can devise a statistical test at size α significance level.

The choice of group G is critical. For example, if $G = \{e\}$, the trivial group, then the theorem says,

$$\mathbb{1}_{T(X)\in B} = \begin{cases} 1 & \text{if } T(X) \in B \\ 0 & \text{if } T(X) \notin B \end{cases} \tag{2.4}$$

2.1.1 Asymptotic Group Invariance

For Corollary 2.1.2 to hold true, the aforementioned condition must be satisfied (Equation 2.2). If this condition is removed, almost sure equality is lost, although it can still be approached asymptotically.

Let $H = \mathbb{R}^\infty$ and for all $n \in \mathbb{N}$, $\mathcal{S}_n$ be the σ-field of G_n-invariant sets in

$\mathbb{R}^n$. Let $T_n : \mathbb{R}^n \to \mathbb{R}$ be c_n-Lipschitz and $\mathcal{T}_n$ be the smallest σ-field on $\mathbb{R}^n$ such that T_n is measurable.

Functions T_n, groups G_n, and sets S_n can be extended to $\mathbb{R}^\infty$. For each set $S_n \in \mathcal{S}_n$, we can express it as $\{S_n \otimes \mathbb{R} \otimes \ldots\} \subset \mathbb{R}^\infty$. This set is invariant to elements of G_n acting on the first n coordinates while leaving the rest unchanged. Define G as the union of all G_n, representing group actions that modify only the first n entries of $x \in \mathbb{R}^\infty$ for all $n \in \mathbb{N}$.

Tychnoff's theorem ensures that as we consider more and more elements, compactness is preserved in the product topology. Simply put, if you have compact groups, their arbitrary products and subgroups are also compact.

A tail set $E \subset \mathbb{R}^\infty$ is a set where if a point x is in E, any point y that differs from x in only a finite number of coordinates is also in E. Because group elements $g \in G$ modify only a finite number of coordinates, if x is in E, then $\pi_g(x)$ (the result of applying g to x) is also in E. This property makes tail sets G-invariant.

Moreover, $\mathcal{S}$ is the σ-field on $\mathbb{R}^\infty$ consisting of sets that remain unchanged under actions of G_n for all $n \in \mathbb{N}$. The function $T : \mathbb{R}^\infty \to \mathbb{R}$ is defined as $T := \lim_{n\to\infty} T_n$, where T_n is defined by projecting $x \in \mathbb{R}^\infty$ onto its first n coordinates.

In the upcoming discussion, the concept of a Lévy family is essential (see [3]; [7]). Consider a family of metric measure spaces $(M^{(n)}, d^{(n)}, \mu^{(n)})$ for $n \geq 1$. The open neighborhood of a set $A \subset M^{(n)}$ for some $t > 0$ is denoted as $A_t = \{x \in M^{(n)} : d^{(n)}(x, A) < t\}$. This collection of metric measure spaces is

termed a normal Lévy family if

$$\sup_{A \subset M^{(n)}} \left\{ \left(1 - \mu^{(n)}(A_t)\right) : \mu^{(n)}(A) \geq \frac{1}{2} \right\} \leq K e^{-knt^2}$$

for some constants $K, k > 0$.

From the preceding paragraph, we have $M^{(n)} = G_n$ treated as a subgroup of G that acts as the identity on all coordinates $i > n$. The measure $\mu^{(n)} = \rho_n$ will be the Haar measure for G_n. The main results below require the family (G_n, d_n, ρ_n) to be a normal Lévy family. This, naturally, encompasses a broad range of groups, including the classical compact groups $\mathbb{SO}(n)$, $\mathbb{SU}(n)$, and $\mathbb{S}p(2n)$ with the Hilbert-Schmidt metric, satisfying this requirement.

The following theorem demonstrates that Theorem 2.1.1 can hold in an asymptotic sense.

Theorem 2.1.3. *Let $X \in \mathbb{R}^\infty$, and $X^{(n)} \in \mathbb{R}^n$ be X projected onto its first n coordinates. Let T_n be c_n-Lipschitz such that for some $p \geq 1$, $E\|X^{(n)}\|_p < \infty$ for all n, and $\sum_{n=1}^\infty c_n^p < \infty$. Furthermore, assume $n^{-1/2} c_n \|X^{(n)}\| \xrightarrow{a.s.} 0$. Lastly, let the collection of G_n be a normal Lévy family. Then,*

$$|\rho(\{g \in G_n : T_n(\pi_g X^{(n)}) > t\}) - \mathbb{P}(\{T_n(X^{(n)}) > t \,|\, \mathcal{S}_n \cap \mathcal{T}_n\})| \to 0$$

$\mathbb{P}$-almost surely and in L^1 as $n \to \infty$.

Corollary 2.1.4. *Under the above condition, $\mathbb{P}(T_n(X^{(n)}) > t_\alpha(X^{(n)})) \leq \alpha$.*

Proof. From Theorem 2.1.3, for any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that

for all $n > N$,

$$\begin{aligned}\mathbb{P}\left\{T_n(X^{(n)}) > t_\alpha(X^{(n)})\right\} &= \mathbb{E}\left[\mathbb{P}\left\{T_n(X^{(n)}) > t_\alpha(X^{(n)}) \,|\, \mathcal{S}_n \cap \mathcal{T}_n\right\}\right] \\ &\leq \mathbb{E}\left[\rho\left\{g \in G_n : T_n(\pi_g X^{(n)}) > t_\alpha(X^{(n)})\right\}\right] + \varepsilon \\ &\leq \alpha + \varepsilon.\end{aligned}$$

Taking $\varepsilon \to 0$ finishes the proof. □

In the realm of statistical testing, where the conventional idea often leans towards assuming normality due to the central limit theorem, the use of group invariance marks a notable shift. As it can be seen, group invariance testing presents a more flexible paradigm, without relying on normality assumptions, and also achieving the correct test size. It is also valid asymptotically, assuring reliability even as sample sizes grow.

However, it is important to check all the assumptions stated in the theorems. Furthermore, contemplating the potential applications of other interesting locally compact groups adds a layer of complexity and richness to statistical methodology. So, group invariance testing gives us a useful and reliable way to test things in statistics without relying too much on the normality idea, opening up new possibilities for different kinds of analyses.

In this research, our focus will be on exploring the potential of employing random rotations in $\mathbb{R}^n$, i.e., the group $\mathbb{SO}(n)$. To utilize random rotations, we require some concentration inequality, and we will draw upon one found in Meckes' book [9]. We state the result without proof.

Theorem 2.1.5. *Given $n_1, \ldots, n_k \in \mathbb{N}$, let $G = G_{n_1} \times \cdots \times G_{n_k}$, where for*

each of the n_i, G_{n_i} is one of $\mathbb{SO}(n_i), \mathbb{SO}^-(n_i), \mathbb{SU}(n_i), \mathbb{U}(n_i)$ or $\mathbb{S}p(n_i)$. Let G be equipped with the L_2-sum of Hilbert-Schmidt metrics on the G_{n_i}. Suppose that $F : G \to \mathbb{R}$ is L-Lipschitz, and that $\{U_j \in G_{n_j} : 1 \leq j \leq k\}$ are independent, Haar-distributed random matrics. Then for each $t > 0$,

$$\mathbb{P}[F(U_1, \ldots, U_k) \geq \mathbb{E}F(U_1, \ldots, U_k) + t] \leq \exp\left\{\frac{-(n-2)t^2}{24L^2}\right\} \tag{2.5}$$

where $n = \min\{n_1, \ldots, n_k\}$.

Corollary 2.1.6. *Let $G = \mathbb{SO}(n)$ be equipped with the Hilbert-Schmidt metric. Suppose that $F : G \to \mathbb{R}$ is L-Lipschitz, and that $U \in G$ be a Haar-distributed random matrix. Then for each $t > 0$,*

$$\mathbb{P}[F(U) \geq \mathbb{E}F(U) + t] \leq \exp\left\{\frac{-(n-2)t^2}{4L^2}\right\} \tag{2.6}$$

2.2 Random Rotations

In the context of permutation and rotation tests, two crucial symmetries to focus on are exchangeability and rotatability. It's important to note a clear difference in the simulation process: simulating permutations is computationally easier, while generating rotation matrices, especially with large datasets, is not feasible.

For example, Environment Canada has collected over 50 years of climate data for Edmonton, but the dataset's size, like a 100-year daily measurement dataset with $36,500$ points, poses computational challenges. A single random rotation matrix acting on this dataset would require approximately 40 GB of

computer memory making it unfeasible to compute with brute force.

The construction of such randomization tests have already been discussed in [8] and [10]. It is based on a group of transformations $\mathcal{O}$ such that,

$$Y \stackrel{d}{=} \mathbb{O}Y \text{ under } H \tag{2.7}$$

for every $n \times n$ matrix $\mathbb{O} \in \mathcal{O}$. The main advantage of this condition is that it provides different observations, $\mathbb{O}Y$ which have the same probability of occurring under H.

Equation 2.7 is referred to as the randomization hypothesis and the transformations $\mathbb{O}$ as null-invariants. They are called null-invariants as the transformation of Y by $\mathbb{O}$ has no effect on the underlying null distribution. In case of permutations, $\mathcal{O}$ consists of all $n \times n$ permutation matrices and in case of rotations, all $n \times n$ rotation matrices. Note that, in contrast to permutations, which are finite in number ($n!$ precisely), rotations are uncountably infinite, thereby making it more challenging to employ random rotations for hypothesis testing.

2.3 The AR(1) Model and the Durbin-Watson Test

We start by defining autoregressive process (of order 1) formally as,

Definition 2.3.1. The time series $\{X_t\}$ is an AR(1) process if X_t has zero

mean and if we can write it as,

$$X_t = \phi X_{t-1} + \varepsilon_t \tag{2.8}$$

where ε_t is white noise with mean zero and $\phi \in \mathbb{R}$ is not zero.

Here, ϕ is the autocorrelation between X_{t-1} and X_t for all $t \in \{2, \ldots, n\}$. Without loss of generality, we assume that the sample mean, $\bar{X}$ is zero. The reason is two fold:

- Firstly, we can always center the data by defining a new time series $Y_t = X_t - \bar{X}$ for all t. Note that this doesn't affect any autocorelation that may exist in the data.

- Secondly, the expectation of X_t is zero. Therefore for large n, the sample mean will converge to the true mean which is zero.

The sample autocorrelation is estimated by the statistic,

$$\hat{\phi} = \frac{1}{n\hat{\sigma^2}} \sum_{i=1}^{n-1} X_i X_{i+1} \tag{2.9}$$

where $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\|X\|^2$, is the sample variance and $X = (X_1, \ldots, X_n)^t$.

For our thesis, we also assume that $\phi \in (-1, 1)$ and the process to be stationary. Note that we can write autocorrelation estimator as,

$$\hat{\phi} = \frac{X^t A X}{\|X\|^2} = \frac{X^t}{\|X\|} A \frac{X}{\|X\|} \tag{2.10}$$

where $A_{n\times n} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$ is the shift matrix. Upon writing $\hat{\phi}$ in this way, we can, without loss of generality, assume that X lies in the n-dimensional unit sphere, i.e., $X \in S^{n-1}$ and $\hat{\phi} = X^t A X$.

There are various statistical tests to determine autocorrelation in the data. One of the most well-known test is the Durbin-Watson Test. The Durbin-Watson test tests for autocorrelations of order 1 among the residuals of a linear model. Considering the linear model

$$Y_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p + r_t \tag{2.11}$$

with $t = 1, \dots, n$, we can compute the least squares estimator $\hat{\beta}$ and then compute the residuals $\hat{r_t} = r_t - \langle \hat{\beta}, (1, t, \dots, t^p) \rangle$. The Durbin–Watson Test assumes the following model for the residuals:

$$\hat{r_t} = \rho \hat{r}_{t-1} + w_t \tag{2.12}$$

where w_t is white noise. Then, it tests the hypotheses $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$. It does this by computing the test statistics,

$$Q_{DW} = \frac{\sum_{t=2}^{n} (\hat{r_t} - \hat{r}_{t-1})^2}{\sum_{t=2}^{n} \hat{r_t}^2}. \tag{2.13}$$

If this test statistic is close to zero, it implies that $\hat{r_t}$ and $\hat{r}_{t-1}$ are close in value

indicating a strong positive autocorrelation of order 1. In contrast, if the test statistic is large (close to the max of 4), then it indicates that there is a strong negative autocorrelation of order 1. Otherwise, a test statistic near 2 indicates no autocorrelation of order 1.

The main purpose of the Durbin-Watson test is to check whether a fitted model is a good fit or not. It examines the presence of autoregressive correlations within the residual vector. Therefore, it is crucial to conduct tests for serial correlation for better model selection.

In the context of this thesis, we are more focused on detecting autocorrelation in any time series data, rather than the idea of whether the model is a good fit. The same can be done using the Durbin-Watson test as well. If X_t is the time series, then we can fit the most simple intercept model,

$$X_t = \mu + \gamma_t \tag{2.14}$$

where γ_t is the white noise and μ is the mean intercept. Now the Durbin-Watson test is essentially testing for autocorrelation in the time series itself.

Chapter 3

Main Results

In this chapter, we consider the group $\mathbb{SO}(n)$, consisting of all orthogonal matrices with determinant 1. For each matrix $M \in \mathbb{SO}(n)$, we define a test statistic,

$$\tilde{\phi}_M = (MX)^t A(MX) \tag{3.1}$$

where $X \in S^{n-1}$ and A is the shift matrix as defined before. We are essentially considering a new test statistic which randomly rotates the sample autocorrelation function. As noted in Equation 2.10, we can, without loss of generality, assume that $X \in S^{n-1}$. Let $M \in \mathbb{SO}(n)$ and $x \in \mathbb{R}^n$. Then we have,

$$\|Mx\|_2^2 = (Mx)^t Mx = x^t M^t Mx = x^t x = \|x\|_2^2 \tag{3.2}$$

i.e., orthogonal matrices preserve the L^2 norm and as a consequence, the operator norm as well. Therefore, we can conclude that MX also belongs to S^{n-1}

3.1 Concentration Inequality

Note that our matrix A is not symmetric. In the context of quadratic forms, using a symmetric matrix significantly streamlines our calculations. The following lemma fulfills this requirement.

Lemma 3.1.1. *Let $B = \frac{A+A^t}{2}$, be the symmetric part of A. Then for all $x \in \mathbb{R}^n$, we have $x^tAx = x^tBx$.*

Proof. We note that,

$$\begin{aligned} x^tBx &= x^t\left(\frac{A+A^t}{2}\right)x \\ &= \frac{x^tAx}{2} + \frac{x^tA^tx}{2} \\ &= \frac{x^tAx}{2} + \frac{(x^tAx)^t}{2} \\ &= \frac{x^tAx}{2} + \frac{x^tAx}{2} \\ &= x^tAx \end{aligned}$$

□

Therefore, we can replace our matrix A with the new matrix B (which is symmetric) without changing our quadratic form. This lemma is useful in case we want to use a result that requires symmetry. It turns out that B is a tridiagonal matrix with 0 in the main diagonal and $\frac{1}{2}$ both in the upper and the lower diagonal.

For a fixed $x \in S^{n-1}$, define $F_x : \mathbb{SO}(n) \to \mathbb{R}$ by,

$$M \longmapsto (Mx)^tA(Mx) \tag{3.3}$$

Note that, $F_X(M) = \tilde{\phi}_M$. In order to use Corollary 2.1.6, we have the following result.

Theorem 3.1.2. *For a fixed $x \in S^{n-1}$, the function F_x is Lipschitz with Lipschitz constant 2, where the group $\mathbb{SO}(n)$ is equipped with the Hilbert-Schmidt metric.*

Proof. Let $g : S^{n-1} \to \mathbb{R}$ be defined by, $g(x) = x^t Ax = x^t Bx$ (using Lemma 3.1.1). It is clear that g is continuously differentiable. Since S^{n-1} is compact and any continuously differentiable function over a compact domain is Lipschitz and the Lipschitz constant equal to the maximum magnitude of the derivative, we can conclude that g is also Lipschitz. Also, $\nabla g(x) = 2Bx$. Therefore, the Lipschitz constant of g is,

$$\max_{x \in S^{n-1}} \|2Bx\| = 2 \max_{x \in S^{n-1}} \|Bx\| = 2\rho(B) \tag{3.4}$$

where $\rho(B)$ is the spectral radius of B. Using [6], we know that $\rho(B) \leq 1$, so, we can conclude that g is 2-Lipschitz.

Now, let $P, Q \in \mathbb{SO}(n)$ and $x \in S^{n-1}$ be fixed. Then, $F_x(P) = g(Px)$ and

$F_x(Q) = g(Qx)$. Therefore,

$$\begin{aligned}
|F_x(P) - F_x(Q)| &= |g(Px) - g(Qx)| \\
&\leq 2\|Px - Qx\|_2 \\
&= 2\|(P-Q)x\|_2 \\
&\leq 2\|(P-Q)\|_{op}\|x\|_2 \\
&\leq 2\|(P-Q)\|_{op} \\
&= 2\sqrt{\lambda_{max}[(P-Q)^t(P-Q)]} \\
&\leq 2\sqrt{\sum_{i=1}^{n} \lambda_i[(P-Q)^t(P-Q)]} \\
&= 2\sqrt{\text{trace}[(P-Q)^t(P-Q)]} \\
&= 2\|(P-Q)\|_{HS}
\end{aligned}$$

where $\|.\|_{HS}$ is the Hilbert-Schmidt norm. Therefore, we can conclude that the function F_x is also 2-Lipschitz. □

In order to apply Corollary 2.1.6, we also need to calculate the expected value of $\tilde{\phi}_M$ over $\mathbb{SO}(n)$. We use the following proposition from [4].

Proposition 3.1.3. *Let $S \in \mathbb{R}^{n\times n}$ be a symmetric matrix with spectrum $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and let $b_S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ be a bilinear form defined by,*

$$b_{S,\mathbb{SO}(n)}(x,y) = \int_{\mathbb{SO}(n)} (Mx)^t S(My) d\rho(M) \tag{3.5}$$

where the integration is taken over $\mathbb{SO}(n)$ with Haar measure ρ. Then, $b_{S,\mathbb{SO}(n)}(x,y)$

is rotationally invariant and furthermore,

$$b_{S,\mathbb{SO}(n)}(x,y) = \bar{\lambda}\langle x, y\rangle \tag{3.6}$$

where $\langle ., . \rangle$ *is the standard Euclidean inner product and* $\bar{\lambda} = n^{-1}\sum_{i=1}^{n}\lambda_i$.

The proof of this proposition follows from the fact that any bilinear form on a real Hilbert Space is of the form $\langle Mx, y\rangle$ for some bounded operator M. As a consequence, we can write $b(x,y) = \sum_{i=1}^{n} cx_iy_i$ for some $c > 0$ by rotational invariance. The subsequent calculations follow directly from the computation of c, which is determined to be $n^{-1}\sum_{i=1}^{n}\lambda_i$. Using this proposition, we get the following corollary.

Corollary 3.1.4. *The expectation of* $\tilde{\phi}_M$ *over* $\mathbb{SO}(n)$ *is* 0*, i.e.,* $\mathbb{E}_{\mathbb{SO}(n)}[\tilde{\phi}_M] = 0$

Proof. By definition,

$$\mathbb{E}_{\mathbb{SO}(n)}[\tilde{\phi}_M] = b_{B,\mathbb{SO}(n)}(X,X) = \int_{\mathbb{SO}(n)} (MX)^t B(MX) d\rho(M) \tag{3.7}$$

where B is the symmetric part of our shift matrix A. Now the proof directly follows from Proposition 3.1.3, noting that B is symmetric with sum of eigenvalues zero, since trace of B is zero. □

Having established the Lipschitz continuity of F_x and computed the expected value, we can now combine these results to derive the desired concentration inequality required for our hypothesis test.

Theorem 3.1.5. *Let $M \in \mathbb{SO}(n)$. Then for each $t > 0$,*

$$\mathbb{P}(\tilde{\phi}_M \geq t) \leq \exp\left\{\frac{-(n-2)t^2}{16}\right\} \tag{3.8}$$

where $\tilde{\phi}_M$ is as defined above.

Proof. The proof follows directly from Corollary 2.1.6, 3.1.4 and Theorem 3.1.2. □

Note that we test the hypothesis, $H_0 : \phi = 0$ vs $H_1 : \phi > 0$, i.e., we start by detecting if the there is any positive autocorrelation in our data. Based on our notation, $\tilde{\phi}_I$ is based on the original data, where I is the identity matrix. Then the p-value of our hypothesis test is $\mathbb{P}(\tilde{\phi}_M \geq \tilde{\phi}_I)$. One way to approximate this is by randomly generating huge number of random rotations from $\mathbb{SO}(n)$. However, computations of such large order may be impractical as explained in Section 2.2. Therefore, Theorem 3.1.5 allows us to avoid relying on simulation-based approximations and provides a solid sub-Gaussian bound for the p-value.

3.2 The Beta Adjustment

A significant challenge in employing inequalities, including concentration inequalities, for statistical testing is the substantial loss of power to reject the null hypothesis caused by universal constants that are often too big for applications. Motivated by the approach in [1], we want to develop a transformation based on the beta distribution to rectify our p-values and restore the lost statistical power.

We begin this section by calculating the mean and variance of the autocorrelation estimator, $\hat{\phi}$, under H_0. We start by observing that under H_0, $\phi = 0$, i.e, under the null hypothesis, our data behaves like white noise. Therefore, we have for all $t \in \{1, \ldots, n\}$,

$$X_t = \varepsilon_t \tag{3.9}$$

where ε_t is white noise with mean zero. Hence, $\mathbb{E}[X_t] = \mathbb{E}[\varepsilon_t] = 0$ for all $t \in \{1, \ldots, n\}$. Hence, for $i \in \{1, \ldots, n-1\}$

$$\mathbb{E}[X_i X_{i+1}] = \mathbb{E}[\varepsilon_i \varepsilon_{i+1}] = \mathbb{E}[\varepsilon_i]\mathbb{E}[\varepsilon_{i+1}] = 0 \tag{3.10}$$

The above equation holds true since ε_i's are uncorrelated. Therefore, $\mathbb{E}[\hat{\phi}] = 0$ under H_0.

In order to calculate the variance, we will need the following lemma from [11]. We state the result without proof.

Lemma 3.2.1. *Let x have a spherical distribution. Then for symmetric matrices A and B, we have,*

$$\mathbb{E}[x^t Ax \cdot x^t Bx] = \mu_{22}(tr(A) \cdot tr(B) + 2tr(AB)) \tag{3.11}$$

where $\mu_{22} = \mathbb{E}[X_1^2 X_2^2]$.

Here spherical distribution means that $\frac{x}{||x||}$ is distributed uniformly over the surface of the unit sphere, independently of $||x||$. This result is crucial for our work, as it allows us to easily calculate expectations of product of quadratic forms, aiding us in the computation of the variance.

Proposition 3.2.2. *Under H_0, the variance of $\hat{\phi}$ is $\frac{n-1}{n(n+2)}$.*

Proof. We start by seeing that, $\text{Var}(\hat{\phi}) = \mathbb{E}[\hat{\phi}^2] - \mathbb{E}[\hat{\phi}]^2 = \mathbb{E}[\hat{\phi}^2]$, since $\mathbb{E}[\hat{\phi}] = 0$ under H_0. Now, $\hat{\phi}^2 = (X^tBX)^2 = X^tBX \cdot X^tBX$, where X is uniformly distributed in the sphere S^{n-1}. Using Lemma 3.2.1, since trace of B is zero, we have,

$$\mathbb{E}[\hat{\phi}^2] = \mathbb{E}[X^tBX \cdot X^tBX] = \mu_{22}(2\text{tr}(B^2)) \tag{3.12}$$

Now $\mu_{22} = \mathbb{E}[X_1^2X_2^2]$. To calculate μ_{22}, we note that,

$$X_1^2X_2^2 = (1 - X_2^2 - \cdots - X_n^2)X_2^2 = X_2^2 - X_2^4 - X_3^2X_2^2 - \cdots - X_n^2X_2^2 \tag{3.13}$$

By symmetry, $\mu_{22} = \mathbb{E}[X_i^2X_j^2]$ for any $i \neq j$. Therefore, taking expectations,

$$\mu_{22} = \mathbb{E}[X_2^2] - \mathbb{E}[X_2^4] - (n-2)\mu_{22} \tag{3.14}$$

Upon rearranging, we get,

$$\mu_{22} = \frac{1}{n-1}\left(\mathbb{E}[X_2^2] - \mathbb{E}[X_2^4]\right) \tag{3.15}$$

Now, $\sum_1^n X_i^2 = 1$. Taking expectation, $\mathbb{E}[X_2^2] = \frac{1}{n}$ by symmetry. Also, since $X_2^2 \overset{d}{=} \text{Beta}\left(\frac{1}{2}, \frac{(n-1)}{2}\right)$, $\mathbb{E}[X_2^4]$ is equal to the second moment of the beta distribution, which is $\frac{3}{n(n+2)}$. Therefore,

$$\mu_{22} = \frac{1}{n-1}\left(\frac{1}{n} - \frac{3}{n(n+2)}\right) = \frac{1}{n(n+2)}\,. \tag{3.16}$$

Finally, we calculate the trace of B^2. From [6], we know the eigenvalues of B

are,

$$\lambda_k = \cos\left(\frac{k\pi}{n+1}\right) \tag{3.17}$$

where $k = 1, 2, \cdots, n$. Since the trace of B^2 is the sum of its eigenvalues, we have,

$$\mathrm{tr}(B^2) = \sum_{k=1}^{n} \lambda_k^2 = \sum_{k=1}^{n} \cos^2\left(\frac{k\pi}{n+1}\right) \tag{3.18}$$

To simplify our calculations, we express cos in terms of complex numbers, i.e.,

$$\cos\left(\frac{k\pi}{n+1}\right) = \frac{1}{2}\left(e^{i\frac{k\pi}{n+1}} + e^{-i\frac{k\pi}{n+1}}\right) \tag{3.19}$$

Hence,

$$\cos^2\left(\frac{k\pi}{n+1}\right) = \frac{1}{4}\left(e^{2i\frac{k\pi}{n+1}} + e^{-2i\frac{k\pi}{n+1}} + 2\right) \tag{3.20}$$

Therefore,

$$\sum_{k=1}^{n} \cos^2\left(\frac{k\pi}{n+1}\right) = \frac{1}{4}\sum_{k=1}^{n}\left(e^{2i\frac{k\pi}{n+1}} + e^{-2i\frac{k\pi}{n+1}} + 2\right) \tag{3.21}$$

We have two geometric progressions, which are straightforward to calculate.

$$1 + \sum_{k=1}^{n} e^{2i\frac{k\pi}{n+1}} = \sum_{k=0}^{n}\left(e^{2i\frac{\pi}{n+1}}\right)^k = \frac{1 - (e^{2i\frac{\pi}{n+1}})^{n+1}}{1 - e^{2i\frac{\pi}{n+1}}} = 0 \tag{3.22}$$

$$1 + \sum_{k=1}^{n} e^{-2i\frac{k\pi}{n+1}} = \sum_{k=0}^{n}\left(e^{-2i\frac{\pi}{n+1}}\right)^k = \frac{1 - (e^{-2i\frac{\pi}{n+1}})^{n+1}}{1 - e^{-2i\frac{\pi}{n+1}}} = 0 \tag{3.23}$$

Plugging this in equation 3.18, we get,

$$\mathrm{tr}(B^2) = \frac{1}{4}\left(-1 - 1 + 2n\right) = \frac{n-1}{2} \tag{3.24}$$

Thus, finally we have our variance,

$$\mathrm{Var}(\hat{\phi}) = \frac{n-1}{n(n+2)} \tag{3.25}$$

□

Therefore for large n, the variance is approximately equal to $\frac{1}{n}$. The $\frac{1}{n}$ term signifies that, with larger sample sizes n, the estimation of autocorrelation becomes more precise. Finally, combining the above results, we have the following theorem.

Theorem 3.2.3. *Under the setting of Theorem* 3.1.5 *with n sufficiently large,*

$$\mathbb{P}\left(\exp\left\{\frac{-(n-2)(\tilde{\phi}_I)^2}{16}\right\} \leq u\right) \leq C_0 I\left(u; \frac{8n(n+2)}{(n-1)(n-2)}, \frac{1}{2}\right) \tag{3.26}$$

where $I(u;\alpha,\beta)$ is the regularized incomplete beta function and,

$$C_0 = \left(\frac{8n(n+2)}{(n-1)(n-2)}\right)^{1/2} \Gamma\left(\frac{8n(n+2)}{(n-1)(n-2)}\right) \Gamma\left(\frac{1}{2} + \frac{8n(n+2)}{(n-1)(n-2)}\right)^{-1}$$

Proof. We begin by noting that $\hat{\phi} = \tilde{\phi}_I$. By central limit theorem,

$$\left(\frac{\tilde{\phi}_I}{\sqrt{\frac{n-1}{n(n+2)}}}\right)^2 = \frac{n(n+2)(\tilde{\phi}_I)^2}{n-1} \tag{3.27}$$

is approximately $\chi^2(1)$. Now for $Z \sim \chi^2(1)$, some $c > 0$ and $u \in (0,1)$, we

have,

$$\begin{aligned}
\mathbb{P}(e^{-Z/c} \leq u) &= \mathbb{P}(Z \geq -c \log u) \\
&= \sqrt{2\pi} \int_{-c \log u}^{\infty} x^{-1/2} e^{-x/2} dx \\
&= \sqrt{\frac{c}{2\pi}} \int_0^u (-\log y)^{-1/2} y^{c/2-1} dy \\
&\leq \sqrt{\frac{c}{2\pi}} \int_0^u (1-y)^{-1/2} y^{c/2-1} dy \\
&= \frac{(c/2)^{1/2}\Gamma(c/2)}{\Gamma((c+1)/2)} I(u; c/2, 1/2)
\end{aligned}$$

where the inequality $-\log y \geq 1 - y$ for $y \in (0, 1)$ is used. Replacing our constant c with $\frac{16n(n+2)}{(n-1)(n-2)}$, we get that,

$$\mathbb{P}\left(\exp\left\{\frac{-(n-2)(\tilde{\phi}_M)^2}{16}\right\} \leq u\right) \leq C_0 I\left(u; \frac{8n(n+2)}{(n-1)(n-2)}, \frac{1}{2}\right) \tag{3.28}$$

where $C_0 = \left(\frac{8n(n+2)}{(n-1)(n-2)}\right)^{1/2} \Gamma\left(\frac{8n(n+2)}{(n-1)(n-2)}\right) \Gamma\left(\frac{1}{2} + \frac{8n(n+2)}{(n-1)(n-2)}\right)^{-1}$.

□

Theorem 3.2.3 allows us to improve our p-values from Theorem 3.1.5 in order to recover the lost statistical power. The improved bound is,

$$\mathbb{P}(\tilde{\phi}_M \geq t) \leq C_0 I\left(\exp\left\{\frac{-(n-2)t^2}{16}\right\}; \frac{8n(n+2)}{(n-1)(n-2)}, \frac{1}{2}\right) \tag{3.29}$$

Note that since $t > 0$, this test essentially tests for positive autocorrelation instead of any autocorrelation in the data. Hence, this is a one-sided test. In order to make it a two-sided test, we need a bound for $\mathbb{P}(|\tilde{\phi}_M| \geq t)$. We note

that, for $t > 0$,

$$\begin{aligned}\mathbb{P}(|\tilde{\phi}_M| \geq t) &= \mathbb{P}(\tilde{\phi}_M \geq t) + \mathbb{P}(\tilde{\phi}_M \leq -t) \\ &= \mathbb{P}(\tilde{\phi}_M \geq t) + \mathbb{P}(-\tilde{\phi}_M \geq t)\end{aligned}$$

By similar argument, $-\tilde{\phi}_M$ is also 2-Lipschitz and we have the same concentration inequality for $-\tilde{\phi}_M$ as well. Therefore, we can conclude that,

$$\mathbb{P}(|\tilde{\phi}_M| \geq t) \leq 2C_0 I\left(\exp\left\{\frac{-(n-2)t^2}{16}\right\}; \frac{8n(n+2)}{(n-1)(n-2)}, \frac{1}{2}\right) \tag{3.30}$$

results in a two-sided test.

3.3 Simulation Study

We conducted simulation studies to better understand how our rotation test is performing when compared to the Durbin-Watson test.

3.3.1 Normal Errors

In Figure 3.1, we compared the mean $\log_2$ p-values for three results for a sample size of 1000, repeated 100 times as the autocorrelation is varied from 0 to 0.20. We have considered a two sided test. For this example, we have assumed that the errors are $\mathcal{N}(0, 1)$. We considered the original p-value from Theorem 3.1.5, the improved p-value from 3.30 and the p-value from the Durbin-Watson test. As it can be seen, our original p-values are larger than the DW test while

the beta adjusted bound is tighter giving a close approximation to the DW p-value.

In Figure 3.2, we have considered a higher sample size of 10000. The other conditions remain the same. As it can be seen, as the sample size increases, the beta adjusted and the DW p-values almost become indistinguishable.

3.3.2 t-Distributed Errors

In this context, our framework remains consistent with that of normal errors, with the sole distinction being the adoption of t-distributed errors. Our analysis is based on a sample size of 1000. Specifically, in Figure 3.3, we set the degrees of freedom to 2, and in Figure 3.4, the degrees of freedom are set to 3. The sharp corners present in Figure 3.3 makes sense since the t-distribution with degrees of freedom equal to 2 has infinite variance. As observed previously, our original p-values are larger than those obtained from the Durbin-Watson test. However, the application of the beta-adjusted bound produces a more constrained range, closely mirroring the DW p-value.

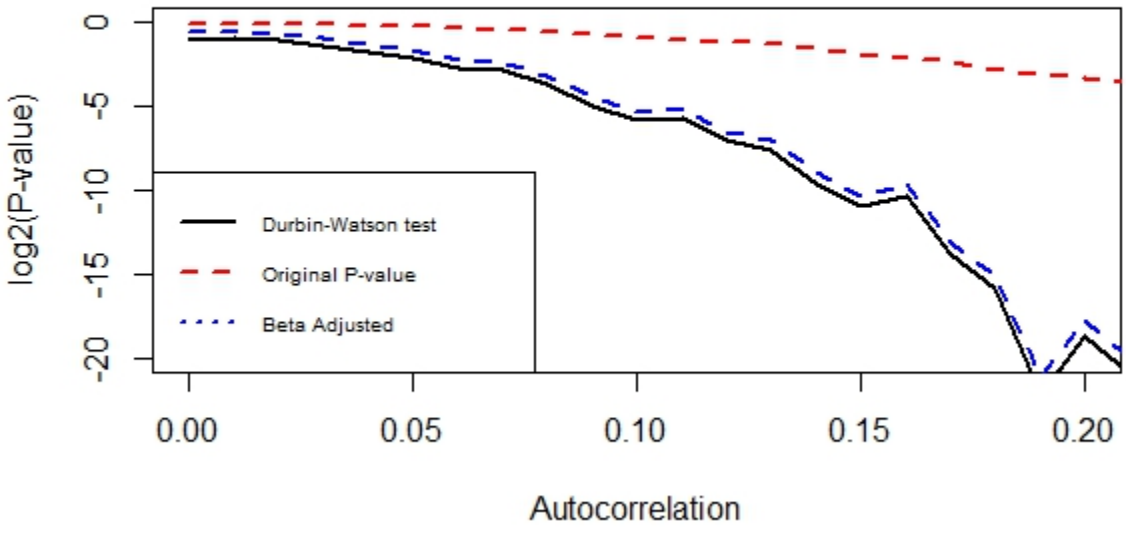

Figure 3.1: Normal errors with $n = 1000$

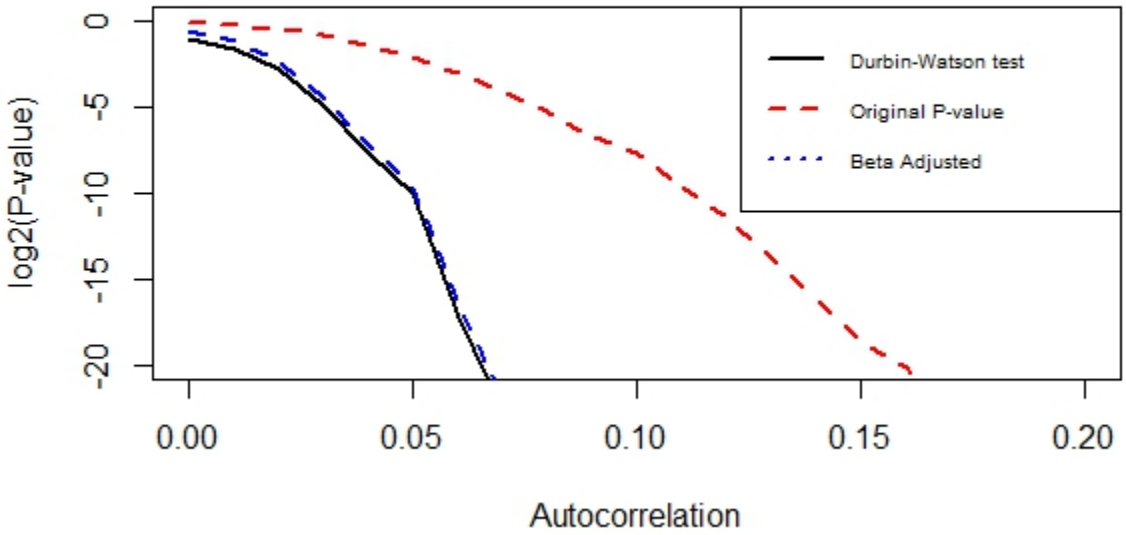

Figure 3.2: Normal errors with $n = 10000$

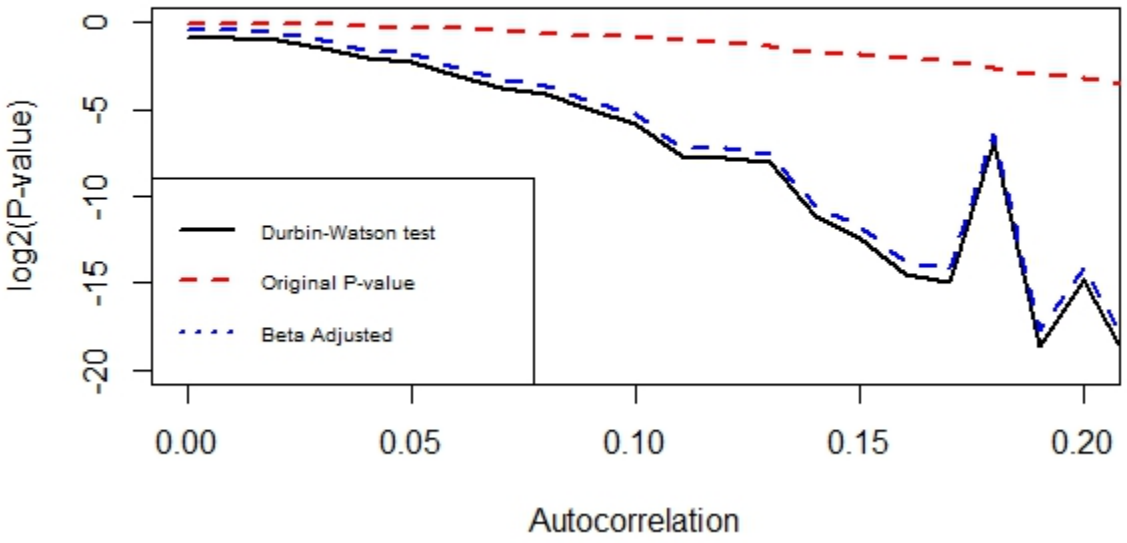

Figure 3.3: t-Distributed errors with df = 2

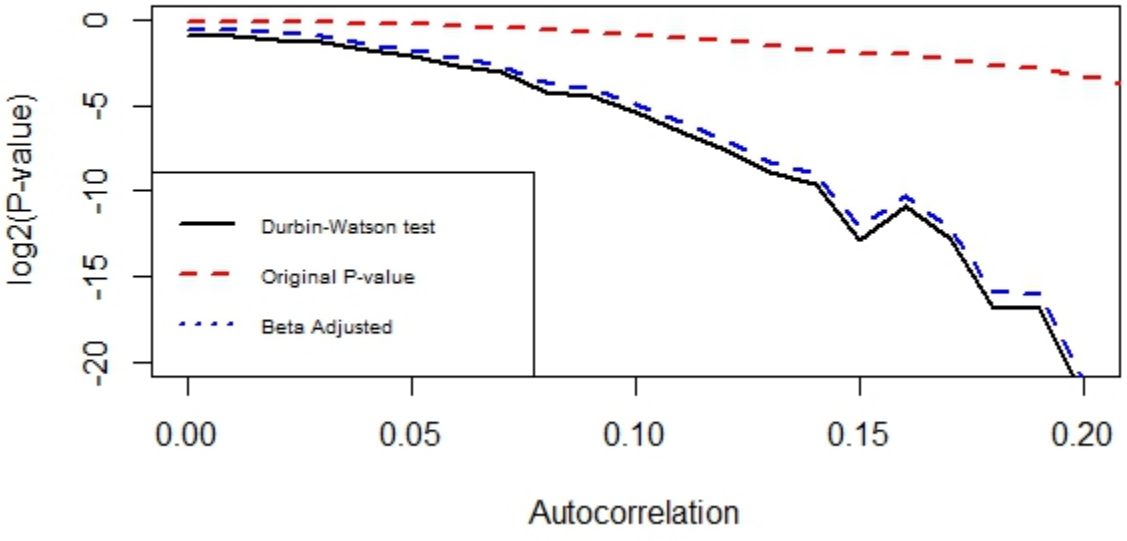

Figure 3.4: t-Distributed errors with df = 3

3.4 Discussion and Future Work

Our rotation test currently works well for finding autocorrelation with lag of order 1, similar to the Durbin-Watson test. The advantage of our test is that it can be extended to work with higher lags, unlike the Durbin-Watson test, where the DW statistic gets too complicated as the lag increases.

To handle higher lags, let's say for k, greater than 1, we can use a new shift matrix A^k where A is the original shift matrix. This leads to our autocorrelation estimator:

$$\hat{\phi}_k = X^t A^k X \tag{3.31}$$

While the basic math stays the same, we should be able to adjust how we deal with the eigenvalues of A^k when calculating the variance of $\hat{\phi}_k$. This tweak is important for keeping the accuracy of our method.

Looking ahead, we aim to make our approach work for higher lag orders, showing our commitment to improving and expanding its usefulness. Also, it's worth noting that our method can be applied to any general quadratic form, making it useful for other types of analyses.

Bibliography

[1] Sergii Myroshnychenko Adam B. Kashlak and Susanna Spektor. Analytic permutation testing for functional data anova. *Journal of Computational and Graphical Statistics*, 32(1):294–303, 2023.

[2] Phillip Good. *Permutation Tests*. Springer New York, NY, 2013. https://doi.org/10.1007/978-1-4757-2346-5.

[3] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *American Journal of Mathematics*, 105(4):843–854, 1983.

[4] Adam B. Kashlak. Asymptotic invariance and robustness of randomization tests. https://arxiv.org/abs/2211.00144, 2023.

[5] Adam B. Kashlak and Weicong Yuan. Computation-free nonparametric testing for local spatial association with application to the us and canadian electorate. *Spatial Statistics*, 48:100617, 2022.

[6] Devadatta Kulkarni, Darrell Schmidt, and Sze-Kai Tsui. Eigenvalues of tridiagonal pseudo-toeplitz matrices. *Linear Algebra and its Applications*, 297(1):63–80, 1999.

[7] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2001.

[8] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. New York: Springer, 2005.

[9] Elizabeth S. Meckes. *The Random Matrix Theory of the Classical Compact Groups*. Cambridge Tracts in Mathematics. Cambridge University Press, 2019.

[10] Goeman JJ Solari A, Finos L. Rotation-based multiple testing in the multivariate linear model. *Biometrics*, 70(4):954–61, 2014 Dec.

[11] Douglas P. Wiens†. On moments of quadratic forms in non-spherically distributed variables. *Statistics*, 23(3):265–270, 1992.