Bugni and Horowitz (2021) Permutation Tests for the Equality of Distributions of Functional Data

Master's Thesis presented to the

Department of Economics at the

Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of

Master of Science (M.Sc.)

Supervisor: Prof. Dr. Dominik Liebl

Submitted in July 2022 by:

Jakob R. Juergens

Matriculation Number: 2996491

Contents

1 Introduction

In modern economics, it is becoming more and more common to use data measured at a very high frequency. As the frequency of observing a variable increases, it often becomes natural to view the data not as a sequence of distinct observations but as a smooth curve describing the variable. This idea, to think of observations as measurements of a continuous process, is the motivating thought behind functional data analysis. Functional data analysis is a branch of statistics that has its beginnings in the 1940s and 1950s in the works of Ulf Grenander and Kari Karhunen. It gained traction during the following decades and focused more on possible applications during the 1990s. Functional data analysis is still a relatively exotic field in economics, but it is beginning to gain traction. A typical question in economics is whether observations from two or more data sets, e.g., data generated by treatment and control groups, are systematically different across groups. In statistical terms, this can be formulated as whether the same stochastic process generated observations in both data sets. This question also occurs in functional data analysis, where each observation in a data set is itself a smooth curve. The main focus of this thesis is a test developed in Bugni and Horowitz 2021 trying to answer this question.

This thesis aims to present the aforementioned test while giving the necessary theoretical context. Therefore, Section 2 introduces the necessary concepts from functional data analysis. Section 3 explores Cramér-von Mises tests in a scalar setting and Section 4 introduces the necessary background in Permutation Testing. After explaining these concepts, section 5 focuses on the test developed in Bugni and Horowitz 2021 for the case of a two-sample test. The main contribution of this thesis is a variant of the original test with the goal to identify specific violations of the null hypothesis relating to the persistence of the data generating processes. This variant of the original test is presented in Section 6. Section 7 presents a set of simulations exploring the properties of the original test and shows a heuristic simulation to motivate the proposed extension. Section 8 applies the test from Bugni and Horowitz 2021 to half-hourly electricity demand data from Adelaide. Finally, Section 9 gives an Outlook on possible further extensions, addresses some problems and shortcomings of the presented results and outlines simulations that could be performed to better understand the properties of the procedures described in this thesis.

Two additional aspects, that I want to shine light on are the following. First, all code and data that was used in this thesis is available in the following public GitHub repository: `https://github.com/JakobJuergens/Masters_Thesis`. Second, Appendix 11.1 gives a very informal description of the test scenario and the underlying idea of the test. It shall serve as a primer for readers without prior experience in functional data analysis and permutation testing.

2 Functional Data Analysis

The overarching concept of functional data analysis is to analyze data that are functional in nature. Functional observations can often be understood as smooth curves and functional data sets, thus, consist of realizations of some process that generates smooth curves. A classical example for a functional data set is shown in Figure 1. It depicts growth curves of 93 humans up to the age of 18 provided in the R package fda[1].

Figure 1: Human Growth Curves up to the Age of 18

Even though measurements are only taken at discrete ages, it is clear that each human has a height at every point in time. Thus, discrete data points are only measurements of this underlying continuous curve. The higher the measurement frequency, the closer we get to data that resembles the curve itself. In many cases, functional data analysis restricts its scope to subsets of the functions $f : \mathbb{R} \to \mathbb{R}$. As these are inherently infinite-dimensional, it is necessary to introduce theory to deal with their unique properties. Sections 2.1 and 2.2 closely follow the corresponding sections of Hsing and Eubank 2015. More information on specific topics such as potential applications or the implementation in statistical software is given in Ramsay and Silverman 2005 and Kokoszka and Reimherr 2021.

2.1 Hilbert Space of Square Integrable Functions

For many methods, such as functional linear regression, it is necessary to put restrictions on the analyzed functions. A typical assumption is that curves are square-integrable. The space of square-integrable functions is a so-called separable Hilbert space, and its properties often simplify theoretical considerations in functional data analysis. However, in many cases there is potential to relax this assumption and generalize methods to less restrictive function spaces. As Bugni and Horowitz 2021 assume square-integrability, the following section introduces the necessary concepts around Hilbert spaces in general and the space of square-integrable functions in particular.

---

[1] Ramsay, Graves, and Hooker 2021.

To define Hilbert spaces, it is necessary to first define inner product spaces as Hilbert spaces are special cases of this larger category of spaces.

Definition 2.1 (Inner Product)
A function $\langle \cdot, \cdot \rangle : \mathbb{V}^2 \to \mathbb{F}$ on a vector space $\mathbb{V}$ over a field $\mathbb{F}$ is called an inner product if the following four conditions hold for all $v, v_1, v_2 \in \mathbb{V}$ and $a_1, a_2 \in \mathbb{F}$.

1. $\langle v, v \rangle \geq 0$

2. $\langle v, v \rangle = 0$ if $v = 0$

3. $\langle a_1 v_1 + a_2 v_2, v \rangle = a_1 \langle v_1, v \rangle + a_2 \langle v_2, v \rangle$

4. $\langle v_1, v_2 \rangle = \overline{\langle v_2, v_1 \rangle}$

As this thesis is limited to the case $\mathbb{F} = \mathbb{R}$, property 4 can be restated as $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$, since the complex conjugate of a real number is the number itself.

Definition 2.2 (Inner Product Spaces and Hilbert Spaces)
A vector space with an associated inner product is called an inner product space. Two elements v_1 and v_2 of an inner product space are orthogonal if $\langle v_1, v_2 \rangle = 0$. An inner product space that is complete with respect to the distance induced by the norm $\|v\| = \sqrt{\langle v, v \rangle}$ is called a Hilbert space.

As previously mentioned, Hilbert spaces play an important role in functional data analysis and many methods like, for example, functional linear regression make extensive use of their properties. In the context of this thesis, they are especially useful to describe objects such as random functions. Analogously to vector spaces, it is useful to express elements of a Hilbert space as linear combinations of a basis. However, as elements of Hilbert spaces can be infinite dimensional, the classical idea of a finite basis used to express every element has to be extended. Hsing and Eubank 2015 define the closed span and orthonormal sequences in the following ways, leading to subsequent definition of bases for Hilbert spaces.

Definition 2.3 (Closed Span)
The closed span of a subset A of some normed space, e.g. a normed vector space or a Hilbert space, is the closure of $span(A)$ with respect to the distance induced by the norm of the space. In the following it is denoted by $\overline{span(A)}$.

Definition 2.4 (Orthonormal Sequence in a Hilbert Space)
Let $\{x_n\}$ be a countable collection of elements in a Hilbert space such that every finite subcollection of $\{x_n\}$ is linearly independent. Define $e_1 = \frac{x_1}{\|x_1\|}$ and $e_i = \frac{v_i}{\|v_i\|}$ for

$$v_i = x_i - \sum_{j=1}^{i-1} \langle x_i, e_j \rangle e_j.$$

Then, $\{e_n\}$ is an orthonormal sequence and $\overline{span(\{x_n\})} = \overline{span(\{e_n\})}$

Using these definitions leads to a natural analogon of the basis in finite-dimensional vector spaces for the case of potentially infinite-dimensional Hilbert spaces.

Definition 2.5 (Orthonormal Basis of a Hilbert Space)

An orthonormal sequence $\{e_n\}$ in a Hilbert space $\mathbb{H}$ is called an orthonormal basis of $\mathbb{H}$ if $\overline{span\left(\{e_n\}\right)} = \mathbb{H}$.

A special class of Hilbert spaces is separable Hilbert spaces. Their properties often allow for significant simplifications in the derivation of theoretical properties of methods in functional data analysis. One such case is the calculation of specific test statistics that are at the core of Bugni and Horowitz 2021.

Definition 2.6 (Separable Hilbert Space)

A Hilbert space that possesses a countable complete orthonormal basis is called separable Hilbert space.

The most prevalent Hilbert space in functional data analysis is the space of square-integrable functions. However, it is often interesting to consider whether square-integrability is a necessary assumption. Due to its prevalence in functional data analysis, square-integrability is sometimes assumed even though generalizations to less restrictive function spaces are possible.

Definition 2.7 (Hilbert Space of Square Integrable Functions)

The space of square-integrable functions on a closed interval $\mathcal{I} \subset \mathbb{R}$ together with the norm $\langle f, g \rangle = \int_{\mathcal{I}} f(t)g(t)\mathrm{d}t$ is a Hilbert space, denoted by $\mathbb{L}_2(\mathcal{I})$.

A function $f : \mathcal{I} \to \mathbb{R}$ is called square-integrable if the following condition holds.

$$\int_{\mathcal{I}} [f(t)]^2 \, \mathrm{d}t < \infty \tag{1}$$

To give the space the properties that are typically desired, it is defined as a space of equivalence classes, where two functions are equivalent if they differ at most on a set of Lebesgue-measure zero. Without loss of generality we reduce our treatment to the case of $\mathcal{I} = [0, 1]$.

Atypically, Bugni and Horowitz 2021 define two square-integrable functions to be distinct if they differ on a non-empty set of Lebesgue-measure zero. This detail makes theoretical considerations, such as deriving the exact properties of the test, more complicated. However, as using this definition in the context of Bugni and Horowitz 2021 only adds alternatives against which the test cannot have any power by construction, this thesis is limited to the classical definition of $\mathbb{L}_2[0, 1]$. However, the distinction can be important for proving some aysmptotic properties of the test and should therefore be kept in mind. An overview of the differences between the definitions is provided in Appendix 11.2.

2.2 Bases of $\mathbb{L}_2[0,1]$

One commonly used orthonormal basis of $\mathbb{L}^2[0,1]$ is the Fourier Basis. It consists of a series of functions $\left(\psi_i^F(x)\right)_{i\in\mathbb{N}}$ taken from the sine-cosine form of the Fourier series.

$$\psi_i^F(x) = \begin{cases} 1 & \text{if} \quad i = 1 \\ \sqrt{2}\cos(\pi i x) & \text{if} \quad i \quad \text{is even} \\ \sqrt{2}\sin(\pi(i-1)x) & \text{otherwise} \end{cases} \tag{2}$$

Figure 2 shows the first seven Fourier basis functions on $[0,1]$.
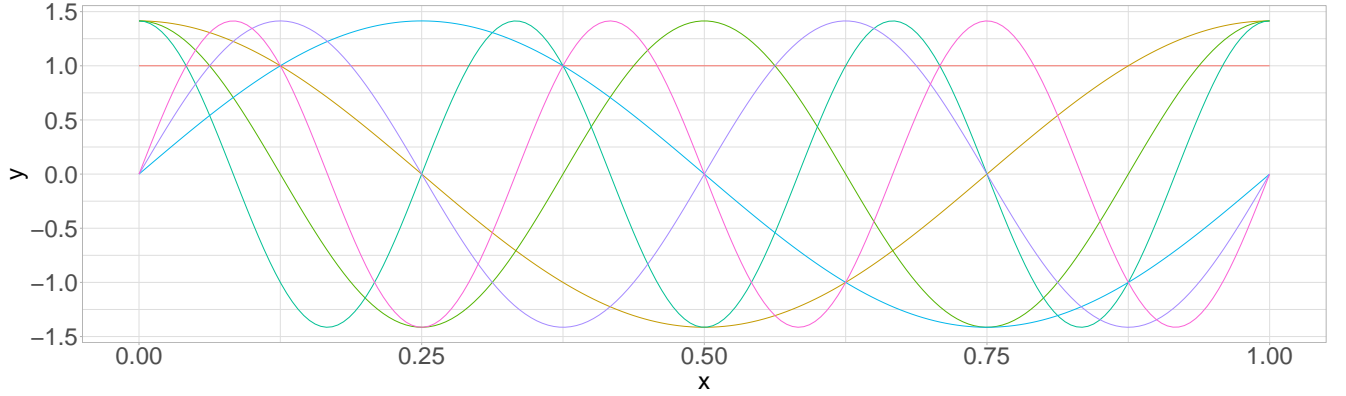
Figure 2: The first seven Fourier basis functions

A proof that the Fourier basis is an orthonormal basis of $\mathbb{L}^2[0,1]$ is given in section 2.4 of Hsing and Eubank 2015. As the Fourier basis is a countable orthonormal basis of $\mathbb{L}^2[0,1]$, it follows that $\mathbb{L}^2[0,1]$ is a separable Hilbert space. Many other bases are often used in functional data analysis, such as, for example, b-splines and orthogonal polynomials. The choice of the basis is typically motivated by the desired application, and the Fourier basis is therefore often chosen when the underlying curves are periodical.

2.3 Random Functions

In statistics, it is common to interpret data as realizations of random variables. In functional data analysis, the same is true. However, since the observations are now curves, it is relevant to consider the kinds of random variables that generate functional realizations. These are called random functions, and they are a special case of random variables. Paraphrasing from Bauer 2011 a general random variable is defined in the following way.

Definition 2.8 (Random Variable)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\Omega', \mathcal{A}')$ be a measure space. Then every $\mathcal{A}$-$\mathcal{A}'$-measurable function $X : \Omega \to \Omega'$ is called a $(\Omega', \mathcal{A}')$-random variable.

Since random functions are a special case, we can define them by specializing the objects in Definition 2.8.

Definition 2.9 (Random Function)

A random variable that realizes in a function space, e.g. $\mathbb{L}^2[0,1]$, is called a random function.

2.4 Probability Measures on $\mathbb{L}_2[0,1]$

For the test developed in Bugni and Horowitz 2021, it is important to evaluate expectations of a functional on $\mathbb{L}_2[0,1]$ with respect to some probability measure μ. Thus, it is necessary to explore probability measures on $\mathbb{L}_2[0,1]$. However, due to the way measures are constructed in the paper, this description is limited to probability measures induced by existing random functions. To understand how a random variable induces a measure, it suffices to look at a relatively basic fact from probability theory. Paraphrasing from Bauer 2011, let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\Omega', \mathcal{A}')$ be a random variable, then a probability measure $\mathbb{P}_X(B)$ on $\mathcal{A}'$ is induced by X as shown in Equation 3.

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) \quad \forall B \in \mathcal{A}' \tag{3}$$

This idea specializes to the case of random functions realizing in $\mathbb{L}_2[0,1]$ and thereby introduces the option to use measures induced by existing random functions. Therefore, assume that $Z(t)$ is a random function in $\mathbb{L}_2[0,1]$ inducing a measure μ. As previously explained, $Z(t)$ can be expressed in terms of a functional basis.

$$Z(t) = \sum_{k=1}^{\infty} b_k \psi_k(t) \quad \text{s.t.} \quad \sum_{k=1}^{\infty} b_k^2 \leq \infty \quad \text{a.s.} \tag{4}$$

The latter property is needed to ensure the square-integrability of the random function. The random function is, therefore, fully described by the distribution of a countably infinite, almost surely square-summable sequence of scalar random variables b_k. The theory of random sequences is out of the scope of this thesis. However, it is interesting to look at the relationship between the measure μ induced by the infinite sum and the measure μ_K induced by its finite counterpart shown in Equation 5.

$$Z_K(t) = \sum_{k=1}^{K} b_k \psi_k(t) \tag{5}$$

Understanding the relationship between these measures allows the justification of using a finite number of basis functions in the implementation of the test. Bugni, Hall, et al. 2009 show that the measure μ_K induced by $Z_K(t)$ converges to the measure μ induced by $Z(t)$ in the following sense. Let A be a μ-measurable subset of $\mathbb{L}_2[0,1]$. Then $\forall a \in A$, there is a unique countably infinite sequence of Fourier coefficients $b(a) = (b_k(a))_{k \in \mathbb{N}}$ given by $b_k(a) = \int_0^1 a(t)\psi_k(t)\mathrm{d}t$. As $a \in \mathbb{L}_2[0,1]$ these coefficients fulfill $\sum_{k=1}^{\infty} b_k^2 < \infty$.

To do so, Bugni, Hall, et al. 2009 define the following objects.

- $\mathbf{B} = \{b(a) \mid a \in A\}$

- $\mathbf{B}_K = \{(b_1, \ldots, b_K) \mid \exists b' \in B \; \forall k = 1, \ldots, K \quad b_k = b'_k\}$

- $\mathbf{A}_K = \{\sum_{i=1}^{\infty} b_k \psi_k(t) \mid (b_1, \ldots, b_K) \in \mathbf{B}_K \quad \text{and} \quad \sum_{k=1}^{\infty} b_k^2 < \infty\}$

The authors then show that for any A in the Borel sigma field of subsets of $\mathbb{L}_2[0,1]$, the following condition holds, where $\mathbb{P}$ denotes the probability measure associated with μ.

$$\lim_{K \to \infty} \mathbb{P}(\mathbf{A}_K) = \mathbb{P}(\mathbf{A}) \tag{6}$$

This argument allows using a truncated basis in the construction of a measure at a reasonably high truncation parameter K as integrals with respect μ_K converge to their counterparts with respect to μ as the truncation parameter K goes to infinity.

$$\lim_{K \to \infty} \int_{\mathbb{L}_2[0,1]} \mathcal{F}(z) \mathrm{d}\mu_K = \int_{\mathbb{L}_2[0,1]} \mathcal{F}(z) \mathrm{d}\mu \tag{7}$$

A more general treatment of these aspects would be mathematically very involved. Therefore, this section shall primarily serve to give intuition. For readers that are interested in a more detailed mathematical analysis, Gihman and Skorokhod 2004 and Skorohod 1974 give a rigorous treatment of the necessary theory on measures, probabilities, and integration in Hilbert spaces.

2.5   Functional Integration on $\mathbb{L}_2[0,1]$

To evaluate the expected value of the functional mentioned at the beginning of Section 2.4, it is necessary to integrate over a function space. Therefore, it is necessary to explore the ideas of functional integration and integration on separable Hilbert spaces. The formal definition of a functional integral over the functional $G[f]$ over $\mathbb{L}_2[0,1]$ is given as follows.

$$\int_{\mathbb{L}_2[0,1]} G\,[f]\,[Df] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G\,(f_1, f_2, \ldots) \prod_n \mathrm{d}f_n \tag{8}$$

As in the previous section, an in-depth treatment of integration on Hilbert spaces is out of the scope of this thesis. However, a detailed overview of integration on Hilbert spaces is given in Skorohod 1974 and could be used to derive theoretical properties of the test in a more general setting. In many cases, functional integrals over Hilbert spaces do not have closed solutions. Instead, it is often necessary to evaluate them using methods from perturbation theory. Because of the nature of the test in Bugni and Horowitz 2021, actually using these methods to approximate the functional integrals would be inadvisable due to the computational cost. Nevertheless, for theoretical considerations on the properties of Cramér-von Mises type tests in a functional setting, perturbation theory could be an interesting approach. An introduction to the necessary theory is given in Jeribi 2021.

In Bugni and Horowitz 2021, the test statistic is instead calculated using Monte-Carlo integration. Therefore, it is more useful to introduce Monte-Carlo integration with respect to some measure induced by a random variable. Like the more general concept of Monte-Carlo simulations, Monte-Carlo integration relies on randomness to approximate an object that might be difficult or impossible to evaluate algebraically.

To illustrate the principle, assume that the following integral has a fixed value in $\mathbb{R}$ and that μ is induced by a continuous, scalar-valued random variable Z realizing in $[0, 1]$.

$$V = \int_0^1 g(x)\mathrm{d}\mu = \int_0^1 g(x)f_Z(x)\mathrm{d}x \quad \text{where} \quad g : [0, 1] \to \mathbb{R} \tag{9}$$

Assuming that it is possible to draw realizations of Z, an intuitive approach to approximate V is given by the following two equations.

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n} g(Z_i) \quad \text{where} \quad Z_i \sim_{\text{i.i.d.}} Z \quad i = 1, \ldots, n \tag{10}$$

Assuming that $Var(g(Z)) < \infty$, this approach is justified because $\hat{V} \to_p V$ as shown by the following equations.

$$\mathbb{E}[\hat{V}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i)\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[g(Z_i)\right] \overset{Z_i \text{ i.i.d.}}{=} \mathbb{E}\left[g(Z)\right] = V \tag{11}$$

$$Var\left(\hat{V}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i)\right) \overset{Z_i \text{ i.i.d.}}{=} \frac{1}{n} Var\left(g(Z)\right) \to 0 \quad \text{as} \quad n \to \infty \tag{12}$$

Therefore, assuming that expectation and variance are finite, this approach will give approximations of the expectation that converge to the correct value as the number of realizations of Z used in the Monte-Carlo integration goes to infinity. As the principle of this method does not depend on the domain of g, it can analogously be used in the context of function spaces. It gives a convenient alternative to the theoretically complex methods mentioned before and is, therefore, a useful alternative in the context of this thesis. More information on Monte-Carlo methods in general and Monte-Carlo integration in particular can be found in Shonkwiler and Mendivil 2009.

3 Cramér-von Mises Tests

The second prerequisite to understanding the test presented in Bugni and Horowitz 2021 is some knowledge about Cramér-von Mises type tests. To understand their working principle and to gain some intuition about how they interact with different scenarios, it is useful to start in a simple setting, which is why this section is limited to the case of scalar random variables.

In economics, it is often interesting to ask whether the same stochastic process generated the observations in two distinct data sets. In an experimental setting, we could ask whether a treatment assigned at random to a subset of agents changed the distribution of an outcome variable. One approach to answering this question is given by the two-sample Cramér-von Mises test.

3.1 Empirical Distribution Functions

The idea of Cramér-von Mises type tests is to compare the distribution functions of the underlying random variables. However, as these are typically unknown, it is necessary to define a sample analog to construct a suitable test statistic. Gibbons and Chakraborti 2021 define the order statistic of an observation in a sample and the empirical distribution function in the following way, allowing the construction of the test.

Definition 3.1 (Order Statistic)
Let $\{x_i \mid i = 1, \ldots, n\}$ be a random sample from a population with continuous cumulative distribution function F_X. Then there almost surely exists a unique ordered arrangement within the sample.

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

$X_{(r)} \quad r \in \{1, \ldots, n\}$ is called the rth-order statistic.

Definition 3.2 (Empirical Distribution Function)
Assuming a continuously distributed random variable where there are no binds with probability one, we define the Empirical Distribution Function as the following step function.

$$\hat{F}_n(x) = \begin{cases} 0 & \text{if} \quad x < x_{(1)} \\ \frac{r}{n} & \text{if} \quad x_{(r)} \leq x < x_{(r+1)} \\ 1 & \text{if} \quad x \geq x_{(n)} \end{cases} \tag{13}$$

An example plot is given in Figure 11 in Appendix 11.9 illustrating the empirical distribution function for a selection of samples drawn from a standard normal. As this plot illustrates, the empirical distribution function gets increasingly close to the theoretical distribution function of the underlying random variable. In fact, using the strong law of large numbers it is easy to show that $\hat{F}_n(x) \overset{\text{a.s.}}{\to} F(x) \quad \forall x$ implying that $\hat{F}_n(x)$ is a consistent estimator of $F(x)$.

3.2 Null Hypothesis

As previously mentioned, the null hypothesis of Cramér-von Mises tests relates to distributional equality of random variables. This is different from tests such as the t-test which only tests for differences in specific properties of random variables like their mean.

To formulate the precise null hypothesis, let $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ be two data sets generated by random variables $X \sim_{\text{i.i.d.}} F(t)$ and $Y \sim_{\text{i.i.d.}} G(t)$. The null hypothesis that both samples were independently generated by random variables following the same cumulative distribution function is then formulated as follows.

$$
\begin{aligned}
H_0 &: F(t) = G(t) \quad \forall t \in \mathbb{R} \\
H_1 &: \exists t \in \mathbb{R} \quad \text{s.t.} \quad F(t) \neq G(t)
\end{aligned}
\tag{14}
$$

3.3 Two-Sample Cramér-von Mises Statistic

The Cramér-von Mises test is motivated by the idea that differences between the empirical distribution functions should be small under the null hypothesis. Therefore, the test statistic is based on the integrated squared difference of the empirical distribution functions to quantify the difference between two functions. Adapting a definition from Büning and Trenkler 2013 to the notation used in this thesis, it is defined as shown in Equation 15.

$$
C_{m,n} = \frac{nm}{n+m} \int_{-\infty}^{\infty} \left(\hat{F}_m(t) - \hat{G}_n(t) \right)^2 \mathrm{d} \left(\frac{m\hat{F}_m(t) + n\hat{G}_n(t)}{m+n} \right)
\tag{15}
$$

Darling 1957 gives an overview on the Cramér-von Mises test and some related tests in both the one-sample goodness-of-fit and the two-sample setting. Anderson 1962 explores the small sample distribution of the two-sample test statistic and provides a comparison to the limiting distribution derived by Rosenblatt 1952 and Fisz 1960 which is given in Section 11.3 in the Appendix. These distributions can be used to perform tests with appropriate critical values and have been implemented in many packages for statistical computing.

It is possible to generalize this test by introducing a weight function $w(x)$ in the following way.

$$
T_{m,n}^{EDF} = \frac{nm}{n+m} \int_{-\infty}^{\infty} \left(\hat{F}_m(t) - \hat{G}_n(t) \right)^2 w(t)\mathrm{d} \left(\frac{m\hat{F}_m(t) + n\hat{G}_n(t)}{m+n} \right)
\tag{16}
$$

The Cramér-von Mises test corresponds to $w(t) = 1$. It is advisable to think about the desired properties of the test when choosing a specific weight function as it can affect the power against different types of alternatives. A well-known weight function for the one-sample case was proposed by Anderson and Darling 1952 and Anderson and Darling 1954 and adapted to the two-sample setting by Pettitt 1976. The latter is given by $w(t) = \frac{m\hat{F}_m(t) + n\hat{G}_n(t)}{m+n}$ and has been shown to have better properties in some scenarios.

4 Permutation Tests

The third and final component to understanding Bugni and Horowitz 2021 is to understand the principles of permutation testing. In layman's terms, the idea of a permutation test

is the following: if two samples show distinctly different properties, that will lead to differences in an appropriately chosen summary statistic. If one were to permute the elements of the samples randomly, one would expect these differences in the test statistic to disappear. Permutation tests formalize this intuition. The following section closely follows chapter 15 from Lehmann and Romano 2005 with some additional input from Vaart and Wellner 1996.

4.1 Functional Principle of Permutation Tests

One of the defining features of a test is its null hypothesis. For the case of randomization tests, it is often formulated in the sense of distributional equality making the approach particularly interesting for the questions studied in this thesis. Therefore, the following section gives an overview of the principles of randomization testing in general and permutation testing in particular.

Let X be a random variable realizing in a sample space $\mathcal{X}$. Then, a typical null hypothesis in randomization testing is that the distribution P generating X belongs to a family of distributions Ω_0. For this very general setting, Lehmann and Romano 2005 define an assumption called the randomization hypothesis that allows for the construction of randomization tests.

Assumption 4.1 (Randomization Hypothesis)
Let G be a finite group of transformations $g : \mathcal{X} \to \mathcal{X}$. Under the null hypothesis of the randomization test, the distribution of X is invariant under the transformations $g \in G$. In other words, gX and X have the same distribution whenever X has distribution $P \in \Omega_0$.

One setting where this hypothesis holds is the question the Cramér-von Mises test covered in Section 3 tries to answer. If both samples were generated by the same random variable as the null hypothesis states, the distribution of the permuted samples would be identical to the distribution of the original samples.

Definition 4.1 (Randomization Distribution)
The randomization distribution of T is defined by the following distribution function.

$$\hat{R}(t) = \frac{1}{M} \sum_{g \in G} \mathbb{1}\left[T(gX) \leq t\right]$$

It describes the distribution of the test statistic T under the assumption of a uniformly distributed choice of transformation g. As X is itself a random variable, $\hat{R}(t)$ is random too.

Under the randomization hypothesis one can construct a permutation test based on any test statistic $T : \mathcal{X} \to \mathbb{R}$ that is suitable to test the null hypothesis under consideration. Suppose that G has M elements, then given $X = x$, let

$$T_{(1)}(x) \leq T_{(2)}(x) \leq \cdots \leq T_{(M)}(x)$$

be the ordered values of the test statistic $T(gx)$ as described in Definition 3.1 as g varies over G. For a fixed nominal level $\alpha \in (0, 1)$, define $k = M - \lfloor M\alpha \rfloor$ and the following objects.

$$M^+ = \sum_{m=1}^{M} \mathbb{1}\left[T_{(m)}(x) > T_{(k)}(x)\right] \qquad M^0 = \sum_{m=1}^{M} \mathbb{1}\left[T_{(m)}(x) = T_{(k)}(x)\right] \qquad (17)$$

Using these objects, we can define a test function that decides whether the null hypothesis is rejected based on the value of the test statistic calculated on the non-permuted samples T and a chosen size α for the test.

Definition 4.2 (Randomization Test Function)

We define the Randomization Test Function as a function $\phi : \mathcal{X} \to \mathbb{R}$ in the following way.

$$\phi(x) = \begin{cases} 1 & \text{if } T > T_{(k)}(x) \\ a & \text{if } T = T_{(k)}(x) \\ 0 & \text{if } T < T_{(k)}(x) \end{cases} \quad \text{where} \quad a = \frac{M\alpha - M^+(x)}{M^0(x)}$$

This function can be interpreted as a decider of whether the corresponding null hypothesis is rejected. If $\phi(X) = 1$, the null hypothesis is rejected, if $\phi(X) = 0$ it cannot be rejected, and if $\phi(X) = a$, the decision is randomized using a Bernoulli variable that is 1 with probability a. Under Assumption 4.1, it is possible to show that, given a test statistic $T = T(X)$, the resulting test ϕ has size α.

$$\mathbb{E}_P\left[\phi(X)\right] = \alpha \quad \forall P \in \Omega_0 \qquad (18)$$

This effectively means, that a permutation test constructed in the described way, always has the correct specified size. However, the power against alternatives can be highly sensitive to both the sample sizes and the number of permutations that was used to derive the permutation distribution. In a similar way, the p-value of a randomization test can be calculated as shown in Equation 19.

$$p\text{-value} = \frac{1}{M} \sum_{g \in G} \mathbb{1}\left[T(gX) \geq T(X)\right] \qquad (19)$$

Lehmann and Romano 2005 explore the example of testing for the equality of the generating probability laws of two independent samples. This is precisely the relevant application for the permutation variant of the two-sample Cramér-von Mises test that Bugni and Horowitz 2021 extend to the setting of functional data.

Definition 4.3 (Permutation)

Let S be a set, then a permutation of S is a bijective function $\pi : S \to S$.

If S is a finite set with N elements, there are $N!$ different permutations. If we apply this idea to the setting of two samples with n and m observations respectively, there are $(n + m)!$ permutations in the combined set of observations. One way of describing the corresponding group of transformations G is shown in Equation 20.

$$\Pi_N = \{\pi : \{1, \ldots, N\} \to \{1, \ldots, N\} \mid \pi \text{ is bijective}\}$$
$$G = \{g : \mathbb{R}^N \to \mathbb{R}^N \mid \exists \pi \in \Pi_N \ \forall x \in \mathbb{R}^N \ g(x) = (x_{\pi(1)}, \ldots, x_{\pi(N)})\}$$
(20)

Using G as described in Equation 20 in the construction of a randomization test leads to the concept of permutation tests. These are special cases of the testing procedure introduced before and rely on the specific choice of G. Assuming that the test statistic is invariant over intra-sample permutations, it is possible to obtain identical results using combinations instead of permutations. As the number of combinations is $\binom{m+n}{m}$ and thereby considerably smaller than the number of permutations, this approach can be beneficial in some scenarios.

In reality, it is often infeasible to calculate a test statistic on M permuted samples. Therefore, it is typical to randomly sample a chosen number B of permutations used to approximate the corresponding values. These stochastic approximations are justified for large numbers of samples permutations, as the corresponding p-values converge in probability to the p-value associated with the test using all permutations.

For the implementation of the test described in Bugni and Horowitz 2021 used in this thesis, distinct combinations were chosen if it was feasible to calculate the test statistic for all combinations. If this approach is infeasible due to the number of combinations, the implementation resorts to randomly sampling permutations with replacement instead.

4.2   Asymptotic Properties

One important question concerning the properties of randomization tests is their asymptotic properties when the number of observations goes to infinity. This section follows 15.2.2 from Lehmann and Romano 2005. Concerning these asymptotic properties, it is necessary to consider not only a sequence of samples increasing in size but a sequence of settings, where the prerequisite objects change accordingly. Consider therefore a sequence of settings with $X = X^N$, $P = P_N$, $\mathcal{X} = \mathcal{X}_N$, $G = G_N$, $T = T_N$ etc. for an increasing sequence of total number of observations N. Defining the randomization distribution of T_N analogously to Definition 4.1 as $\hat{R}_N(t) = \frac{1}{M_N} \sum_{g \in G_N} \mathbb{1}\left[T_N(gX^N) \leq t\right]$ it is possible to analyze the limiting behavior of $\hat{R}_N(t)$ as $N \to \infty$ under the null hypothesis.

As previously mentioned, the value of the randomization distribution function is itself a random variable. Thus, observe that for a random variable $\mathcal{G}_N$ that is uniformly distributed on G_N, the following holds under the randomization hypothesis.

$$\mathbb{E}[\hat{R}_N(t)] = \mathbb{P}\left[T_N(\mathcal{G}_N X^N) \leq t\right] = \mathbb{P}\left[T_N(X^N) \leq t\right]$$
(21)

Assuming that $T_N(X^N)$ converges in distribution to a stable limiting distribution with cumulative distribution function $R(t)$ continuous at t, it follows that $\mathbb{E}[\hat{R}_N(t)] \to R(t)$. In addition it is possible to show that under the null hypothesis $\hat{R}_N(t) \to_P R(t)$ at all continuity points t of $R(t)$. The proof of the latter statement given in Lehmann and Romano 2005 courtesy of Hoeffding 1952 is given in Appendix 11.4.

It should be noted that to make use of the stable limiting distribution of, for example, the Cramér-von Mises test, as described in Appendix 11.3, it is necessary to let the ratio of observations in the two samples, n and m, go to a constant λ as the total number of observations N increases. A similar argument is necessary in the case of the Cramér-von Mises type test in Bugni and Horowitz 2021.

5 Test by Bugni and Horowitz (2021)

Similar to the idea of the Cramér-von Mises tests in a scalar setting, Bugni and Horowitz 2021 devise a test for the equality of the distributions of the data generating processes of two independent samples of functional data. To define the exact hypothesis, they define a distribution function for random variables realizing in $\mathbb{L}_2[0,1]$ as follows.

Definition 5.1 (Distribution Function of a Random Function)
Let $X : \Omega \to \mathbb{L}_2[0,1]$ be a random function realizing in the square-integrable functions. Then its distribution function is defined as the following object.

$$F_X(z) = \mathbb{P}\left[X(t) \leq z(t) \quad \forall t \in [0,1]\right] \quad z \in \mathbb{L}_2[0,1]$$

As previously mentioned, the authors assume that two functions $z_1, z_2 \in \mathbb{L}_2[0,1]$ are distinct even if they only differ on a set of Lebesgue-measure zero. However, as this convention only adds alternatives against which the test cannot have any power by construction, this thesis will only cover the case created by the typical definition of $\mathbb{L}_2[0,1]$. Appendix 11.2 gives a short overview of how this convention influences the theoretical objects used by the test statistic.

Tests of this nature exist in some forms already and the following list shall give a non-exhaustive overview of other procedures that could be applied.

- Hall and Tajvidi 2002

- Hall and Van Keilegom 2007

- Pomann, Staicu, and Ghosh 2016

- Andere Tests auffuehren

5.1 Assumptions

As for most tests, there are assumptions that have to be fulfilled for the test to work correctly. The authors explicitly make four assumptions, some of which can be relaxed to allow for a more general setting.

Assumption 5.1   1. $X(t)$ and $Y(t)$ are separable, μ-measurable stochastic processes.

2. $\{X_i(t) \mid i = 1, \ldots, n\}$ is an independent random sample of the process $X(t)$. $\{Y_i(t) \mid i = 1, \ldots, m\}$ is an independent random sample of $Y(t)$ and is independent of $\{X_i(t) \mid i = 1, \ldots, n\}$.

As this assumption contains multiple points, it is useful to address them individually.

- The assumptions from the scalar Cramér-von Mises test, independence of observations in each sample and independence between the samples, carry over to allow for a construction of a Cramér-von Mises type test in a functional setting. Without these, the idea of using an empirical distribution function to detect differences between the samples fails without additional assumptions that allow for the estimation of the distribution functions.

- The Cramér-von Mises type test hinges on the measurability of the processes with respect to the chosen measure μ. Otherwise, the underlying idea to use functions generated according to μ to identify differences between the functions is not theoretically justified. Therefore, measurability with respect to μ has to be fulfilled for the test to be applicable.

- Separability of the stochastic processes is a less common assumption. Paraphrasing from Gihman and Skorokhod 2004 a separable stochastic process is defined as shown in Definition 5.2.

Definition 5.2 (Separable Stochastic Process)
A real-valued random function $X(t, \omega)$ – here denoted explicitly as a function of ω – with an associated probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is separable if:

1. There is a dense countable subset $I \subset T$ of its index set T

2. There is a set $\Omega_0 \subset \Omega$ of probability zero, so $\mathbb{P}(\Omega_0) = 0$, such that for an arbitrary open set $G \subset T$ and an arbitrary closed set $F \subset \mathbb{R}$ the two sets $\{\omega \mid X(t, \omega) \in F \quad \forall t \in G\}$ and $\{\omega \mid X(t, \omega) \in F \quad \forall t \in G \cap I\}$ differ from each other at most on Ω_0

In less theoretical terms this means that the properties of the process are determined by its behavior on a countable subset of points of its index set. This assumption is useful to prove some results on the asymptotic properties of the permutation test based on the Cramér-von Mises statistic. However, this assumption could potentially be relaxed.

Assumption 5.2

$\mathbb{E}\left[X(t)\right]$ and $\mathbb{E}\left[Y(t)\right]$ exist and are finite for all $t \in [0,1]$.

Assumption 5.2 is necessary for one of the two tests used in Bugni and Horowitz 2021. The test uses the squared difference between the sample mean functions to detect differences in the respective mean functions. Therefore it relies on the existence and finiteness of the mean functions in every point of the domain.

Assumption 5.3

$X_i(t)$ and $Y_i(t)$ are observed for all $t \in [0,1]$.

Assumption 5.3 can be relaxed and a similar test can be constructed for the case of discretely observed processes. This variation of the test will not be addressed in this thesis. However, Bugni and Horowitz 2021 provide a description of how to extend their idea to this common scenario.

5.2   Null Hypothesis

As with most permutation tests, the null hypothesis of the test presented by Bugni and Horowitz 2021 is distributional in nature. In the following let $F(z)$ denote the distribution function of the random function $X(t)$ and $G(z)$ denote the distribution function of the random function $Y(t)$. Then the null hypothesis can be formulated in the following way.

$$
\begin{aligned}
H_0: & \quad F(z) = G(z) \quad \forall z \in \mathbb{L}_2[0,1] \\
H_1: & \quad \mathbb{P}_\mu\left[F(Z) \neq G(Z)\right] > 0
\end{aligned}
\tag{22}
$$

Here, μ is a probability measure on $\mathbb{L}_2[0,1]$ and Z is a random function with probability distribution μ. As can be seen by the formulation of the hypothesis, this test does not intend to have power against alternatives that differ only on a set of functions that have μ-measure zero. Thereby, the choice of the measure is an important factor to consider when trying to use this test against a specific suspected alternative.

5.3   Cramér-von Mises type Test

Similar to the Cramér-von Mises test described in Section 3, the test constructed by Bugni and Horowitz 2021 relies on the empirical distribution functions describing the samples. Therefore, it is necessary to define empirical distribution functions for the functional setting which is immediately inspired by the distribution functions described in Definition 5.1.

$$
\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left[X_i(t) \leq z(t) \ \forall t \in [0,1]\right] \qquad \hat{G}_m(z) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}\left[Y_i(t) \leq z(t) \ \forall t \in [0,1]\right]
\tag{23}
$$

As in the scalar setting described in Section 3, the idea of the test is that the differences between these empirical distribution functions are expected to be comparatively small under the null hypothesis.

$$\tau = \int_{\mathbb{L}_2[0,1]} [F(z) - G(z)]^2 \, d\mu(z) \tag{24}$$

Sample analog:

$$\tau_{n,m} = (n+m) \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z) \right]^2 d\mu(z) \tag{25}$$

As algebraic functional integration is not suitable for this testing procedure due to the structure of the underlying objects, the authors use Monte-Carlo integration as explained in Section 2.5 for the approximation of this integral. To perform the test, it is therefore necessary to choose a parameter L that determines the number of functions $\{Z_l \mid l = 1, \ldots, L\}$ that are used to approximate the value of $\tau_{n,m}$ by the following equation.

$$\hat{\tau}_{n,m} = \frac{n+m}{L} \sum_{l=1}^{L} \left[\hat{F}_n(Z_l) - \hat{G}_m(Z_l) \right]^2 \tag{26}$$

These functions are drawn from a random function associated with the chosen measure μ, and their construction is explained in more detail in Section 5.5. At this point, it is interesting to ask whether Monte-Carlo integration is a suitable approach in this specific setting; in other words, whether the mean and variance of $\left[\hat{F}_n(z) - \hat{G}_m(z) \right]^2$ with respect to μ exist and are finite. Assuming existence, which depends on the chosen measure, an argument for finiteness is given in Appendix 11.5.

Using the theory from Section 4.3 the critical values for the permutation test statistic are then calculated in the following way. Here $\tau_{n,m,q}$ denotes the test statistic calculated on the samples permuted using permutation q assuming that Q permutations were considered overall.

$$t_{n,m}^{*,\tau}(1-\alpha) = \inf \left\{ t \in \mathbb{R} \quad | \quad \frac{1}{Q} \sum_{i=1}^{Q} \mathbb{1}\left[\tau_{n,m,q} \leq t \right] \geq 1 - \alpha \right\} \tag{27}$$

With these values it is now possible to perform the test at a chosen significance level.

5.4   Asymptotics for the Cramér-von Mises type Test

Similar to the case presented in Bugni, Hall, et al. 2009, it is possible to derive an asymptotic distribution for the Cramér-von Mises type test. Even though this is not necessary to perform the described permutation test, it could be an interesting benchmark to compare the test procedure in future large sample simulations. The authors mention that under the null hypothesis and mild regularity conditions, $\frac{1}{\sqrt{n+m}} \left[\hat{F}_n(z) - \hat{G}_m(z) \right]$ converges to a mean zero Gaussian process $\Upsilon(z) \quad z \in \mathbb{L}_2[0,1]$. Here, $G(z)$ again denotes the distribution function corresponding to the second sample and $\hat{G}_m(z)$ its empirical equivalent. $\Upsilon(z)$ has the covariance function shown in Equation 28.

$$cov\left[\Upsilon(z), \Upsilon(\tilde{z})\right] = \left[\frac{(1+\lambda)^2}{\lambda}\right]\{G(\min(z, \tilde{z})) - G(z)G(\tilde{z})\}$$

$$\text{where} \quad \min(z, \tilde{z})(t) = \min(z(t), \tilde{z}(t)) \quad \forall t \in [0, 1] \tag{28}$$

Here, as in the case of the limiting distribution of the scalar Cramér-von Mises test described in Appendix 11.3, λ denotes the limiting ratio of observations in the two samples as $N = n + m \to \infty$, meaning $\frac{n}{m} \to \lambda$. Thus, the following limiting distribution can be obtained for $\tau_{n,m}$.

$$\tau_{n,m} \to_d \int_{\mathbb{L}_2[0,1]} \Upsilon^2(z) \, \mathrm{d}\mu(z) \tag{29}$$

The derivation of this limiting distribution follows a similar argument as the proof of Theorem 3.1 given in Bugni, Hall, et al. 2009. It is presented in Appendix 11.6. Using the arguments presented in Appendix 11.6 and Section 4.2, it is possible to show that this limiting distribution also applies to the permutation variant of the Cramér-von Mises type test described in this thesis.

In the context of the limiting distribution, it is necessary to consider the implications of the definition of $\mathbb{L}_2[0, 1]$ as given by Bugni and Horowitz 2021. Where previously, it was sufficient to consider the classical definition of $\mathbb{L}_2[0, 1]$, it can make a difference in the asymptotic considerations. Even though this distinction will not be considered further in this thesis, for theoretical considerations of this test, it is important that, for example, $\Upsilon(z)$ is defined with respect to the definition of $\mathbb{L}_2[0, 1]$ from Bugni and Horowitz 2021.

5.5 Construction of the Measure μ

As can be seen in the fact that the null hypothesis is formulated with respect to a measure μ on $\mathbb{L}_2[0, 1]$, this measure significantly impacts the test's performance. Compared to the scalar Cramér-von Mises test described in Section 3, it fulfills a similar function to the weight function in that it determines against which alternatives the test has comparatively high power. It is, therefore, crucial to choose the measure according to the kind of violation of the null hypothesis that is suspected. If no specific type of violation is expected, a choice similar to a constant weight function can be made.

As hinted at in Section 2.4, Bugni and Horowitz 2021 approach this problem by constructing a random function $Z(t)$ that induces a suitable probability measure μ as described in Section 2.4.

$$Z(t) = \sum_{k=1}^{\infty} b_k \psi_k(t) \quad \text{s.t.} \quad \sum_{k=1}^{\infty} b_k^2 < \infty \quad \text{a.s.} \tag{30}$$

By truncating the random function to use a finite amount of basis functions as shown in Equation 31, it is possible to draw realizations from it given distributions for the Fourier

coefficients. As explained in Section 2.4, the measure induced by this truncated random variable converges to the measure induced by the full random variable as the truncation parameter K goes to infinity. This fact justifies the approach theoretically.

$$Z_K(t) = \sum_{k=1}^{K} b_k \psi_k(t) \tag{31}$$

The construction now hinges on specifying appropriate distributions for the Fourier coefficients. To obtain a reasonable degree of power it is important to specify the distribution of the random function $Z(t)$ such that its realizations are able to distinguish potential differences in the samples. This could for example be done by choosing a function $w(t) : [0,1] \to \mathbb{R}$ that is used to specify the expectation of $Z(t)$ dependent on the data. Specifying the expected values of the Fourier coefficients as shown in Equation 32 then leads to the expected value of the random function $Z_K(t)$.

$$\mathbb{E}\left[b_k\right] = \int_0^1 w(t)\psi_k(t)\ \mathrm{d}t \quad \text{leading to} \quad \mathbb{E}\left[Z_K(t)\right] = \sum_{k=1}^{K} \mathbb{E}\left[b_k\right] \psi_k(t) \tag{32}$$

Lastly, a distribution around the expected value has to be specified. In Equation 33 this takes the form of random errors U_k and square-summable factors $(\rho_k)_{k \in \mathbb{N}}$.

$$b_k = \mathbb{E}\left[b_k\right] + \rho_k U_k \quad \text{s.t.} \quad \sum_{k=1}^{\infty} \rho_k^2 < \infty \tag{33}$$

The authors suggest choosing $w(t)$ to be comparatively large where possible differences between processes in the two samples are expected to be large. The following two ways of choosing the mean function are shown in Figure 3.

1. In their application, Bugni and Horowitz 2021 set the distribution of the Fourier coefficients manually, as $b_1 \sim \mathcal{N}(\mu_1, \frac{1}{K})$ and $b_k \sim \mathcal{N}(0, \frac{1}{K}) \quad k = 2, \ldots, K$ where

$$\mu_1 = median_i \left\{ \max_{t \in [0,1]} X_i(t) \mid i = 1, \ldots, n, \right\}$$

   This method is equivalent to setting $w(t) = \mu_1 \quad \forall t \in [0,1]$, as the inner product of any constant function with Fourier basis functions of order two or higher will be zero due to their cyclical nature.

2. Another idea is to choose a non-constant weight function based on the reference sample. One option is to use specific point-wise quantiles of the reference sample: $w(t) = quantile_q\{X_i(t) \mid i = 1, \ldots n\}$. This allows the generated functions for the approximation of the Cramér-von Mises statistic to resemble the reference data more closely, which could improve the properties of the test.
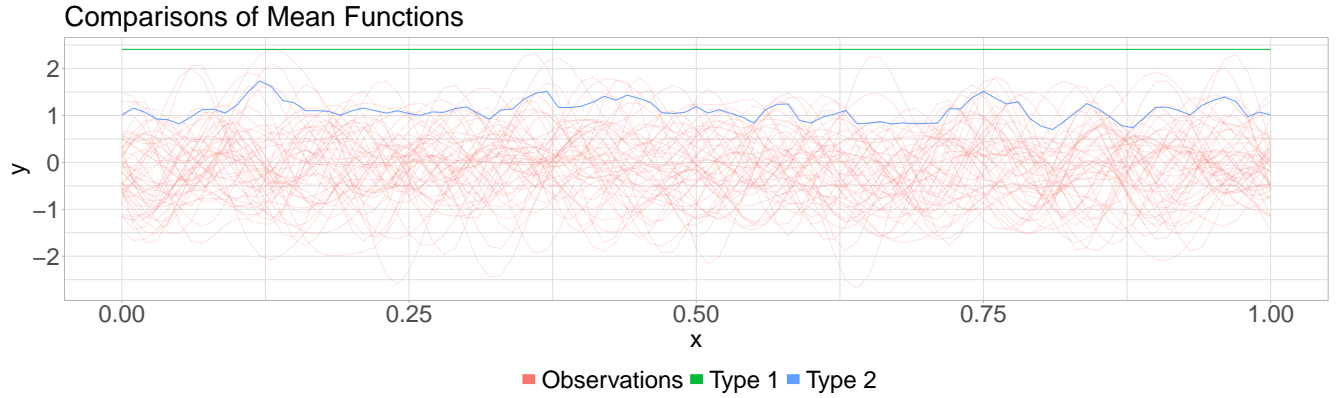
19

Figure 3: Mean Functions calculated for a Reference Sample as described in Section 7 for the 95% Quantile.

In addition to the mean function $w(t)$, the choice of the sequence $(\rho_k)_{k \in \mathbb{N}}$ and the distribution of the error terms U_k is of importance. The authors do not recommend a procedure for choosing $(\rho_k)_{k \in \mathbb{N}}$. However, it seems reasonable to choose parameters that lie in the vicinity of the standard deviations for the Fourier coefficients of the reference sample when fitting a Fourier basis to the original observations. Other methods, such as specifying a sequence without consideration of the data, are possible. In the application part of Bugni and Horowitz 2021, the authors, for example, use $\frac{1}{K}$ for every Fourier coefficient. Even though this sequence would not be square-summable in the infinite case, for a finite variant, this should not pose a large problem.

Concerning the choice of distribution for U_k, the authors recommend different choices depending on the properties of the samples. For samples that indicate a thin tailed distribution of $X(t)$, the authors recommend one of the following choices.

- $U_k \sim \mathcal{N}(0, 1)$
- $U_k \sim Unif\left[-3^{\frac{1}{3}}, 3^{\frac{1}{3}}\right]$

If instead, the sample hints at a heavy-tailed distribution, one should consider a heavy-tailed distribution for the Fourier coefficients, such as a t-distribution with a low number of degrees of freedom.

5.6 Mean-focused Test

In addition to the Cramér-von Mises type test, the authors use a second test statistic aimed at detecting violations due to a mean shift in the data generating process. This test is added due to low power against pure mean-shift alternatives in the original paper's simulations and is supposed to complement this weakness of the other test statistic.

The variable μ on which this second test is based is the distance between the mean functions of the data generating processes as induced by the natural norm on $\mathbb{L}_2[0, 1]$.

$$\nu = \int_0^1 \left(\mathbb{E}\left[X(t)\right] - \mathbb{E}\left[Y(t)\right] \right)^2 \mathrm{d}t \tag{34}$$

Replacing the expected values by their sample equivalents leads to a test statistic that

can be used to perform a permutation test. The sample equivalents are defined as shown in Equation 35 and the resulting test statistic is shown in Equation 36.

$$\hat{\mathbb{E}}\left[X(t)\right] = \frac{1}{n}\sum_{i=1}^{n} X_i(t) \qquad\qquad \hat{\mathbb{E}}\left[Y(t)\right] = \frac{1}{m}\sum_{i=1}^{m} Y_i(t) \qquad (35)$$

$$\nu_{n,m} = (n+m)\int_0^1 \left[\hat{\mathbb{E}}\left[X(t)\right] - \hat{\mathbb{E}}\left[Y(t)\right]\right]^2 dt \qquad (36)$$

In the same way the critical values for the Cramér-von Mises test are derived using the techniques described in Section 4.3, the critical values for the mean-based test are calculated as shown in Equation 37. Analogous to the previous case, $\nu_{n,m,q}$ denotes the test statistic calculated on the permuted samples corresponding to permutation q.

$$t_{n,m}^{*,\nu}(1-\alpha) = \inf\left\{t \in \mathbb{R} \quad | \quad \frac{1}{Q}\sum_{i=1}^{Q} \mathbb{1}\left[\nu_{n,m,q} \le t\right] \ge 1-\alpha\right\} \qquad (37)$$

5.7 Combined Permutation Test

Using the test statistics described in Sections 5.3 and 5.6, it is possible to combine both into one test trying to harness the advantageous properties of both test statistics. Therefore, define for the two underlying tests the following permutation test functions as described in Definition 4.2 for the general case of a randomization test.

$$\phi_{n,m} = \begin{cases} 1 & \text{if } \tau_{n,m} > t_{n,m}^*(1-\alpha_\tau) \\ a_\tau & \text{if } \tau_{n,m} = t_{n,m}^*(1-\alpha_\tau) \\ 0 & \text{if } \tau_{n,m} < t_{n,m}^*(1-\alpha_\tau) \end{cases} \qquad \tilde{\phi}_{n,m} = \begin{cases} 1 & \text{if } \nu_{n,m} > t_{n,m}^*(1-\alpha_\nu) \\ a_\nu & \text{if } \nu_{n,m} = t_{n,m}^*(1-\alpha_\nu) \\ 0 & \text{if } \nu_{n,m} < t_{n,m}^*(1-\alpha_\nu) \end{cases} \qquad (38)$$

Here, a_τ and a_ν are once again given by the following equations to ensure that the expected values of ϕ and $\tilde{\phi}$ have the desired values.

- $a_\tau = \frac{Q\alpha_\tau - Q_\tau^+}{Q_\tau^0}$
- $Q_\tau^+ = \sum_{q=1}^{Q} \mathbb{1}\left[\tau_{n,m,q} > t_{n,m}^*(1-\alpha_\tau)\right]$
- $Q_\tau^0 = \sum_{q=1}^{Q} \mathbb{1}\left[\tau_{n,m,q} = t_{n,m}^*(1-\alpha_\tau)\right]$

- $a_\nu = \frac{Q\alpha_\nu - Q_\nu^+}{Q_\nu^0}$
- $Q_\nu^+ = \sum_{q=1}^{Q} \mathbb{1}\left[\nu_{n,m,q} > t_{n,m}^*(1-\alpha_\nu)\right]$
- $Q_\nu^0 = \sum_{q=1}^{Q} \mathbb{1}\left[\nu_{n,m,q} = t_{n,m}^*(1-\alpha_\nu)\right]$

By the theoretical considerations of Section 4.3, it is clear that under Assumption 5.1 and 5.2 and under the null hypothesis and any $\alpha_\tau, \alpha_\mu \in (0,1)$, the following statements hold. This means, that regardless of most choices that had to be made during the construction of the test, the size of the individual tests will be equal to the corresponding α.

$$\mathbb{E}_P\left(\phi_{n,m}\right) = \alpha_\tau \qquad\qquad \mathbb{E}_P\left(\tilde{\phi}_{n,m}\right) = \alpha_\nu \qquad (39)$$

To combine the two tests, the authors make use of the Bonferroni inequality. Under H_0 this approach leads to the following relationship motivating a choice of $\alpha_\tau + \alpha_\nu = \alpha$ for a chosen overall size for the combined test of α. More information on this approach is given in Appendix 11.7.

$$\max(\alpha_\tau, \alpha_\nu) \leq \mathbb{P}\left[(\phi_{n,m} > 0) \cup (\tilde{\phi}_{n,m} > 0)\right] \leq \alpha_\tau + \alpha_\nu \tag{40}$$

This opens the option to, for example, choose a size of 0.03 for the Cramér-von Mises type test and a level of 0.02 for the mean-based test, as the combined test will at most have a level of 0.05.

5.8 Asymptotic Properties under the Alternative

As alluded to in Sections 3 and 5.3, for the asymptotic properties of the test, it is useful to make another assumption on the sequence of tests that is considered as the number of observations goes to infinity.

Assumption 5.4
As $N = m + n \to \infty$ let $\frac{n}{m} \to \lambda$ for some $\lambda \in \mathbb{R}_{>0}$.

This assumption allows the derivation of a stable limiting distribution and helps with the asymptotic properties of the test if the null hypothesis is incorrect.

Under Assumptions 5.1 and 5.4, Bugni and Horowitz 2021 show that the following statement holds.

$$\int_{\mathbb{L}_2[0,1]} [F(z) - G(z)]^2 \, d\mu(z) > 0 \implies \lim_{N\to\infty} \mathbb{P}\left[\tau_{n,m} > t_{n,m}^{*,\tau}(1 - \alpha_\tau)\right] = 1 \quad \forall \alpha_\tau \in (0,1) \tag{41}$$

If additionally Assumption 5.2 holds, then the analogous statement holds for the mean-focused test.

$$\int_0^1 [\mathbb{E}[X(t)] - \mathbb{E}[Y(t)]]^2 \, dt > 0 \implies \lim_{N\to\infty} \mathbb{P}\left[\nu_{n,m} > t_{n,m}^{*,\nu}(1 - \alpha_\nu)\right] = 1 \quad \forall \alpha_\nu \in (0,1) \tag{42}$$

The combination of these statements can be used to show that the combined test is consistent.

5.9 Problematic Curve Fitting

One crucial question in implementing the method is whether to employ curve-fitting using a functional basis to the samples. It seems natural to fit the same basis to the data that is used for constructing the measure. Then the comparison between functions that is necessary for the approximation of the Cramér-von Mises type statistic would reduce to a problem of finding the zeroes of a finite Fourier-series, which could be approached with

methods such as those presented in Boyd 2006. However, this can lead to problems, one of which is illustrated by Figure 4, which shows curves that were fitted to samples generated by two data generating processes.
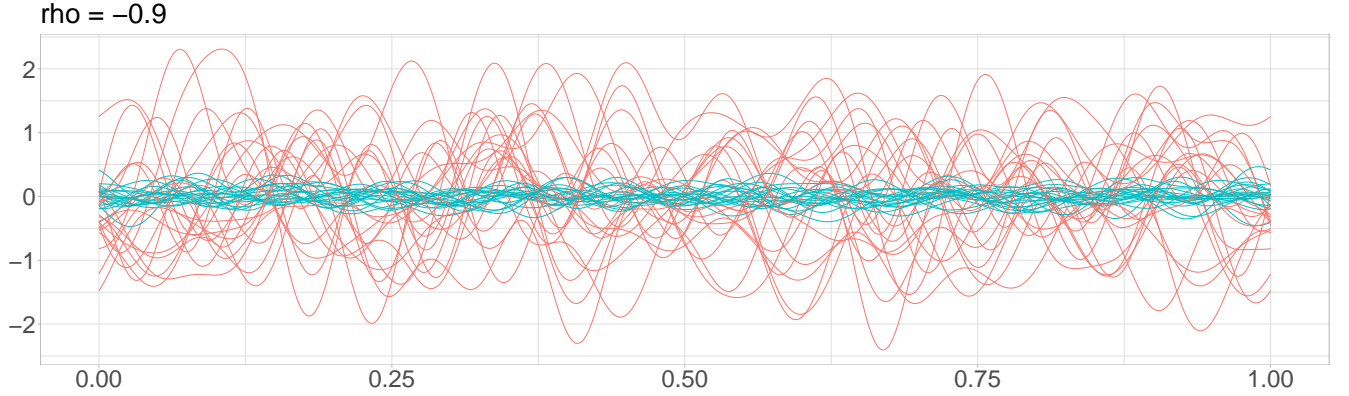

Figure 4: Artifact of the Curve Fitting Procedure

These processes are constructed as described in Section 7 for a choice of $\rho_2(t) = -0.9$. This implies that the data generating process generates data with a point-wise variance of one. However, due to the structure of the data, the curve fitting procedure fails to conserve this property. Therefore, artifacts, such as the one shown in Figure 4, can heavily influence the power of the constructed test, meaning that it could be advisable to change the approach to one of the following.

1. **Comparisons at the discrete data points.** Suppose observations are discrete but dense enough to be interpreted as smooth curves. In that case, one could use the points where data is observed to compare the functions generated for the approximation of the Cramér-von Mises test to the original samples. This approach is close to the one described in Bugni and Horowitz 2021 for the case of discrete observations.

2. **Linear Interpolation.** Instead, one could use linear interpolation of the original curves to compare them to the generated curves. This would preserve the functional nature while not generating the fitting artifacts shown in Figure 4.

The simulations in Sections 7 use the first approach, as it limits the computational cost of the procedure while staying close to the original paper. The same problem occurs for the mean-focused test. In this case, linear interpolation between the discrete points is used to approximate the mean function.

6   Test for Persistence Alternatives

One particular application for the Cramér-von Mises type test presented in Bugni and Horowitz 2021 is a slight modification leading to the following test for differences in the persistence structure of the data generating processes. There are already many tests

focusing on differences in the mean functions of processes. A non-exhaustive list of examples can be found in the following publications.

- J. Lee, D. Cox, and Follen 2015

- D. D. Cox and J. S. Lee 2008

- Hier tests auflisten

There is also a large number of tests for differences in the variance functions of processes such as for example the tests presented in the following papers.

- Hier tests auflisten

However, there are few tests that address differences in the persistence of processes and the proposed procedure adds to the limited arsenal.

6.1 Null-Hypothesis

Let $\{X_i(t) \mid i = 1, \ldots, n\}$ and $\{Y_i(t) \mid i = 1, \ldots, m\}$ again denote the samples under consideration. Assume that the samples consist of i.i.d. realizations of random functions $X(t)$ and $Y(t)$. To focus on the persistence properties of the random functions, define the standardized counterparts to these random functions in the following way.

$$
\begin{aligned}
\mu_X(t) &= \mathbb{E}\left[X(t)\right] & \mu_Y(t) &= \mathbb{E}\left[Y(t)\right] \\
\sigma_X(t) &= \sqrt{\mathbb{E}\left[X(t) - \mu_X(t)\right]^2} & \sigma_Y(t) &= \sqrt{\mathbb{E}\left[Y(t) - \mu_Y(t)\right]^2}
\end{aligned}
\tag{43}
$$

Then the standardized random variables can be defined as follows.

$$
\tilde{X}(t) = \frac{X(t) - \mu_X(t)}{\sigma_X(t)} \qquad\qquad \tilde{Y}(t) = \frac{Y(t) - \mu_Y(t)}{\sigma_Y(t)} \tag{44}
$$

Their distribution functions can be defined analogously to Section 5 and shall be denoted by $F_{\tilde{X}}(z)$ and $F_{\tilde{Y}}(z)$ respectively in the following. Then a modified null hypothesis similar to Bugni and Horowitz 2021 can be formulated with respect to the distribution functions of the standardized random variables.

$$
\begin{aligned}
H_0: &\quad F_{\tilde{X}}(z) = F_{\tilde{Y}}(z) \quad \forall z \in \mathbb{L}_2[0.1] \\
H_1: &\quad \mathbb{P}_\mu\left[F_{\tilde{X}}(z) \neq F_{\tilde{Y}}(z)\right] > 0
\end{aligned}
\tag{45}
$$

As the standardization eliminates all differences in these processes' mean and variance patterns, one of the distinct remaining features that might differentiate these processes is their persistence properties. Therefore, as the more general form is supposed to find very general alternatives, this focused approach could be tailored by choice of μ to detect differences in the persistence structure of the data generating processes.

6.2 Test-Procedure

The modified test is functionally nearly identical to the Cramér-von Mises type test described in Bugni and Horowitz 2021. However, it adds a standardization step in the calculation of the test statistic to eliminate differences in the mean functions of the processes and their variance structure. In each permutation step, the calculation of the test statistic is thus described by the following algorithm.

Let $\mathcal{W} = \{W_i(t) \mid i = 1, \ldots, n\}$ be sample one and $\mathcal{V} = \{V_i(t) \mid i = 1, \ldots, m\}$ be sample two. These samples are generated by randomly permuting the original samples $\{X_i(t) \mid i = 1, \ldots, n\}$ and $\{Y_i(t) \mid i = 1, \ldots, m\}$.

1. Calculate the sample mean functions for the permuted samples:
   $\bar{W}(t) = \frac{1}{n} \sum_{i=1}^{n} W_i(t)$ and $\bar{V}(t) = \frac{1}{m} \sum_{i=1}^{m} V_i(t)$

2. Center the sample curves to obtain centered permutation samples:
   $\bar{\mathcal{W}} = \{W_i(t) - \bar{W}(t) \mid i = 1, \ldots, n\}$ and $\bar{\mathcal{V}} = \{V_i(t) - \bar{V}(t) \mid i = 1, \ldots, m\}$

3. Calculate the point-wise standard deviations:
   $\sigma_{\mathcal{W}}(t) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[W_i(t) - \bar{W}(t)\right]^2}$ and $\sigma_{\mathcal{V}}(t) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left[V_i(t) - \bar{V}(t)\right]^2}$

4. Standardize the observations in the permuted samples:
   $\tilde{\mathcal{W}} = \{\frac{W_i(t) - \bar{W}(t)}{\sigma_{\mathcal{W}}(t)} \mid i = 1, \ldots, n\}$ and $\tilde{\mathcal{V}} = \{\frac{V_i(t) - \bar{V}(t)}{\sigma_{\mathcal{V}}(t)} \mid i = 1, \ldots, m\}$

5. Calculate $\hat{\tau}_{n,m}$ for the standardized permutation samples $\tilde{\mathcal{W}}$ and $\tilde{\mathcal{V}}$ as described in Section 5 for a chosen measure μ.

The overarching test procedure is again given by the derivation of a permutation distribution and a permutation test decision according to the rules stated in Section 4.

Due to the introduction of the standardization step, the problem described in Section 5.9 is potentially of even greater importance than in the previous section. As the test relies on standardizing the objects used for the calculation of the test statistic, it would be of immense importance whether curve fitting is employed and at what step of the procedure it is introduced. For example, if standardization would be applied to discrete observations, which are then fitted using a functional basis, the resulting curves would generally not be standardized in the desired way. This can be seen in Figures 12 and 13 in Appendix 11.9 which show samples that were generated by processes with a point-wise mean of 0 and a point-wise variance of 1 and their fitted counterparts. Using curve-fitting in this way could create simulations indicating a strong power against persistence alternatives. However, this power might entirely be based on the artifacts created by the curve fitting procedure. Therefore, it is advisable to either perform standardization on the fitted curves if curve fitting is desired because of the data itself, or to rely on comparisons between the discrete

observations and the curves created for the approximation of the Cramér-von Mises type test statistic. For the purposes of this thesis the latter approach will be used. This can be seen as analogous to the variant of the Cramér-von Mises type test for discrete data described in Bugni and Horowitz 2021.

7 Simulation Study

To learn more about the potential practical applications of this method, it is helpful to study its properties in a series of simulations. Bugni and Horowitz 2021 studied an array of different setups that give an idea of the performance in a collection of settings, and this thesis will replicate some of these results and explore some where the method struggles. The simulations presented as part of this thesis have been conducted on *bonna*[2]. *bonna* is the high performance computing cluster provided by the University of Bonn. However, slight modifications of the provided code suffice to run it on personal computers.

All analyses in this thesis have been conducted with R[3]. As a supplement to this thesis, the two-sample variant of the test taken from Bugni and Horowitz 2021 is implemented in an R package called *PermFDATest*. The R package and all code that has been used to produce the following results is publicly available as part of a GitHub repository[4] that complements this thesis. The package makes extensive use of the implementations by Ramsay, Graves, and Hooker 2021 and Wickham et al. 2019.

7.1 Simulation Setup

The simulations presented in this thesis, are inspired by the simulation study from Bugni and Horowitz 2021 and shine more light on specific settings that offer insight into the potential benefits and problems of the method. To describe the specific processes used in the simulation, this Section introduces the notation used in Bugni and Horowitz 2021 shortly to then describe the data generating processes mathematically.

Deviating from Bugni and Horowitz 2021, these simulations focus on the two-sample variant of the test. Additionally, the data are generated on the closed interval $[0, 1]$ and use a smaller sample size of 20 observations per sample. Therefore, let $\mathcal{T} = \{0, 0.01, 0.02, \ldots, 1\}$ be the index set of the discrete points on the unit interval, then the observations are constructed in the following way.

---

[2]https://www.dice.uni-bonn.de/de/hpc/hpc-a-bonn/infrastruktur
[3]R Core Team 2022.
[4]https://github.com/JakobJuergens/Masters_Thesis

1. Draw random variables $\{\xi_{i,s} \mid (i,t) \in \{1,\ldots,20\} \times \mathcal{T}; s = 1,2\}$ independently from the $\mathcal{N}(0,1)$ distribution.

2. For all $i = 1,\ldots,20$ and $s = 1,2$, set $\tilde{X}_{i,s}(0) = \xi_{i,s}(0)$.

3. For all $i = 1,\ldots,20$; $s = 1,2$ and $t = 0.01,\ldots,1$, set $\tilde{X}_{i,s}(t) = \rho_s(t)\tilde{X}_{i,s}(t - 0.01) + \xi_{i,s}(t)\sqrt{1 - \rho_s^2(t)}$ where $\rho_s(t)$ is a parameter as defined below.

4. For all $i = 1,\ldots,20$; $s = 1,2$ and $t \in \mathcal{T}$, set $X_{i,s}(t) = \mu_s(t) + \sigma_s(t)\tilde{X}_{i,s}(t)$ where $\mu_s(t)$ and $\sigma_s(t)$ are parameters as defined below.

The resulting random variables $\{X_{i,s}(t) \mid t \in \mathcal{T}; \ i = 1,\ldots,20; \ s = 1,2\}$ are normally distributed with the following properties.

1. $\mathbb{E}\left[X_s(t)\right] = \mu_s(t)$

2. $\mathrm{Var}\left[X_s(t)\right] = \sigma_s^2(t)$

3. $\mathrm{Corr}\left[X_s(t), X_s(t - 0.01)\right] = \rho_s(t)$
$\forall t \in \mathcal{T}$ with $t > 0.01$

This thesis focuses on three distinct violations of the null hypothesis described in the following. For each of the following settings sample 1 is generated from the same process as a reference point to compare the different settings. Therefore, the parameters for sample 1 are therefore chosen as follows.

1. $\mu_1(t) = 0 \quad \forall t = 0.01,\ldots,1$

2. $\sigma_1(t) = 1 \quad \forall t = 0.01,\ldots,1$

3. $\rho_1(t) = 0.5 \quad \forall t = 0.01,\ldots,1$

For the second sample, three different violations of the null hypothesis and a benchmark case where the null hypothesis is correct are generated.

1. **Identical Data Generating Processes**
   The first setting is the benchmark where both samples are generated using the same random function. This setting is meant to support the theoretical points arguing that the permutation tests described in this thesis have the correct size.

2. **Mean Shift**
   The second setting introduces a shifted mean function. This is one possible violation where Bugni and Horowitz 2021 describe weak power of the Cramer-von Mises type test. Deviating from sample 1, sample 2 has mean function

$$\mu_2(t) = x(x - 1) \quad t \in \mathcal{T}$$

3. **Correlation Shift**
   The third setting violates the null hypothesis by changing the correlation structure

between neighboring observations. Here, observations are less persistent due to a lower, but still constant choice of the correlation parameter.

$$\rho_2(t) = 0 \quad \forall t = 0.01, \ldots, 1$$

This setting can therefore be seen as testing for alternatives consisting of variations of the persistence structure of a random function.

4. **Variance Shift**

   In the fourth simulation setting, the variance point-wise function is altered in the following way.

   $$\sigma_2(t) = 1 + 0.5t \qquad t \in \mathcal{T}$$

   Therefore, towards the end of the interval a higher variance in the realized curves is generated by the process.
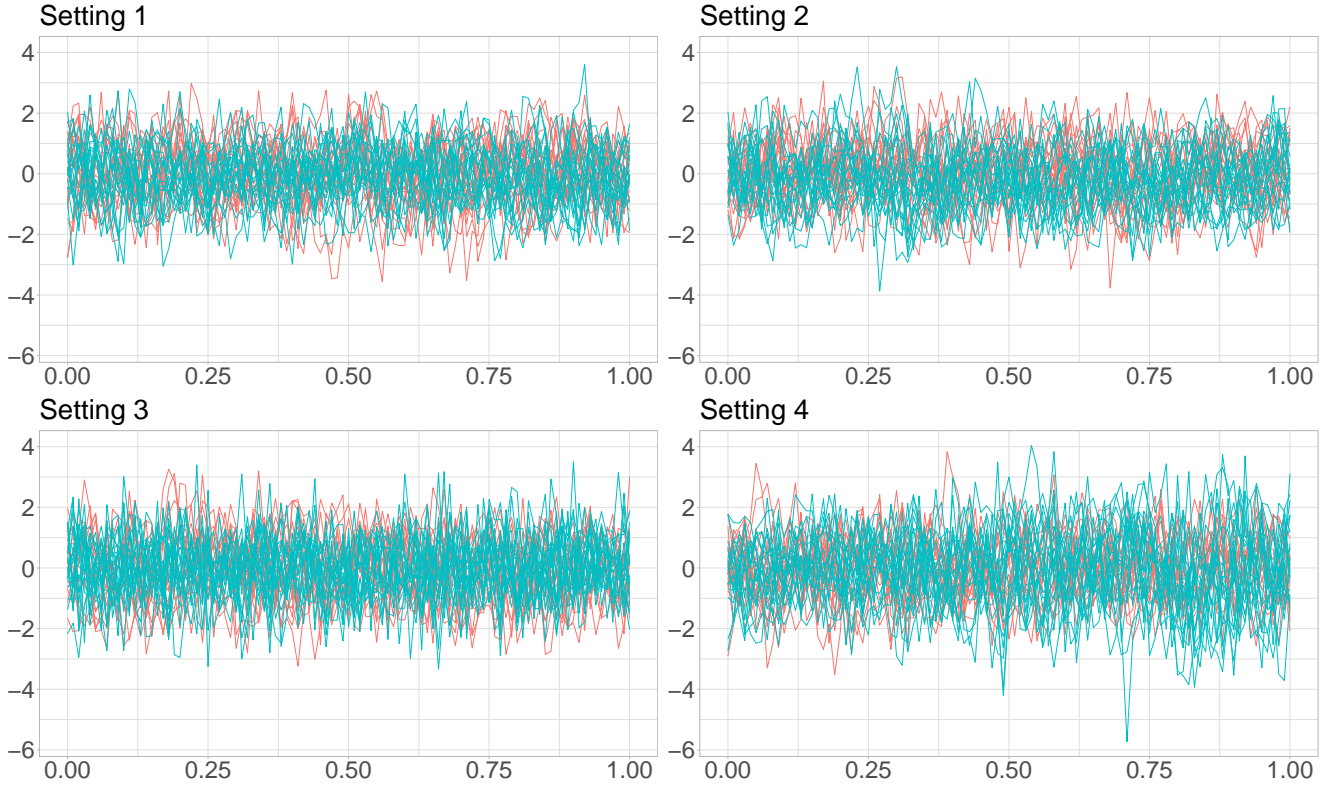

Figure 5: Samples generated for the four settings.
Sample 1 in red, Sample 2 in blue

As shown in Figure 5, the difference between the samples is optically quite subdued. Therefore, a high power against these alternatives is not expected. However, any nontrivial power shown in these settings is a good indicator for the qualities of the test.

As described in the sections describing the test procedure theoretically, there is a number of parameter choices that has to be made to perform the simulation.

1. As shown in the mathematical construction of the data, these simulations compare samples consisting of a smaller number of observations per sample $n = 20$ and focus on the two-sample version of the test. Additionally, each observation is constructed from a lower number of discrete observations (101) than in the original paper.

2. Concerning the choice of the truncation parameter K, which is used in the construction of the measure μ as shown in Equation 31, the authors argue that their simulations showed that the power of the test became flat at about $K = 20$ and chose $K = 25$. For the sake of reproducing their results, the same parameter was used in this thesis.

3. Another parameter that has to be chosen is L, determining the number of functions used in the approximation of the test statistic $\hat{\tau}_{n,m}$ by Monte-Carlo integration as shown in Equation 26. Bugni and Horowitz 2021 chose $L = 4000$ and preliminary testing confirmed that this choice seems reasonable in many settings when combined with reasonable choices for the construction of the measure μ. However, problems such as the one described in Section 5.9 might be closely connected to this choice.

4. For the number of permutations Q used for the approximation of the permutation distribution of both test statistics the authors chose a value of $Q = 500$. This thesis uses the same parameter. However, absent strong computational limitations a higher choice for Q could benefit the properties of the tests.

5. For the construction of the measure μ, the weight function $w(t)$ has to be chosen. In this simulation, the weight function is chosen as in the empirical application in Bugni and Horowitz 2021. This approach is described in Section 5.5 as the first approach listed for the choice of $w(t)$.

6. Deviating from the original authors' choice due to preliminary tests, the sequence $(\rho_i)_{i=1,\ldots,K}$ is determined by checking the sample standard deviation of the corresponding Fourier coefficients in the first sample obtained by fitting the same basis that is used in the construction of μ to the observations. These are then multiplied by a constant factor to increase variability in the functions drawn from μ. This higher variability seemed to improve performance in preliminary tests, but should be studied in further experiments.

7. Using the same choice as in the original paper, the simulations use independent standard normal error terms $U_k \quad k = 1, \ldots, K$ for the Fourier coefficients shown in Equation 33 for the construction of the measure. As the processes under consideration are Gaussian by construction and thereby have thin tails, this choice should be reasonable following the authors' argumentation.

7.2 Results

Taking a look at the results from the simulation study, it is interesting to compare the empirical rejection probabilities for different types of violations of the null hypothesis.

Test	$(\alpha_\tau, \alpha_\nu)$	Setting 1	Setting 2	Setting 3	Setting 4
Means-Test	$\alpha_\nu = 0.05$	0.058	0.5	0.057	0.057
CvM-Test	$\alpha_\tau = 0.05$	0.03775	0.55675	0.09938171	0.1350278
Combined	(0.04, 0.01)	0.04	0.577	0.09063807	0.1272009
	(0.03, 0.02)	0.043	0.587	0.08540022	0.121842
	(0.025, 0.025)	0.045	0.576	0.08476733	0.1137233
	(0.02, 0.03)	0.044	0.564	0.07644395	0.1093622
	(0.01, 0.04)	0.046	0.558	0.06362453	0.1001554

Table 1: Empirical Rejection Probabilities

Looking first at setting 1, which is a simulation under the null hypothesis, the rejection probabilities seem to be compatible with the chosen size of 0.05. However, the size of the rejection frequency for the Cramér-von Mises test seems slightly low, which could be an artifact caused by the relatively low number of simulation runs. A larger simulation using 10000 runs for the Cramér-von Mises test resulted in a rejection frequency of 0.0469 and therefore seems to confirm this hypothesis.

For setting 2, which is a mean shift between the processes the power is relatively high considering the rather low sample size of 20. Especially considering the fact that the authors mentioned that the power of the Cramér-von Mises type test in pure mean shift settings is somewhat lacking. However, this can be remedied by the fact, that even thought the mean shift is rather subdued visually, it is stronger than the mean shift used in the simulation study of Bugni and Horowitz 2021.

Due to the change in setting 3, picking up the differences in the persistence structure might be difficult with a measure induced by a random variable that was truncated at a low truncation parameter K. However, for a choice of $K = 25$, the power shows Hier weiterschreiben

7.3 Simulation - Test for Persistence Alternatives

Using simulations akin to setting 3, it is possible to test the proposed test for persistence alternatives from Section 6. However, due to computational constraints, this simulation currently cannot follow the real test's structure and shall instead serve as a heuristic for the real test's performance.

Primarily, the simulation differs from the proposed test in the following way to reduce computational cost while staying reasonably accurate in the proposed simulation setting. The standardization step makes it necessary to perform the comparison between

functions that is performed in the calculation of the Cramér-von Mises test statistic in each permutation step. This massively increases the computational cost. Therefore, as a heuristic for the real test, these simulations are performed without a standardization step. However, a data generating process with a mean function $\mu(t) = 0$ and a point-wise variance function $\sigma(t) = 1$ is used. Therefore, in expectation, sample equivalents will share these characteristics. However, individual samples may not.

Due to this concession to limited computational resources, the following simulation results could potentially overstate the power of the test against alternatives as described above. This is caused by potential differences in the sample mean and point-wise sample variance present in the permuted samples that would have been removed by standardization. However, in the proposed test structure the possibility to specifically test for alternatives in the persistence structure is a significant advantage.

The results for different correlation parameters are given in the following tables. Again, the same reference process with $\rho_1(t) = 0.5$ generated the reference sample.

α_τ	$\rho_2(t) = -0.9$	$\rho_2(t) = -0.5$	$\rho_2(t) = 0$	$\rho_2(t) = 0.5$	$\rho_2(t) = 0.9$
0.05	0.198	0.858	0.640	0.048	0.977
0.01	0.044	0.623	0.393	0.009	0.913
0.001	0.016	0.385	0.197	0.002	0.734

Table 2: Empirical Rejection Probabilities

These results show multiple things. First, as expected the size of the test, approximated by the case $\rho_2(t) = 0.5$ is close to the chosen nominal level. Due to the properties of permutation tests, this is not a surprise as it holds by construction. Second, the other cases, which approximate power against specific alternatives, are interesting as they are in part counter-intuitive.

In most settings, one would typically assume that the rejection frequency increases as the parameter that specifies the deviation from the null hypothesis changes away from its equivalent in the reference sample. These simulations, however, show that for the case of $\rho_2(t) = -0.9$ the opposite happens. Even though the persistence parameter is even further away from 0.5 than for the case of $\rho_2(t) = -0.5$, the rejection frequency decreases. This could be linked to the measure that is used for the calculation of the test statistic. As the parameter decreases, the curves are less persistent, meaning that Fourier functions of higher frequency would be needed to approximate them. However, if due to the truncation, it is not possible to use appropriately high-frequent functions in the Cramér-von Mises test, the performance might suffer. This is just a first hypothesis that could be studied in more detail using further simulations with different persistence parameters and different truncation parameters for the basis used to construct the measure.

Hier weiterschreiben.

8 Application

In order to give an example for the real-world merits of the method, this section will compare electricity demand data from Adelaide. This data is provided as part of the *fds*[5] package for R. It presents half-hourly energy demand in megawatts and was originally used by Magnano and Boland 2007 and Magnano, Boland, and R. J. Hyndman 2008. It contains electricity demand curves for 3556 days from 6/7/1997 to 31/3/2007. Some of these curves, specifically a selection of curves observed on Wednesdays and Saturdays fitted using a Fourier basis consisting of 25 functions, is shown in Figure 6.
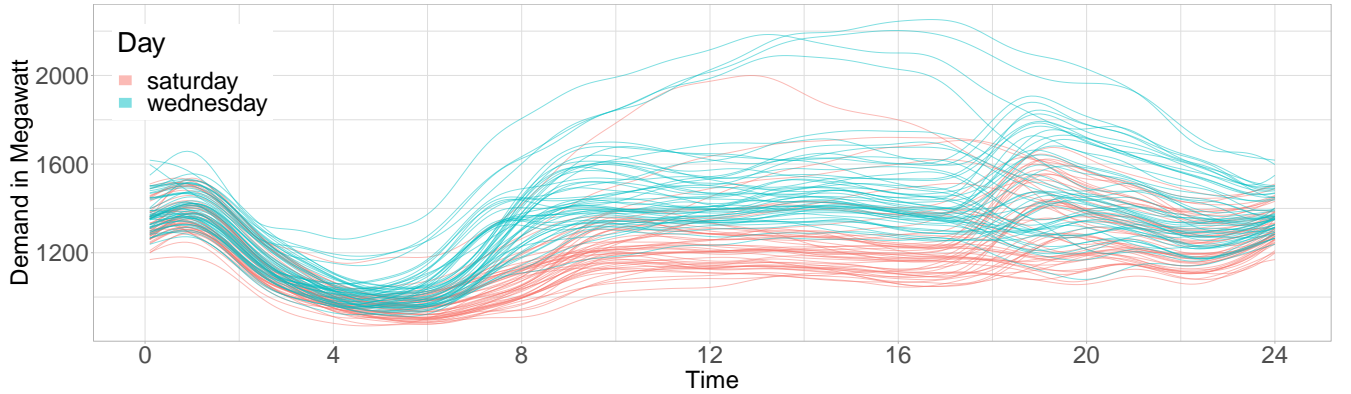

Figure 6: Electricity Demand in Adelaide

The specific question studied with the method presented in this thesis is whether electricity demand on working days and weekends can be seen as if they were generated by the same stochastic process. However, due to information obtained during the data cleaning step, that indicates a violation of the i.i.d. assumption, the observations for the weekend will be limited to Saturdays.

8.1 Data Cleaning and Preprocessing

Before actually applying the test from Bugni and Horowitz 2021, it is important to prepare the data for the procedure. As this problem does not have the structure that an experiment as described in Bugni and Horowitz 2021 possesses, some problems have to be addressed before using the procedure. These adjustments are described below and some details are substantiated in Appendix 11.8.

8.1.1 Removal of Mondays and Fridays

One potential problem of this procedure is the question whether observations on different weekdays or days of the weekend can be seen as independent and identically distributed. While it seems reasonable to assume that demand may be similar on Saturdays and Sundays, it is questionable whether the same can be said for working days. One potential problem is the ramping up and down of industrial production and commercial activity on

---

[5]Shang and Rob J Hyndman 2018.

Mondays and Fridays. Therefore, I decided to exclude these days from my analysis and only compare the weekend with Tuesdays, Wednesdays and Thursdays.

8.1.2 Holidays

Furthermore, holidays could appear more regularly on specific weekdays than others. Whereas on weekends, a holiday would not significantly influence the electricity demand due to the already reduced economic activity, this is different for weekdays. Therefore if holidays would occur systematically more often on specific days - such as for example Thursdays for the case of Germany - this could create problems. Therefore public holidays in Southern Australia, the Australian federal territory containing Adelaide, were excluded in the analysis. A list of holidays that were excluded is given in Appendix 11.8. Additionally, days immediately before and after holidays are excluded due to the same reason as Mondays and Fridays. This procedure of eliminating holidays from the data set removed 299 out of 3556 curves from the data set which was used in further steps of the analysis.

8.1.3 Curve Fitting

To employ methods that are suitable to remove trend and seasonal components from the data, it is useful to fit a functional basis to the underlying discrete observations. However, as described in Section 5.9, this can lead to problems, if the truncated functional basis is unable to capture the features of the discrete data. However, in the setting explored in this application, visual inspection of the resulting curves showed no signs of artifacts created by the curve fitting process. This makes sense from a theoretical perspective, as the curves are more persistent than those explored in the simulations. Therefore, the truncated basis, consisting of trigonometric functions of lower frequencies, should be able to capture the curves' properties appropriately.

This, in turn, makes the removal of trend and seasonal components from the fitted curves more attractive, as it more closely addresses the functional nature of the underlying data generating processes. Therefore, for the purposes of this simulation, a Fourier basis consisting of 25 functions was fitted to the original data to perform the data cleaning step.

8.1.4 Detrending and Deseasoning

A third potential problem of this data set is its functional time series structure. For example, electricity demand might be systematically higher in the summer months due to the added energy consumption of air-conditioning units. Therefore, a simple interpretation of the data as generated by an i.i.d. process might be unsubstantiated and additional steps have to be made before the procedure can be justified.

Additionally, it might be the case that electricity demand has a trend component that has to be removed before this method can reasonably be applied to this data. To combat this low frequency seasonal component due to the seasons and a potential long-term trend I

specify a model as follows and demean the data as described below. For the estimation of this model, holidays and days immediately before and after holidays are excluded. Mondays and Fridays are included and removed from the data set after the estimation for the creation of the samples that are used to apply the method described in this thesis.

$$
\begin{aligned}
f_{demand} = & f_{mean} + f_{trend}(year - 1997) + \sum_{j=2}^{12} \mathbb{1}_{[month = j]} f_{month,j} \\
& + \sum_{k=2}^{7} \mathbb{1}_{[day = k]} f_{day,k} + f_{random}
\end{aligned}
\tag{46}
$$

This is estimated with the usual theory for function-on-scalar regression which is described for example in Ramsay and Silverman 2005. Then, the following objects are used for the further treatment.

$$
\tilde{f} = f_{mean} + \sum_{k=2}^{7} \mathbb{1}_{[day = k]} \hat{f}_{day,k} + \hat{f}_{random}
\tag{47}
$$

The function-on-scalar regression described in Equation 46 was performed in R using the fda[6] package and gave the following results. As these are the results of a function-on-scalar regression, it is convenient to plot the resulting estimates as they are functions instead of giving the estimated Fourier coefficients. The estimated coefficient functions for the different weekdays are of special interest as they are directly linked to the problem that is to be studied using the method from Bugni and Horowitz 2021. In this case Sunday is the baseline and these curves describe the deviation relative to it.
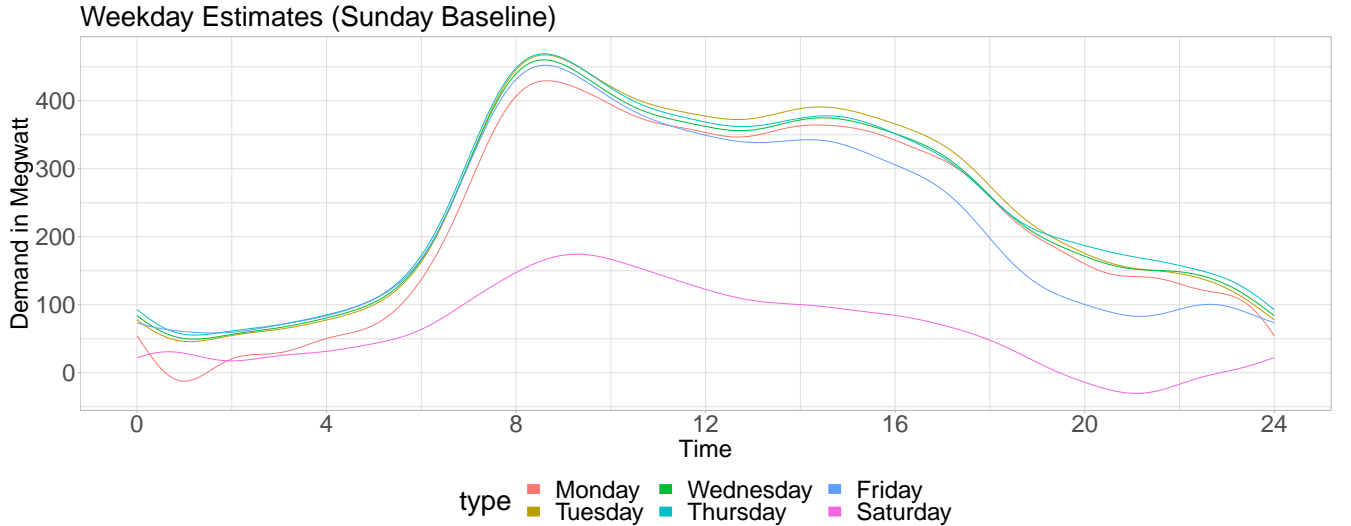

Figure 7: Estimates for the weekdays (Sunday as baseline)

This diagram already hints at a considerable mean shift between the working days and the weekend. As these estimates also hint at a considerable difference between Saturdays and Sundays, it could be advisable to treat them as non-identically distributed. Therefore,the application is limited to a comparison of working days and Saturdays as the results of the

---

[6]Ramsay, Graves, and Hooker 2021.

regression provide clear evidence against the assumption that Saturday and Sunday can be seen as if they were generated by the same random variable. Further results of the regression including estimates of the constant and the coefficient functions for different months are shown in Appendix 11.8.2.

8.2 Test from Bugni and Horowitz 2021

After cleaning the data as described in the previous section, the method from Bugni and Horowitz 2021 can now be applied to the cleaned data. To give a visualization of what these cleaned curves look like, Figure 8 shows randomly chosen curves observed on Wednesdays and Saturdays that were prepared with the procedure described above.
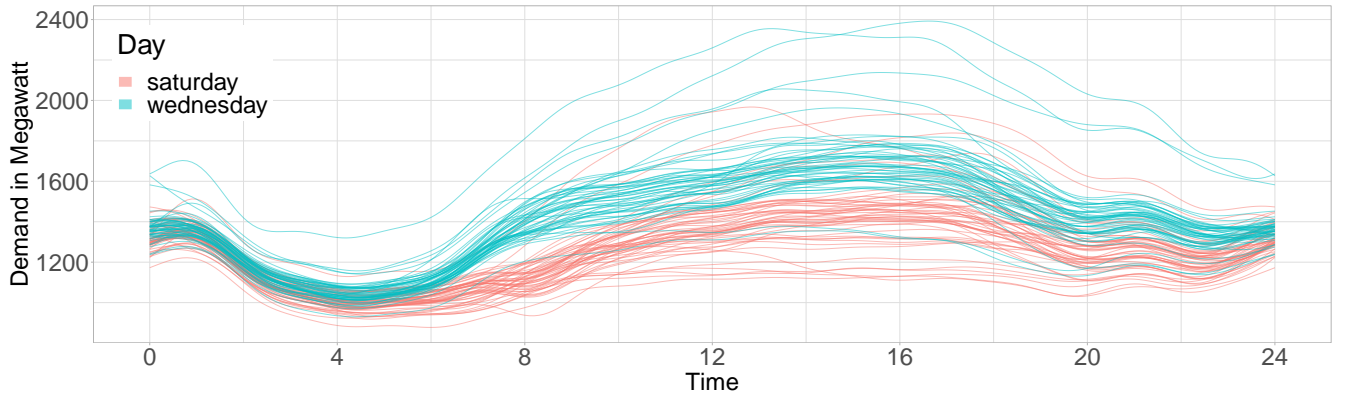

Figure 8: Pre-Processed Electricity Demand in Adelaide

As Figure 8 already shows a considerable mean shift, that is more extreme than those shown in the simulation, it is expected that the test will pick up on these differences. Additionally, as the sample size is considerably bigger in this scenario with 478 observations and 1404 observations corresponding to Saturdays and working days respectively. As the power of permutation tests is typically quite sensitive to the sample size, this is another reason to assume strong power against the apparent violation of the null hypothesis. Using the same decision rules as in the simulation study leads to the following rejection decisions.

Test	$(\alpha_\tau, \alpha_\nu)$	Rejection of the Null?
Means-Test	$\alpha_\nu = 0.05$	✓
CvM-Test	$\alpha_\tau = 0.05$	✓
Combined	(0.04, 0.01)	✓
	(0.03, 0.02)	✓
	(0.025, 0.025)	✓
	(0.02, 0.03)	✓
	(0.01, 0.04)	✓

Table 3: Results of the Test from Bugni and Horowitz 2021

As mentioned before, these rejections of the null hypothesis are unsurprising when looking

at the curves generated for the different days in the data cleaning step. There, it was obvious that the difference in mean alone was stark between workdays and Saturdays. Therefore, it was expected for the mean-based test to find the violation of the null hypothesis in this comparatively large sample problem. Due to the large differences in mean alone, it is also unsurprising that the Cramér-von Mises type test picked up on the violation. This application can therefore serve more as a real-world proof of concept for the test. In reality, most scenarios where applying this test will be interesting will be less obvious in the violation of the null hypothesis. Especially violations created by differing persistence structures will typically be difficult to identify. Therefore, applying this test to more challenging settings would be interesting. Section 9.4 presents a small selection of settings that might be interesting for further application of the procedures described in this thesis.

9 Outlook

After giving the necessary background information and taking an in-depth look at the ideas of Bugni and Horowitz 2021 in this thesis, this Outlook section will present some problems that came up during the process of writing this thesis that might be interesting to study in further research. Additionally, it presents some ideas for simulations that could improve the understanding of methods such as Cramér-von Mises type tests in a functional data setting and presents real-world scenarios that might be interesting applications for the methods introduced in this thesis.

9.1 Potential Problems

The results presented in this thesis give a good impression of the possibilities of the method developed in Bugni and Horowitz 2021. However, there are also potential problems that should not go unnamed.

Implementation

One source of problems that might influence the results of the simulations is the specific implementation that was used to perform the simulation. As addressed in Section 5.9, seemingly small choices such as whether to use a functional basis to represent the observations can have a significant impact on the performance and reliability of the simulation and its results.

Definition of $\mathbb{L}_2[0,1]$

As mentioned multiple times in this thesis, Bugni and Horowitz 2021 use an unconventional definition of $\mathbb{L}_2[0,1]$ that makes some theoretical considerations of the methods considerably harder. Therefore, questions such as how this convention interacts with the type of measures that is used in the Cramér-von Mises type test should be studied further.

Potential Gap in a Proof

Parts of the proofs presented in the supplementary material to Bugni and Horowitz 2021 hinge on the assumption that the space of square-integrable functionals with respect to the chosen measure μ is a separable Hilbert space. In many cases square-integrability is a property that easily allows to show the properties of a separable Hilbert space. However, as it is not shown in the proofs to be a general property or follow from the type of measure the authors construct, they seem to be incomplete and further consideration should be given to fill these gaps.

9.2   Theoretical Extensions

Variants of the Test based on different Norms

In the classical scalar setting one common alternative to the Cramér-von Mises test is the Kolmogorov-Smirnov test. Instead of using the squared distance between to objects, it is based on the supremum norm. It could therefore be used to construct an alternative test statistic in the following way.

$$\tau_{m,n}^{KS} = \sup_{z \in \mathbb{L}_2[0,1]} |\hat{F}_n(z) - \hat{G}_m(z)| \tag{48}$$

Similar to the approximation by Monte-Carlo integration in the main case, this would have to be approximated in an actual implementation. One obvious approach being the following, where $Z_i \overset{\text{i.i.d.}}{\sim} Z$ and Z is a random function as described in Section 5.5.

$$\hat{\tau}_{m,n}^{KS} = \max_{i=1,\ldots,H} \left\{ |\hat{F}_n(Z_i) - \hat{G}_m(Z_i)| \right\} \tag{49}$$

Other variants could be used in a similar way leading to a plethora of possible test statistics. It would be interesting to compare them both from a theoretical perspective and using simulations to assess their power in different scenarios.

9.3   Possible Further Simulations

After studying the method developed by Bugni and Horowitz 2021 in detail there are a few points that might be interesting for further research. In the following I name a few that could be very interesting extension to this thesis.

Comparing different Functional Bases

For the construction of the measure and potentially also for fitting the observations a functional basis is used. Even in settings, where observations are not fitted using methods like this, the basis that is used to construct the measure might influence the properties of the test. From a theoretical perspective, the test is only justified for complete orthonormal bases, but it would be interesting to study its properties in simulations even for bases that do not have these properties, such as for example a b-spline basis.

Comparing choices of $w(t)$ and $(\rho_i)_{i=1,the...,K}$

In this thesis I only perform simulations for one intuitive choice of $w(t)$ and $(\rho_i)_{i=1,...,K}$. It seems reasonable to assume that these parameter choices could have a significant impact on the performance of the method and it would therefore be interesting to compare different choices in a structured framework.

Less restrictive Distributions of Fourier Coefficients

The measures constructed as part of the Cramér-von Mises type test described in this thesis drew their Fourier coefficients using independently distributed error terms. Even though this leads to a convenient implementation, using measures induced by random variables generated using error terms with a more general dependence structure might be beneficial to increase the power against specific alternatives. It would therefore be interesting to study the influence on the measure on the power of the test in more general settings both from a theoretical perspective and using simulations.

Simulations for Unbalanced Sample Sizes

The simulations in this thesis only dealt with settings in which both the reference sample and the second sample contained the same number of curves. Even though there is no obvious theoretical reason for unbalanced sample sizes to influence the power of the test meaningfully, it would be interesting to study settings with unbalanced sample sizes in further simulations.

Comparing Implementations using different orders of Operations

As described in Sections 6.2 and 7.3, the actual implementation could be highly important for the properties of the proposed test for persistence alternatives. This thesis only provides two heuristic simulations to compare the performance of the proposed options due to computational limitations. Therefore, it would be interesting to actually compare these options in further full-scale simulations.

9.4   Potential Applications

As explained in Section 8, the chosen setting is suitable to apply the test, but it is more of a showcase then what this method would be used for in reality. Therefore, it would be interesting to apply the method and the variant for persistence alternatives to data that is more fitting. Two examples that could provide interesting applications for this method are listed in the following.

- **Audio-Curves** and specifically speech recordings are a typical scenario for functional data. The tests described in this thesis could be used to compare samples of recordings of spoken words. One interesting application could be to determine whether a set of recordings was manipulated when there is an appropriate reference sample that is known to be authentic. Especially in the context of computer generated speech

imitation, sometimes known as voice deep-fakes, a test that could reliably distinguish between original and imitated speech recordings could be highly relevant.

- **Movement-Curves** such as for example curves describing the movement of a runner's legs during a 100m sprint might also be an interesting field for potential applications. Given a reference sample of movement curves of runners with a specific alteration to their movement due to e.g. a sickness, the test might be useful in the context of medical diagnostics.

10   Bibliography

Anderson, T. W. (1962). "On the Distribution of the Two-Sample Cramer-von Mises Criterion". In: *The Annals of Mathematical Statistics* 33.3. Publisher: Institute of Mathematical Statistics, pp. 1148–1159. DOI: 10.1214/aoms/1177704477.

Anderson, T. W. and D. A. Darling (1952). "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes". In: *The Annals of Mathematical Statistics* 23.2, pp. 193–212. DOI: 10.1214/aoms/1177729437.

— (Dec. 1954). "A Test of Goodness of Fit". In: *Journal of the American Statistical Association* 49.268, pp. 765–769. DOI: 10.1080/01621459.1954.10501232.

Bauer, Heinz (2011). *Probability Theory*. De Gruyter. ISBN: 978-3-11-081466-8. DOI: 10.1515/9783110814668.

Boyd, John P. (2006). "Computing the zeros, maxima and inflection points of Chebyshev, Legendre and Fourier series: solving transcendental equations by spectral interpolation and polynomial rootfinding". en. In: *Journal of Engineering Mathematics* 56.3, pp. 203–219. ISSN: 1573-2703. DOI: 10.1007/s10665-006-9087-5.

Bugni, Federico A., Peter Hall, et al. (2009). "Goodness-of-fit tests for functional data". In: *The Econometrics Journal* 12.S1, S1–S18. ISSN: 1368-4221. URL: https://www.jstor.org/stable/23116593.

Bugni, Federico A. and Joel L. Horowitz (2021). "Permutation tests for equality of distributions of functional data". en. In: *Journal of Applied Econometrics* 36.7, pp. 861–877. DOI: 10.1002/jae.2846.

Büning, Herbert and Götz Trenkler (2013). *Nichtparametrische statistische Methoden*. De Gruyter. ISBN: 978-3-11-090299-0. DOI: 10.1515/9783110902990. URL: https://www.degruyter.com/document/doi/10.1515/9783110902990/html?lang=en.

Cox, Dennis D. and Jong Soo Lee (2008). "Pointwise Testing with Functional Data Using the Westfall-Young Randomization Method". In: *Biometrika* 95.3, pp. 621–634. ISSN: 0006-3444. URL: https://www.jstor.org/stable/20441490 (visited on 05/26/2022).

Darling, D. A. (1957). "The Kolmogorov-Smirnov, Cramer-von Mises Tests". In: *The Annals of Mathematical Statistics* 28.4, pp. 823–838. DOI: 10.1214/aoms/1177706788.

Dunn, Olive Jean (1961). "Multiple Comparisons among Means". In: *Journal of the American Statistical Association* 56.293. Publisher: Taylor & Francis, pp. 52–64. ISSN: 0162-1459. DOI: 10.1080/01621459.1961.10482090.

Fisz, M. (1960). "On a Result by M. Rosenblatt Concerning the Von Mises-Smirnov Test". In: *The Annals of Mathematical Statistics* 31.2. Publisher: Institute of Mathematical Statistics, pp. 427–429. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177705905.

Gibbons, Jean Dickinson and Subhabrata Chakraborti (2021). *Nonparametric statistical inference*. 6th edition. Boca Raton: CRC Press. ISBN: 978-1-138-08744-6.

Gihman, Iosif Il'ich and Anatoliĭ Vladimirovich Skorokhod (2004). *The Theory of Stochastic Processes I.* Classics in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-20284-4. DOI: `10.1007/978-3-642-61943-4`.

Hall, Peter and Nader Tajvidi (June 2002). "Permutation tests for equality of distributions in high-dimensional settings". In: *Biometrika* 89.2, pp. 359–374. DOI: `10.1093/biomet/89.2.359`. (Visited on 05/26/2022).

Hall, Peter and Ingrid Van Keilegom (2007). "TWO-SAMPLE TESTS IN FUNCTIONAL DATA ANALYSIS STARTING FROM DISCRETE DATA". In: *Statistica Sinica* 17.4, pp. 1511–1531. ISSN: 1017-0405. URL: `https://www.jstor.org/stable/24307686`.

Hoeffding, Wassily (June 1952). "The Large-Sample Power of Tests Based on Permutations of Observations". In: *The Annals of Mathematical Statistics* 23.2, pp. 169–192. DOI: `10.1214/aoms/1177729436`.

Hsing, Tailen and Randall L. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley series in probability and statistics. John Wiley and Sons, Inc. ISBN: 978-0-470-01691-6.

Jeribi, Aref (Jan. 2021). *Perturbation Theory for Linear Operators.* Springer Singapore Pte. Limited. ISBN: 9789811625275.

Kokoszka, Piotr and Matthew Reimherr (2021). *Introduction to functional data analysis.* First issued in paperback. Texts in statistical science series. CRC Press. ISBN: 978-1-03-209659-9 978-1-4987-4634-2.

Lee, Jong, Dennis Cox, and Michele Follen (Mar. 2015). "A Two Sample Test for Functional Data". In: *Communications for Statistical Applications and Methods* 22, pp. 121–135. DOI: `10.5351/CSAM.2015.22.2.121`.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses.* en. Springer Texts in Statistics. Springer New York. ISBN: 978-0-387-98864-1 978-0-387-27605-2. DOI: `10.1007/0-387-27605-X`.

Magnano, L. and J. W. Boland (Nov. 2007). "Generation of synthetic sequences of electricity demand: Application in South Australia". en. In: *Energy* 32.11, pp. 2230–2243. DOI: `10.1016/j.energy.2007.04.001`.

Magnano, L., J. W. Boland, and R. J. Hyndman (2008). "Generation of synthetic sequences of half-hourly temperature". en. In: *Environmetrics* 19.8, pp. 818–835. DOI: `10.1002/env.905`.

Pettitt, A. N. (1976). "A Two-Sample Anderson–Darling Rank Statistic". In: *Biometrika* 63.1, pp. 161–168. DOI: `10.2307/2335097`.

Pomann, Gina-Maria, Ana-Maria Staicu, and Sujit Ghosh (2016). "A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 65.3, pp. 395–414. ISSN: 0035-9254. URL: `https://www.jstor.org/stable/24773028`.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Ramsay, J. O., Spencer Graves, and Giles Hooker (2021). *fda: Functional Data Analysis*. R package version 5.5.1. URL: `https://CRAN.R-project.org/package=fda`.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer New York. ISBN: 978-0-387-40080-8 978-0-387-22751-1. DOI: `10.1007/b98888`.

Rosenblatt, M. (1952). "Limit Theorems Associated with Variants of the Von Mises Statistic". In: *The Annals of Mathematical Statistics* 23.4. Publisher: Institute of Mathematical Statistics, pp. 617–623. ISSN: 0003-4851. URL: `https://www.jstor.org/stable/2236587`.

Shang, Han Lin and Rob J Hyndman (2018). *fds: Functional Data Sets*. R package version 1.8. URL: `https://CRAN.R-project.org/package=fds`.

Shonkwiler, Ronald W. and Franklin Mendivil (Aug. 2009). *Explorations in Monte Carlo Methods*. en. Springer Science & Business Media. ISBN: 978-0-387-87836-2.

Skorohod, A. V. (1974). *Integration in Hilbert Space*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-65634-7. DOI: `10.1007/978-3-642-65632-3`.

Vaart, Aad W. van der and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer New York. ISBN: 978-1-4757-2547-6 978-1-4757-2545-2. DOI: `10.1007/978-1-4757-2545-2`.

Wickham, Hadley et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: `10.21105/joss.01686`.

11 Appendix

11.1 Informal Intuition

For potential readers unfamiliar with functional data analysis and permutation testing, the initial hurdle of reading this thesis without any intuitive understanding might be high. Therefore, this short informal introduction should serve as a primer that gives an intuitive idea about the following questions.

1. What setting are we in?

2. What do we want to test?

3. How do we want to test it?

4. Why should I care?

By choice, this section will **not** be formal and it will **not** be precise so as not to introduce too much detail that could hinder an intuitive understanding.

Answering question one is relatively simple. This thesis deals with a statistical problem where observations are continuously observed curves. But what does that mean intuitively? One scenario that is easy to imagine is a curve over time for some variable such as speed: When driving a car at the 24-hour race at Le-Mans, it has a speed at every point in time during the race - for example, 155.3 km/h at 2 hours, 33 minutes and 7 seconds after the race has started. Observations in the sense of this thesis are similar to this curve: at every value of x in some closed interval $[a, b] \subset \mathbb{R}$, we observe a value y.

Question two is also simple. We want to test if two samples of curves are similar in a specific way. As in many statistical problems, we interpret the observations as realizations of some random variable. The only difference is that these observations are curves. So we can ask the question of whether the curves in the two samples are dissimilar enough for us to say confidently that they were not generated by the same random variable. Staying with the example of Le-Mans, we could have two identical cars with two equally skillful drivers collect 24-hour speed curves for the next ten years. We do this to determine if two different types of fuel change the way the cars act on the track. We now have two samples, each containing 3650 speed curves in this hypothetical scenario. We want to test statistically whether changing the fuel made any difference.

Question three is more difficult. Therefore, let's give the correct intuition by answering a simpler question instead. Let us return to the scalar setting for a moment - each observation is just a real number $x \in \mathbb{R}$ again. We want to determine whether two samples are different enough for us to confidently say: "These samples are too dissimilar to be generated by the same random variable." If you have some statistical knowledge, your first intuition might be to perform a Kolmogoroff-Smirnov, Cramér-von Mises or Anderson-Durbin test. However, let us take one step back first and think about a more general idea.

From an intuitive point of view, if the two samples were generated by the same random variable, the effect of randomly switching observations between the samples should be small. Formalizing this idea leads to the concept of permutation tests. By randomly permuting the samples and calculating a chosen test statistic on the permuted samples, one could derive a distribution for the test statistic that can be used for testing. If the test statistic calculated on the non-permuted samples is comparatively extreme, this could be an indicator that the samples were not generated by the same random variable.

Why you should care about this is the hardest question. I care about it because it is cool from a mathematical perspective. Nevertheless, if you care about real-life problems, there are also good reasons to be interested. Economic data is being observed at increasing frequencies as technology improves. Many problems that were previously low dimensional due to data constraints or suitable for classical time series methods are becoming more complex as methodological challenges such as high-dimensionality and extreme correlation of neighboring observations arise. Functional data analysis is a comparatively new approach to many of these problems that transform some of these problems into strengths by acknowledging the functional structure of the underlying processes. Permutation tests are also comparatively new, at least in our ability to apply them on a larger scale.

11.2 $\mathbb{L}_2[0,1]$ as defined in Bugni and Horowitz 2021

To distinguish between the typical case presented in the previous section, let $\mathbb{L}_2[0,1]$ denote the Hilbert space of square-integrable functions and $\mathbb{L}_2^*[0,1]$ the square-integrable functions under the the convention from Bugni and Horowitz 2021.

Using $\mathbb{L}_2^*[0.1]$ creates some interesting theoretical challenges, as the resulting object is in fact not a Hilbert space. To understand the theoretical problems that can occur, I first introduce some additional concepts to illustrate the challenges.

Definition 11.1 (Norm and Seminorm)
A function $p : \mathbb{V} \to \mathbb{F}$ on a vector space $\mathbb{V}$ over a field $\mathbb{F}$ is called a norm if the following four conditions hold for all $v, u \in \mathbb{V}$ and $s \in \mathbb{F}$.

1. $p(v + u) \leq p(v) + p(u)$
2. $p(sv) = |s|p(v)$
3. $p(v) \geq 0$
4. $p(v) = 0 \implies v = 0$

If $p : \mathbb{V} \to \mathbb{F}$ fulfills only properties (1) to (3) it is called a seminorm.

In the same way a norm induces a distance on its corresponding normed vectorspace, a seminorm p induces a so-called pseudometric d. It is given by $d(v, u) = p(u - v)$.
This is from Wikipedia!!!

Definition 11.2 (Pseudometric Space)

A pseudometric space (X, d) is a set X together with a function $d : X \times X \to \mathbb{R}_{\geq 0}$, such that $\forall x, y, z \in X$ the following properties hold.

1. $d(x, x) = 0$

2. $d(x, y) = d(y, x)$

3. $d(x, z) \leq d(x, y) + d(y, z)$

Therefore, deviating from a metric space, two distinct points in a pseudometric space can have a distance of zero $d(x, y) = 0$ for $x \neq y$.

That $\mathbb{L}_2^*[0, 1]$ is not a Hilbert space becomes clear, when checking the for the properties of the norm induced by the inner product $\|v\| = \sqrt{\langle v, v \rangle}$. One of the properties that has to be fulfilled by a norm is $\|v\| = 0 \iff v = 0$. Let $f : [0, 1] \to \mathbb{R}$ be given by $f(x) = \mathbb{1}\,[x = 0.5]$. Then we can evaluate the following expression to create a contradiction to the norm properties.

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_0^1 [f(t)]^2\, \mathrm{d}t} = 0 \tag{50}$$

As f is not the zero element of this space, this is a violation of positive definiteness. Positive definiteness applied to the case at hand, states that $\forall v \in \mathbb{L}_2[0, 1] \|v\| = 0 \implies v(x) = 0 \quad \forall x \in [0, 1]\ \forall v \in$. Instead, $\|v\| = \sqrt{\langle v, v \rangle}$ is a seminorm and the defined space should more correctly be treated as a pseudometric space.

This is from Wikipedia!!!

Definition 11.3 (Hausdorff Space)

A Hausdorff space is a topological space where for any two distinct points x and y, there exist a neighborhood U of x and a nieghborhood V of y such that U and V are disjoint. This property is also called neighborhood-separability.

One problem of $\mathbb{L}_2^*[0, 1]$ is, that if we give the space the topology induced by the obvious seminorm, the resulting space would not be Hausdorff. Thus, limits in the later part of Bugni and Horowitz 2021 would not be defined. A second problem is, that it is not clear how a Schauder basis would be defined for a pseudometric space such as $\mathbb{L}_2^*[0, 1]$ and that typical existence results for orthonormal bases might not be available.

11.3 Asymptotic Distribution of the Cramér-von Mises Test

As shown by Rosenblatt 1952 and Fisz 1960, under the null hypothesis that both samples were independently generated by random variables sharing the same distribution function,

we can find the following limiting distribution of $C_{m,n}$.

$$C_{m,n} \xrightarrow{d} \int_0^1 \left(Z(u) + (1+\lambda)^{-\frac{1}{2}} f(u) - \left[\frac{\lambda}{1+\lambda} \right]^{\frac{1}{2}} g(u) \right)^2 du \tag{51}$$

$$\text{as} \quad n \to \infty, \quad m \to \infty, \quad \frac{n}{m} \to \lambda \in \mathbb{R}$$

Here, $f(u)$ and $g(u)$ are the probability density functions of the corresponding random variables and $Z(u)$ is a Gaussian stochastic process with the following properties.

- $\mathbb{E}[Z(u)] = 0 \quad \forall u \in [0,1]$

- $Cov(Z(u), Z(v)) = \min(u,v) - uv \quad \forall u,v \in [0,1]$

11.4 Proof from Hoeffding 1952

Theorem as stated in Lehmann and Romano 2005

Suppose X^N has distribution P_N in $\mathcal{X}_N$, and G_N is a finite group of transformations from $\mathcal{X}_N$ to $\mathcal{X}_N$. Let $\mathcal{G}_N$ be a random variable that is uniform on G_N. Also let $\mathcal{G}'_N$ have the same distribution as $\mathcal{G}_N$, with X^N, $\mathcal{G}_N$, and $\mathcal{G}'_N$ mutually independent. Suppose, under P_N,

$$\left(T_N(\mathcal{G}_N X^N), T_N(\mathcal{G}'_N X^N) \right) \to_d (T, T'), \tag{52}$$

where T and T' are independent, each with common cumulative distribution function $R(t)$. Then, under P_N,

$$\hat{R}_N(t) \to_P R(t) \tag{53}$$

for every t which is a continuity point of $R(t)$.

Proof as presented in Lehmann and Romano 2005

Let t be a continuity point of $R(t)$. Then,

$$\mathbb{E}_{P_N}\left[\hat{R}_N(t) \right] = P_N\left[T_N(\mathcal{G}_N X^N) \le t \right] \to R(t), \tag{54}$$

by the convergence hypothesis shown in Equation 52. It therefore suffices to show that $Var_{P_N}\left[\hat{R}_N(t) \right] \to 0$ or, equivalently, that

$$\mathbb{E}_{P_N}\left[\hat{R}_N^2(t) \right] \to R^2(t). \tag{55}$$

But,

$$\mathbb{E}_{P_N}\left[\hat{R}_N^2(t) \right] = \frac{1}{M_N^2} \sum_{g \in G_N} \sum_{g' \in G_N} P_N\left[T_N(g X^N) \le t \ \wedge \ T_N(g' X^N) \le t \right]$$
$$= P_N\left[T_N(\mathcal{G}_N X^N) \le t \ \wedge \ T_N(\mathcal{G}'_N X^N) \le t \right] \to R^2(t), \tag{56}$$

again by the convergence hypothesis. Hence $\hat{R}_N(t) \to R(t)$ in P_N probability.

11.5 Argument for Finiteness

The following argument can be made concerning the finiteness of the mean.

$$\hat{F}_n(z) \in [0,1] \wedge \hat{G}_m(z) \in [0,1] \implies \left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \in [0,1] \quad \forall z \in \mathbb{L}_2[0,1]$$

$$\int_{\mathbb{L}_2[0,1]} 1 \, \mathrm{d}\mu(z) = 1 \wedge \int_{\mathbb{L}_2[0,1]} 0 \, \mathrm{d}\mu(z) = 0 \implies \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \mathrm{d}\mu(z) \in [0,1]$$

(57)

The same argument can be made concerning the variance.

$$\left(\left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \in [0,1] \quad \forall z \in \mathbb{L}_2[0,1]\right) \wedge \left(\int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \mathrm{d}\mu(z) \in [0,1]\right)$$

$$\implies \left[\left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 - \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \mathrm{d}\mu(z)\right]^2 \in [0,1] \; \forall z \in \mathbb{L}_2[0,1]$$

$$the \int_{\mathbb{L}_2[0,1]} 1 \, \mathrm{d}\mu(z) = 1 \wedge \int_{\mathbb{L}_2[0,1]} 0 \, \mathrm{d}\mu(z) = 0$$

$$\implies \int_{\mathbb{L}_2[0,1]} \left[\left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 - \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z)\right]^2 \mathrm{d}\mu(z)\right]^2 \mathrm{d}\mu(z) \in [0,1]$$

(58)

11.6 Derivation of Asymptotic Distribution

This section presents an adapted proof for the limiting distribution of the Cramér-von Mises permutation test as presented in the supporting information of Bugni and Horowitz 2021, additional information was taken from the proof of Theorem 3.1 in Bugni, Hall, et al. 2009.

Theorem 11.1
Let Assumptions 5.1 and 5.2 be fulfilled and and define $\Upsilon(z)$ as in Section 5.4. Then the following statement holds.

$$\tau_{n,m} \to_d \int_{\mathbb{L}_2[0,1]} \Upsilon^2(z) \mathrm{d}\mu(z) \tag{59}$$

Additionally, it is useful to prove the following Lemmas taken from the supporting information of Bugni and Horowitz 2021 before proving the main point. Define the following objects

$$W_i(t) = \mathbb{1}\left[i \leq n\right] X_i(t) + \mathbb{1}\left[i > n\right] Y_{i-n}(t) \tag{60}$$

$$U_i = \mathbb{1}\left[i \leq n\right] \frac{N}{n} + \mathbb{1}\left[i > n\right] \frac{N}{m} \tag{61}$$

$$F_{nX}(z) = G(z) + \frac{1}{N} D(z) \tag{62}$$

for a chosen function $D(z)$ fulfilling $\int_{\mathbb{L}_2[0,1]} D^2(z)\mathrm{d}\mu(z) < \infty$.

Theorem 11.2 (Lemma A.3)
content...

Proof. content... □

Theorem 11.3 (Lemma A.4)
Let Assumptions 5.1 and 5.2 be fulfilled. Then,

$$\frac{1}{\sqrt{N}} \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_n(z) - \hat{G}_m(z)\right] \{\psi_k(z)\}_{k=1}^K \mathrm{d}\mu(z) \to_d \mathcal{N}(\Xi, \Sigma_K) \tag{63}$$

where

$$\Xi = \int_{\mathbb{L}_2[0,1]} D(z) \{\psi_k(z)\}_{k=1}^K \mathrm{d}\mu(z) \tag{64}$$

and

$$(\Sigma_K)_{i,j} = \frac{(1+\lambda)^2}{\lambda} \int_{\mathbb{L}_2[0,1]} \int_{\mathbb{L}_2[0,1]} [G(\min(z,\tilde{z})) - G(z)G(\tilde{z})]\, \psi_i(z)\psi_j(\tilde{z})\mathrm{d}\mu(z)\mathrm{d}\mu(\tilde{z}) \tag{65}$$

where $\min(z, \tilde{z})$ is defined as in Section 5.4.

Proof. We have

$$\frac{1}{\sqrt{N}} [F(z) - G(z)] = \frac{1}{\sqrt{N}} \left[\sum_{i=1}^N U_i \mathbb{1}\left[W_i(t) \le z(t)\ \forall t \in [0,1]\right]\right] \tag{66}$$

Hier Zwischenschritt ergaenzen
For any $K = 1, 2, 3, \ldots$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N U_i \left\{ \int_{\mathbb{L}_2[0,1]} \mathbb{1}\left[W_i(t) \le z(t)\ \forall t \in [0,1]\right] \psi_k(z)\mathrm{d}\mu(z) - \mu_{i,N,k} \right\}_{k=1}^K \to \mathcal{N}(0_K, \Sigma_K) \tag{67}$$

where

$$\mu_{i,N,k} = \mathbb{1}\left[i \le n\right] \int_{\mathbb{L}_2[0,1]} F_{nX}(z)\psi_k(z)\mathrm{d}\mu(z) + \mathbb{1}\left[i > n\right] \int_{\mathbb{L}_2[0,1]} G(z)\psi_k(z)\mathrm{d}\mu(z) \tag{68}$$

Additionally,

$$\frac{1}{\sqrt{N}} \left[\sum_{i=1}^N U_i \mu_{i,N,k}\right] = \frac{1}{\sqrt{N}} \int_{\mathbb{L}_2[0,1]} D(z)\psi_k(z)\mathrm{d}\mu(z) \tag{69}$$

The Lemma holds by Equations 66, 67 and 69. □

Theorem 11.1. Let $(\psi_k)_{k\in\mathbb{N}}$ be a complete orthonormal basis of $\mathbb{L}_2(\mu)$ where $\mathbb{L}_2(\mu)$ denotes the Hilbert space of functionals on $\mathbb{L}_2[0,1]$ that are square-integrable with respect to μ. Using a countable basis is only possible if $\mathbb{L}_2(\mu)$ is a separable Hilbert space, which the

original authors do not show.

For any $z(t) \in \mathbb{L}_2[0,1]$ define:

$$H_N(z) = \frac{1}{N} \left[\sum_{i=1}^{n} \frac{N}{n} \mathbb{1}\left[X_i(t) \le z(t) \ \forall t \in [0,1]\right] + \sum_{i=1}^{m} \frac{N}{m} \mathbb{1}\left[Y_i(t) \le z(t) \ \forall t \in [0,1]\right] \right] \quad (70)$$

Then,

$$\tau_{n,m} = \int_{\mathbb{L}_2[0,1]} H_N^2(z) \mathrm{d}\mu(z) \quad (71)$$

For any positive integer K, define the following objects.

$$H_{N,K1}(z) = \sum_{k=1}^{K} c_{N,k} \psi_k(z) \qquad\qquad H_{N,K2}(z) = \sum_{k=K=1}^{\infty} c_{N,k} \psi_k(z) \quad (72)$$

where

$$c_{N,k} = \int_{\mathbb{L}_2[0,1]} H_N(z) \psi_k(z) \mathrm{d}\mu(z) \quad (73)$$

and therefore

$$H_N(z) = \sum_{k=1}^{\infty} c_{N,k} \psi_k(z) = H_{N,K1}(z) + H_{N,K2}(z). \quad (74)$$

Similarly, define

$$\tau_{N,K1} = \int_{\mathbb{L}_2[0,1]} H_{N,K1}^2(z) \mathrm{d}\mu(z) = \sum_{k=1}^{K} c_{N,k}^2 \qquad \tau_{N,K2} = \int_{\mathbb{L}_2[0,1]} H_{N,K2}^2(z) \mathrm{d}\mu(z) = \sum_{k=K+1}^{\infty} c_{N,k}^2 \quad (75)$$

and

$$\Upsilon_{N,K1}(z) = \int_{\mathbb{L}_2[0,1]} b_k \psi_k(z) \mathrm{d}\mu(z) = \sum_{k=1}^{K} c_{N,k}^2 \qquad \Upsilon_{N,K2}(z) = \int_{\mathbb{L}_2[0,1]} b_k \psi_k(z) \mathrm{d}\mu(z) = \sum_{k=K+1}^{\infty} c_{N,k}^2 \quad (76)$$

where

$$b_k = \int_{\mathbb{L}_2[0,1]} \Upsilon(z) \psi_k(z) \mathrm{d}\mu(z) \quad (77)$$

and thus

$$\Upsilon(z) = \Upsilon_{N,K1}(z) + \Upsilon_{N,K2}(z). \quad (78)$$

Additionally, define

$$\tau_{K1} = \int_{\mathbb{L}_2[0,1]} \Upsilon_{N,K1}^2(z) \mathrm{d}\mu(z) \qquad\qquad \tau_{K2} = \int_{\mathbb{L}_2[0,1]} \Upsilon_{N,K2}^2(z) \mathrm{d}\mu(z). \quad (79)$$

Now, it is necessary to show the following three points.

1. $\tau_{K1} \to \tau$ as $K \to \infty$

2. $\tau_{N,K1} \to \tau_{K1}$ as $N \to \infty$ $\forall K = 1, 2, 3, \ldots$

3. $(\tau_N - \tau_{N,K1}) = \tau_{N,K2} \to_p 0$ as $N \to \infty$ followed by $K \to \infty$

Concerning point one, the following is easy to see.

$$\mathbb{E}[\tau - \tau_{K1}] = \mathbb{E}[\tau_{K2}] = \sum_{k=K+1}^{\infty} \mathbb{E}[b_k^2] \to 0 \quad \text{as} \quad K \to \infty \quad \text{because} \quad \mathbb{E}[\Upsilon(z)] \in \mathbb{L}_2(\mu) \quad (80)$$

For point two, define the following objects.

$$C_{N,K} = \{c_{N,k}\}_{k=1}^{K} \qquad\qquad B_K = \{b_k\}_{k=1}^{K} \qquad (81)$$

By Lemma A.4, it follows that $C_{N,K} \to \mathcal{N}(0, \Sigma_K) \sim B_K$ as $N \to \infty$ which shows the point.

Concerning point three, by Lemma A.3 the following holds.

$$
\begin{aligned}
\mathbb{E}[\tau_{N,K2}] &= \sum_{k=K+1}^{\infty} \mathbb{E}[c_{N,k}^2] \\
&= \sum_{k=K+1}^{\infty} \Bigg\{ \frac{1}{\sqrt{N}} \left(1 + \frac{m}{n}\right) \int_{z_1 \in \mathbb{L}_2[0,1]} \int_{z_2 \in \mathbb{L}_2[0,1]} D\left[\min(z_1, z_2)\right] \psi_k(z_1)\psi_k(z_2)\mathrm{d}\mu(z_1)\mathrm{d}\mu(z_2) \\
&\quad - \frac{2}{\sqrt{N}} \left(1 + \frac{m}{n}\right) \left[\int_{z \in \mathbb{L}_2[0,1]} D(z)\psi_k(z)\mathrm{d}\mu(z)\right] \left[\int_{z \in \mathbb{L}_2[0,1]} G(z)\psi_k(z)\mathrm{d}\mu(z)\right] \\
&\quad + (\Sigma_K)_{k,k} \left(2 + \frac{n}{m} + \frac{m}{n}\right) \frac{\lambda}{(1+\lambda)^2} + \frac{1}{N^2} \left[\int_{z \in \mathbb{L}_2[0,1]} D(z)\psi_k(z)\mathrm{d}\mu(z)\right] \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[U_i U_j\right] \Bigg\} \\
&\hspace{12cm} (82)
\end{aligned}
$$

The last expression is bounded as $n \to \infty$ for every positive integer K, which implies that $\lim_{K \to \infty} \lim_{N \to \infty} \mathbb{E}\left[\tau_{N,K2}\right] = 0$. Thus, by Markov's inequality, point three holds. $\qquad\square$

11.7 Multiple Testing

When testing statistical hypotheses, it is often helpful or even necessary to test multiple hypotheses independently of each other. One setting where this could be useful is when we want to combine the desirable properties of two tests, as is done by Bugni and Horowitz 2021. If the tests do not perfectly depend on each other, this creates a problem relating to the size of the combined test.

Definition 11.4 (Family-wise Error Rate)
The family-wise error rate is the probability of making at least one type-1 error when performing multiple hypothesis tests.

The most straightforward correction for this multiple testing problem is the so-called Bonferroni Correction. Introduced by Dunn 1961, it is based on Boole's Inequality, which

is sometimes referred to as the Bonferroni Inequality.

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}\left[A_i\right] \tag{83}$$

for a countable set of events $A_1, A_2, \ldots$.

11.8 Application Data Cleaning

11.8.1 Excluded Holidays

Holidays that were excluded in the analysis are the following[7]: New Year's Day, Australia Day, March Public Holiday, Good Friday, Holy Saturday, Easter Monday, Anzac Day, Queen's Birthday, Labour Day, Christmas Day, Christmas Eve, Christmas Day, Proclamation Day and New Year's Eve. Sunday is nominally a public holiday in South Australia. Easter Sunday is therefore not a special public holiday, but due to its prominence it was excluded in this analysis.

As Australia replaces some holidays that fall on weekends with substitute holidays on the next working day, these were also excluded. For the South Australia, this can occur for New Year's Day, Australia Day, ANZAC Day, Christmas Day, Proclamation Day and New Year's Eve.

11.8.2 Detrending and Deseasoning Results

Figures 9 and 10 show the estimates for the coefficient functions as specified in Equation 46. These curves were subtracted from the respective observations to eliminate the influence of, for example, the month the data was observed.
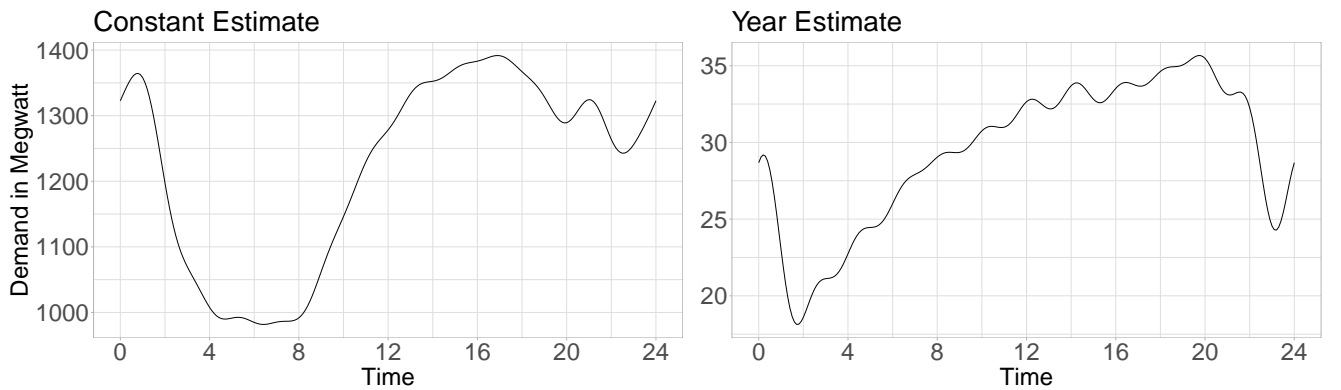

Figure 9: Estimates for the constant and year

[7]These were taken from `https://www.australia.gov.au/public-holidays` accessed on 20.05.2022.
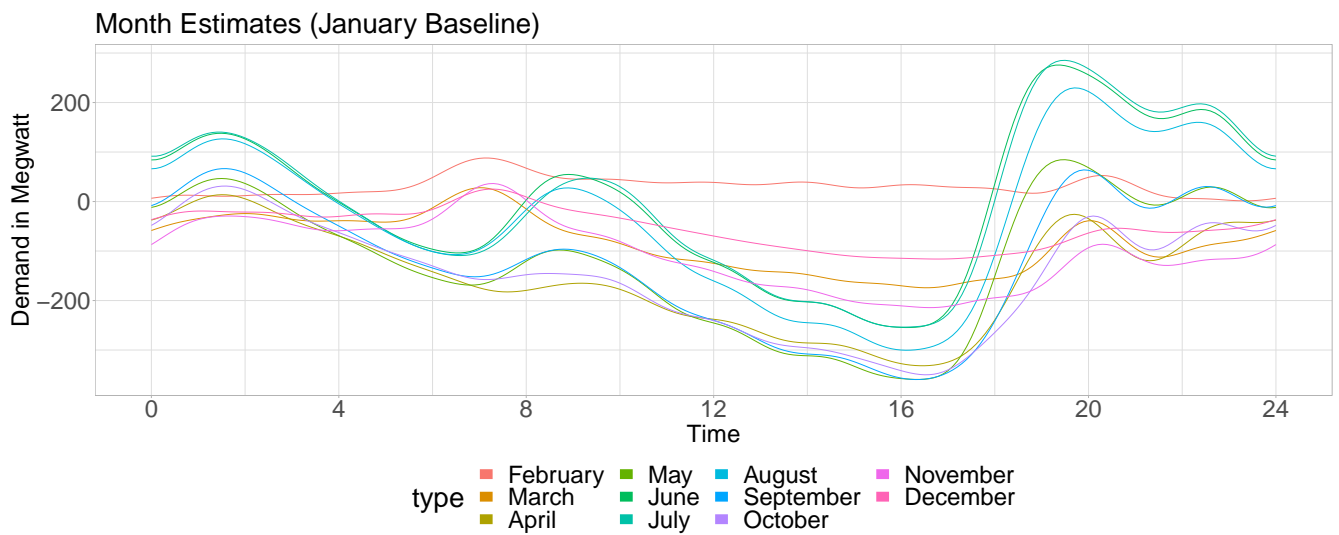
Figure 10: Estimates for the months (January as baseline)
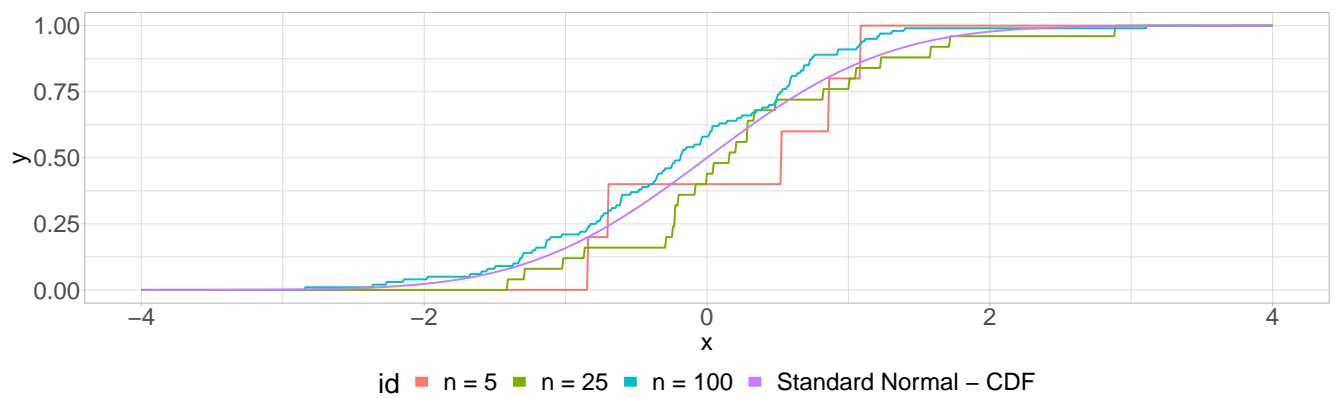
11.9 Additional Figures

Figure 11: Empirical Distribution Functions for different samples drawn from a Standard Normal Distribution
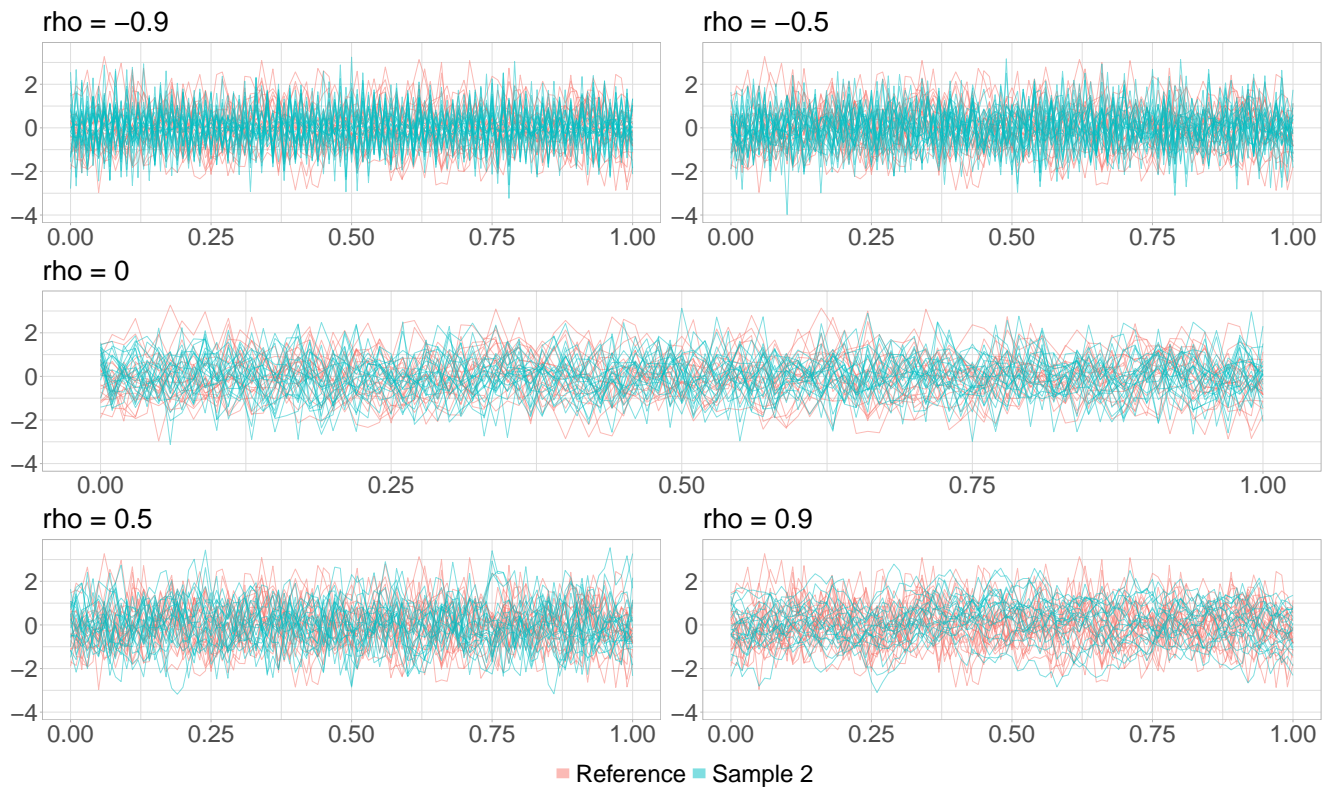
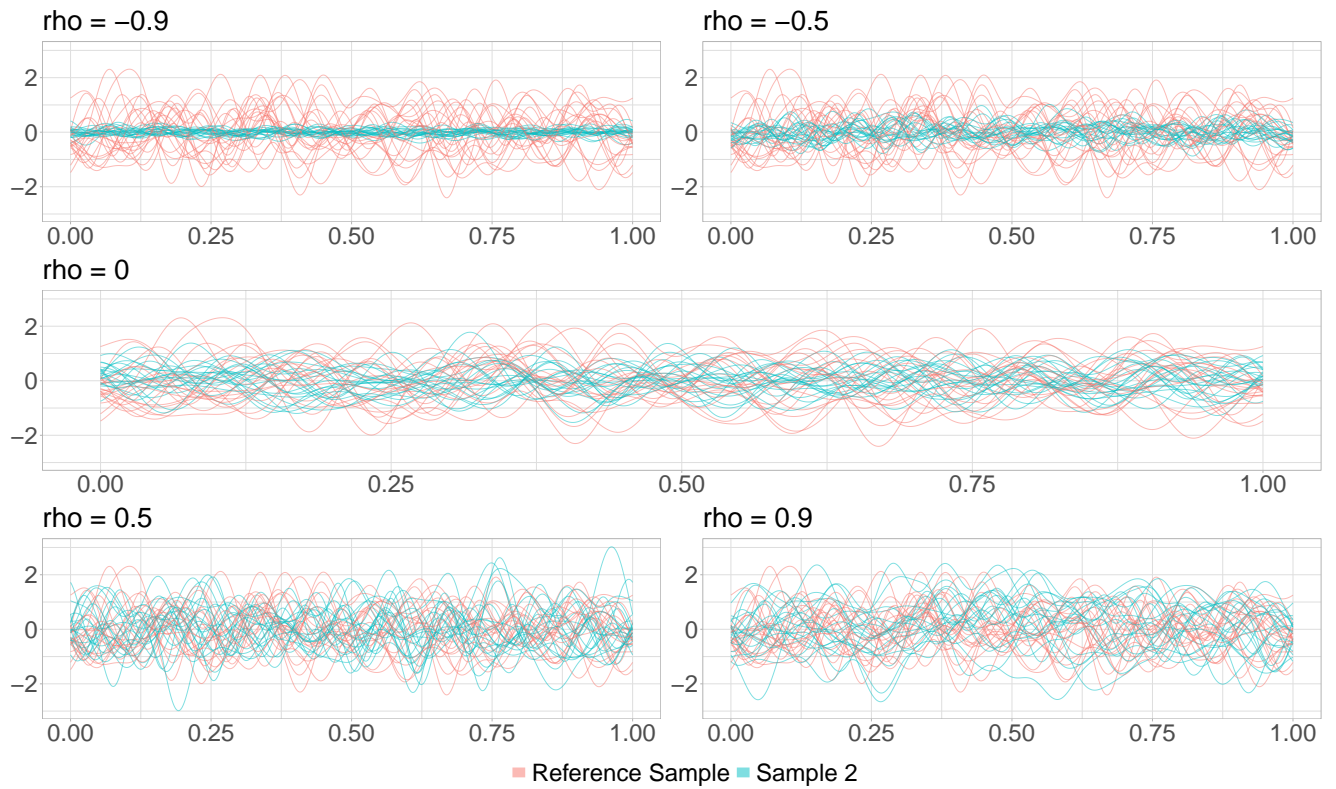Figure 12: Original Samples for Persistence Test

Figure 13: Samples for Persistence Test fitted using a Fourier Basis with 25 functions

Versicherung an Eides statt

Ich versichere hiermit, dass ich die vorstehende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass die vorgelegte Arbeit noch an keiner anderen Hochschule zur Prüfung vorgelegt wurde und dass sie weder ganz noch in Teilen bereits veröffentlicht wurde. Wörtliche Zitate und Stellen, die anderen Werken dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall kenntlich gemacht.

Bonn, XX.07.2021
Jakob R. Juergens