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 June 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 of the underlying idea, addresses some problems and shortcomings of the presented results and outlines potential 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. Therefore, functional observation can often be understood as smooth curves. A classical example of this 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].
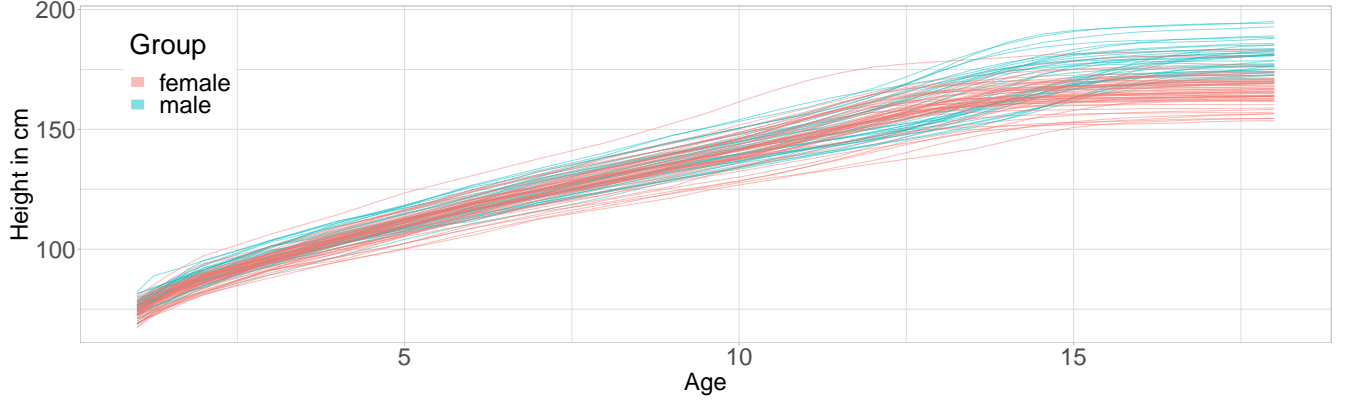
Figure 1: Human Growth Curves up to the Age of 18

Even though measurements are only taken at discrete ages, it is clear that each human has a height at every point in time. Thus, the data are only measurements of this 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 additional theory to deal with their unique properties appropriately. Sections 2.1 and 2.2 closely follow Hsing and Eubank 2015 who provide a detailed introduction into the theory of functional data analysis.

- Ramsay and Silverman 2005

- 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 further 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, which often simplifies theoretical considerations in functional data analysis. However, there is often 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.

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 such as for example functional linear regression make extensive use of their properties. In the context of this thesis, they are necessary to adequately describe objects such as random functions or probability measures on Hilbert spaces.

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 this definition only adds alternatives against which the test cannot have any power by construction, this thesis is limited to the classical definition of $\mathbb{L}_2[0, 1]$. 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 $[0,1]$.
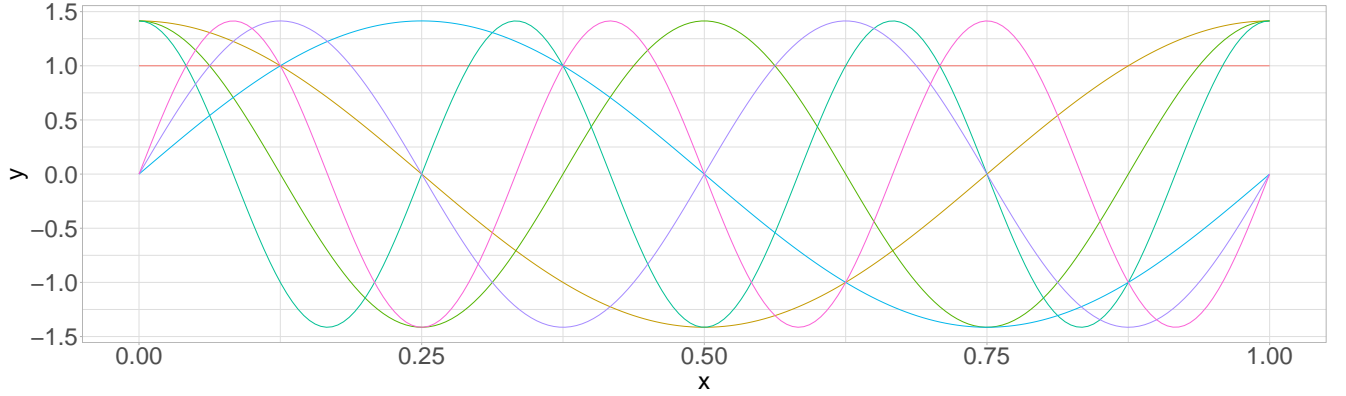
Figure 2: The first seven Fourier basis functions

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 $(\Omega, \mathcal{A}, \mathcal{P})$ 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 $X : (\Omega, \mathcal{A}, \mathcal{P}) \to (\Omega', \mathcal{A}')$ be a random variable, then a probability measure $\mathcal{P}_X(B)$ on $\mathcal{A}'$ is induced by X as shown in Equation 3.

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

This idea specializes to the case of random functions realizing in $\mathbb{L}_2[0,1]$ and allows us to use measures induced by existing random variables. 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 us to justify 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 $\sum_{k=1}^{\infty} b_k^2 < \infty$.

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 $\mathcal{P}$ denotes the probability measure associated with μ.

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

This argument allows us to use 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.

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

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

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

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

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.

- Darling 1957

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 11.7 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.

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

Anderson 1962 explores the small sample distribution of this test statistic and provides a comparison to the limiting distribution derived by Rosenblatt 1952 and Fisz 1960. The latter 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 we were to permute the elements of

the groups randomly, we would expect these differences to disappear. Permutation tests formalize this intuition. The following section closely follows chapter 15 from Lehmann and Romano 2005 with some input from Vaart and Wellner 1996.

4.1 Functional Principle of Permutation Tests

One on the defining features of a test is its null hypothesis. For the case of randomization tests, it is typically very general. Let X be data taking values in a sample space $\mathcal{X}$. Then, the hypothesis is that the probability law P generating X belongs to a family of distributions Ω_0. Lehmann and Romano 2005 define an assumption called the randomization hypothesis that allows for the construction of randomization tests. As permutation tests are a special case of randomization tests, we can specialize this definition to the case under consideration.

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 $P \in \Omega_0$.

Under this assumption 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 $k = M - \lfloor M\alpha \rfloor$. Additionally define the following two objects.

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

Definition 4.1 (Randomization Test Function)
We define the Randomization Test Function as the following function $\phi : \mathcal{X} \to \mathbb{R}$.

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

This function can be interpreted as a decider of whether a null hypothesis is rejected. If $\tau(X) = 1$, the null hypothesis is rejected, if $\tau(X) = 0$ it cannot be rejected, and if $\tau(X) = a$, 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.

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 4.2 (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 19.

$$
\begin{aligned}
\Pi_N &= \{\pi : \{1, \ldots, N\} \to \{1, \ldots, N\} \mid \pi \text{ is bijective}\} \\
G &= \{g : \mathbb{R}^N \to \mathbb{R}^N \mid \exists \pi \in \Pi_N \; \forall x \in \mathbb{R}^N \; g(x) = \left(x_{\pi(1)}, \ldots, x_{\pi(N)}\right)\}
\end{aligned}
\tag{19}
$$

Number of Combinations: $\binom{m+n}{m}$

For my implementation of the test described in Bugni and Horowitz 2021, I chose the latter variant and focused on distinct combinations if it was feasible to calculate the test statistic for all combinations. If this approach was infeasible due to the number of combinations, I resorted to randomly sampling permutations.

4.2 Size and Power

4.3 Asymptotic Properties

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 permutation test for the equality of the distributions of the data generating processes of two independent samples of functional data. To define the exact hypothesis, it is therefore necessary to define a distribution function for a random variable realizing in $\mathbb{L}_2[0,1]$.

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.

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

Deviating from the norm in functional data analysis, 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.

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 are necessary 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
Contains two assumptions

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 $\{X_i(t) \mid i = 1, \ldots, n\}$.

Definition 5.2 (Separable Stochastic Process)
From Wikipedia: A real-valued continuous time stochastic process X with a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is separable if its index set T has a dense countable subset $U \subset T$ and there is a set $\Omega_0 \subset \Omega$ of probability zero, so $\mathcal{P}(\Omega_0) = 0$, such that for every open set $G \subset T$ and every closed set $F \subset \mathbb{R}$ the two events $\{X_t \in F \quad \forall t \in G \cap U\}$ and $\{X_t \in F \quad \forall t \in G\}$ differ from each other at most on a subset Ω_0.

In less theoretical terms this means that the process is determined by its values on a countable subset of points of its index set.

Assumption 5.2
$\mathbb{E}X(t)$ and $\mathbb{E}Y(t)$ exist and are finite for all $t \in [0,1]$.

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 and is formulated by the authors as follows.

$$
\begin{aligned}
H_0: \quad & F_X(z) = F_Y(z) \quad \forall z \in \mathbb{L}_2[0, 1] \\
H_1: \quad & \mathbb{P}_\mu \left[F_X(Z) \neq F_Y(Z) \right] > 0
\end{aligned}
\tag{20}
$$

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.

$$
\hat{F}_X(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \left[X_i(t) \leq z(t) \ \forall t \in [0, 1] \right] \qquad \hat{F}_Y(z) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1} \left[Y_i(t) \leq z(t) \ \forall t \in [0, 1] \right]
\tag{21}
$$

Test statistic

$$
\tau = \int_{\mathbb{L}_2[0,1]} \left[F_X(z) - F_Y(z) \right]^2 \mathrm{d}\mu(z)
\tag{22}
$$

Sample analog:

$$
\tau_{n,m} = (n + m) \int_{\mathbb{L}_2[0,1]} \left[\hat{F}_X(z) - \hat{F}_Y(z) \right]^2 \mathrm{d}\mu(z)
\tag{23}
$$

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.

$$\hat{\tau}_{n,m} = \frac{n+m}{L} \sum_{l=1}^{L} \left[\hat{F}_X(Z_l) - \hat{F}_Y(Z_l) \right]^2 \tag{24}$$

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. Hier weiterschreiben Critical values for Permutation Test Statistic

$$t_{n,m}^*(1 - \alpha) = \inf \left\{ t \in \mathbb{R} \quad | \quad \frac{1}{Q} \sum_{i=1}^{Q} \mathbb{1} \left[\tau_{n,m,q} \leq t \right] \geq 1 - \alpha \right\} \tag{25}$$

5.4 Asymptotics for the Cramér-von Mises type Test

Similar to the case presented in Bugni, Hall, et al. 2009, we can derive an asymptotic distribution for the Cramér-von Mises type test. Even though this is not necessary to perform the described permutation test, it is an interesting benchmark to compare the test procedure.

Under the null hypothesis and mild regularity conditions, it is possible to show that

$$(n+m)^{\frac{1}{2}} \left[\hat{F}_X(z) - \hat{F}_Y(z) \right] \tag{26}$$

converges to a mean-zero Gaussian process. For the derivation look in Appendix 11.5.

5.5 Construction of the Measure μ

Another component of the test that has to be chosen is the probability measure with respect to which the hypothesis is formulated. This probability measure has a similar function to the weight function described in Section 3.3 in that it determines against which alternatives the test has comparatively high power. It is therefore important 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.

For the calculation of the Cramér-von Mises type test, we need to construct one such probability measure that is suitable to detect the kind of alternative we expect to find. As hinted at in Section 2.4, Bugni and Horowitz 2021 approach this problem by constructing a random function that induces a suitable probability measure. Its construction is shown below.

$$Z(t) = \sum_{k=1}^{\infty} b_k \psi_k(t) \quad \text{s.t.} \quad \sum_{k=1}^{\infty} b_k^2 < \infty \quad \text{a.s.} \tag{27}$$

$$Z_K(t) = \sum_{k=1}^{K} b_k \psi_k(t) \tag{28}$$

$$\mathbb{E}\left[Z_K(t)\right] = \sum_{k=1}^{K} \mathbb{E}\left[b_k\right] \psi_k(t) \quad \text{where} \quad \mathbb{E}\left[b_k\right] = \int_0^1 w(t)\psi_k(t)\mathrm{d}t \tag{29}$$

$$b_k = \mathbb{E}\left[b_k\right] + \rho_k U_k \quad \text{s.t.} \quad \sum_{k=1}^{\infty} \rho_k^2 < \infty \tag{30}$$

To obtain a reasonable degree of power it is important to set the expected value of the random function $Z(t)$ in a region, where data is observed and optimally in a region where realizations of the random function inducing the measure μ are able to distinguish potential differences in the samples. This could for example be done by choosing $w(t)$ dependent on the data. The authors suggest to choose $w(t) : [0,1] \to \mathbb{R}$ to be comparatively large in the parts of $[0,1]$ where possible differences between the empirical distribution functions are expected to be large. Taking the option used for the application in the original paper and one further possibility, the following two ways of choosing the mean function might come to mind.

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 $\mu_1 = median_i \left(max_t \left\{X_i(t) \mid i = 1, \ldots, n_1, t \in [0,1]\right\}\right)$. This method is identical to setting $w(t) = \mu_1(t)$, 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 for this choice is to use specific point-wise quantiles of the reference sample: $w(t) = quantile_q\{X_i(t) \mid i = 1, \ldots n_1\}$. This allows the generated functions for the Monte-Carlo integration in the approximation of the Cramér-von Mises statistic to resemble the reference data more closely, which could improve the properties of the test.
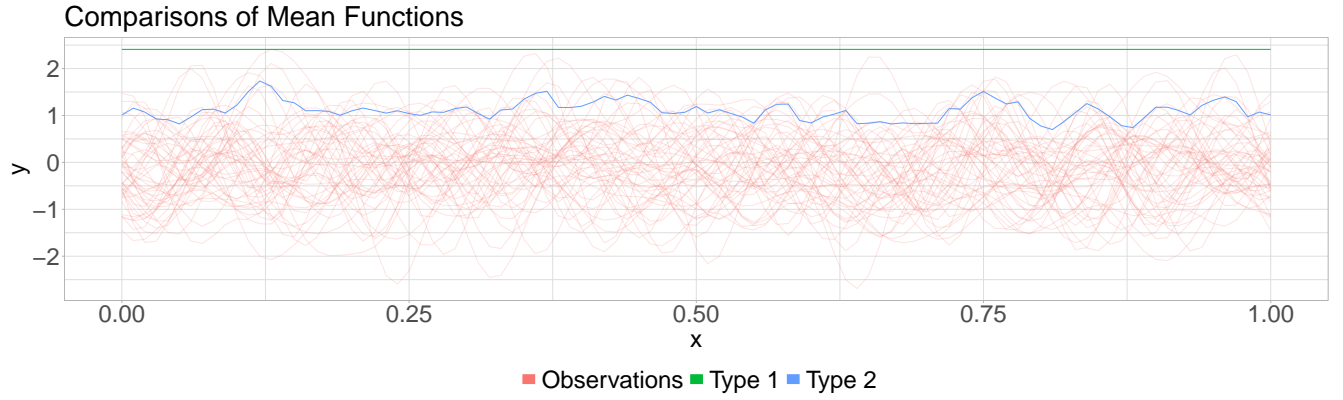
Figure 3: Mean Functions calculated for a Reference Sample as described in Section 7 for the 95% Quantile.

Additionally, the choice of the sequence $(\rho_k)_{k \in \mathbb{N}}$ is of importance. The authors do not recommend a procedure for choosing these coefficients, but it seems reasonable to choose parameters that lie in the vicinity of the standard deviations for the Fourier coefficients of the reference sample. This method might be a reasonable possibility to choose these parameters dependent on the data.

In Bugni, Hall, et al. 2009, the authors show that the approximation of the probability measure, which is induced by the approximation of the random variable in Equation 28, converges appropriately to the probability measure induced by the random variable in Equation 27. Hier weiterschreiben

At this point it is again interesting to look at the convention that two elements of $\mathbb{L}_2[0,1]$ are distinct even if they differ only on a nonempty set of Lebesgue-measure zero. One problem that this convention entails is the fact that the Fourier basis is an almost everywhere basis of $\mathbb{L}_2[0,1]$ as shown by Carleson 1966. However, in many cases point-wise convergence of a sum as shown in Equation 27 is not fulfilled. This is simple to see for any function that does not have identical values on both sides of its domain. Hier weiterschreiben

In more basic terms, this implies that using the Fourier basis, it is impossible to construct some functions in $\mathbb{L}_2^*[0,1]$ even using a non-truncated representation as shown in Equation 27. In a typical scenario where we would use $\mathbb{L}_2[0,1]$, so only their equivalence classes of the functions, this does not pose a problem. This in turn implies that a probability measure that is constructed as shown in the previous section cannot give positive weight to many functions in the space $\mathbb{L}_2[0,1]$ if we use the convention used by the authors. Namely, for any function in $\mathbb{L}_2[0,1]$ for which the Fourier series fails to converge point-wise, we cannot assign a positive probability. This in turn could be a potential problem for the method described in the paper, if this is necessary for its working principle.

17

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

$$\nu = \int_0^1 \left(\mathbb{E}\left[X(t)\right] - \mathbb{E}\left[Y(t)\right] \right)^2 \mathrm{d}t \tag{31}$$

Mean Estimators

$$\hat{\mathbb{E}}\left[X(t)\right] = \frac{1}{n}\sum_{i=1}^n X_i(t) \qquad\qquad \hat{\mathbb{E}}\left[Y(t)\right] = \frac{1}{m}\sum_{i=1}^m Y_i(t) \tag{32}$$

Sample Analog

$$\nu_{n,m} = (n+m)\int_0^1 \left[\hat{\mathbb{E}}X(t) - \hat{\mathbb{E}}Y(t)\right]^2 \mathrm{d}t \tag{33}$$

Critical values for Permutation Test Statistic

$$t_{n,m}^*(1-\alpha) = \inf\left\{ t \in \mathbb{R} \quad | \quad \frac{1}{Q}\sum_{i=1}^Q \mathbb{1}\left[\nu_{n,m,q} \le t\right] \ge 1-\alpha \right\} \tag{34}$$

5.7 Combined Permutation Test

Define for the two underlying tests the following permutation test functions as described in Definition 4.1 for the general case of a randomization test.

$$\phi_{n,m} = \begin{cases} 1 & \text{if } \tau_{n,m} > t_{n,m}^*(1-\alpha_\tau) \\ a_\tau & \text{if } \tau_{n,m} = t_{n,m}^*(1-\alpha_\tau) \\ 0 & \text{if } \tau_{n,m} < t_{n,m}^*(1-\alpha_\tau) \end{cases} \qquad \tilde{\phi}_{n,m} = \begin{cases} 1 & \text{if } \nu_{n,m} > t_{n,m}^*(1-\alpha_\nu) \\ a_\nu & \text{if } \nu_{n,m} = t_{n,m}^*(1-\alpha_\nu) \\ 0 & \text{if } \nu_{n,m} < t_{n,m}^*(1-\alpha_\nu) \end{cases} \tag{35}$$

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

- $a_\tau = \frac{Q\alpha_\tau - Q_\tau^+}{Q_\tau^0}$ \qquad\qquad • $a_\nu = \frac{Q\alpha_\nu - Q_\nu^+}{Q_\nu^0}$

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

- $Q_\tau^0 = \sum_{q=1}^Q \mathbb{1}\left[\tau_{n,m,q} = t_{n,m}^*(1-\alpha_\tau)\right]$ \quad • $Q_\nu^0 = \sum_{q=1}^Q \mathbb{1}\left[\nu_{n,m,q} = t_{n,m}^*(1-\alpha_\nu)\right]$

Bonferroni inequality under H_0 leads to

$$\max(\alpha_\tau, \alpha_\nu) \le \mathbb{P}\left[(\phi_{n,m} > 0) \cup (\tilde{\phi}_{n,m} > 0)\right] \le \alpha_\tau + \alpha_\nu \tag{36}$$

5.8 Finite Sample Properties under the null hypothesis

For any distribution P that satisfies the null hypothesis and any $\alpha_\tau, \alpha_\mu \in (0,1)$, we have

$$\mathbb{E}_P\left(\phi_{n,m}\right) = \alpha_\tau \qquad\qquad \mathbb{E}_P\left(\tilde{\phi}_{n,m}\right) = \alpha_\nu \qquad (37)$$

5.9 Asymptotic Properties under the Alternative

6 Test for Persistence Alternatives

One particularly interesting potential 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. We assume that the samples consist of i.i.d. realizations of random functions $X(t)$ and $Y(t)$ respectively. To focus on the persistence properties of the random functions, we can define the standardized counterparts to these random functions in the following way. Define the following objects.

$$\begin{aligned} \mu_X(t) &= \mathbb{E}X(t) & \mu_Y(t) &= \mathbb{E}Y(t) \\ \sigma_X(t) &= \sqrt{\mathbb{E}\left[X(t) - \mu_X(t)\right]^2} & \sigma_Y(t) &= \sqrt{\mathbb{E}\left[Y(t) - \mu_Y(t)\right]^2} \end{aligned} \qquad (38)$$

Then the standardized random variables can be defined as follows.

$$\tilde{X}(t) = \frac{X(t) - \mu_X(t)}{\sigma_X(t)} \qquad\qquad \tilde{Y}(t) = \frac{Y(t) - \mu_Y(t)}{\sigma_Y(t)} \qquad (39)$$

Their distribution functions are defined analogously to Section 5 and denoted by $F_{\tilde{X}}(z)$ and $F_{\tilde{Y}}(z)$ respectively. Then we can formulate a Null-hypothesis as in Bugni and Horowitz 2021 with respect to the standardized random functions.

$$
\begin{aligned}
H_0 : \quad & F_{\tilde{X}}(z) = F_{\tilde{Y}}(z) \quad \forall z \in \mathbb{L}_2[0.1] \\
H_1 : \quad & \mathbb{P}_\mu\left[F_{\tilde{X}}(z) \neq F_{\tilde{Y}}(z)\right] > 0
\end{aligned}
\qquad (40)
$$

As all differences in the mean and variance patterns of these processes is eliminated by the standardization, one of the distinct remaining features that might differentiate these processes are their persistence properties. Therefore, as this test in its 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 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 1 and $\mathcal{V} = \{V_i(t) \mid i = 1, \ldots, m\}$ be sample 2. These samples are generated by randomly permuting the original samples $\{X_i(t) \mid i = 1, \ldots, n\}$ be sample 1 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 a derivation of a permutation distribution and a permutation test decision according to the rules stated in Section 4.

Due to the introduction of a standardization step after the permutation of the samples, the precise order of operations chosen for the implementation is now of considerable importance if the sample does not actually consist of continuously observed curves but of sequences of discrete observations sharing the points where observations are made. This could for example mean a setting where at distinct points in time, measurements are made for a group of individuals.

1. The intuitive approach to standardizing observations like this would be to standardize the discrete observation points in each permutation step and to fit a functional basis to the discrete sequence of standardized observations.

2. Another approach would be to fit the functional basis to the non-standardized samples and to standardize on the level of the fitted curves.

This distinction can become important as the resulting curves that are compared in the derivation of the Cramér-von Mises type test statistic might look considerably different depending on the chosen approach. In Section 7.4, two heuristic simulations are performed to compare the difference in power between these approaches.

7 Simulation Study

To learn more about the potential practical applications of this method, it is useful to study its properties in a simulation. Bugni and Horowitz 2021 studied an array of different setups that give an idea of the performance in different settings and this thesis will replicate some of these results and explore some where the method struggles.

7.1 Implementation as an R package

All analyses in this thesis have been conducted with R[2]. I implemented the two-sample variant of the test presented taken from Bugni and Horowitz 2021 in an R package called *PermFDATest*. The R package and all code that has been used to produce the following results are publicly available as part of a GitHub repository[3] that complements this thesis.

- Ramsay, Graves, and Hooker 2021

- Wickham et al. 2019

- Goldsmith et al. 2021

---

[2]R Core Team 2022.
[3]https://github.com/JakobJuergens/Masters_Thesis

7.1.1 Use of High-Performance Computing

The simulations presented as part of this thesis have been conducted on *bonna*[4]. *bonna* is the high performance computing cluster provided by the University of Bonn. The implementation is heavily parallelized and makes use of a SLURM scheduling system. However, slight modifications of the provided code suffice to run it on personal computers.

7.1.2 Particularities of the Implementation

For my implementation in the package that is provided in the GitHub repository corresponding to this thesis, I chose some slightly unusual methods to increase the package's performance. I want to present some in the following as they present opportunities for performance gains that are somewhat interesting on their own.

- For the evaluation of the empirical distribution function, I make use of a result Boyd 2006 that simplifies the problem of finding the zeroes of a finite Fourier series to an eigenvalue problem for a matrix that is defined in terms of the Fourier coefficients. If the chosen basis is the Fourier basis this is more rigorous than the grid based approach used for the other basis types and faster then potential numerical methods to approximate the zeroes in a more general setting.

7.2 Simulation Setup

As part of this thesis, I take inspiration from 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, I introduce the notation used in Bugni and Horowitz 2021 shortly to then describe the data generating processes mathematically. I generate the data on the closed interval $[0, 1]$ by discretely generating observations and fitting a Fourier basis to the discrete data to obtain functional observations. Let $\mathcal{T} = \{0, 0.01, 0.02, \ldots, 1\}$ be the index set of the discrete points on the unit interval, then we construct the observations in the following way.

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.

---

[4]https://www.dice.uni-bonn.de/de/hpc/hpc-a-bonn/infrastruktur

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

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

3. $\mathrm{Corr}\left[X_s(t), X_s(t-1)\right] = \rho_s(t)$

$\forall t \in [0,1]$ with $t > 1$

As can be seen in the mathematical description above, deviating from the original paper, I focus on the two sample setting and use a sample size of 20 observations per sample. I focus on three distinct violations of the null hypothesis described in the following. For each of the following settings I construct sample 1 identically as a reference point to compare the different settings. 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, I chose three different violations of the null hypothesis and a benchmark case where the null hypothesis is correct.

1. **Identical Data Generating Processes**

   The first setting is the benchmark where both samples are generated using the same random function.

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.

   $$\rho_2(t) = 0.2 \quad \forall t = 0.01, \ldots, 1$$

4. **Variance Shift**

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

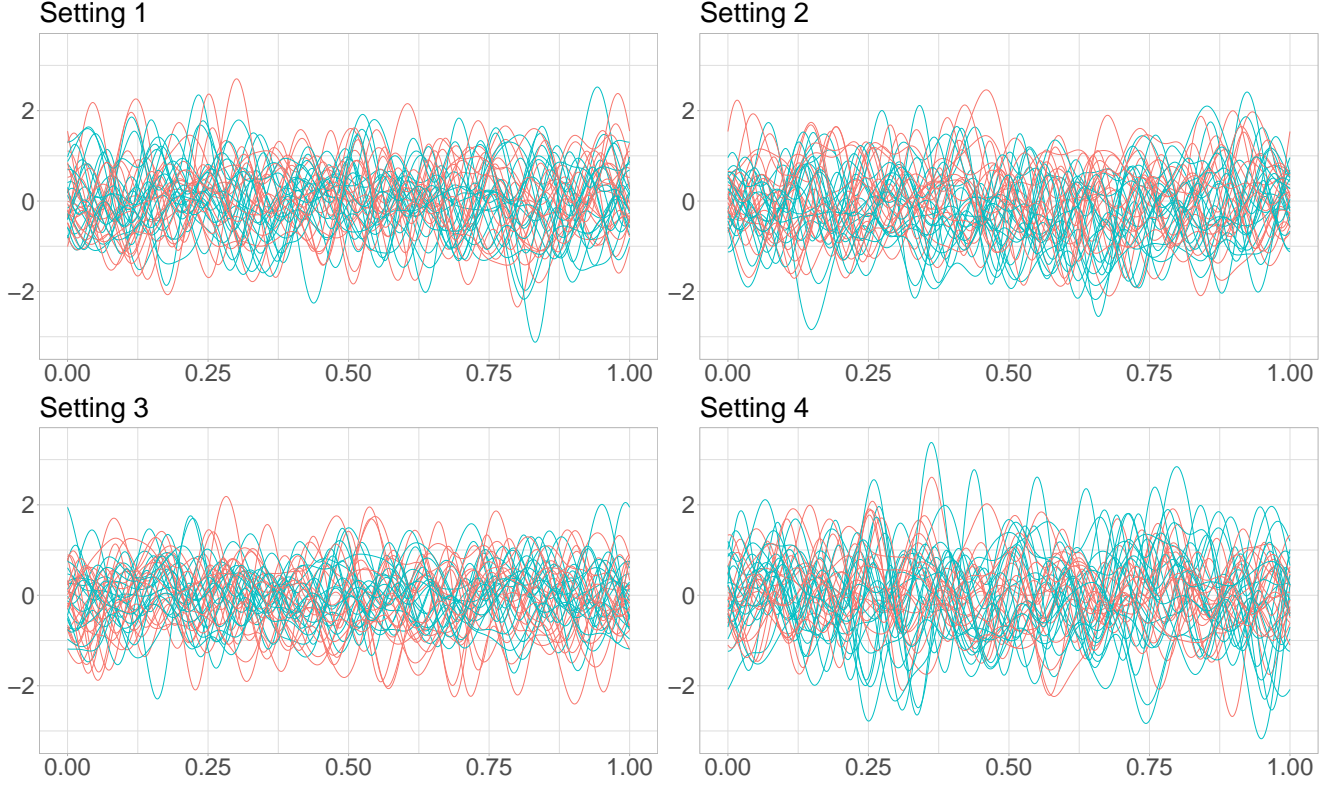 $$\sigma_2(t) = 1 + 0.5t \quad t \in \mathcal{T}$$

Figure 4: Samples generated for the four settings.
Sample 1 in red, Sample 2 in blue

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, I chose 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. The truncation parameter K for the construction of the random function inducing the measure μ as shown in Equation 28. The original 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 I used the same choice.

3. The parameter L determining the number of functions drawn according to μ in the approximation of the test statistic $\hat{\tau}_{n,m}$ of the Cramér-von Mises type test. The authors of the original paper chose $L = 4000$ and preliminary testing confirmed that this choice seems to be reasonable in many settings when combined with reasonable choices for the construction of the measure μ.

4. The number of permutations Q used for the approximation of the permutation distribution of both test statistics. The authors chose a value of $Q = 500$ which I copied for the sake of reproducing their results.

5. For the construction of the measure μ, the weight function $w(t)$ has to be chosen. For the purposes of this simulation, the weight function is given as the point-wise 95% quantile of the observations of the reference sample. This is a potentially significant deviation from the approach chosen by the original authors.

6. Deviating from the choice of the original authors due to the results of preliminary testing, I chose the sequence $(\rho_i)_{i=1,\dots,K}$ by checking the sample standard deviation of the corresponding Fourier coefficients in the first sample. I then doubled these values 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, I chose to use independent standard normal error terms $U_k \quad k = 1, \dots, K$ for the Fourier coefficients shown in Equation 30 for the construction of the measure. As the processes under consideration are Gaussian by construction, this choice is reasonable. For distributions with heavier tails, the authors recommend using different error terms and it would be interesting to study the effects of different choices.

7.3   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	0.519	0.057	0.057
CvM-Test	$\alpha_\tau = 0.05$	0.04147313	0.1842943	0.09938171	0.1350278
Combined	(0.04, 0.01)	0.04353598	0.3652623	0.09063807	0.1272009
	(0.03, 0.02)	0.0527092	0.4280618	0.08540022	0.121842
	(0.025, 0.025)	0.05417252	0.4456977	0.08476733	0.1137233
	(0.02, 0.03)	0.05297339	0.4585891	0.07644395	0.1093622
	(0.01, 0.04)	0.05927385	0.4860989	0.06362453	0.1001554

Table 1: Empirical Rejection Probabilities

Even though these rejection frequencies seem low at first, one has to keep in mind the magnitude of the violation of the null hypothesis and the small sample size. Looking at Figure 4, differences between the processes are not particularly stark visually and a low power therefore somewhat expected. Additionally, the power of permutation tests is highly dependent on the sample size of the underlying samples. Therefore, a rather small sample size of 20 as used in these simulations justifies the rejection frequencies. It would be interesting to repeat these simulations for a number of different sample sizes as this could give a more precise idea of the context in which this procedure has higher power.

7.4 Simulation - Test for Persistence Alternatives

Using simulation setting 3, it is possible to also test the proposed test for persistence alternatives from Section **??**. But due to computational constraints, this simulation does not follow the real test structure and shall instead serve as a heuristic for the real test's performance. Mainly 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	0.830	-	0.250	0.0415	-
0.01	0.783	-	0.119	0.0104	-
0.001	0.735	-	0.0492	0.00261	-

Table 2: Empirical Rejection Probabilities

The apparent power of this test comes in part from the basis that is used to fit the data. Even though standardization of the original observations as described in approach one shown in Section 6.2 ensures that the discrete data itself has a point-wise sample variance of one, the method that is used to fit the observations can lead to curves that are highly dissimilar between samples once the data points are fitted. Figure 5 shows a rather extreme case for the case of $\rho_1(t) = 0.5$ and $\rho_2(t) = -0.9$. In Appendix 11.7 Figure 12 shows some original data sets generated for these simulations.These are contrasted by Figure 13 which shows the same samples fitted using a Fourier basis containing 25 functions. Even though the underlying data generating processes have an identical point-wise variance, the method that is used to fit the discrete observations, creates potentially distorting artifacts.
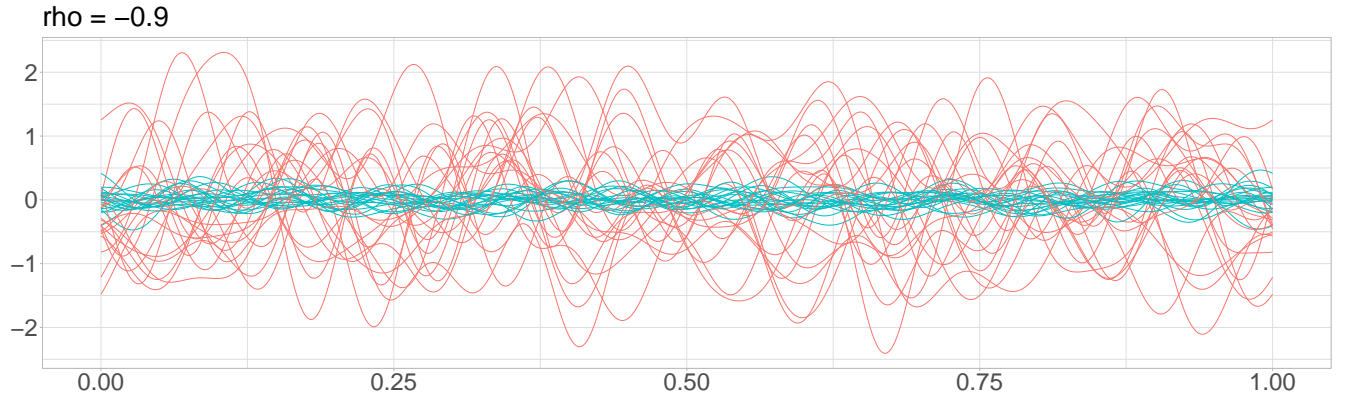
Figure 5: Artifact of the Curve Fitting Procedure

In some settings this could be a desirable property of the test. However, it is difficult to predict how this property could influence the test's performance in real world settings. Therefore, it could be advisable to modify the test in one of two ways to mitigate these artifacts.

1. Standardize the samples of fitted curves to have a point-wise mean of zero and a point-wise variance of one.

2. Using methods to compare the functions that do not rely on fitting a functional basis to the original observations.

The first option The latter option could for example be implemented in the following ways.

1. Compare only the actual discrete observations with their corresponding values on the functions generated for Monte-Carlo integration.

2. Use linear interpolation on the observations to compare points on a chosen grid with their corresponding values.

Variant one can be seen as close to approach two from Section 6.2 as the standardization is applied to every point that is actually considered during the comparison of the functions for the Craér-von Mises test. The point-wise functional standardization mentioned in 6.2 is computationally intensive and requires a more advanced method of comparison between functions. However, as variant one described above can be seen as a heuristic for its performance, I performed a second simulation using this approach, leading to the results presented in Table 3. These could potentially be a better approximation of the power of the test procedure as the rejections do not rely on artifacts created by the process of fitting a functional basis to the discrete observations.

8 Application

To test the real-world merits of the method, I will compare electricity demand data from Adelaide which is provided as part of the *fds*[5] package for R. This data set presents

---

[5]Shang and Rob J Hyndman 2018.

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	-	-	-	-	
0.01	-	-	-	-	
0.001	-	-	-	-	

Table 3: Empirical Rejection Probabilities

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, is shown in Figure 6.
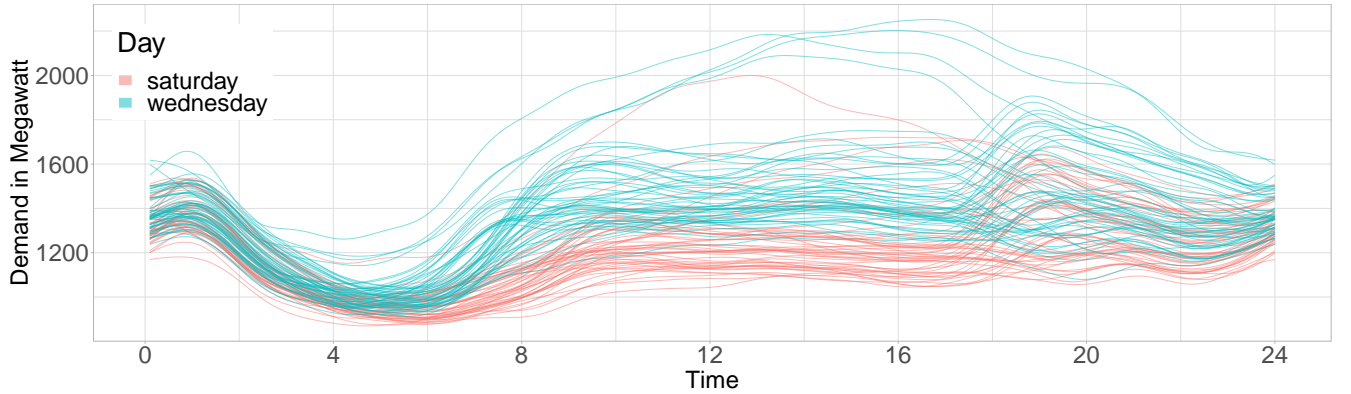

Figure 6: Electricity Demand in Adelaide

The question I want to study 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.

As this problem does not have the structure that an experiment as described in Bugni and Horowitz 2021 possesses, a few problems have to be addressed before using the procedure. These adjustments are described below and some details are substantiated in Appendix 11.6.

8.1 Data Cleaning and Preprocessing

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 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 11.6. 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 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.

$$
\begin{aligned}
f_{demand} =& f_{mean} + f_{trend}(year - 1997) + \sum_{j=2}^{12} \mathbb{1}_{[month = j]} f_{month,j} \\
&+ \sum_{k=2}^{7} \mathbb{1}_{[day = k]} f_{day,k} + f_{random}
\end{aligned}
\tag{41}
$$

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.

$$
\tilde{f} = f_{mean} + \sum_{k=2}^{7} \mathbb{1}_{[day = k]} \hat{f}_{day,k} + \hat{f}_{random}
\tag{42}
$$

The function-on-scalar regression described in Equation 41 was performed in R using the fda[6] package and gave the following results. As these are the results of a function-on-scalar

---
[6]Ramsay, Graves, and Hooker 2021.

regression, it is convenient to plot the resulting estimates as they are functions instead of giving the estimated Fourier coefficients.
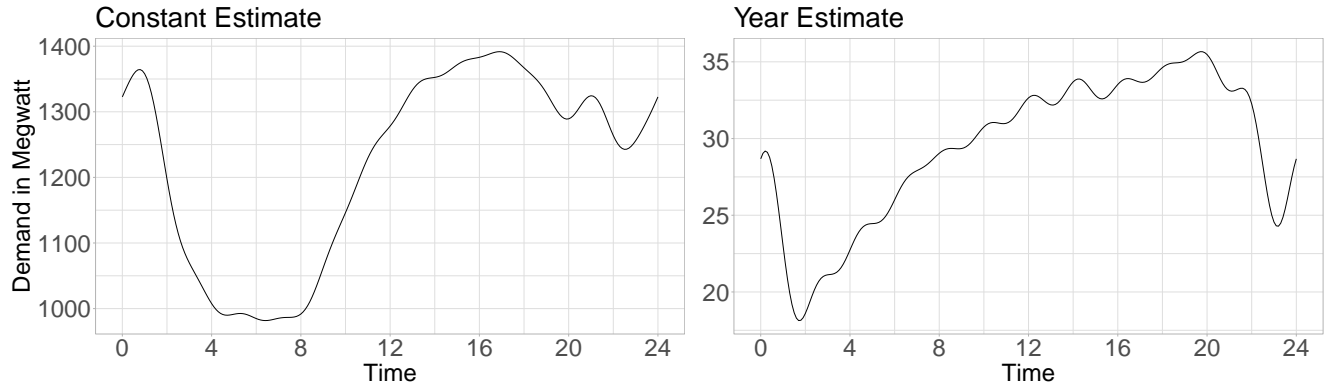

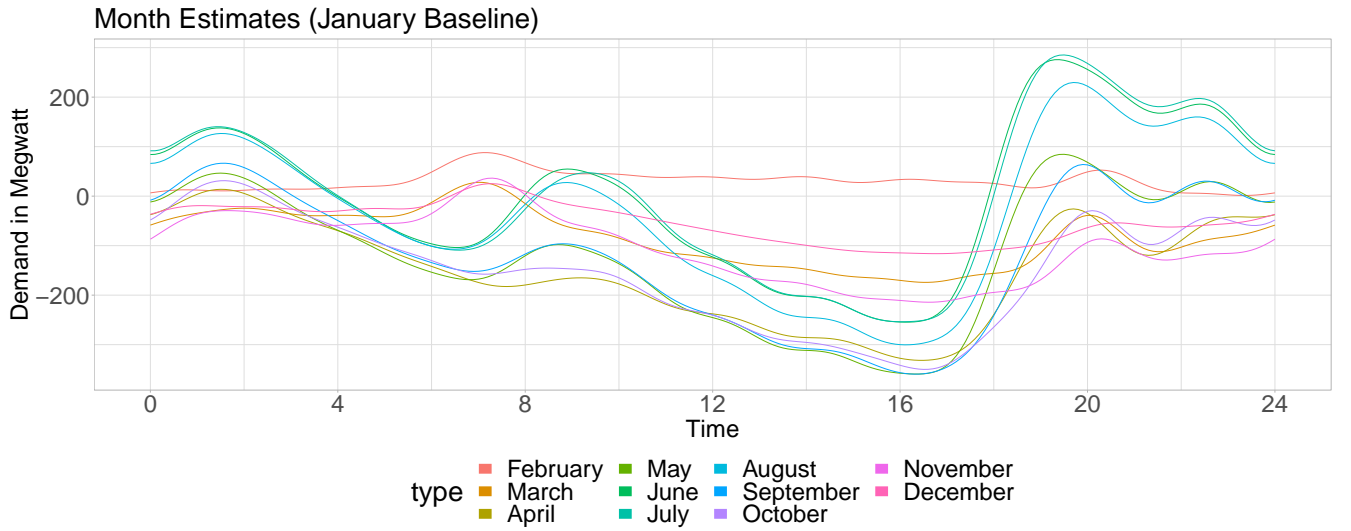Figure 7: Estimates for the constant and year

Figure 8: Estimates for the months (January as baseline)

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.
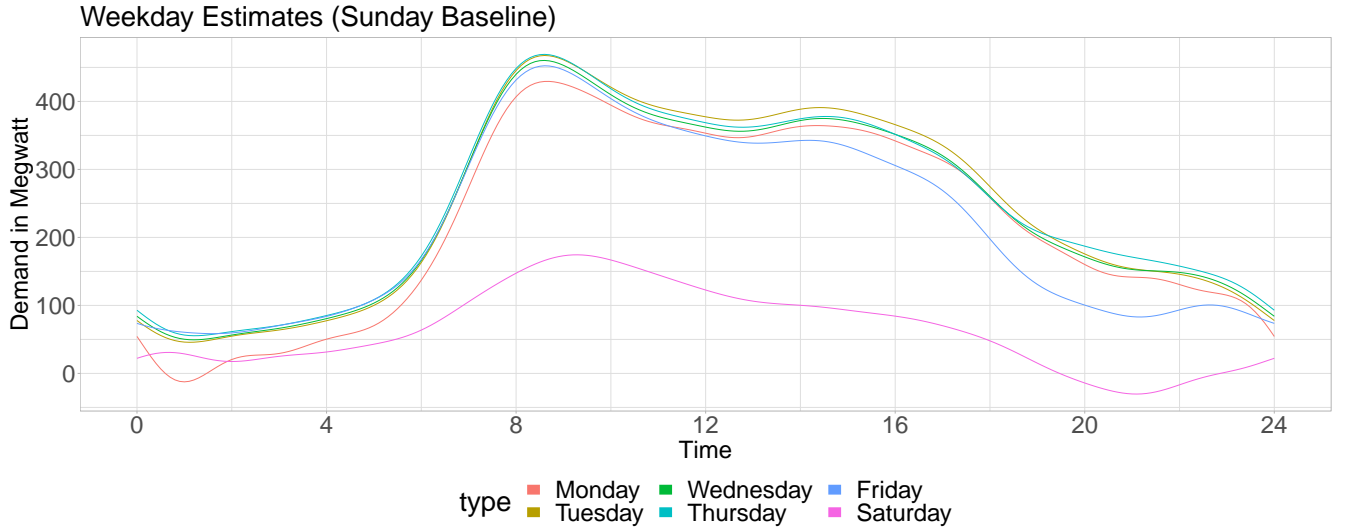
Figure 9: Estimates for the weekdays (Sunday as baseline)

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 regression provide clear evidence against the assumption that Saturday and Sunday can be seen as if they were generated by the same random variable.

8.2   Test from Bugni and Horowitz 2021

After cleaning the data as described in the previous section, we can now apply the method from Bugni and Horowitz 2021. To give a visualization of what these cleaned curves look like, Figure 10 shows randomly chosen curves observed on Wednesdays and Saturdays that were prepared with the procedure described above.
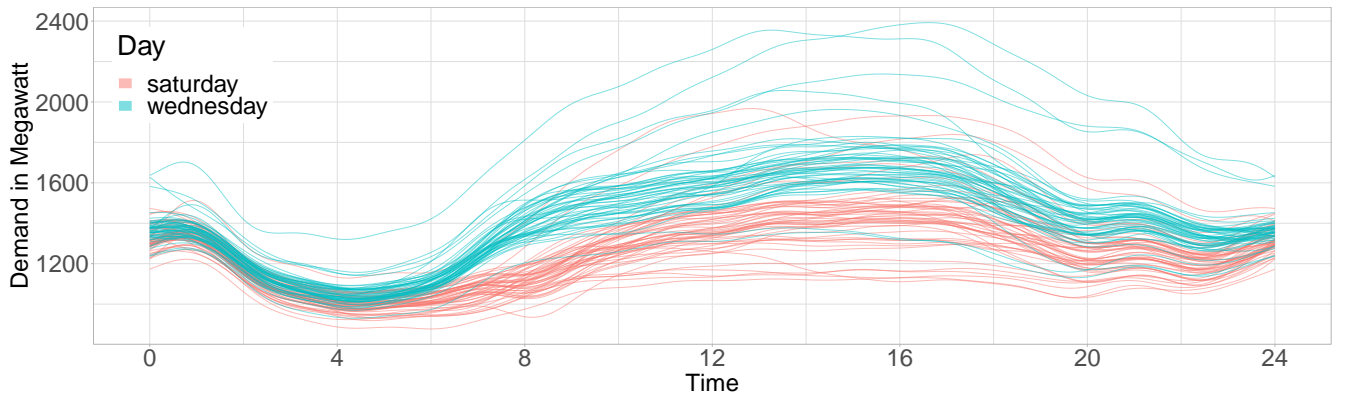

Figure 10: Pre-Processed Electricity Demand in Adelaide

Applying this test to the cleaned data leads to the following p-values for the two underlying tests.

- $p_\tau =$ hier p-value ergaenzen
- $p_\nu =$ hier p-value ergaenzen

Therefore using the same decision rules as in the simulation study, we come to the following rejection decisions. These rejections of the null hypothesis are unsurprising when looking

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 4: Results of the Test from Bugni and Horowitz 2021

at the curves that were generated for the different days in the data cleaning step. There, it was already obvious that the difference in mean alone was stark between workdays and Saturdays. Therefore, it was expected for the means based test to find the violation of the null hypothesis in this comparatively large sample problem. Due to the large differences in mean, 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, it would be interesting to apply this test to some more challenging settings. A small selection of settings is therefore presented in Section 9.

9 Outlook

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.

9.1 Possible Further Simulations

Comparing different Functional Bases

For the construction of the measure and fitting the observations. Hier weiterschreiben.

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

In this thesis I only perform simulations for one intuitive choice of $w(t)$ and $(\rho_i)_{i=1,\ldots,K}$. 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 7.4, the actual implementation could be highly important for the properties of the proposed test for persistence alternatives. This thesis only provided 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 simulations.

Kolmogorov-Smirnov Variant of the test

Hier weiterschreiben

9.2 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`. (Visited on 05/24/2022).

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`.

Carleson, Lennart (Jan. 1966). "On convergence and growth of partial sums of Fourier series". In: *Acta Mathematica* 116.none. Publisher: Institut Mittag-Leffler, pp. 135–157. DOI: `10.1007/BF02392815`.

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.

Goldsmith, Jeff et al. (2021). *refund: Regression with Functional Data.* R package version 0.1-24. URL: https://CRAN.R-project.org/package=refund.

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.

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 that are unfamiliar with the idea of 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 shall 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 as to not 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. What we want to test is 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 being that these observations are curves. So we can ask the question 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 10 years. We do this to find out if two different types of fuel change the way the cars act on the track. In this hypothetical scenario, we now have two samples each containing 3650 speed curves. We want to test statistically whether changing the fuel made any difference.

Question three is more difficult. Therefore, I want to give the right intuition by answering simpler question instead. Let's return to the scalar setting for a moment - each observation is just a real number $x \in \mathbb{R}$ again. We want to find out whether two samples are different enough for us to say confidently: "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. But let's 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 said 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 should you care about this is the hardest question. I care about it, because it is cool from a mathematical perspective. But 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 transforms 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.

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_0^1 [f(t)]^2 \, dt} = 0 \tag{43}$$

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 $C_{m,n}$.

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

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

(44)

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.

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

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

11.4 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.

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

(45)

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

11.5 Derivation of Asymptotic Distribution

Here I will derive the asymptotic distribution

11.6 Application Data Cleaning

11.6.1 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,

---

[7]These were taken from `https://www.australia.gov.au/public-holidays` accessed on 20.05.2022.

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.

11.6.2 Detrending and Deseasoning Results
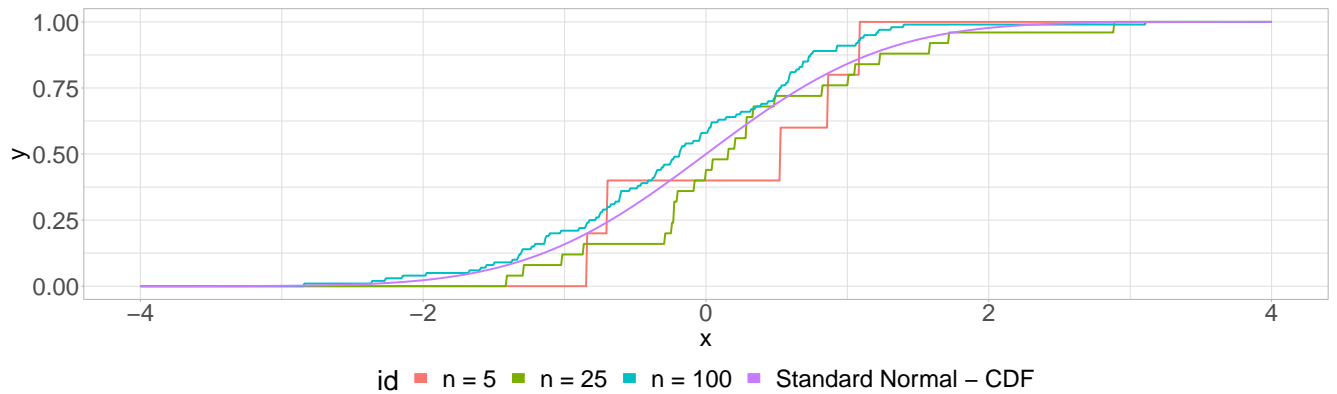
11.7 Additional Figures

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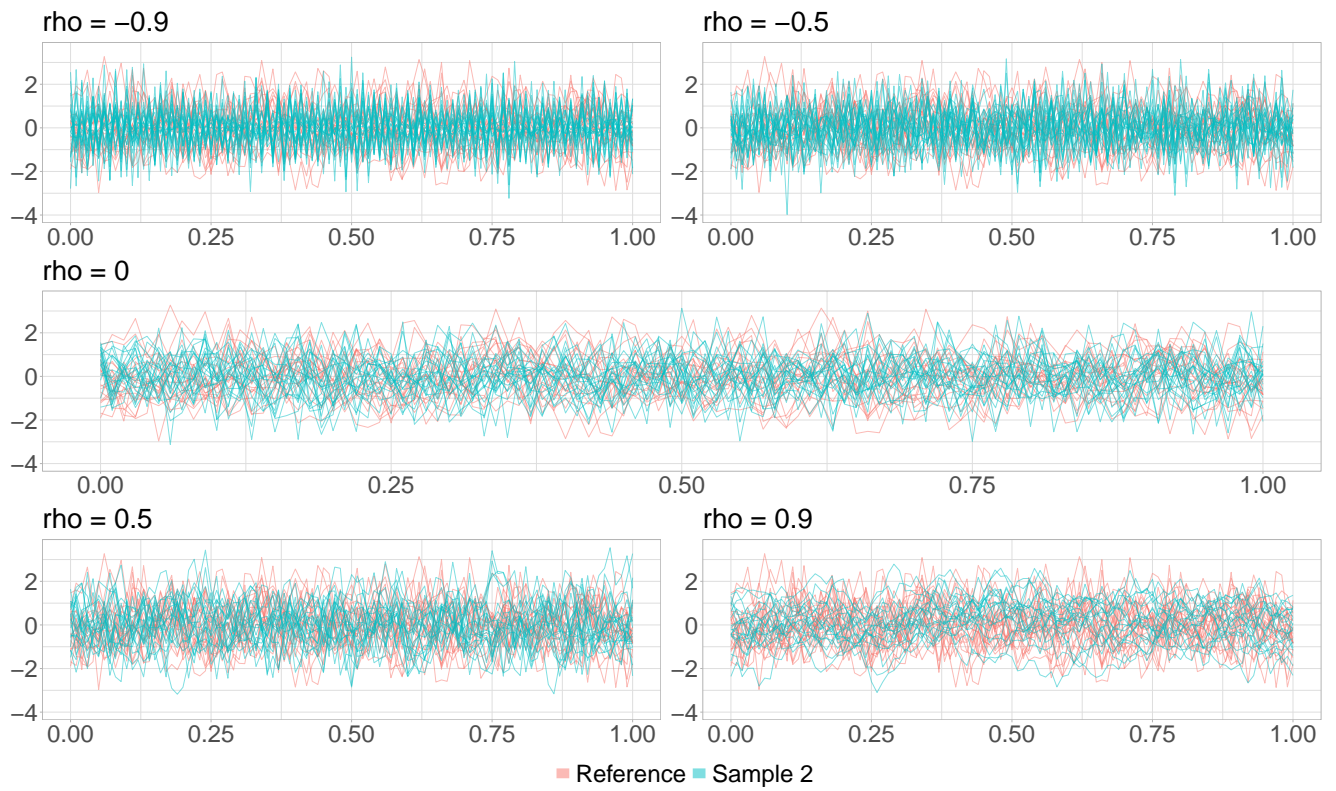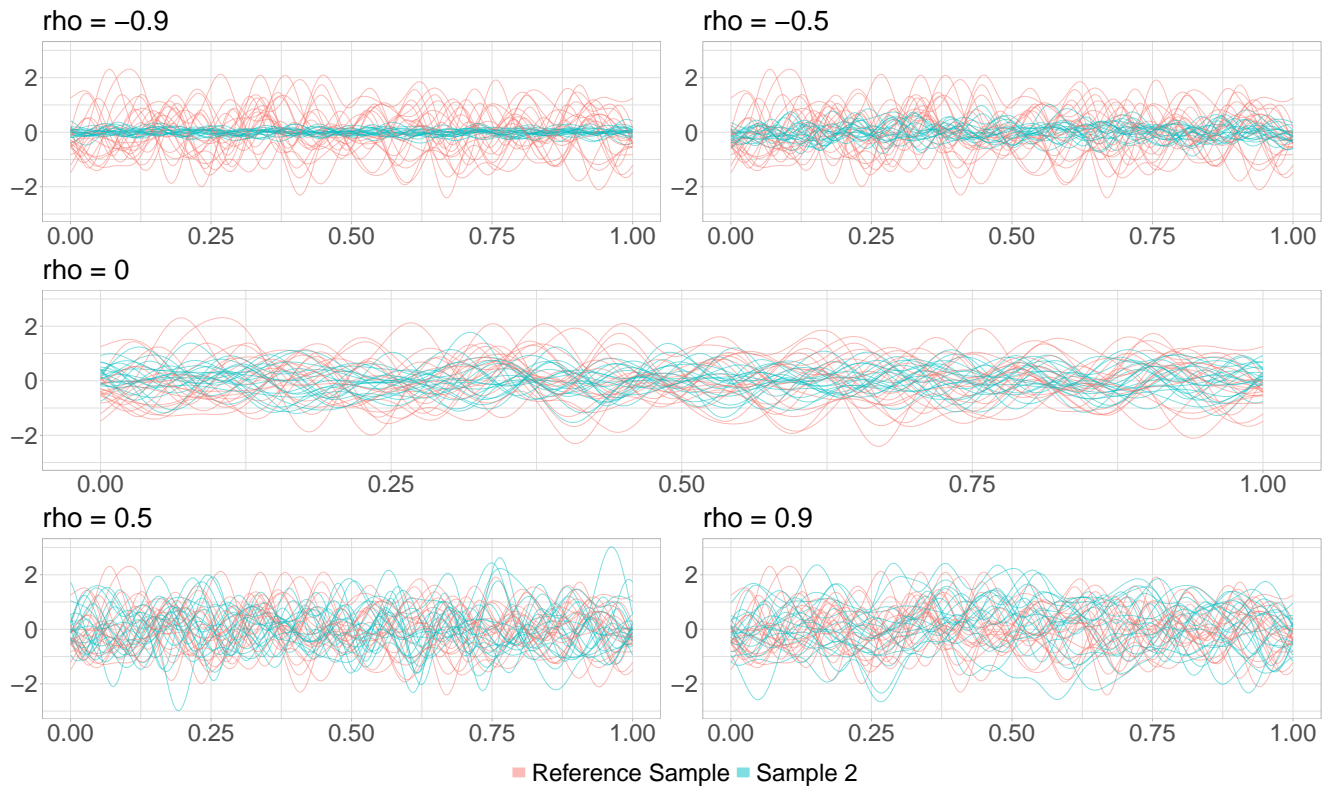
Figure 12: Original Samples for Persistence Test

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