

DELFT UNIVERSITY OF TECHNOLOGY

MASTER'S THESIS

**RECOVERING LOW-DIMENSIONAL STRUCTURE IN
ESTIMATED CONDITIONAL KENDALL'S TAU
CURVES VIA FUNCTIONAL PCA**

Author:

S. H. H. Groen

Supervisor:

Dr. A. F. F. Derumigny



RECOVERING LOW-DIMENSIONAL STRUCTURE IN ESTIMATED CONDITIONAL KENDALL'S TAU CURVES VIA FUNCTIONAL PCA

by

Stan GROEN

to obtain the degree of Master of Science in Applied Mathematics at the faculty of
Electrical Engineering, Mathematics and Computer Science
at the Delft University of Technology.

This thesis is to be defended publicly on Wednesday June 24, 2026.

Student number: 5314380
Thesis duration: October 2025 - June 2026
Thesis committee: Dr. A. F. F. Derumigny, TU Delft, supervisor
Dr. K. Kurowicka, TU Delft

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.



ABSTRACT

Functional Principal Component Analysis (FPCA) is a powerful method for identifying the dominant modes of variation within a large collection of curves, thereby revealing their underlying low-dimensional structure. However, the existing FPCA framework assumes that the functional data are directly observed. This thesis develops a new theoretical framework for applying FPCA when the functional objects are estimated rather than observed, focusing specifically on estimated conditional Kendall's tau (CKT) curves. The analysis is supported by additional theoretical work on conditional U -statistics, which form the basis for establishing the asymptotic properties of the proposed method. In the theoretical setting where the CKT curves are estimated over the entire domain, we prove the consistency of the associated Hilbert-Schmidt integral operator and under suitable regularity and asymptotic conditions, we show that the spectral decomposition of the estimated operator converges to the true one with an error rate of $\mathcal{O}(h^2 + (nh^d)^{-1/2})$. We then extend the analysis to the practically relevant setting in which the curves are observed on a finite grid of N points and establish convergence of the spectral properties with a rate of $\mathcal{O}(N^{-1/d} + h^2 + (nh^d)^{-1/2})$. This illustrates the bias-variance trade-off with respect to the choice of the bandwidth h , as well as the numerical bias introduced by the discretization step, $N^{-1/d}$. We support our theoretical results with an empirical study of global financial assets, applying ARMA-GARCH filtering and conditioning on different markets to recover the most dominant modes of variation within a large dataset of CKT curves. We find that all CKT curves can be approximately summarized by only two degrees of freedom: a constant baseline level, representing the average conditional Kendall's tau, and a parabolic component that contrasts extreme market conditions with more typical periods, capturing the contagion effects. Ultimately, this framework provides a foundation for future applications, including evaluating the simplifying assumption.

ACKNOWLEDGEMENTS

This thesis marks the end of my time as a Master's student in Applied Mathematics. The past year has been an incredible journey, filled with a few challenges but far more rewards. I have learned an immense amount over the last few months, and I am proud to conclude my student career here in Delft.

I began my time in Delft with a double Bachelor's degree in Applied Mathematics and Applied Physics. Although I thoroughly enjoyed physics, I realized that my true passion lies in mathematics, specifically in financial modeling. When I approached Alexis, whose Financial Mathematics course had sparked my enthusiasm, to discuss a Master's thesis, he immediately came with very interesting ideas for potential directions. The path forward quickly became clear, and I embarked on this nine-month journey.

I would like to start by tremendously thanking Alexis for his guidance and support during my project. I truly appreciate your dedication and the substantial amount of time you invested in meeting with me each week. You always took the time to clarify complex concepts and steer me in the right direction, while still giving me freedom to explore my own ideas. I really enjoyed the discussion we had about unravelling the problems we faced and also the many talks not related to the project. This thesis genuinely feels like a joint effort, and I could not have wished for a better supervisor. Furthermore, I would also like to thank Dr. D. Kurowicka for taking the time to read my report and attend the presentation.

Finally, I want to thank my friends and family who supported me throughout my entire studies.

*Stan Groen
Uitgeest, June 2026*

CONTENTS

Abstract	v
Acknowledgements	vii
Index of Assumptions	x
1 Introduction	1
2 Preliminaries	3
2.1 Copulas	3
2.2 Measures of Association.	6
2.3 Simplifying Assumption.	8
2.4 Estimation Methods	9
2.4.1 Kernel Density Estimation	9
2.4.2 Kernel Regression and CKT Estimation	10
2.4.3 Bias-variance Trade-off	13
2.5 Stochastic o and $\mathcal{O}$ Notation	14
2.6 Linear Algebra	15
2.6.1 Hilbert-Schmidt Integral Operators	17
2.6.2 Spectra of Operators	21
2.6.3 Modes of Convergence.	23
2.6.4 Auxiliary Results	25
2.7 U -statistics	27
2.7.1 Conditional U -statistics	28
2.8 Principal Component Analysis	29
2.8.1 PCA in Practice.	31
2.9 Functional Principal Component Analysis	32
2.9.1 FPCA in Practice	34
2.9.2 Regularized FPCA	37
2.10 ARMA-GARCH Models	38
3 Statistical Model	41
3.1 FPCA for the True Conditional Kendall's Tau	41
3.2 FPCA for the Estimated Conditional Kendall's Tau	44
3.3 The Simplifying Assumption and its Spectral Characterization	46
4 FPCA on the Estimated Conditional Kendall's Tau	49
4.1 Auxiliary Results in Operator Theory	49
4.2 Modes of Convergence for Random Operators	56
4.3 Consistency Results.	57
4.3.1 Convergence of the Operators	58

4.3.2	Convergence of the Eigenvalues and Eigenfunctions.	59
4.4	Convergence Rate.	60
5	PCA on the Estimated Conditional Kendall's Tau	63
5.1	Theoretical Framework	64
5.2	Properties of $\tilde{T}_{\Gamma^{(N)}}$	68
5.3	Convergence of Discretized Operators	74
5.4	Consistency of the Eigenvalues	80
5.5	Consistency of the Eigenvectors.	82
5.6	Convergence Rate.	85
6	On Conditional U-statistics	91
6.1	Bias of a Conditional U -statistic.	96
6.1.1	Proof of Lemma 6.1.2	101
6.1.2	Proof of Lemma 6.1.4	103
6.1.3	Proof of Lemma 6.1.3	107
6.2	Variance of a Conditional U -statistic	112
6.2.1	Proof of Lemma 6.2.2	118
6.2.2	Proof of Lemma 6.2.3	121
6.2.3	Proof of Lemma 6.2.4	125
6.3	Expectation and Variance of the CKT Estimator.	128
6.4	Moments of the CKT Estimator	131
6.5	Convergence of the Integral Operators	135
6.5.1	Proof of Theorem 4.4.1	138
7	Application to Financial Data	141
7.1	Description and Visualization of the Dataset	141
7.2	ARMA-GARCH Filtering of Returns	143
7.3	Unconditional Dependence.	144
7.4	FPCA Conditional to the Amsterdam Market	146
7.5	FPCA on all Asset Triplets	150
7.5.1	Deflation of the First Principal Component	153
7.6	Smoothed FPCA.	155
7.6.1	Pre-smoothing the Estimated Curves	155
7.6.2	Smoothing Embedded in FPCA	157
7.6.3	Post-smoothing of the Principal Components	157
7.6.4	Empirical Comparison of the Three Smoothing Approaches	158
8	Conclusion	160
	Bibliography	163
A	Subset Analysis of Highly Dependent Markets	167

INDEX OF ASSUMPTIONS

The following table provides a quick reference for the structural and statistical assumptions used throughout this thesis.

Assumption	Description	Page
Global assumption	Measurability of the conditional Kendall's tau	41
Global assumption	Compactness of $\mathcal{Z}$	41
Assumption 1	Properties of the kernel K	57
Assumption 2	Continuity of $f_{\mathbf{Z}}$ and the conditional Kendall's tau	57
Assumption 3	Relation between the weights and the partition cells	68
Assumption 4	Properties of the diameter of the partition set	74
Assumption 5	On the numerical quadrature rule	76
Assumption 6	Properties of the partition cells	76
Assumption 7	Differentiability of $f_{\mathbf{X},\mathbf{Z}}$	86
Assumption 8	Properties of $f_{\mathbf{Z}}$	86
Assumption 9	Finite integral	96
Assumption 10	Finite integral	112
Assumption 11	Properties of the kernel K	135
Assumption 12	Finite integral	135

1

INTRODUCTION

In an increasingly integrated global economy, understanding the dependencies among financial assets remains a vital challenge in financial economics. Copula models provide a general framework for capturing dependencies between variables and have been widely adopted across various disciplines including empirical finance [1]. Originally formalized by Sklar [2], this approach has gained widespread traction due to its ability to decouple the modelling of univariate marginal distributions from the joint dependence structure. Copulas therefore provide a flexible framework for capturing complex, non-linear relationships in the global market. In the field of dependence modelling, it is common to quantify these associations using rank correlation coefficients, such as Kendall's tau [3]. However, the joint dependence structure among markets may vary conditionally on other variables such as the market volatility or the performance of other assets. An effective non-parametric tool to trace this conditional dependence is the conditional Kendall's tau (CKT). The CKT maps the evolution of the rank-based association between asset pairs as a continuous function of a chosen covariate.

The practical utility of conditional copulas becomes particularly important through vine copula models. These models break the high-dimensional dependence structures into a sequence of bivariate conditional copulas [4]. To maintain statistical and computational tractability, the vast majority of vine copula literature relies on the *simplifying assumption*, which asserts that the conditional copula itself remains invariant to changes in the conditioning variable. Under this assumption, the corresponding CKT curves are perfectly flat. While the simplifying assumption is a powerful tool for achieving computational efficiency, it is often restrictive and potentially polarizing in empirical applications [5]. Consequently, a lot of research has emerged to test the validity of this assumption [6, 7, 8], alongside recent frameworks designed to formally quantify the degree of non-simplifyingness within conditional structures [9]. Within the context of this thesis, analysing the structure of a vast collection of CKT curves becomes essential to determining whether a simplified structure is valid.

To systematically evaluate a large collection of curves, we require a framework capable of extracting the main structural modes present among these functional data. Functional Principal Component Analysis (FPCA) serves as a powerful tool to analyse and isolate the most important modes of variation within a large set of curves [10]. The existing mathematical framework of FPCA is built on the fundamental assumption that the functional objects are directly observed [11, 12]. However, in financial econometrics and other dynamic fields, the conditional dependence curves are intrinsically latent and are never directly observed. Instead, we estimate them and obtain estimated dependence paths. Consequently, shifting the FPCA framework to accommodate estimated conditional Kendall's tau curves introduces an additional layer of statistical variation. The total error no longer originates solely from drawing a finite sample of curves, instead, it also consists of the pointwise estimation error inherent to the estimators. This additional layer of variation distorts the standard asymptotic properties, meaning the standard consistency proofs can no longer be assumed to hold.

In this thesis, we establish a new mathematical framework for conducting FPCA on estimated functional objects, focussing specifically on the conditional Kendall's tau. We formalize an estimator for the functional covariance operator based on the pre-estimated CKT curves and derive explicit asymptotic convergence rates for its spectral properties. We extend these proofs from an idealized setting, in which the estimated curves are fully known, to the realistic scenario where the trajectories are observed only across a finite grid of N points. Through this framework, we characterize the asymptotic interplay between the grid resolution N , the smoothing bandwidth h , and the sample size n . Ultimately, this method provides a systematic foundation for evaluating the simplifying assumption within the FPCA framework.

The remainder of this thesis is organized as follows. In Chapter 2 we provide the necessary background for this thesis. Chapter 3 lays the foundation of our statistical model by formally introducing the required functional operators and outlining the Functional Principal Component Analysis (FPCA) framework. Building upon this baseline, Chapter 4 establishes the methodology in an idealized theoretical setting where the conditional Kendall's tau curves are assumed to be known continuously over their entire domain. We define the empirical covariance operators, analyse their spectral properties and compute their asymptotic convergence rates. Chapter 5 subsequently extends this framework to the practically realistic scenario where the CKT curves are observed only across a finite grid of discretization points. We again provide proofs for the asymptotic consistency and convergence behaviour of both the operators and their spectral decompositions under grid constraints. In order to compute the convergence rates in both Chapters 4 and 5, Chapter 6 develops additional theory for conditional U -statistics needed in these convergence proofs. This rather technical material is made into a separate chapter for the convenience of the reader. Finally, the thesis concludes with an empirical implementation in Chapter 7, which applies the developed framework to global financial market data by estimating the principal components of the CKT curves of all combinations of asset triplets. This allows us to find the most important structural modes with this collection of curves.

2

PRELIMINARIES

In this chapter, we provide the necessary mathematical background for this thesis. We start with discussing the concept of (conditional) copulas, which are the core of multivariate dependence modelling. We next introduce the concept of rank correlation coefficients which are measures to quantify the dependence between random variables. Specifically we introduce the (conditional) Kendall's tau measure and discuss its properties. After briefly touching upon the simplifying assumption, we discuss our estimation methods and introduce the concepts of Kernel Density Estimation and Kernel Regression and introduce our CKT estimator. Next, we consider some linear algebra preliminaries which serve as the foundation for much of the subsequent proofs. We also briefly discuss the notion of (conditional) U -statistics which will become important for proving the convergence rate of our operators. Furthermore, we consider (Functional) Principal Component Analysis and discuss its working and its properties. We will end the chapter with considering ARMA-GARCH filtering.

2.1. COPULAS

Suppose we are given two random variables, X_1 and X_2 , with marginal distribution functions F_1 and F_2 , respectively, and we wish to analyze their mutual dependence. To isolate the dependency structure from the effects of the individual marginal distributions, it is helpful to standardize the random variables using the *probability integral transform*:

Lemma 2.1.1 (Probability Integral Transform (PIT)). *If $X \sim F$ is a continuous random variable, then $F(X) \sim U(0, 1)$.*

Consequently, any continuous random variable can be transformed into a standard uniform random variable, removing the influence of its original marginal distribution. The converse property also holds true:

Lemma 2.1.2 (Quantile Transformation). *Let $U \sim U(0, 1)$ and let F be any distribution function. Then $F^{-1}(U) \sim F$, where F^{-1} is the quantile function of F defined by*

$$F^{-1}(y) := \inf \{x \in \mathbb{R} : F(x) \geq y\}.$$

Using these transformations, we can now define the joint distribution function of $(F_1(X_1), F_2(X_2))$ as the *copula* of (X_1, X_2) .

Definition 2.1.3 (Copula). *A function $C : [0, 1]^p \rightarrow [0, 1]$ for $p \geq 2$ is called a copula if it satisfies the following conditions:*

1. $C(u_1, u_2, \dots, u_p) = 0$ if $u_j = 0$ for any $j \in \{1, \dots, p\}$.
2. $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for all $u_j \in [0, 1]$ and all $j \in \{1, \dots, p\}$.
3. C is p -increasing, that is, for any hypercube $A = \prod_{i=1}^p [a_i, b_i] \subseteq [0, 1]^p$, the C -volume of A is non-negative:

$$\int_A dC(\mathbf{u}) \geq 0.$$

In probabilistic terms, a copula is a multivariate distribution function defined on the unit cube with standard uniform margins. While this formal definition may not immediately reveal the connection between copulas and dependence modeling, the key insight comes from Sklar's Theorem. This fundamental result in copula theory shows how copulas can be used to link the joint cumulative distribution function (CDF) of any multivariate random variable to its univariate marginal distributions. The proof can be found in [13]. As we continue, given a univariate distribution function F , $\text{ran}(F) := \{F(x) : x \in \mathbb{R}\}$ denotes the range of F .

Theorem 2.1.4 (Sklar's Theorem, Theorem 2.3.3 [14]).

1. For any p -dimensional distribution function H with univariate margins $F_1, \dots, F_p$ there exists a p -dimensional copula C such that

$$H(\mathbf{x}) = C(F_1(x_1), \dots, F_p(x_p)), \quad \mathbf{x} \in \mathbb{R}^p. \quad (2.1)$$

If $F_1, \dots, F_p$ are continuous, then C is unique. Otherwise C is uniquely defined on $\prod_{i=1}^p \text{ran } F_i$. The copula is given by

$$C(\mathbf{u}) = H(F_1^-(u_1), \dots, F_p^-(u_p)), \quad \mathbf{u} \in \prod_{i=1}^p \text{ran}(F_i). \quad (2.2)$$

2. Conversely, given a copula C and the univariate margins $F_1, \dots, F_p$ then H is defined as in Equation (2.1).

Theorem 2.1.4 can be expanded if the margins $F_1, \dots, F_d$ have densities.

Corollary 2.1.5 (Remark 2.3.2 [15]). *The distribution function H from Equation (2.1) is absolutely continuous if and only if C and $F_1, \dots, F_p$ are absolutely continuous. In that case we can write the density h of H as*

$$h(\mathbf{x}) = c(F_1(x_1), \dots, F_p(x_p)) \prod_{i=1}^p f_i(x_i), \quad \mathbf{x} \in \prod_{i=1}^p \text{ran}(X_i)$$

where, for any $j \in \{1, \dots, p\}$, $\text{ran } X_j := \{x \in \mathbb{R} : f_j(x) > 0\}$, f_i denotes the density of F_i and c denotes the density of C . It follows that

$$c(\mathbf{u}) = \frac{\partial^p C(\mathbf{u})}{\partial u_1 \cdots \partial u_p} = \frac{h(F_1^-(u_1), \dots, F_p^-(u_p))}{f_1(F_1^-(u_1)) \cdots f_p(F_p^-(u_p))}, \quad \mathbf{u} \in (0, 1)^p$$

An important property of copulas is the invariance principle, which plays a crucial role when defining distribution-free measures of association.

Lemma 2.1.6 (Invariance Principle, Theorem 2.4.7 [15]). *Let $\mathbf{X} \sim H$ be a random vector with continuous univariate marginal distributions $F_1, \dots, F_d$ and copula C . If T_j is a strictly increasing transformation on $\text{ran}(X_j)$ for each $j \in \{1, \dots, p\}$, then the transformed random vector*

$$(T_1(X_1), \dots, T_p(X_p))$$

also possesses the identical copula C .

We now extend this framework to the conditional setting by introducing a conditioning random variable $\mathbf{Z}$. By adapting the previous unconditional concepts, we state the formal definition of a conditional copula as follows.

Definition 2.1.7 (Conditional Copula). *Consider a conditional random variable $\mathbf{Z}$ taking values in $\mathcal{Z} \subset \mathbb{R}^d$. A measurable function $C : [0, 1]^p \times \mathcal{Z} \rightarrow [0, 1]$ for $p \geq 2$ is called a conditional copula if $\mathbb{P}_{\mathbf{Z}}$ -almost surely for every $\mathbf{z} \in \mathcal{Z}$:*

1. $C(u_1, u_2, \dots, u_p | \mathbf{Z} = \mathbf{z}) = 0$ if $u_j = 0$ for some $j = 1, \dots, p$
2. For any $j = 1, \dots, p$, $C(1, \dots, 1, u_j, 1, \dots, 1 | \mathbf{Z} = \mathbf{z}) = u_j$ for all $u_j \in [0, 1]$
3. C is p -increasing, i.e., if for any hypercube $A = \prod_{i=1}^p [a_i, b_i] \subseteq [0, 1]^p$:

$$\int_A dC(\mathbf{u} | \mathbf{Z} = \mathbf{z}) \geq 0.$$

Sklar's Theorem can also be adjusted accordingly. The proof can be found in [16].

Theorem 2.1.8 (Conditional Sklar's Theorem, Theorem 3 [16]).

1. Consider a conditional random variable $\mathbf{Z}$ taking values in $\mathcal{Z}$. For any p -dimensional conditional distribution function $H_{\mathbf{Z}}$ with univariate margins $F_{1|\mathbf{z}}, \dots, F_{p|\mathbf{z}}$ there exists a p -dimensional copula $C_{\mathbf{Z}}$ such that

$$H_{\mathbf{Z}}(\mathbf{x}) = C_{\mathbf{Z}}(F_{1|\mathbf{z}}(x_1), \dots, F_{p|\mathbf{z}}(x_p)), \quad \mathbf{x} \in \mathbb{R}^p \text{ and } \mathbf{z} \in \mathcal{Z}. \quad (2.3)$$

If $F_{1|\mathbf{z}}, \dots, F_{p|\mathbf{z}}$ are continuous, then $C_{\mathbf{Z}}$ is unique. Otherwise $C_{\mathbf{Z}}$ is uniquely defined on $\prod_{i=1}^p \text{ran}(F_{i|\mathbf{z}})$. The copula is given by

$$C_{\mathbf{Z}}(\mathbf{u}) = H_{\mathbf{Z}}(F_{1|\mathbf{z}}^-(u_1), \dots, F_{p|\mathbf{z}}^-(u_p)), \quad \mathbf{u} \in \prod_{i=1}^p \text{ran}(F_{i|\mathbf{z}}) \text{ and } \mathbf{z} \in \mathcal{Z}. \quad (2.4)$$

2. Conversely, given a copula $C_{\mathbf{Z}}$ and the univariate margins $F_{1|\mathbf{z}}, \dots, F_{p|\mathbf{z}}$ then $H_{\mathbf{Z}}$ is defined as in Equation (2.3).

Again, the concept of conditional copula densities follows naturally.

2

Corollary 2.1.9. *The distribution function $H_{\mathbf{Z}}$ from Equation (2.3) is absolutely continuous if and only if $C_{\mathbf{Z}}$ and $F_{1|\mathbf{z}}, \dots, F_{p|\mathbf{z}}$ are absolutely continuous. In that case we can write $\mathbb{P}_{\mathbf{Z}}$ -almost surely for every $\mathbf{z} \in \mathcal{Z}$ the density $h_{\mathbf{z}}$ of $H_{\mathbf{Z}}$ as*

$$h_{\mathbf{z}}(\mathbf{x}) = c_{\mathbf{z}}(F_{1|\mathbf{z}}(x_1), \dots, F_{p|\mathbf{z}}(x_p)) \prod_{i=1}^p f_{i|\mathbf{z}}(x_i), \quad \mathbf{x} \in \prod_{i=1}^p \text{ran}(X_i)$$

where, for any $j \in \{1, \dots, p\}$, $\text{ran}(X_j) = \{x \in \mathbb{R} : f_{j|\mathbf{z}}(x) > 0\}$, $f_{i|\mathbf{z}}$ denotes the density of $F_{i|\mathbf{z}}$ and $c_{\mathbf{z}}$ denotes the density of $C_{\mathbf{Z}}$. It follows that

$$c(\mathbf{u}|\mathbf{Z} = \mathbf{z}) = \frac{\partial^p C(\mathbf{u}|\mathbf{Z} = \mathbf{z})}{\partial u_1 \cdots \partial u_d} = \frac{h_{\mathbf{z}}(F_{1|\mathbf{z}}^{-1}(u_1), \dots, F_{p|\mathbf{z}}^{-1}(u_p))}{f_{1|\mathbf{z}}(F_{1|\mathbf{z}}^{-1}(u_1)) \cdots f_{p|\mathbf{z}}(F_{p|\mathbf{z}}^{-1}(u_p))}, \quad \mathbf{u} \in (0, 1)^p$$

2.2. MEASURES OF ASSOCIATION

As established in previous sections, our primary objective is to model the dependence structure between random variables. A classical and widely used approach to quantify dependence is the *Pearson correlation coefficient*.

Definition 2.2.1 (Pearson Correlation Coefficient). *Let X_1 and X_2 two real-valued random variables with finite second moments. The Pearson correlation coefficient is defined by*

$$\rho := \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)}\sqrt{\text{Var}(X_2)}}.$$

Though widely used, the Pearson correlation coefficient possesses notable limitations. Most importantly, it only measures the linear dependency between variables, ignoring any other type of dependency. In addition, it is also not invariant under monotone increasing transformations. To overcome these deficiencies, we consider *rank correlation coefficients*, which measure the similarity between the orderings of data when ranked by their respective marginal distributions. The building block for these invariant measures is the geometric concept of concordance and discordance between pairs of points.

Definition 2.2.2. *A pair of observations (X_1, X_2) and (X'_1, X'_2) is called an concordance pair if*

$$\text{sign}(X_1 - X'_1) = \text{sign}(X_2 - X'_2)$$

else it is called an discordance pair. Here, $\text{sign}(\cdot)$ denotes the standard signum function:

$$\text{sign}(t) = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ +1 & \text{if } t > 0. \end{cases}$$

By evaluating the difference between the probability of concordance and the probability of discordance of two independent pairs drawn from the same joint distribution, we arrive at Kendall's tau.

Definition 2.2.3 (Kendall's tau). *Let (X_1, X_2) be a bivariate random vector and let (X'_1, X'_2) be an independent copy of (X_1, X_2) . Then Kendall's tau is defined as*

$$\tau_{X_1, X_2} = \tau(X_1, X_2) = \mathbb{E} [\text{sign}(X_1 - X'_1) \text{sign}(X_2 - X'_2)] \quad (2.5)$$

$$= \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0). \quad (2.6)$$

By construction, $\tau_{X_1, X_2} \in [-1, 1]$, where the boundaries represent perfect negative and positive rank dependence, respectively. If X_1 and X_2 are independent, $\tau_{X_1, X_2} = 0$, however, the converse does not hold in general. When the marginal distributions of X_1 and X_2 are continuous, ties occur with probability zero, meaning $\mathbb{P}((X_1 - X'_1)(X_2 - X'_2) = 0) = 0$. Under this continuous regime, Sklar's Theorem allows us to decouple the marginal behaviour from the underlying copula, revealing that Kendall's tau is completely distribution-free.

Corollary 2.2.4. *Let (X_1, X_2) be a bivariate random vector with continuous marginal distributions and underlying copula C . Then*

$$\tau_{X_1, X_2} = 4 \int_{[0,1]^2} C(\mathbf{u}) dC(\mathbf{u}) - 1. \quad (2.7)$$

Proof.

$$\begin{aligned} \tau_{X_1, X_2} &= \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0) \\ &= 2\mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0) - 1 \\ &= 4\mathbb{P}(X_1 < X'_1, X_2 < X'_2) - 1 \\ &= 1 - 4\mathbb{P}(X_1 < X'_1, X_2 > X'_2) \\ &= 4 \int_{[0,1]^2} C(\mathbf{u}) dC(\mathbf{u}) - 1 \end{aligned}$$

□

Corollary 2.2.4 confirms that Kendall's tau is invariant under strictly increasing transformations of the margins. Because it can be fully represented by the copula, it serves as a pure reflection of the underlying dependence structure. To evaluate how this association shifts across different contexts, we generalize this framework to a conditional setting by introducing a covariate vector $\mathbf{Z}$ taking values in a space $\mathcal{Z}$.

Definition 2.2.5 (Conditional Kendall's Tau). *Let $(X_1, X_2, \mathbf{Z})$ be a random vector where $\mathbf{Z} \in \mathcal{Z}$. Let $(X'_1, X'_2, \mathbf{Z}')$ be an independent copy of $(X_1, X_2, \mathbf{Z})$. For any $\mathbf{z} \in \mathcal{Z}$, the conditional Kendall's tau is defined as*

$$\begin{aligned} \tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}} &= \mathbb{E} (\text{sign}((X_1 - X'_1)(X_2 - X'_2)) | \mathbf{Z} = \mathbf{Z}' = \mathbf{z}) \\ &= \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) > 0 | \mathbf{Z} = \mathbf{Z}' = \mathbf{z}) - \mathbb{P}((X_1 - X'_1)(X_2 - X'_2) < 0 | \mathbf{Z} = \mathbf{Z}' = \mathbf{z}). \end{aligned}$$

If the conditional marginal distributions of X_1 and X_2 given $\mathbf{Z} = \mathbf{z}$ are continuous, it can be expressed in terms of the unique conditional copula $C_{\mathbf{z}}$ as:

$$\tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}} = 4 \int_{[0,1]^2} C_{\mathbf{z}}(\mathbf{u}) dC_{\mathbf{z}}(\mathbf{u}) - 1. \quad (2.8)$$

2

2.3. SIMPLIFYING ASSUMPTION

In general, allowing the conditional copula $C_{\mathbf{z}}$ to vary with the conditioning covariate $\mathbf{Z}$ introduces severe practical challenges. As the dimension of $\mathbf{Z}$ increases, estimation of $C_{\mathbf{z}}$ suffers from the curse of dimensionality, while parametric modelling requires specifying functional forms for how the copula evolves across $\mathcal{Z}$. To bypass this problem, a structural restriction known as the *simplifying assumption* is frequently invoked.

Definition 2.3.1 (Simplifying Assumption). *The conditional copula $C_{\mathbf{z}}$ satisfies the simplifying assumption if it remains invariant with respect to the conditioning value, meaning there exists a single, copula C such that*

$$C_{\mathbf{z}}(\mathbf{u}) = C(\mathbf{u}) \quad \text{for all } \mathbf{u} \in [0, 1]^2 \text{ and for all } \mathbf{z} \in \mathcal{Z}.$$

Definition 2.3.1 means that the dependence structure between X_1 and X_2 is constant across all values of $\mathcal{Z}$. Only the marginal conditional distributions may change with $\mathbf{Z}$. It is good to emphasize that the simplifying assumption does not imply that the random variables X_1 and X_2 are conditionally independent of $\mathbf{Z}$. Instead, the assumption asserts that the copula mechanism linking the margins together remains uniform across $\mathcal{Z}$. Consequently, under the simplifying assumption, the conditional Kendall's tau curves degenerate into a constant function over the entire domain:

$$\tau(\mathbf{z}) = \tau \quad \text{for all } \mathbf{z} \in \mathcal{Z}.$$

The following lemma, which is an adjustment of Proposition 4 of [7], also gives more useful intuition about the simplifying assumption.

Lemma 2.3.2. *Let F_1 and F_2 the marginal distribution functions of X_1 and X_2 . Define the vector $(U_1, U_2) := (F_1(X_1 | \mathbf{Z}), F_2(X_2 | \mathbf{Z}))$. Then the simplifying assumption holds if and only if $(U_1, U_2) \perp \mathbf{Z}$.*

Proof. Recall by the Probability Integral Transform (Lemma 2.1.1) that $U_1 | \mathbf{Z} = \mathbf{z} \sim U(0, 1)$ and $U_2 | \mathbf{Z} = \mathbf{z} \sim U(0, 1)$ for every $\mathbf{z}$. By Sklar's Theorem,

$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2 | \mathbf{Z} = \mathbf{z}) = C_{\mathbf{z}}(u_1, u_2).$$

Now if $(U_1, U_2) \perp \mathbf{Z}$ then

$$C_{\mathbf{z}}(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2 | \mathbf{Z} = \mathbf{z}) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2),$$

and the right-hand side does not depend on $\mathbf{z}$. Hence Definition 2.3.1 holds and the simplifying assumption is satisfied.

Conversely suppose that the simplifying assumption holds. Then there exists a single copula C_s such that $C_z(u_1, u_2) = C(u_1, u_2)$. Then

$$\mathbb{P}(U_1 \leq u_1, U_2 \leq u_2 | \mathbf{Z} = \mathbf{z}) = C_z(u_1, u_2) = C(u_1, u_2),$$

which no longer depends on $\mathbf{z}$. It thus follows that (U_1, U_2) must be independent on $\mathbf{Z}$. This proves the claim $\square$

Note that from Lemma 2.3.2 it is not enough for $U_1 = F_1(X_1 | \mathbf{Z})$ and $U_2 = F_2(X_2 | \mathbf{Z})$ to be individually independent on $\mathbf{Z}$. The simplifying assumption really concerns the joint distribution of (U_1, U_2) .

Remark 1 (Remark 1 in [7]). *Note that under the simplifying assumption, the conditional copula C_z is not necessarily equal to C , the usual copula of (X_1, X_2) . Which in turn also implies that it not true that under the simplifying assumption, the conditional Kendall's tau $\tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}$ equals the unconditional Kendall's tau τ_{X_1, X_2} .*

This can be seen in the following example. Let (X_1, X_2, Z) be a trivariate random vector such that, given Z ,

$$X_1 \sim \mathcal{N}(Z, 1) \quad \text{and} \quad X_2 \sim \mathcal{N}(Z, 1)$$

with X_1 and X_2 independent given Z . For convenience suppose $Z \sim \mathcal{N}(0, 1)$. It can be checked that for any couple $(u_1, u_2) \in [0, 1]^2$ and any real number z , the conditional copula equals (p. 9 of [15])

$$C_z(u_1, u_2 | z) = u_1 u_2.$$

This conditional copula does not depend on z hence Definition 2.3.1 is satisfied. But the copula of (X_1, X_2) is not the independence copula, because X_1 and X_2 are not independent.

2.4. ESTIMATION METHODS

In this project, we investigate the asymptotic behaviour of the conditional Kendall's tau, which provides a measure of how the dependence between X_1 and X_2 evolves as $\mathbf{Z}$ varies. Before defining estimators for $\tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}$, we introduce non-parametric estimation methods on which these estimators are based. Non-parametric methods are widely preferred due to their broad applicability and reliance on relatively weak structural assumptions. However, this flexibility typically gives a loss of statistical efficiency. Conversely, parametric estimation techniques yield more efficient estimators but they demand rigid assumptions and precise a priori knowledge of the underlying model.

2.4.1. KERNEL DENSITY ESTIMATION

Kernel Density Estimation (KDE) is a nonparametric technique used to estimate the probability density function of a continuous random variable from observed data. A classic, intuitive way to estimate the density function is using a histogram. Given an independent and identically distributed sample $\{X_1, \dots, X_n\}$, a histogram divides the support of the data into k disjoint bins of width h , centred at locations t_j for $j = 1, \dots, k$.

While histograms provide a baseline visual representation, they suffer from discontinuities and depend heavily on the choice of bin origins. To obtain a smooth and localized estimate that varies continuously with x , we propose the following density estimator

$$\begin{aligned}\hat{f}(x; h) &= \frac{1}{2nh} \sum_{j=1}^n \mathbb{1}_{|x - X_j| \leq h} \\ &= \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right),\end{aligned}\tag{2.9}$$

where $K(u) = \frac{1}{2} \mathbb{1}_{|u| \leq 1}$, denotes the *uniform kernel*. This particular choice of K is simple but not necessarily optimal for all applications. In kernel density estimation, K can be any nonnegative, even, and normalized function, which is then referred to as the *kernel*. Popular choices include the Gaussian kernel, cosine kernel or the Epanechnikov kernel.

Conceptually, rather than assigning discrete weights within fixed bins, KDE centers a continuous probability distribution, the kernel, at each observed data point X_i . The positive parameter h governs the scale or window width of these component distributions and is known as the *bandwidth*.

2.4.2. KERNEL REGRESSION AND CKT ESTIMATION

In this subsection, we introduce the notion of *kernel regression* for the construction of non-parametric conditional estimators. We begin by recalling the general definition of the Nadaraya-Watson estimator, which models the conditional expectation for a random variable Y given a d -dimensional covariate vector $\mathbf{Z} \in \mathcal{Z}$.

For a continuous bivariate random vector $(Y, \mathbf{Z})$ with joint density $f(y, \mathbf{z})$, the conditional expectation is given by

$$\mathbb{E}[Y | \mathbf{Z} = \mathbf{z}] = \int y f(y | \mathbf{z}) dy = \int y \frac{f(y, \mathbf{z})}{f(\mathbf{z})} dy.$$

We can estimate $f(y, \mathbf{z})$ and $f(\mathbf{z})$ using the kernel density estimator based on an i.i.d. sample $\{(Y_j, \mathbf{Z}_j)\}_{j=1}^n$ as derived in Equation (2.9)

$$\hat{f}(y, \mathbf{z}) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h} K_2\left(\frac{y - Y_j}{h}\right) \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_j}{h}\right)\tag{2.10}$$

$$\hat{f}(\mathbf{z}) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_j}{h}\right).\tag{2.11}$$

Hence

$$\hat{\mathbb{E}}[Y | \mathbf{Z} = \mathbf{z}] = \frac{\sum_{j=1}^n \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_j}{h}\right) \int y \frac{1}{h} K_2\left(\frac{y - Y_j}{h}\right) dy}{\sum_{j=1}^n \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_j}{h}\right)}.$$

Assuming that the response kernel K_2 is a symmetric probability density function integrating to 1, yields

$$\int y \frac{1}{h} K_2\left(\frac{y - Y_i}{h}\right) dy = \int (hu + Y_i) K_2(u) du = Y_i.$$

Substituting this simplification back into the expression yields the standard Nadaraya-Watson kernel estimator

$$\hat{\mathbb{E}}[Y | \mathbf{Z} = \mathbf{z}] = \sum_{j=1}^n w_{j,n}(\mathbf{z}) Y_j,$$

where the Nadaraya-Watson weights are defined as

$$w_{j,n}(\mathbf{z}) = \frac{K_h(\mathbf{Z}_j - \mathbf{z})}{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})} \quad (2.12)$$

with $K_h(\cdot) = h^{-d} K_1(\cdot/h)$.

This framework directly motivates our construction of estimators for the conditional Kendall's tau. Let $(X_1, X_2, \mathbf{Z})$ be a pair of random variables and let $(X'_1, X'_2, \mathbf{Z}')$ be an independent copy of $(X_1, X_2, \mathbf{Z})$. Recall that the conditional Kendall's tau can be defined via three equivalent expressions

$$\begin{aligned} \tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}} &= \mathbb{E} \left[\text{sign}((X_1 - X'_1)(X_2 - X'_2)) \mid \mathbf{Z} = \mathbf{Z}' = \mathbf{z} \right] \\ \tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}} &= 4\mathbb{P}(X_1 < X'_1, X_2 < X'_2 \mid \mathbf{Z} = \mathbf{Z}' = \mathbf{z}) - 1 \\ \tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}} &= 1 - 4\mathbb{P}(X_1 < X'_1, X_2 > X'_2 \mid \mathbf{Z} = \mathbf{Z}' = \mathbf{z}) \end{aligned}$$

Given an i.i.d. sample $\{(X_1^{(i)}, X_2^{(i)}, \mathbf{Z}_i)\}_{i=1}^n$, distributed from $(X_1, X_2, \mathbf{Z})$, we can map the conditional probabilities and expectations in these expression to pairwise sample indicators. We define the kernel functions $g^{(k)}$ for $k \in \{1, 2, 3\}$ as

$$\begin{aligned} g^{(1)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}) &:= \mathbb{1}_{(X_1^{(i)} - X_1^{(j)})(X_2^{(i)} - X_2^{(j)}) > 0} - \mathbb{1}_{(X_1^{(i)} - X_1^{(j)})(X_2^{(i)} - X_2^{(j)}) < 0} \\ g^{(2)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}) &:= 4\mathbb{1}_{\{X_1^{(i)} < X_1^{(j)}, X_2^{(i)} < X_2^{(j)}\}} - 1 \\ g^{(3)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}) &:= 1 - 4\mathbb{1}_{\{X_1^{(i)} < X_1^{(j)}, X_2^{(i)} < X_2^{(j)}\}}, \end{aligned} \quad (2.13)$$

where $\mathbf{X}_i^{(1,2)} = (X_1^{(i)}, X_2^{(i)})$. By matching these kernels with pairs of Nadaraya-Watson weights, we propose the following three estimators for $\tau_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}$

$$\begin{aligned} \hat{\tau}_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}^{(1)} &= \sum_{i=1}^n \sum_{j=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) g^{(1)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}) \\ \hat{\tau}_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}^{(2)} &= \sum_{i=1}^n \sum_{j=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) g^{(2)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}) \\ \hat{\tau}_{X_1, X_2 | \mathbf{Z} = \mathbf{z}}^{(3)} &= \sum_{i=1}^n \sum_{j=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) g^{(3)}(\mathbf{X}_i^{(1,2)}, \mathbf{X}_j^{(1,2)}). \end{aligned} \quad (2.14)$$

Each $\hat{\tau}_{j_1, j_2 | \mathbf{Z} = \mathbf{z}}$ can be seen as a smoothed estimator of $\mathbb{E}(g_{j_1 j_2}^{(k)}(\mathbf{X}_1, \mathbf{X}_2) | \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z})$.

Although the estimators of Equations (2.14) are structurally similar, they have distinct support due to the inclusion of diagonal terms in their double summations. To see this, let $s_n(\mathbf{z}) := \sum_{j=1}^n w_{j,n}^2(\mathbf{z})$ denote the sum of the squared Nadaraya-Watson weights. It follows that $\hat{\tau}^{(1)}(\mathbf{z})$ takes values in the interval $[s_n(\mathbf{z}) - 1, 1 - s_n(\mathbf{z})]$, whereas $\hat{\tau}^{(2)}(\mathbf{z})$ and $\hat{\tau}^{(3)}(\mathbf{z})$ are bounded within $[-1, 1 - 2s_n(\mathbf{z})]$ and $[-1 + 2s_n(\mathbf{z}), 1]$, respectively. Despite these differing supports, the estimators share a direct relationship, as established by Derumigny and Fermanian [17].

Corollary 2.4.1. *For any sample size $n \geq 1$, the following algebraic identities hold almost surely:*

$$\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(1)} = \hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(2)} + s_n = \hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(3)} - s_n. \quad (2.15)$$

This suggests that we can define a new estimator taking values on $[-1, 1]$.

$$\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(4)} := \frac{\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(1)}}{1 - s_n} = \frac{\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(2)} + s_n}{1 - s_n} = \frac{\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(3)} - s_n}{1 - s_n}. \quad (2.16)$$

The asymptotic behavior of these estimators is investigated in [17]. While the formal regularity conditions will be detailed later in the thesis, we state their consistency here for completeness.

Lemma 2.4.2 (Consistency of the CKT estimator, Proposition 7 of [17]). *Under Assumptions 1 and 2, if $nh^d \rightarrow +\infty$ and $h \rightarrow 0$ then for $k \in \{1, 2, 3, 4\}$,*

$$\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}^{(k)} \xrightarrow[n \rightarrow \infty]{\text{prob}} \tau_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}.$$

We conclude this subsection with a classical real-analysis result that will become useful when proving the asymptotic properties of kernel-based conditional estimators.

Lemma 2.4.3 (Bochner's Lemma, Lemma 17 of [18]). *Let f a bounded function on $\mathbb{R}^d$ which is continuous in a neighborhood around $\mathbf{z} \in \mathbb{R}^d$. Let K a function in $\mathbb{R}^d$ such that*

$$\int |K(\mathbf{u})| d\mathbf{u} < \infty$$

then

$$\lim_{h \rightarrow 0} \int K(\mathbf{u}) f(\mathbf{z} + h\mathbf{u}) d\mathbf{u} = f(\mathbf{z}) \int K(\mathbf{u}) d\mathbf{u}.$$

Proof. Take $h > 0$ and write

$$\begin{aligned}
 \left| \int K(\mathbf{u})f(\mathbf{z} + h\mathbf{u})d\mathbf{u} - f(\mathbf{z}) \int K(\mathbf{u})d\mathbf{u} \right| &\leq \int |f(\mathbf{z} + h\mathbf{u}) - f(\mathbf{z})| |K(\mathbf{u})| d\mathbf{u} \\
 &\leq \sup_{|\mathbf{u}| \leq h^{-1/2}} |f(\mathbf{z} + h\mathbf{u}) - f(\mathbf{z})| \int |K(\mathbf{u})| d\mathbf{u} \\
 &\quad + \int_{|\mathbf{u}| \geq h^{-1/2}} |f(\mathbf{z} + h\mathbf{u})| |K(\mathbf{u})| d\mathbf{u} \\
 &\leq \sup_{|\mathbf{u}| \leq h^{-1/2}} |f(\mathbf{z} + h\mathbf{u}) - f(\mathbf{z})| \int |K(\mathbf{u})| d\mathbf{u} \\
 &\quad + 2 \sup_{\mathbf{u}} |f(\mathbf{u})| \int_{|\mathbf{u}| \geq h^{-1/2}} |K(\mathbf{u})| d\mathbf{u} \\
 &\leq \sup_{|\mathbf{v}| \leq h^{1/2}} |f(\mathbf{z} + \mathbf{v}) - f(\mathbf{z})| \int |K(\mathbf{u})| d\mathbf{u} \\
 &\quad + 2 \sup_{\mathbf{u}} |f(\mathbf{u})| \int_{|\mathbf{u}| \geq h^{-1/2}} |K(\mathbf{u})| d\mathbf{u}.
 \end{aligned}$$

The result follows when $h \rightarrow 0$. $\square$

2.4.3. BIAS-VARIANCE TRADE-OFF

We conclude this section with the concept known as *the bias-variance trade-off*. This trade-off plays a central role when selecting the bandwidth h in the Nadaraya–Watson estimator. Let $\hat{\theta}$ be an estimator of the parameter θ . The bias of the estimator is defined as

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta. \quad (2.17)$$

A model with high bias fails to capture relevant patterns in the data, resulting in underfitting. Underfitting typically occurs when the model is too simple, for instance, when the bandwidth h is chosen too large. Conversely, reducing the bias by increasing the model's complexity tends to increase the variance of the estimator, defined as

$$\text{Var}(\hat{\theta}) = \mathbb{E} \left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2 \right]. \quad (2.18)$$

A small bandwidth h concentrates the kernel weights on only a few local observations, leading to high variance and overfitting. In this way the estimator mistakenly captures random sampling noise as a real signal. To balance these competing sources of error, people often evaluate performance using the *mean squared error* (MSE)

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left[(\theta - \hat{\theta})^2 \right] = \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta}). \quad (2.19)$$

This decomposition illustrates the fundamental bias-variance trade-off: reducing bias often increases variance, and vice versa. In most statistical problems, it is key to find a balance between the two, yielding the smallest possible MSE. In recent years, studies as [19] have proposed strategies to obtain lower bounds for the bias-variance trade-off. This tells us to what extent the trade-off is an artifact of the estimator, or just simply unavoidable.

2.5. STOCHASTIC o AND $\mathcal{O}$ NOTATION

In non-probabilistic mathematics, the big-O notation is a widely used concept to describe the approximate size of a function. We can extend this concept to the stochastic setting by using the following definitions.

Definition 2.5.1 (Section 2.2 of [20]).

- i) For a sequence of random variables X_n and a corresponding set of constants $a_n \neq 0$, the notation $X_n = o_P(a_n)$ means that for $n \rightarrow \infty$

$$X_n / a_n \xrightarrow[n \rightarrow \infty]{\text{prob}} 0.$$

- ii) For a sequence of random variables X_n and a corresponding set of constants $a_n \neq 0$, the notation $X_n = \mathcal{O}_P(a_n)$ means that the set X_n / a_n is bounded in probability, i.e. for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$ such that

$$\mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon, \quad \forall n > N.$$

The following lemma will become useful when proving the convergence rate of functions.

Lemma 2.5.2. Let $\{X_n\}_{n \geq 1}$ a sequence of random variables and suppose that

$$\mathbb{E}[|X_n| \mid A_n] = \mathcal{O}(a_n) \quad \text{and} \quad \mathbb{P}(A_n) \xrightarrow[n \rightarrow \infty]{} 1,$$

then $X_n = \mathcal{O}_P(a_n)$.

Proof. To prove $X_n = \mathcal{O}_P(a_n)$, we should show that $\frac{X_n}{a_n}$ is bounded in probability, i.e. for every $\epsilon > 0$, there exists a $M > 0$ such that $\sup_n \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon$.

Now fix $M > 0$. Then

$$\begin{aligned} \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) &= \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M, A_n\right) + \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M, A_n^c\right) \\ &\leq \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M \mid A_n\right) \mathbb{P}(A_n) + \mathbb{P}(A_n^c). \end{aligned}$$

Then by Markov's inequality we find that

$$\mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M \mid A_n\right) \leq \frac{\mathbb{E}[|X_n| \mid A_n]}{M|a_n|} + \mathbb{P}(A_n^c).$$

Since $\mathbb{E}[|X_n| \mid A_n] = \mathcal{O}(a_n)$,

$$\mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) \leq \frac{C}{M} + \mathbb{P}(A_n^c).$$

By assumption,

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) \leq \frac{C}{M}.$$

If we now choose M sufficiently large such that $C/M < \epsilon$. Then

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon.$$

Hence $X_n = \mathcal{O}_P(a_n)$. □

2.6. LINEAR ALGEBRA

This chapter aims at providing the essential linear algebra background which is needed for this report. We will mostly be working with *Hilbert spaces*. A Hilbert space is real (or complex) inner product space that is also a complete metric space with respect to the distance function induced by the inner product. We will specifically focus on the Hilbert space of real-valued, square-integrable functions defined on a measurable domain $D \subseteq \mathbb{R}^n$ (equipped with the Lebesgue measure). Strictly speaking, $L^2(D)$ consists of equivalence classes of functions that coincide almost everywhere, defined by

$$L^2(D) := \left\{ f : D \rightarrow \mathbb{R} \mid f \text{ is measurable and } \int_D |f(t)|^2 dt < +\infty \right\},$$

and equipped with the standard inner product

$$\langle f, g \rangle_{L^2(D)} := \int_D f(t)g(t)dt \quad \text{for } f, g \in L^2(D).$$

It should be noted that some papers allow operators that are only defined on a linear subspace of H . In this thesis we only consider operator whose domain is the entire space.

For a linear operator on some Hilbert space H , $A : H \rightarrow H$ with norm $\|\cdot\|_H$, we can define the concept of bounded operators.

Definition 2.6.1 (Bounded Operator). *Let H be a Hilbert space with norm $\|\cdot\|_H$. A linear operator $A : H \rightarrow H$ is said to be bounded if there exists a constant $M \geq 0$ such that*

$$\|Ax\|_H \leq M\|x\|_H \quad \text{for all } x \in H.$$

A consequence of this definition (see Chapter 3, §2 of [21]) is that a linear operator is bounded if and only if it is continuous. This allows us to define the *operator norm* of A as the smallest such bound M , which can be computed as the supremum over the unit ball:

$$\|A\|_{op} := \sup \{ \|Ax\|_H : x \in H, \|x\|_H \leq 1 \}. \quad (2.20)$$

It follows immediately that A is bounded if and only if $\|A\|_{op} < \infty$. Another useful characteristic of operators is compactness.

Definition 2.6.2 (Compact Operator, Chapter 3, §4 of [21]). *Let H a Hilbert space and $A : H \rightarrow H$ a bounded linear operator. Then A is compact if any bounded sequence $\{u_n\} \subset H$ contains a subsequence $\{u_{n_k}\}$ such that $\{Au_{n_k}\}$ converges.*

The following two lemmas can become useful.

Lemma 2.6.3 (Lemma 1.3 of [22]). *Let H be a Hilbert space and $A : H \rightarrow H$ a linear operator. If A is compact, then A is continuous.*

Lemma 2.6.4 (Lemma 1.1 of [22]). *Let H be a Hilbert space and $A : H \rightarrow H$ a linear operator. Then A is continuous if and only if A is bounded.*

Another useful group of operators are the normal and self-adjoint operators. Before we can define them, we need the notion of *closed* operators.

Definition 2.6.5 (Closed Operators, Chapter 3 §5 of [21]). *Let $A : H \rightarrow H$ a linear operator on a Hilbert space H . The operator A is called *closed* when for any sequence $\{x_n\} \in H$ such that $x_n \rightarrow x$ and $Ax_n \rightarrow y$, we have $x \in H$ and $y = Ax$.*

It can be shown that since we are only dealing with operators defined on the entire space H , any bounded operator is necessarily closed (see [23] p. 25).

Definition 2.6.6 (Normal Operators, Chapter 5, §3 of [21]). *Let $A : H \rightarrow H$ a linear operator on a Hilbert space H with inner product $\langle \cdot, \cdot \rangle_H : H \times H \rightarrow \mathbb{R}$. The operator A^* is called the *adjoint operator* of A if*

$$\langle Ax, y \rangle = \langle x, A^* y \rangle, \quad \forall x, y \in H.$$

We call A normal if A is closed and A and A^ commute:*

$$A^* A = A A^*.$$

Definition 2.6.7 (Self-adjoint Operators, Chapter 5, §3 [21]). *Let $A : H \rightarrow H$ a linear operator on a Hilbert space with adjoint operator A^* . We call A *selfadjoint* (or *symmetric*) if $A^* = A$.*

The following lemma describes some properties for the eigenvalues and eigenfunctions of self-adjoint, compact operators.

Lemma 2.6.8 (Lemma's 1.19, 1.20, 1.21 of [22]). *Let H be a Hilbert space and $A : H \rightarrow H$ a compact, self-adjoint linear operator. Then*

- a) the eigenvectors $\xi_1, \xi_2 \in H$ corresponding to distinct eigenvalues $\lambda_1, \lambda_2 \neq 0$ of A are orthogonal.*
- b) the set of eigenvalues of A is either finite or countably infinite. In addition, the only possible limit point of a sequence of distinct eigenvalues is 0.*
- c) the eigenvalues are bounded, and for any eigenvalue λ of A , $|\lambda| \leq \|A\|_{op}$.*

A very useful theorem tells us that the eigenfunctions of any compact, self-adjoint operator forms in fact an orthonormal basis. This theorem is known as the Hilbert-Schmidt Theorem.

Theorem 2.6.9 (Hilbert-Schmidt Theorem, Theorem 8.94 of [24]). *Let $(H, \langle \cdot, \cdot \rangle_H)$ a Hilbert space and let $A : H \rightarrow H$ a compact, self-adjoint linear operator. Then there exists an at most countable subset $\{\xi_i\}_{i \geq 1}$ of H and a corresponding set $\{\lambda_i\}_{i \geq 1}$ of reals such that*

$$Ax = \sum_{i=1}^{\infty} \langle \xi_i, x \rangle_H \lambda_i \xi_i, \quad \forall x \in H.$$

Moreover, for every $i \geq 1$, (ξ_i, λ_i) is an eigenvector-eigenvalue pair of A . The set of eigenvectors $\{\xi_i\}_{i \geq 1}$ is an orthonormal basis for the range of A and all eigenvectors of A are proportional to some ξ_i .

The Hilbert-Schmidt Theorem can be seen as an extension of the fact that symmetric matrices can be diagonalized in an orthonormal basis. Indeed any symmetric matrix $A \in \mathbb{R}^{n \times n}$ can be decomposed as

$$A = Q\Lambda Q^\top$$

where Q is an orthogonal matrix, $QQ^\top = I$ and Λ is a diagonal matrix of the real eigenvalues of A .

The last characteristic we will discuss is positive-definiteness.

Definition 2.6.10 (Positive-definite Operators, p. 250 of [25]). *Let $(H, \langle \cdot, \cdot \rangle)$ a Hilbert space and $A : H \rightarrow H$ a self-adjoint linear operator. A is called a positive-semidefinite operator if*

$$\langle Ax, x \rangle \geq 0, \quad \forall x \in H.$$

The operator is called positive-definite if $\langle Ax, x \rangle > 0$, for every $x \neq 0$ in H

From the definition we immediately obtain this lemma.

Lemma 2.6.11 (Theorem 10.22 in [25]). *Let $A : H \rightarrow H$ a positive-semidefinite operator. Then for any eigenvalue λ of A , we have $\lambda \geq 0$. If A is positive-definite then $\lambda > 0$.*

2.6.1. HILBERT-SCHMIDT INTEGRAL OPERATORS

In this paper we will be working extensively with operators known as *Hilbert-Schmidt Integral Operators*.

Definition 2.6.12 (Hilbert-Schmidt integral operator, p. 264 of [21]). *Let $D \subset \mathbb{R}^n$ a (Lebesgue-)measurable domain. We call a function $k : D \times D \rightarrow \mathbb{R}$ a Hilbert-Schmidt kernel if*

$$\iint_{D \times D} |k(x, y)|^2 dx dy < \infty,$$

that is, $k \in L^2(D \times D)$. The integral operator T_k on $L^2(D)$ is defined for any $f \in L^2(D)$ as

$$(T_k f)(x) = \int_D k(x, y) f(y) dy \quad \text{for a.e. } x \in D.$$

We call this operator a Hilbert-Schmidt integral operator. The Hilbert-Schmidt norm of T_k is defined by

$$\|T_k\|_{HS} := \|k\|_{L^2(D \times D)} = \left(\iint_{D \times D} |k(x, y)|^2 dx dy \right)^{\frac{1}{2}}$$

The following lemma will prove to be very useful.

Lemma 2.6.13. *Let K a Hilbert-Schmidt integral operator on $D \subset \mathbb{R}^n$ with kernel $k \in L^2(D \times D)$, then*

$$\|K\|_{op} \leq \|K\|_{HS}.$$

Proof. Let $f \in L^2(D)$. By applying the Cauchy-Schwarz inequality to the integral definition of the operator, we observe that for almost every $x \in D$:

$$\begin{aligned} |Kf(x)| &= \left| \int_D k(x, y) f(y) dy \right| \\ &\leq \int_D |k(x, y)| |f(y)| dy \\ &\leq \left(\int_D |k(x, y)|^2 dy \right)^{1/2} \|f\|_{L^2(D)}. \end{aligned} \quad (2.21)$$

Squaring both sides of Equation (2.21) and integrating over $x \in D$, Fubini's theorem allows us to write

$$\|Kf\|_{L^2(D)}^2 = \int_D |Kf(x)|^2 dx \leq \left(\iint_{D \times D} |k(x, y)|^2 dx dy \right) \|f\|_{L^2(D)}^2 = \|k\|_{L^2(D \times D)}^2 \|f\|_{L^2(D)}^2.$$

Taking the square root on both sides yields $\|Kf\|_{L^2(D)} \leq \|k\|_{L^2(D \times D)} \|f\|_{L^2(D)}$. Taking the supremum over all f with $\|f\|_{L^2(D)} \leq 1$,

$$\|K\|_{op} = \sup_{\|f\|_{L^2(D)} \leq 1} \|Kf\|_{L^2(D)} \leq \|k\|_{L^2(D \times D)} = \|K\|_{HS}.$$

Which proves the claim. $\square$

Some useful properties of the Hilbert-Schmidt integral operator are grouped in the following lemma.

Lemma 2.6.14. *Let $D \subset \mathbb{R}^n$ be a measurable domain, and let $k : D \times D \rightarrow \mathbb{R}$ be a Hilbert-Schmidt kernel. The associated integral operator T_k defined by $(T_k f)(x) = \int_D k(x, y) f(y) dy$ satisfies the following properties:*

- i) T_k is a bounded linear operator on $L^2(D)$;
- ii) T_k is a compact operator on $L^2(D)$;
- iii) T_k is self-adjoint if and only if k is symmetric almost everywhere on $D \times D$.

Proof.

- i) Using Cauchy-Schwartz,

$$\begin{aligned} \|T_k f\|_{L^2(D)}^2 &= \int_D \left(\int_D k(x, y) f(y) dy \right)^2 dx \\ &\leq \int_D \left(\int_D |k(x, y)|^2 dy \right) \left(\int_D f(y)^2 dy \right) dx \\ &= \|f\|_{L^2(D)}^2 \iint_{D \times D} |k(x, y)|^2 dx dy < +\infty. \end{aligned}$$

ii) See Proposition 3 of [22].

iii) First, assume T_k is self-adjoint. Then for all $f, g \in L^2(D)$, we have $\langle T_k f, g \rangle_{L^2(D)} = \langle f, T_k g \rangle_{L^2(D)}$. Expanding this identity via Fubini's theorem yields:

$$\iint_{D \times D} k(x, y) f(y) g(x) dy dx = \iint_{D \times D} k(y, x) f(y) g(x) dy dx.$$

Since this equality holds for all test functions $f, g \in L^2(D)$, it follows that $k(x, y) = k(y, x)$ for almost every $(x, y) \in D \times D$.

Conversely, suppose $k(x, y) = k(y, x)$ almost everywhere. For any $f, g \in L^2(D)$, by Fubini's Theorem

$$\begin{aligned} \langle T_k f, g \rangle_{L^2(D)} &= \iint_{D \times D} k(x, y) f(y) g(x) dy dx \\ &= \iint_{D \times D} k(y, x) f(y) g(x) dx dy \\ &= \int_D f(y) \left(\int_D k(y, x) g(x) dx \right) dy = \langle f, T_k g \rangle_{L^2(D)}, \end{aligned}$$

which completes the proof. □

Definition 2.6.15 (Positive-semidefinite and positive-definite kernels). *Let $\mathcal{X}$ a set. A symmetric function $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is called a positive-semidefinite kernel if*

$$\sum_{i=1}^M \sum_{j=1}^M c_i c_j k(x_i, x_j) \geq 0$$

for every integer $M > 0$, every finite selection of points $x_1, \dots, x_M \in \mathcal{X}$, and every choice of coefficients $c_1, \dots, c_M \in \mathbb{R}$. Furthermore, the kernel k is called positive-definite if

$$\sum_{i=1}^M \sum_{j=1}^M c_i c_j k(x_i, x_j) > 0$$

for every integer $M > 0$, every finite selection of points $x_1, \dots, x_M \in \mathcal{X}$, and every choice of coefficients $c_1, \dots, c_M \in \mathbb{R}$ where at least one $c_i \neq 0$.

With this definition, the following useful lemma can be proven.

Lemma 2.6.16. *Let $D \subset \mathbb{R}^n$ a measurable domain and let $k : D \times D \rightarrow \mathbb{R}$ a measurable symmetric and positive-semidefinite kernel of the integral operator T_k on $L^2(D)$*

$$(T_k f)(x) = \int_D k(x, y) f(y) dy$$

for any $f \in L^2(D)$. Then T_k is self-adjoint and positive-semidefinite.

Proof. The self-adjointness of T_k follows from the proof of Lemma 2.6.14 iii).

To prove that T_k is positive-semidefinite we must show that $\langle T_k f, f \rangle_{L^2(D)} \geq 0$ for all $f \in L^2(D)$. By definition we have

$$\langle T_k f, f \rangle_{L^2(D)} = \iint_{D \times D} k(x, y) f(x) f(y) dx dy. \quad (2.22)$$

Because k is a positive-semidefinite kernel, for any finite set of point $\{x_1, \dots, x_M\} \subset D$ and any coefficients $c_1, \dots, c_M \in \mathbb{R}$, the discrete quadratic form satisfies

$$\sum_{i=1}^M \sum_{j=1}^M c_i c_j k(x_i, y_j) \geq 0.$$

We can extend this property to the integral in Equation (2.22). First, assume that f is a continuous function with compact support on D . Since k is continuous almost everywhere, the double integral can be approximated arbitrarily closely by its Riemann sums over a fine grid partition $\{A_i\}_{i=1}^M$ of D with representative points $x_i \in A_i$

$$\iint_{D \times D} k(x, y) f(x) f(y) dx dy = \lim_{M \rightarrow \infty} \sum_{i=1}^M \sum_{j=1}^M |A_i| |A_j| f(x_i) f(x_j) k(x_i, x_j).$$

Setting $c_i = f(x_i) |A_i|$, the term inside the limit becomes

$$\sum_{i=1}^M \sum_{j=1}^M c_i c_j k(x_i, y_j) \geq 0.$$

Finally, let f be any arbitrary function in $L^2(D)$. Since the space of compactly supported continuous functions is dense in $L^2(D)$, there exists a sequence of functions $(f_m)_{m=1}^\infty$ in this space such that $\|f_m - f\|_{L^2(D)} \rightarrow 0$ as $m \rightarrow \infty$.

From our previous step, we know that $\langle T_k f_m, f_m \rangle_{L^2(D)} \geq 0$ holds for every $m \geq 1$. Because T_k is a bounded linear operator, the quadratic mapping $f \mapsto \langle T_k f, f \rangle_{L^2(D)}$ is continuous with respect to the L^2 -norm. Taking the limit as $m \rightarrow \infty$,

$$\langle T_k f, f \rangle_{L^2(D)} = \lim_{m \rightarrow \infty} \langle T_k f_m, f_m \rangle_{L^2(D)} \geq 0.$$

Since this holds for any $f \in L^2(D)$, the operator T_k is positive-semidefinite. $\square$

We can decompose any self-adjoint Hilbert-Schmidt integral operator into its spectral properties. This is captured in the following theorem.

Theorem 2.6.17 (Hilbert-Schmidt decomposition theorem (Theorem 7.2.6 of [26])). *Let $D \subset \mathbb{R}^n$ a (Lebesgue-)measurable domain. Let T_k a self-adjoint Hilbert-Schmidt integral operator on $L^2(D)$ with Hilbert-Schmidt kernel function k . Then there exists an orthonormal basis $\{\xi_i\}_{i \geq 1}$ of $L^2(D)$ consisting of eigenfunctions of T_k , such that*

$$\int_{\Omega} k(x, y) \xi_i(y) dy = \lambda_i \xi_i(x), \quad \forall i \geq 1.$$

Moreover, k also admits the following representation

$$k(x, y) = \sum_{i=1}^{\infty} \lambda_i \xi_i(x) \xi_i(y),$$

where this series converges in $L^2(D \times D)$ in the sense that

$$\lim_{m \rightarrow +\infty} \iint_{D \times D} \left(k(x, y) - \sum_{i=1}^m \lambda_i \xi_i(x) \xi_i(y) \right)^2 dx dy = 0$$

Note the distinction between Theorem 2.6.17 and Theorem 2.6.9. The latter provides a spectral decomposition of the Hilbert–Schmidt operator itself, whereas Theorem 2.6.17 yields a decomposition of its integral kernel. For Hilbert–Schmidt integral operators, this kernel representation fully characterizes the operator. In both cases, the eigenvalues and eigenfunctions play a central role, as they encode the structure of the operator through its spectral expansion.

Remark 2. The Hilbert-Schmidt decomposition theorem can be extended if we assume the operator to be positive-semidefinite and the kernel to be continuous. Then there is an orthonormal basis $\{\xi_i(\cdot)\}_{i=1}^{\infty}$ of $L^2(D)$ consisting of eigenfunctions of T_k such that the corresponding sequence of eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ are non-negative. The eigenfunctions are continuous on D and k can be represented as

$$k(x, y) = \sum_{i=1}^{\infty} \lambda_i \xi_i(x) \xi_i(y)$$

where this convergence is now absolute in the sense that for every $(x, y) \in D \times D$

$$\sum_{i=1}^{\infty} |\lambda_i \xi_i(x) \xi_i(y)| < \infty$$

and uniform i.e.

$$\sup_{x, y \in D} \left| k(x, y) - \sum_{i=1}^{\infty} \lambda_i \xi_i(x) \xi_i(y) \right| \xrightarrow{m \rightarrow \infty} 0.$$

This theorem is known as Mercer's Theorem [26].

2.6.2. SPECTRA OF OPERATORS

Before we introduce some useful theorems we will first briefly discuss the spectra of operators. Given a linear operator $A: H \rightarrow H$, we are interested under which conditions the equation $(A - \lambda I)x = y$ can be solved for a given $y \in H$ and $\lambda \in \mathbb{R}$. Recall that when $y = 0$, this equation has a nontrivial solution if $\lambda \in \sigma(A)$, where

$$\sigma(A) := \{ \lambda \in \mathbb{R} \mid \exists x \neq 0, Ax = \lambda x \}.$$

We can alternatively write $\sigma(A) = \mathbb{R} \setminus \rho(A)$ with

$$\rho(A) := \{ \lambda \in \mathbb{R} \mid (A - \lambda I)^{-1} \in B(H) \}$$

where $B(H)$ is the set of all bounded operators. This gives rise to the following definitions.

Definition 2.6.18 (Resolvent, see [21] Chapter 3 §6). *Given a closed linear operator $A : H \rightarrow H$. We call the set*

$$\rho(A) := \{\lambda \in \mathbb{R} \mid (T - \lambda I)^{-1} \in B(H)\}$$

the resolvent set of A . The map $R(\lambda, A) : \rho(A) \mapsto B(H)$ defined by

$$R(\lambda, A) := (A - \lambda I)^{-1}$$

is called the resolvent of operator A .

The following lemma counts as a starting point for this definition.

Lemma 2.6.19 (Lemma 1.7.2 of [23]). *Let B bounded and $\|I - B\| < 1$. Then B is invertible, B^{-1} is bounded and defined as*

$$B^{-1} := \lim_{n \rightarrow \infty} \sum_{i=1}^n (I - B)^i$$

The following useful identities can be proven.

Lemma 2.6.20 (Resolvent Identities, p.40 of [23]).

i) *Let A be a linear operator and let $\lambda_1, \lambda_2 \in \rho(A)$. Then*

$$R(\lambda_1, A) - R(\lambda_2, A) = (\lambda_1 - \lambda_2)R(\lambda_1, A)R(\lambda_2, A).$$

ii) *Let A_1 and A_2 be linear operators with the domain $D(A_1) \subseteq D(A_2)$ and $\lambda \in \rho(A_1) \cap \rho(A_2)$. Then*

$$R(\lambda, A_2) - R(\lambda, A_1) = R(\lambda, A_2)(A_1 - A_2)R(\lambda, A_1).$$

A useful definition we will use in some proofs is the *Riesz projection*.

Definition 2.6.21 (Riesz Projections). *Let $A : H \rightarrow H$ a closed linear operator with resolvent set $\rho(A)$ and let λ_0 an eigenvalue of A . Consider a positively orientated simple closed contour C contained in $\rho(A)$ separating λ_0 from the rest of the spectrum of A . The Riesz projection operator $P_C : H \rightarrow H$ associated with λ_0 is defined by*

$$P_C := -\frac{1}{2\pi i} \int_C R(z, A) dz = -\frac{1}{2\pi i} \int_C (A - zI)^{-1} dz.$$

Lemma 2.6.22 (Properties of Riesz Projection). *Given a closed linear operator $A : H \rightarrow H$ and the Riesz projection operator P defined as in Definition 2.6.21 then*

- i) *P is a bounded linear operator;*
- ii) *P commutes with A ;*
- iii) *P is self-adjoint if and only if A is self-adjoint*
- iv) *P is a projection operator, i.e. $P^2 = P$.*

Proof.

i) Linearity is clear from the fact that A is linear, therefore also $(A - zI)^{-1}$ is linear. Recall that $R(\lambda, A)$ is bounded for any $\lambda \in \rho(A)$ (see Definition 2.6.21) and hence also P is bounded.

ii) See the proof of Theorem 6.17 of [21].

iii) Note that we define

$$P = -\frac{1}{2\pi i} \int_C R(z, A) dz \quad \text{and} \quad P^* = -\frac{1}{2\pi i} \int_C R(z, A)^* dz.$$

By Theorem 6.22 of [21]

$$R(\lambda, A)^* = R(\lambda, A^*), \quad \lambda \in \rho(A)$$

The result now follows.

iv) See p. 39 of [21].

□

2.6.3. MODES OF CONVERGENCE

In this section we give an overview of the possible way of converging for deterministic sequences, random variables and operators.

Definition 2.6.23 (Pointwise convergence). *A deterministic sequence of functions $f_n : X \rightarrow Y$ converges pointwise to f if*

$$f_n(x) \rightarrow f(x) \quad \forall x \in X.$$

Definition 2.6.24 (Uniform convergence). *A deterministic sequence of functions $f_n : X \rightarrow Y$ converges uniformly to f if $\forall \epsilon > 0, \exists N > 0$ such that for all $n \geq N$ and every $x \in X$*

$$\|f_n(x) - f(x)\|_Y < \epsilon.$$

Definition 2.6.25 (L^p convergence). *Assume that $(X, \mathcal{X}, \mu)$ is a measured space. A deterministic sequence of functions $f_n : X \rightarrow Y$ converges in $L^p(X, \mathcal{X}, \mu)$ to f if*

$$\|f_n - f\|_{L^p(X, \mathcal{X}, \mu)} = \left(\int_X \|f_n(x) - f(x)\|_Y^p d\mu(x) \right)^{1/p} \rightarrow 0$$

Definition 2.6.26 (Almost uniform convergence). *Consider a measured space $(X, \mathcal{X}, \mu)$ with $\mu(X) < +\infty$. A sequence of functions $f_n : X \rightarrow Y$ converges μ -almost uniformly to f if $\forall \epsilon > 0, \exists E_\epsilon \subset X$ such that $\mu(E_\epsilon) < \epsilon$ and*

$$f_n \rightarrow f$$

uniformly on $X \setminus E_\epsilon$.

Definition 2.6.27 (Almost everywhere convergence). *A deterministic sequence of functions $f_n : X \rightarrow Y$ converges almost everywhere to f if*

$$f_n(x) \rightarrow f(x) \quad \text{for a.e. } x \in X.$$

We have the following modes for random variables.

Definition 2.6.28 (Convergence in distribution, p. 5 of [20]). *A sequence of random vectors $\{X_n\}_{n \geq 1}$ converges in distribution to X if*

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

at every x where the limit distribution $x \mapsto \mathbb{P}(X \leq x)$ is continuous.

Definition 2.6.29 (Convergence in probability, p. 5 of [20]). *Let $(S, d_S(\cdot, \cdot))$ a metric space. A sequence of random vectors X_n converges to X in probability if $\forall \epsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P}(d_S(X_n, X) > \epsilon) = 0$$

Definition 2.6.30 (Convergence in $L^p(\Omega)$). *A sequence of random variables X_n with values in a normed vector space $(Y, \|\cdot\|_Y)$ converges to X in $L^p(\Omega)$ if*

$$\|X_n - X\|_{L^p(\Omega)} = (\mathbb{E}\|X_n - X\|_Y^p)^{1/p} \rightarrow 0.$$

Definition 2.6.31 (Almost sure convergence, p. 6 [20]). *Let $(S, d_S(\cdot, \cdot))$ a metric space. A sequence of random variables X_n converges to X almost surely if*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} d_S(X_n, X) = 0\right) = 1.$$

Note that convergence in probability and convergence almost surely only make sense if each of X_n and X are defined on the same probability space. For convergence in distribution this is not necessary. We will lastly look at the convergence for deterministic operators.

Definition 2.6.32 (Strong convergence). *A sequence of (deterministic) operators $T_n : X \rightarrow Y$ with X, Y normed spaces converges to $T : X \rightarrow Y$ strongly if*

$$\|T_n x - T x\|_Y \rightarrow 0 \quad \forall x \in X.$$

Definition 2.6.33 (Uniform or in norm convergence). *A sequence of (deterministic) operators $T_n : X \rightarrow Y$ with X, Y normed spaces converges to $T : X \rightarrow Y$ in norm (or uniformly) if*

$$\|T_n - T\|_{op} \rightarrow 0.$$

2.6.4. AUXILIARY RESULTS

In this section we will cover a wide range of useful theorems from linear algebra which will become useful throughout this thesis.

Theorem 2.6.34 (Courant-Fisher Theorem, X.4 Theorem 8 of [27]). *Let $A : H \rightarrow H$ a compact, self-adjoint linear operator on the Hilbert space H with eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \dots$, then*

$$\lambda_{k+1} = \min_{y_1, \dots, y_k \in H} \max_{\substack{\langle x, y_l \rangle = 0 \\ \|x\|=1}} \langle Ax, x \rangle, \quad k \geq 1,$$

$$\lambda_k = \max_{y_1, \dots, y_k \in H} \min_{\substack{\langle x, y_l \rangle = 0 \\ \|x\|=1}} \langle Ax, x \rangle, \quad k \geq 1.$$

Lemma 2.6.35 (Section 8.2 of [28]). *Let p be a positive integer and $A, B \in \mathbb{R}^{p \times p}$ symmetric matrices with eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_p(A)$ and $\lambda_1(B) \geq \lambda_2(B) \geq \dots \geq \lambda_p(B)$ then*

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_2 \quad \forall i = 1, \dots, N.$$

In this theorem, $\|\cdot\|_2$ denotes the spectral norm

$$\|A\|_2 = \max\{\|Av\|_2 : \|v\|_2 \leq 1\} = \sqrt{\rho(A^\top A)} = \sigma_1(A)$$

where $\rho(A^\top A)$ denotes the largest eigenvalue of $A^\top A$ and $\sigma_1(A)$ the largest singular value of A . A similar lemma can be proven for the eigenvalues of operators.

Lemma 2.6.36. *Let $A : H \rightarrow H$ and $B : H \rightarrow H$ compact, self-adjoint operator with eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \dots$ and $\lambda_1(B) \geq \lambda_2(B) \geq \dots$, then*

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_{op}, \quad \forall i = 1, \dots, N.$$

Proof. Define $E := B - A$ and fix $i \geq 1$. By Cauchy-Schwartz and linearity for some $x \in H$ with $\|x\| = 1$

$$\begin{aligned} \langle Bx, x \rangle &= \langle (E + A)x, x \rangle \\ &= \langle Ex, x \rangle + \langle Ax, x \rangle \\ &\leq \langle Ax, x \rangle + \|E\|_{op}. \end{aligned}$$

Now applying Theorem 2.6.34 yields

$$\lambda_i(B) \leq \lambda_i(A) + \|E\|_{op}.$$

Replacing E by $-E$ gives the inequality

$$\lambda_i(A) \leq \lambda_i(B) + \|E\|_{op}.$$

Combining these gives

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_{op}$$

which proves the result. $\square$

For proving consistency of the eigenvalues we need a variant of the Davis-Kahan Theorem. First, consider two vectors $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^n$. We define the angle between $\mathbf{z}_1$ and $\mathbf{z}_2$ as

$$\angle(\mathbf{z}_1, \mathbf{z}_2) := \arccos |\mathbf{z}_1^\top \mathbf{z}_2|. \quad (2.23)$$

2

Theorem 2.6.37 (Davis-Kahan Theorem, Corollary 3 in [29]). *Let $A, B \in \mathbb{R}^{p \times p}$ symmetric matrices with eigenvalues $\lambda_1(A) \geq \dots \geq \lambda_p(A)$ and $\lambda_1(B) \geq \dots \geq \lambda_p(B)$ respectively. Fix $j \in \{1, 2, \dots, p\}$, and assume $\min(\lambda_{j-1}(A) - \lambda_j(A), \lambda_j(A) - \lambda_{j+1}(A)) > 0$ where we let $\lambda_0 = \infty$ and $\lambda_{p+1} = -\infty$. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ are unit-length and satisfy*

$$A\mathbf{x} = \lambda_j(A)\mathbf{x} \quad \text{and} \quad B\mathbf{y} = \lambda_j(B)\mathbf{y}$$

then

$$\sin \angle(\mathbf{x}, \mathbf{y}) \leq \frac{2\|A - B\|_2}{\min(\lambda_{j-1}(A) - \lambda_j(A), \lambda_j(A) - \lambda_{j+1}(A))}.$$

Since unit-length eigenvectors are defined up to a sign, we may reverse the sign of $\mathbf{y}$ if necessary, so there is a choice of orientation of $\mathbf{y}$ for which $\mathbf{y}^\top \mathbf{x} \geq 0$. In that case,

$$\|\mathbf{x} - \mathbf{y}\| \leq \frac{2^{\frac{3}{2}} \|A - B\|_2}{\min(\lambda_{j-1}(A) - \lambda_j(A), \lambda_j(A) - \lambda_{j+1}(A))}.$$

We will end this chapter with a two important lemma's about convergence.

Lemma 2.6.38 (Proposition 2.1 of [30]). *Suppose $\{B_n\}_{n \geq 1}$ and $\{C_n\}_{n \geq 1}$ are two sequences of bounded operators satisfying*

$$B_n \xrightarrow{n \rightarrow \infty} B, \quad \text{and} \quad C_n \xrightarrow{n \rightarrow \infty} C$$

strongly (in the sense of Definition 2.6.32). Then if A is compact

$$\|B_n \circ A \circ C_n - B \circ A \circ C\|_{op} \xrightarrow{n \rightarrow \infty} 0.$$

Lemma 2.6.39. *Let $\{\mathbf{X}_n\}_{n \geq 1}$ a sequence of random vectors and suppose that*

$$\mathbb{E}[\|\mathbf{X}_n\| | A_n] \xrightarrow{n \rightarrow \infty} 0 \quad \text{and} \quad \mathbb{P}(A_n) \xrightarrow{n \rightarrow \infty} 1,$$

then $\mathbf{X}_n \xrightarrow[n \rightarrow \infty]{prob} 0$.

Proof. Let $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}(\|\mathbf{X}_n\| > \epsilon) &= \mathbb{P}(\|\mathbf{X}_n\| > \epsilon, A_n) + \mathbb{P}(\|\mathbf{X}_n\| > \epsilon, A_n^c) \\ &\leq \mathbb{P}(\|\mathbf{X}_n\| > \epsilon | A_n) \mathbb{P}(A_n) + \mathbb{P}(A_n^c). \end{aligned}$$

By Markov,

$$\mathbb{P}(\|\mathbf{X}_n\| > \epsilon | A_n) \leq \frac{\mathbb{E}[\|\mathbf{X}_n\| | A_n]}{\epsilon}.$$

Therefore

$$\mathbb{P}(\|\mathbf{X}_n\| > \epsilon) \leq \frac{\mathbb{E}[\|\mathbf{X}_n\| | A_n]}{\epsilon} + \mathbb{P}(A_n^c) \xrightarrow{n \rightarrow \infty} 0$$

by assumption. Thus $\mathbf{X}_n \xrightarrow[n \rightarrow \infty]{prob} 0$ as desired. $\square$

2.7. *U*-STATISTICS

In many mathematical applications, it is useful to view a statistic as a member of a general class of estimators. One such class is that of *U-statistics*, first introduced by Hoeffding [31].

Following [32], let $\mathbf{X}_1, \mathbf{X}_2, \dots \in \mathbb{R}^p$ be i.i.d. observations with distribution F . We call a functional $\theta = \theta(F)$ *regular* if it admits an unbiased estimator of finite degree, that is, there exists a measurable function $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ such that

$$\theta(F) = \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] = \int \cdots \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) dF(\mathbf{x}_1) \cdots dF(\mathbf{x}_k). \quad (2.24)$$

The smallest such k is called the *degree* of θ .

Definition 2.7.1 (*U-statistic* [33]). *Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ be a measurable function, and let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a i.i.d. sample of size $n \geq k$ from F . The *U-statistic* associated with g is defined by*

$$U_n = U(\mathbf{X}_1, \dots, \mathbf{X}_n) = |\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}), \quad (2.25)$$

where $\mathfrak{I}_{k,n}$ denotes the set of injective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$.

By construction, U_n is an unbiased estimator of $\theta(F)$. Moreover, since each unordered k -tuple appears exactly $k!$ times in the above sum, we have

$$|\mathfrak{I}_{k,n}| = k! \binom{n}{k}.$$

Note that we may assume w.l.o.g. that g is symmetric. Indeed, define the symmetrized kernel

$$g^{\text{sym}}(\mathbf{x}_1, \dots, \mathbf{x}_k) = \frac{1}{k!} \sum_{\pi \in \mathfrak{I}_{k,n}} g(\mathbf{x}_{\pi(1)}, \dots, \mathbf{x}_{\pi(k)}),$$

where $\mathfrak{I}_{n,n}$ denotes the set of all permutations of $\{1, \dots, k\}$. In that case

$$|\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) = |\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} g^{\text{sym}}(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}),$$

since the sum over injective maps averages uniformly over all permutations of the arguments.

U-statistics are optimal among unbiased estimators in the sense of minimum variance. More precisely, any unbiased estimator can be improved, yielding a *U*-statistic.

Lemma 2.7.2 ([33, p. 176]). *Let $S = S(\mathbf{X}_1, \dots, \mathbf{X}_n)$ be an unbiased estimator of $\theta(F)$. Then the corresponding *U*-statistic U_n is also unbiased and satisfies*

$$\text{Var}(U_n) \leq \text{Var}(S),$$

with equality if and only if $\mathbb{P}(U_n = S) = 1$.

In fact we can explicitly compute the variance of a U -statistic. Following Section 5.2.1 of [33], assume that g is symmetric of order k and satisfies $\mathbb{E}[g^2(\mathbf{X}_1, \dots, \mathbf{X}_k)] < \infty$. Define for $1 \leq c \leq k-1$,

$$g_c(\mathbf{x}_1, \dots, \mathbf{x}_c) = \mathbb{E}[g(\mathbf{x}_1, \dots, \mathbf{x}_c, \mathbf{X}_{c+1}, \dots, \mathbf{X}_k)]$$

where we set $g_k := g$. Then

$$\mathbb{E}[g_c(\mathbf{X}_1, \dots, \mathbf{X}_c)] = \theta, \quad 1 \leq c \leq m.$$

Now define $\zeta_0 = 0$ and for $1 \leq c \leq m$,

$$\zeta_c = \text{Var}(g_c(\mathbf{X}_1, \dots, \mathbf{X}_c))$$

Lemma 2.7.3 (Lemma A (p. 183) of [33]). *If $\mathbb{E}[g^2(\mathbf{X}_1, \dots, \mathbf{X}_k)] < \infty$ and g is symmetric, the variance of U_n is given by*

$$\text{Var}(U_n) = \binom{n}{k}^{-1} \sum_{c=1}^k \binom{k}{c} \binom{n-k}{k-c} \zeta_c.$$

The next lemma is an important probability inequalities for U -statistics. For the proof we refer to [33] p.201.

Lemma 2.7.4. *Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ be a measurable function such that $a \leq g(\mathbf{x}_1, \dots, \mathbf{x}_k) \leq b$ and let U_n the U -statistic of order k . Set $\theta = \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]$ and $\sigma^2 = \text{Var}(g(\mathbf{X}_1, \dots, \mathbf{X}_k))$. Then for $t > 0$ and $n \geq k$,*

$$\mathbb{P}(U_n - \theta \geq t) \leq \exp\left(-\frac{[n/k] t^2}{2(\sigma^2 + 1/3(b - \theta)t)}\right).$$

2.7.1. CONDITIONAL U -STATISTICS

In this section, we extend the U -statistic framework to the conditional setting, as first proposed by Stute [34]. Following the approach taken in [32], we let $(\mathbf{X}_1, \mathbf{Z}_1), (\mathbf{X}_2, \mathbf{Z}_2), \dots$ be a sequence of i.i.d. observations in $\mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $F_{\mathbf{X}, \mathbf{Z}}$. We call a functional $\theta = \theta(F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}, \dots, F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k})$ *conditionally regular* of order k if there exists a measurable function $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \theta(F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}, \dots, F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}) &= \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k) | \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k] \\ &= \int \cdots \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) dF_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}(\mathbf{x}_1) \cdots dF_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}(\mathbf{x}_k). \end{aligned}$$

For convenience, we introduce the shorthand

$$\theta(\mathbf{z}_1, \dots, \mathbf{z}_k) := \theta(F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}, \dots, F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}).$$

We are using random design, allowing the continuous conditioning variables $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ to be random. Consequently, for any fixed target points $\mathbf{z}_1, \dots, \mathbf{z}_k$, the event $\{\mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k\}$

$\mathbf{z}_k\}$ has probability zero. While the conditional expectation $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ is well-defined almost surely, it cannot be estimated using a naive sample counterpart, as the sample contains no observations matching the exact conditioning values. Instead, it must be approximated by selecting observations for which the $\mathbf{Z}_i$'s are close to the target points $(\mathbf{z}_1, \dots, \mathbf{z}_k)$ by kernel weighting. This gives rise to the following definition.

Definition 2.7.5 (Conditional U -statistics). *Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ be a measurable function and let $\{(\mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} F_{\mathbf{X}, \mathbf{Z}}$ a sample with $n \geq k$. The conditional U -statistic estimator of $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ is defined as*

$$U_n(\mathbf{z}_1, \dots, \mathbf{z}_k) = \frac{\sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}, \quad (2.26)$$

where $\mathcal{I}_{k,n}$ denotes the set of injective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$, and $K_h(\cdot) = h^{-d} K(\cdot/h)$ is a kernel function with bandwidth $h > 0$.

Note that the numerator of Equation (2.26) corresponds to a localized U -statistic of order k , while the denominator acts as a normalization factor estimating the joint density of $(\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ at $(\mathbf{z}_1, \dots, \mathbf{z}_k)$, ensuring that U_n targets the conditional quantity rather than an unconditional one. In a fixed design setting, the denominator is not needed as a separate random normalization term, since the distribution of the design points is known and deterministic. Any normalization can then be incorporated directly into the kernel function g .

This class of estimators can be viewed as a natural kernel-weighted generalization of U -statistics, and in particular extends the Nadaraya–Watson estimator. Indeed, for $k = 1$ and $g(\mathbf{x}) = \mathbf{x}$, we recover the conditional mean

$$\theta(\mathbf{z}_1) = \mathbb{E}[\mathbf{X}_1 | \mathbf{Z}_1 = \mathbf{z}_1].$$

The corresponding estimator becomes

$$U_n(\mathbf{z}_1) = \frac{\sum_{i=1}^n \mathbf{X}_i K_h(\mathbf{Z}_i - \mathbf{z}_1)}{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z}_1)},$$

which is precisely the Nadaraya–Watson estimator from Section 2.4.2.

2.8. PRINCIPAL COMPONENT ANALYSIS

Throughout this project, we work with high dimensional datasets, which are often hard to visualize and interpret. We therefore aim to reduce the dimensionality of the data while retaining as much of original variation as possible. This is achieved by transforming the data into a new set of uncorrelated variables called *principal components* (PCs). These components are ordered such that the first few retain most of the variability present in the original variables.

In this section we follow along the lines of [35]. Suppose we are interested in the variances and correlation structure among p random variables collected in the vector $\mathbf{X} =$

$(X_1, \dots, X_p)$. When p is large, it is neither practical nor informative to examine all individual variances and pairwise correlations. Instead, we seek a linear combination $\xi_1^\top \mathbf{X}$ of the original variables that has the maximum possible variance. Next, we find a second linear combination $\xi_2^\top \mathbf{X}$ that is uncorrelated with $\xi_1^\top \mathbf{X}$ and that also has the largest possible variance subject to this constraint. This procedure can be repeated up to p times, yielding a sequence of new variables $\xi_k^\top \mathbf{X}$, where the k -th such variable is called the k -th principal component. The goal of this approach is that most of the variation in $\mathbf{X}$ can be captured by the first $m \ll p$ principal components, thereby reducing the dimensionality and complexity of the data while retaining most of its information.

The question still remains how to find these components. For that let us first define the covariance matrix of $\mathbf{X}$:

$$\text{Cov}(\mathbf{X}) = \Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_1, X_p) & \cdots & \cdots & \text{Var}(X_p) \end{bmatrix}. \quad (2.27)$$

To derive the PCs, consider the first linear function $\xi_1^\top \mathbf{X}$ which maximizes $\text{Var}(\xi_1^\top \mathbf{X}) = \xi_1^\top \Sigma \xi_1$. As it stands the maximum will not be attained for finite ξ_1 so we introduce the constrain that $\xi_1^\top \xi_1 = 1$. The standard approach to maximize the variance under this constrain is to use Lagrange multipliers to maximize

$$\xi_1^\top \Sigma \xi_1 - \lambda(\xi_1^\top \xi_1 - 1).$$

Differentiating with respect to ξ_1 yields

$$(\Sigma - \lambda \mathbf{I}_p) \xi_1 = 0. \quad (2.28)$$

This implies that λ is an eigenvalue and ξ_1 an eigenvector of Σ . To see that this eigenvector indeed maximizes the variance note that in that case

$$\xi_1^\top \Sigma \xi_1 = \xi_1^\top \lambda \xi_1 = \lambda.$$

So indeed λ is the largest eigenvalue of Σ and ξ_1 the corresponding eigenvector. The second component maximizes $\text{Var}(\xi_2^\top \mathbf{X}) = \xi_2^\top \Sigma \xi_2$ given that $\text{Cov}(\xi_1^\top \mathbf{X}, \xi_2^\top \mathbf{X}) = \lambda_1 \xi_2^\top \xi_1 = 0$. So by using the Lagrange multipliers again, we need to maximize

$$\xi_2^\top \Sigma \xi_2 - \lambda(\xi_2^\top \xi_2 - 1) - \mu \xi_2^\top \xi_1.$$

Differentiating yields

$$\Sigma \xi_2 - \lambda \xi_2 - \mu \xi_1 = 0. \quad (2.29)$$

If we multiply both sides by $\xi_1^\top$ we get

$$\xi_1^\top \Sigma \xi_2 - \lambda \xi_1^\top \xi_2 - \mu \xi_1^\top \xi_1 = 0. \quad (2.30)$$

Since $\text{Cov}(\xi_1^\top \mathbf{X}, \xi_2^\top \mathbf{X}) = \xi_1^\top \Sigma \xi_2 = \lambda_1 \xi_1^\top \xi_2$, the first two terms are zero. Since $\xi_1^\top \xi_1 = 1$ the above equation shows that $\mu = 0$. Therefore Equation (2.29) leads to

$$(\Sigma - \lambda \mathbf{I}_p) \xi_2 = 0. \quad (2.31)$$

Hence, λ is an eigenvalue and ξ_1 an eigenvector of Σ . Again $\xi_2^\top \Sigma \xi_2 = \lambda$ so λ must be as large as possible. We will not show the derivation of the other principal components here but they can be found in a very similar fashion.

Altogether we have shown that the vectors of coefficients $\xi_1, \xi_2, \dots, \xi_p$ are the eigenvectors of Σ corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p$ with $\text{Var}(\xi_k^\top \mathbf{X}) = \lambda_k$ and where the projection $\xi_k^\top \mathbf{X}$ is called the k -th principal component.

2.8.1. PCA IN PRACTICE

The derivation above was carried out under the (unrealistic) assumption that the random vector $\mathbf{X}$ has a known covariance matrix. In practice, this matrix is unknown, and we instead rely on data to estimate this covariance matrix. Suppose we observe n independent realizations of the p -dimensional random vector $\mathbf{X}$, denoted by $\{\mathbf{X}_1, \dots, \mathbf{X}_n\} \in \mathbb{R}^p$. We can define the sample covariance matrix as

$$\mathbf{S} = \frac{1}{n-1} \mathbf{M}^\top \mathbf{M} \quad (2.32)$$

where $\mathbf{M} \in \mathbb{R}^{n \times p}$ is the centered data matrix with entries

$$(\mathbf{M})_{ij} = X_{ij} - \bar{X}_j \quad \text{and} \quad \bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}.$$

To find the first sample principal component, we look for a direction vector ξ_1 that maximizes the sample variance of the projected data points. Let $z_{i1} = \xi_1^\top \mathbf{X}_i$ for $i = 1, \dots, n$. We want to find ξ_1 to maximize the sample variance

$$\frac{1}{n-1} \sum_{i=1}^n (z_{i1} - \bar{z}_1)^2$$

subject to the normalization constraint $\xi_1^\top \xi_1 = 1$. We denote the optimal solution of this problem as $\hat{\xi}_1$.

To obtain the second principal component, we define the next set of projections as $z_{i2} = \xi_2^\top \mathbf{X}_i$ for $i = 1, \dots, n$. We then find a vector ξ_2 that maximizes the sample variance of z_{i2} , subject to the constraints that $\xi_2^\top \xi_2 = 1$ and that the sample scores z_{i1} and z_{i2} are uncorrelated. Repeating this procedure up to p times yields a sequence of orthogonal sample vectors $\hat{\xi}_1, \dots, \hat{\xi}_p$. This results in the *sample principal components* $\hat{\xi}_k^\top \mathbf{X}$ for $k = 1, \dots, p$. The specific value $z_{ik} = \hat{\xi}_k^\top \mathbf{X}_i$ is called the *score* of the i -th observation on the k -th principal component.

By following the same reasoning as in Section 2.8, it can be shown that the sample variance of the k -th principal component scores equals $\hat{\lambda}_k$, the k -th largest eigenvalue of the sample covariance matrix $\mathbf{S}$, and that $\hat{\xi}_k$ is the corresponding eigenvector.

A considerable amount of work has gone into deriving the asymptotic distributions and consistency properties of the sample eigenvalues and eigenvectors. Many of these classical results require a priori knowledge about the distribution of the data (see, e.g., [36, 35]). Some more recent papers extended these properties to a much broader class of distributions, see for example [37].

2.9. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

In some cases the data we have are functions instead of points. The idea of Functional Principal Component Analysis (FPCA) is to carry over the basic approaches from PCA towards this continuous, infinite-dimensional case.

Suppose we are interested in analysing the variance and correlation structure of a centred stochastic process $X = (X(t))_{t \in [a,b]}$ with values in $L^2([a,b])$. If the process is not centred, we can subtract its mean function $\mu(t) = \mathbb{E}[X(t)]$. Analogous to multivariate PCA, we seek a function $\xi_1(t)$ that captures the dominant mode of variation in X . Successive components are again determined sequentially: the k -th function $\xi_k(t)$ is chosen to maximize the remaining variation in X subject to being orthogonal to all previously determined functions $\xi_1(t), \dots, \xi_{k-1}(t)$. Through this procedure, we obtain a sequence of orthonormal functions $\{\xi_k(t)\}_{k \geq 1}$, known as the functional principal components (or eigenfunctions). Ideally, the vast majority of the total variation is captured by the first m components, thereby effectively reducing the dimensionality and complexity of the functional data.

To formalize the derivation of these functions $\xi_k(t)$, we aim to apply the spectral decomposition theorem (Theorem 2.6.17) to the covariance kernel $\Gamma(s, t) = \text{Cov}(X(s), X(t))$. This requires the associated integral operator T_Γ to be a self-adjoint Hilbert-Schmidt integral operator. By construction, we know that this covariance kernel is symmetric and positive-semidefinite. By Lemma 2.6.16 the associated integral operator T_Γ is then self-adjoint and positive-semidefinite. We can prove that Γ is a Hilbert-Schmidt kernel if $X \in L^2(\Omega \times L^2([a,b]))$.

Lemma 2.9.1. *Let $(X_t)_{t \in [a,b]}$ a measurable stochastic process in $L^2(\Omega \times L^2([a,b]))$. Define the covariance functions of X for $s, t \in [a, b]$ as*

$$\Gamma(s, t) = \text{Cov}(X(s), X(t)) = \mathbb{E} \left[(X(s) - \mathbb{E}(X(s))) (X(t) - \mathbb{E}(X(t))) \right].$$

Then the integral operator

$$(T_\Gamma f)(x) = \int_a^b \Gamma(x, y) f(y) dy$$

is Hilbert-Schmidt.

Proof. Assume without loss of generality that $\mathbb{E}[X(t)] = 0$ for all $t \in [a, b]$. By Definition 2.6.12, the operator T_Γ is Hilbert-Schmidt if and only if its Hilbert-Schmidt norm is finite

$$\|T_\Gamma\|_{HS} < \infty.$$

Recall that the Hilbert-Schmidt norm of an integral operator is equivalent to the L^2 -norm of its kernel. For any $s, t \in [a, b]$, applying the Cauchy-Schwarz inequality yields

$$|\Gamma(s, t)|^2 = |\mathbb{E}(X(s)X(t))|^2 \leq \mathbb{E}(X(s)^2)\mathbb{E}(X(t)^2).$$

Integrating the squared kernel over the domain $[a, b]^2$ and applying Fubini's Theorem to exchange the order of expectation and integration, we obtain

$$\begin{aligned}\|T_\Gamma\|_{HS}^2 &= \|\Gamma\|_{L^2([a,b]^2)}^2 \\ &= \int_a^b \int_a^b |\Gamma(s, t)|^2 ds dt \\ &\leq \left(\int_a^b \mathbb{E}(X(t)^2) dt \right)^2 \\ &= \left(\mathbb{E} \left(\int_a^b X(t)^2 dt \right) \right)^2 \\ &= \left(\mathbb{E} \|X\|_{L^2([a,b])}^2 \right)^2\end{aligned}$$

Since X has finite second moments by assumption, $\mathbb{E}[\|X\|_{L^2([a,b])}^2] < \infty$, which implies $\|T_\Gamma\|_{HS}^2 < \infty$, completing the proof. $\square$

By Lemma 2.9.1, Γ is a Hilbert-Schmidt kernel and it defines therefore a compact, self-adjoint, positive semidefinite operator (Lemma 2.6.14). We can now apply Theorem 2.6.17, which guarantees that Γ admits the spectral decomposition

$$\Gamma(s, t) = \sum_{i=1}^{\infty} \lambda_i \xi_i(s) \xi_i(t) \quad (2.33)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the ordered eigenvalues corresponding to the set of orthonormal eigenfunctions $\{\xi_i(\cdot)\}_{i=1}^{\infty}$. This sum converges in the L^2 sense. If Γ is continuous, the convergence is absolute and uniform. Hence for all i , for almost every $s \in [a, b]$, we have the following eigenequation

$$\int_a^b \Gamma(s, t) \xi_i(t) dt = \lambda_i \xi_i(s). \quad (2.34)$$

To simplify notions we will use the operator notion and write Equation (2.34) as

$$T_\Gamma \xi_i = \lambda_i \xi_i, \quad \forall i \geq 1.$$

We can now apply the main theorem behind FPCA, Karhunen-Loève theorem.

Theorem 2.9.2 (Karhunen-Loève theorem, Theorem 7.2.7 and 7.2.8 [26]). *Let $X = (X(t))_{t \in [a,b]}$ be a measurable stochastic process satisfying $\mathbb{E}[\|X\|_{L^2([a,b])}^2] < \infty$ with mean function $\mu(t) = \mathbb{E}[X(t)]$ and a square-integrable covariance kernel Γ . Then for almost every $t \in [a, b]$, X admits the following representation*

$$X(t) - \mu(t) = \sum_{i=1}^{\infty} \zeta_i \xi_i(t) \quad (2.35)$$

where $\{\xi_i\}_{i=1}^{\infty}$ are the orthonormal eigenfunctions of the covariance operator T_Γ corresponding to the sequence of eigenvalues λ_i . The functional principal component (FPC)

scores ζ_i are random variables defined by the projection

$$\zeta_i = \int_a^b (X(t) - \mu(t)) \xi_i(t) dt = \langle X - \mu, \xi_i \rangle \quad (2.36)$$

with $i, k \in \mathbb{N}$

$$\mathbb{E}(\zeta_i) = 0, \quad \text{Var}(\zeta_i) = \lambda_i, \quad \mathbb{E}(\zeta_i \zeta_k) = \lambda_i \delta_{ik}. \quad (2.37)$$

The series in Equation (2.35) converges in $L^2(\Omega \times [a, b])$. If Γ is continuous, the convergence is uniform in $t \in [a, b]$ and holds $\mathbb{P}$ -almost surely.

By Theorem 2.9.2 we can decompose our stochastic process in terms of eigenfunctions of the covariance operator. To see that these eigenfunctions are indeed the functions depicting the dominant mode of variation in X , recall that we want to maximize

$$\arg\max_{\|\xi\|=1, \langle \xi, \xi_j \rangle = 0 \text{ for } j=1, \dots, k-1} \left[\text{Var} \left(\int_a^b (X(t) - \mu(t)) \xi(t) dt \right) \right].$$

Equations (2.36) and (2.37) confirm that this maximum is uniquely attained when $\xi(t) = \xi_k(t)$, the eigenfunction associated with the k -th largest eigenvalue λ_k .

Consequently, solving the integral eigenvalue problem yields an orthonormal basis of functions where the variance of the k -th FPC score is exactly λ_k . The core principle of FPCA is to project the infinite-dimensional curves onto a finite-dimensional subspace spanned by the first m dominant components. The truncated approximation is given by

$$X^{(m)}(t) = \mu(t) + \sum_{i=1}^m \zeta_i \xi_i(t),$$

where the expected total squared truncation error is captured entirely by the residual eigenvalues:

$$\mathbb{E} \left(\int_a^b (X(t) - X^{(m)}(t))^2 dt \right) = \sum_{j=m+1}^{\infty} \lambda_j \xrightarrow{m \rightarrow \infty} 0.$$

2.9.1. FPCA IN PRACTICE

The derivation above assumes that the true covariance kernel of the stochastic process $X(t) \in L^2(\Omega \times [a, b])$ is known. In practice, this kernel must be estimated from a sample of n observed functional curves, denoted $X_1(t), \dots, X_n(t)$. Throughout this section, we assume these curves have been centered by subtracting the empirical sample mean function, $\bar{X}_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t)$. The sample covariance kernel $\hat{\Gamma}_n(s, t)$ is then defined as

$$\hat{\Gamma}_n(s, t) := \frac{1}{n-1} \sum_{i=1}^n X_i(s) X_i(t).$$

Analogous to the multivariate setting discussed in Section 2.8.1, the k -th sample functional principal component weight function $\hat{\xi}_k(t)$ is chosen to be the solution that maximizes the empirical sample variance of the projections

$$\frac{1}{n-1} \sum_{i=1}^n \left(\int_a^b \xi(t) X_i(t) dt \right)^2 \quad (2.38)$$

subject to the normalization constraint $\int_a^b \xi^2(t) dt = 1$ and the mutual orthogonality constraints $\int_a^b \xi(t) \hat{\xi}_l(t) dt = 0$ for all $l < k$.

By applying the spectral decomposition theorem to this empirical framework, the sample covariance operator $T_{\hat{\Gamma}_n} : L^2([a, b]) \rightarrow L^2([a, b])$, which is induced by the kernel $\hat{\Gamma}_n(s, t)$, can be expressed via its eigenfunctions. Theorem 2.9.2

$$\hat{\Gamma}_n(s, t) = \sum_{i=1}^{\infty} \hat{\lambda}_i \hat{\xi}_i(s) \hat{\xi}_i(t).$$

Because the sample covariance operator is symmetric and positive semi-definite, its corresponding estimated eigenvalues are non-negative and can be ordered such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$. The estimated weight functions $\hat{\xi}_i$ are precisely the solutions to the empirical eigenvalue problem

$$(T_{\hat{\Gamma}_n} \hat{\xi}_i)(t) = \hat{\lambda}_i \hat{\xi}_i(t) \quad \forall i \geq 1. \quad (2.39)$$

As noted by Ramsay and Silverman [10], an exact analytical solution to Equation (2.39) is generally unattainable because functional data are collected at discrete time grids rather than continuously. Instead, they derived computational methods to approximate its solution.

The most straightforward approach would be to discretize the continuous variable t into equally spaced points $a = t_1 < \dots < t_p = b$ to create an $(n \times p)$ matrix $\mathbf{X}$ where row k contains the discrete evaluations of the k -th centered sample curve, $[X_k(t_1), \dots, X_k(t_p)]$.

To connect this matrix structure back to the functional operator setting, we evaluate the left-hand side of Equation (2.39) at a specific grid point t_j . Approximating the integral via a right Riemann sum with equal step size $\Delta t = \frac{b-a}{p}$ yields:

$$\begin{aligned} (T_{\hat{\Gamma}_n} \xi)(t_j) &= \int_a^b \hat{\Gamma}_n(t_j, s) \xi(s) ds \\ &\approx \frac{b-a}{p} \sum_{i=1}^p \hat{\Gamma}_n(t_j, t_i) \xi(t_i) \\ &= \frac{b-a}{(n-1)p} \sum_{i=1}^p \sum_{k=1}^n X_k(t_j) X_k(t_i) \xi(t_i), \end{aligned}$$

where we have substituted the empirical covariance function $\hat{\Gamma}_n(t_j, t_i) = \frac{1}{n-1} \sum_{k=1}^n X_k(t_j) X_k(t_i)$. Collecting these evaluations for all $j = 1, \dots, p$ into vector notation allows us to express the functional eigenvalue problem compactly as the matrix system and rewrite Equation (2.39) as

$$\frac{b-a}{(n-1)p} \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\lambda}} \hat{\boldsymbol{\xi}}, \quad (2.40)$$

where $\hat{\boldsymbol{\xi}} = [\hat{\xi}(t_1), \dots, \hat{\xi}(t_p)]^\top \in \mathbb{R}^p$. Introduce $\mathbf{u} = \sqrt{\frac{b-a}{p}} \hat{\boldsymbol{\xi}}$ and $\rho = \frac{p}{b-a} \hat{\boldsymbol{\lambda}}$ such that Equation (2.40) becomes

$$\frac{1}{n-1} \mathbf{X}^\top \mathbf{X} \mathbf{u} = \rho \mathbf{u}.$$

This is now a standard eigenvalue problem in which $\mathbf{u}^\top \mathbf{u} = 1$ which can be solved. The corresponding eigenvectors $\hat{\xi}$ and eigenvectors $\hat{\lambda}$ can then be found from $\mathbf{u}$ and ρ respectively. The final eigenfunctions $\xi(t)$ should then be approximated by using a suitable interpolation method on $\hat{\xi}$.

A more sophisticated approach is based on basis functions. In this framework, we suppose each curve $X_i(t)$ can be expressed as a linear combination m of basis functions $\{\phi_1(t), \dots, \phi_m(t)\}$. This means at almost any $t \in [a, b]$ we write

$$X_i(t) = \sum_{k=1}^m c_{i,k} \phi_k(t) \quad \forall i = 1, 2, \dots, n$$

or more compactly as

$$\mathbf{X}(t) = \mathbf{C}\Phi(t), \quad (2.41)$$

where $\mathbf{C} \in \mathbb{R}^{n \times m}$ with $(\mathbf{C})_{i,j} = c_{i,j}$, $\mathbf{X}(t) = [X_1(t), \dots, X_n(t)]^\top$ and $\Phi(t) = [\phi_1(t), \dots, \phi_m(t)]^\top$. With these new variables we can express the sample covariance for all $s, t \in [a, b]$ as

$$\hat{\Gamma}_n(s, t) = \frac{1}{n-1} \mathbf{X}^\top(s) \mathbf{X}(t) = \frac{1}{n-1} \Phi^\top(s) \mathbf{C}^\top \mathbf{C} \Phi(t).$$

Now suppose that also $\hat{\xi}(t)$ can be expressed in terms of basis functions in the sense

$$\hat{\xi}(t) = \sum_{k=1}^m b_k \phi_k(t) = \Phi^\top(t) \mathbf{b} \quad (2.42)$$

with $\mathbf{b} \in \mathbb{R}^m$, we have that

$$\int \hat{\Gamma}_n(s, t) \hat{\xi}(t) dt = \frac{1}{n-1} \Phi^\top(s) \mathbf{C}^\top \mathbf{C} \left(\int_a^b \Phi(t) \Phi(t)^\top dt \right) \mathbf{b} = \frac{1}{n-1} \Phi^\top(s) \mathbf{C}^\top \mathbf{C} \mathbf{W} \mathbf{b},$$

where the matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$ is defined as $(\mathbf{W})_{i,j} = \int_a^b \phi_i(t) \phi_j(t) dt$. Therefore, for any $j = 1, 2, \dots, m$ Equation (2.39) can be rewritten as

$$\frac{1}{n-1} \Phi^\top(s) \mathbf{C}^\top \mathbf{C} \mathbf{W} \mathbf{b}_j = \hat{\lambda}_j \Phi^\top(s) \mathbf{b}_j. \quad (2.43)$$

Since the set $\{\phi_i\}_{i=1}^m$ is a set of linearly independent functions and Equation (2.43) must hold for all s , it can be reduced to

$$\frac{1}{n-1} \mathbf{C}^\top \mathbf{C} \mathbf{W} \mathbf{b}_j = \hat{\lambda}_j \mathbf{b}_j.$$

Recall that we wanted the functions $\hat{\xi}(t)$ to be normalized, therefore, by Equation (2.42), we get

$$1 = \int \hat{\xi}_j^2(t) dt = \int \mathbf{b}_j^\top \Phi(t) \Phi(t)^\top \mathbf{b}_j dt = \mathbf{b}_j^\top \mathbf{W} \mathbf{b}_j.$$

Let us introduce a new variable $\mathbf{u}_j = \mathbf{W}^{\frac{1}{2}} \mathbf{b}_j$ ¹. Since $\mathbf{u}_j^\top \mathbf{u}_j = \mathbf{b}_j^\top \mathbf{W} \mathbf{b}_j = 1$, we can convert Equation (2.43) into a standard symmetric eigenproblem as

$$\frac{1}{n-1} \mathbf{W}^{\frac{1}{2}} \mathbf{C}^\top \mathbf{C} \mathbf{W}^{\frac{1}{2}} \mathbf{u}_j = \hat{\lambda}_j \mathbf{u}_j. \quad (2.44)$$

¹ $\mathbf{W}$ is the matrix $(\mathbf{W})_{i,j} = \int \phi_i(t) \phi_j(t) dt$. Hence it is symmetric and positive semidefinite. Since the basis functions $\{\phi_i\}_{i=1}^m$ are linearly independent, $\mathbf{W}$ is positive definite and $\mathbf{W}^{\frac{1}{2}}$ is well defined.

We can solve this standard eigenproblem for $\hat{\lambda}_j$ and $\mathbf{u}_j$. Then we can recover the function $\hat{\xi}_j$ by $\hat{\xi}_j(t) = \Phi(t)^\top \mathbf{W}^{-\frac{1}{2}} \mathbf{u}_j$.

Remark 3. *Ideally, we would like to choose our set of basis functions as an orthonormal basis. Indeed, then $\int \phi_i(t) \phi_j(t) dt = \delta_{i,j}$ and so $\mathbf{W} = \mathbf{I}_m$. In that case Equation (2.44) reduces to*

$$\frac{1}{n-1} \mathbf{C}^\top \mathbf{C} \mathbf{b}_j = \hat{\lambda}_j \mathbf{b}_j$$

which simplifies the problem.

2.9.2. REGULARIZED FPCA

A common challenge in FPCA is the presence of noisy principal components making interpretation harder. This roughness suggests that incorporating smoothing is necessary during estimation. To address this, this section introduces *regularized FPCA*, which integrates the smoothing process directly into the principal component analysis itself. Our approach follows the derivation established by Ramsay and Silverman [10].

Our objective is twofold: we aim to find the principal components that both (i) capture as much variance in the data as possible and (ii) are sufficiently smooth. To enforce smoothness, we use the roughness penalty

$$\text{RP}(f) := \int \left(\frac{d^2 f}{dx^2} \right)^2 dx.$$

Unlike normal FPCA, where our goal was to maximize $\text{Var}(\int \xi(t) X_i(t) dt)$ under the constraint that $\|\xi\|^2$, we now also wish for $\text{RP}(\xi)$ to be small. To balance these two objectives, we introduce the penalized optimization problem as

$$\max_{\|\xi\|^2=1} \frac{\text{Var}(\int \xi(t) X_i(t) dt)}{\|\xi\|^2 + \eta \cdot \text{RP}(\xi)} \quad (2.45)$$

with $\eta \geq 0$ the usual smoothing parameter. Following the same steps as in the derivation of standard FPCA, we can obtain the regularized FPCA formulation.

Suppose again that $X_i(t)$ can be expressed as a linear combination of basis functions $\{\theta_1(t), \dots, \theta_m(t)\}$. This means that $X_i(t) = \mathbf{c}_i^\top \Theta(t)$, with $\mathbf{c}_i$ the vector of coefficients in the basis of $\{\theta_j\}$. Similarly, we assume that also the eigenfunctions can be written in this basis, $\xi(t) = \mathbf{b}^\top \Theta(t)$. In this framework, the numerator of Equation (2.45) can be expressed as

$$\int \xi(t) X_i(t) dt = \mathbf{b}^\top \left(\int \Theta(t) \Theta(t) dt \right) \mathbf{c}_i := \mathbf{b}^\top \mathbf{G} \mathbf{c}_i$$

with $\mathbf{G} \in \mathbb{R}^{m \times m}$ the matrix whose elements are $\int \theta_i \theta_j$. Hence, we have

$$\text{Var} \left(\int \xi(t) X_i(t) dt \right) = \text{Var}(\mathbf{b}^\top \mathbf{G} \mathbf{c}_i) := \mathbf{b}^\top \mathbf{G} \mathbf{V} \mathbf{G}^\top \mathbf{b}$$

where $\mathbf{V}$ is the covariance matrix of the vectors $\mathbf{c}_i$. With a similar derivation we can rewrite the denominator as

$$\|\xi\|^2 = \int \xi^2(t) dt = \mathbf{b}^\top \mathbf{G} \mathbf{b}$$

and

$$\eta \int \xi''(t) dt := \eta \mathbf{b}^\top \mathbf{K} \mathbf{b}$$

where $\mathbf{K}$ the matrix with elements $\int \theta_i''(t) \theta_j''(t) dt$. Combining this, Equation (2.45) becomes

$$\max_{\mathbf{b}} \frac{\mathbf{b}^\top \mathbf{G} \mathbf{V} \mathbf{G}^\top \mathbf{b}}{\mathbf{b}^\top \mathbf{G} \mathbf{b} + \eta \mathbf{b}^\top \mathbf{K} \mathbf{b}} = \max_{\mathbf{b}} \frac{\mathbf{b}^\top \mathbf{G} \mathbf{V} \mathbf{G}^\top \mathbf{b}}{\mathbf{b}^\top (\mathbf{G} + \eta \mathbf{K}) \mathbf{b}}. \quad (2.46)$$

From linear algebra (see also [38]), we know that the Equation (2.46) corresponds to the eigenvalue-equation

$$\mathbf{G} \mathbf{V} \mathbf{G}^\top = \lambda (\mathbf{G} + \eta \mathbf{K}) \mathbf{b}. \quad (2.47)$$

If we now let $\mathbf{G} + \eta \mathbf{K} = \mathbf{L}^\top \mathbf{L}$ by Cholesky and let $\mathbf{S} := \mathbf{L}^{-1}$ then Equation (2.47) becomes

$$(\mathbf{S} \mathbf{G} \mathbf{V} \mathbf{G}^\top \mathbf{S}^\top) (\mathbf{L}^\top \mathbf{b}) = \lambda \mathbf{L}^\top \mathbf{b}. \quad (2.48)$$

Since $\text{Cov}(\mathbf{S} \mathbf{G} \mathbf{c}_i) = \mathbf{S} \mathbf{G} \mathbf{V} \mathbf{G}^\top \mathbf{S}^\top$, solving Equation (2.48) is equivalent to applying normal FPCA to the vector $\mathbf{S} \mathbf{G} \mathbf{c}_i$, the original data smoothed by the smoothing operator $\mathbf{S}$. To find the vector of coefficients $\mathbf{b}$ we need to apply the smoothing operator $\mathbf{S}$ again on the resulting eigenvectors $\mathbf{u}$. This demonstrates that this method can be regarded as a 'half-smoothing' approach: first smooth the data by $\mathbf{S}$, then perform PCA, then smooth the principal components by operating on $\mathbf{S}$ again. The functions $\xi(t) = \mathbf{b}^\top \Theta(t)$ we then find are indeed a smoothed version of the eigenfunctions we find with normal FPCA.

2.10. ARMA-GARCH MODELS

When analyzing time series, observations frequently exhibit a high degree of temporal dependence. To mitigate the complexities introduced by this structure, it is advisable to apply a filtering procedure that removes both linear and conditional variance dependencies. Following the framework of [39], we base this filtering on the foundational concept of white noise.

Definition 2.10.1 (White noise, Definition 1.3 [39]). *The process (ϵ_t) is called weak white noise if for some $\sigma^2 > 0$:*

$$i) \mathbb{E}(\epsilon_t) = 0, \quad \forall t \in \mathbb{Z}$$

$$ii) \mathbb{E}(\epsilon_t^2) = \sigma^2, \quad \forall t \in \mathbb{Z}$$

$$iii) \text{Cov}(\epsilon_t, \epsilon_{t+h}) = 0, \quad \forall t, h \in \mathbb{Z}, h \neq 0.$$

If property (iii) is replaced by

$$iv) \text{ The variables } \epsilon_t \text{ and } \epsilon_{t+h} \text{ are i.i.d}$$

the process (ϵ_t) is called strong white noise.

The class of Autoregressive Moving Average (ARMA) models is the standard way for modeling and predicting second-order stationary processes².

Definition 2.10.2 (ARMA(p, q) process, Definition 1.4 [39]). A second-order stationary process $(X_t)_{t \in \mathbb{Z}}$ is called ARMA(p, q), where p and q are integers, if there exist coefficients $c, a_1, \dots, a_p, b_1, \dots, b_q$ such that

$$\forall t \in \mathbb{Z} \quad X_t + \sum_{i=1}^p a_i X_{t-i} = c + \epsilon_t + \sum_{i=1}^q b_i \epsilon_{t-i},$$

where $(\epsilon_t)_{t \in \mathbb{Z}}$ is the linear innovation process of $(X_t)_{t \in \mathbb{Z}}$, that is

$$\epsilon_t = X_t - \mathbb{E}(X_t | \mathcal{H}_{t-1})$$

where $\mathcal{H}_{t-1}$ denotes the Hilbert space generated by $X_{t-1}, X_{t-2}, \dots$.

The ARMA model captures linear serial dependence through autoregressive and moving-average components, while the remaining innovations (ϵ_t) are a weak white noise. However, financial return series typically exhibit an additional stylized feature known as *volatility clustering*, where periods of high volatility are followed by periods of relative calmness. This phenomenon implies that the variance of the innovations changes over time, violating the constant variance assumption of white noise. More precisely, $\text{Var}(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) \neq \text{constant}$. Assuming that these innovations can be decomposed multiplicatively, we define the GARCH framework.

Definition 2.10.3 (GARCH(r, s) process, Definition 2.2 [39]). Let $(\eta_t)_{t \in \mathbb{Z}}$ be an i.i.d. sequence satisfying

$$\mathbb{E}(\eta_t) = 0, \quad \text{Var}(\eta_t) = 1.$$

The process $(\epsilon_t)_{t \in \mathbb{Z}}$ is called GARCH(r, s) if

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^s \omega_i \epsilon_{t-i}^2 + \sum_{j=1}^r \gamma_j \sigma_{t-j}^2 \end{cases}$$

where $\omega_0 > 0$, $\omega_i \geq 0$, and $\gamma_j \geq 0$ are constants. The variable $\sigma_t > 0$ represents the volatility of the innovation ϵ_t .

To simultaneously capture the conditional mean dynamics and the conditional variance structure of our financial time series, we unify Definitions 2.10.2 and 2.10.3 into a joint ARMA(p, q)-GARCH(n, m) model:

$$\begin{cases} X_t = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i} + \epsilon_t + \sum_{j=1}^q \beta_j \epsilon_{t-j} \\ \epsilon_t = \sigma_t \eta_t, \quad \eta_t \stackrel{i.i.d.}{\sim} (0, 1) \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^n \omega_i \epsilon_{t-i}^2 + \sum_{j=1}^m \gamma_j \sigma_{t-j}^2. \end{cases} \quad (2.49)$$

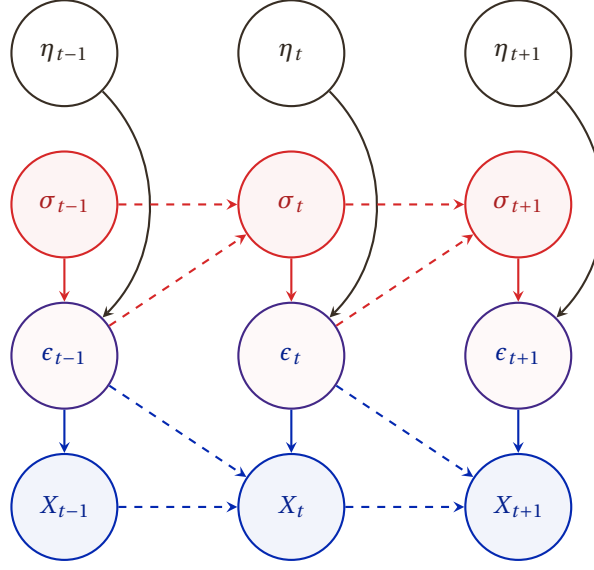


Figure 2.1: Schematic representation of the ARMA(1,1)-GARCH(1,1) model dynamics. The blue lines represent the time series dynamics and the red lines represent the volatility dynamics. The dashed lines indicate the influence by a previous timepoint.

Here, ϵ_t is known as the raw *innovations*, η_t as the *standardized innovations* and σ_t tracks the time-varying *volatility* of ϵ_t and X_t . The working on the ARMA(1,1)-GARCH(1,1) model is visualized in Figure 2.1.

For the dependence analysis developed in the remainder of this thesis, the standardized innovations $(\eta_t)_{t \in \mathbb{Z}}$ are of primary interest. After estimating and filtering the ARMA-GARCH model, the resulting *standardized residuals* $\hat{\eta}_t$ behave approximately as an i.i.d. sequence. Consequently, the temporal dependence structure is largely removed, allowing the subsequent analysis to focus exclusively on the contemporaneous dependence between the underlying financial assets.

²A process $(X_t)_{t \in \mathbb{Z}}$ is second-order stationary if $\forall t, h \in \mathbb{Z}$: (i) $\mathbb{E}(X_t^2) < \infty$, (ii) $\mathbb{E}(X_t) = m$, and (iii) $\text{Cov}(X_t, X_{t+h}) = \gamma_X(h)$, where $\gamma_X(h)$ is the autocovariance function.

3

STATISTICAL MODEL

In this chapter, we introduce the statistical framework which forms the basis of this thesis. Our primary objective is to study the dependence structure between components of a multivariate random vector, conditional on an additional set of covariates. In particular, we focus on the conditional Kendall's tau and analyse their variation as functions of the conditioning variable. We operate in a classical random design setting, where observations are assumed to be independent and identically distributed.

3.1. FPCA FOR THE TRUE CONDITIONAL KENDALL'S TAU

Let $(\mathbf{X}, \mathbf{Z})$ be a random vector taking values in $\mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $P_{\mathbf{X}, \mathbf{Z}}$. We observe an i.i.d. sample $\{(\mathbf{X}_i, \mathbf{Z}_i)\}_{i=1}^n$, where each $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p}) \in \mathbb{R}^p$ and $\mathbf{Z}_i \in \mathbb{R}^d$. We denote the support of $\mathbf{Z}$ by $\mathcal{Z} \subset \mathbb{R}^d$ and assume throughout this thesis that $\mathcal{Z}$ is compact. This assumption ensures, in particular, that integrals over $\mathcal{Z}$ are well-defined and finite.

For notational convenience, we collect the observations in matrix form as

$$\mathbf{T} = \left[\begin{array}{c|c} \mathbf{X}_1^\top & \mathbf{Z}_1^\top \\ \mathbf{X}_2^\top & \mathbf{Z}_2^\top \\ \vdots & \vdots \\ \mathbf{X}_n^\top & \mathbf{Z}_n^\top \end{array} \right].$$

Based on $\mathbf{T}$, we can compute the estimators $\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ of the conditional Kendall's tau $\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ for each pair (j_1, j_2) such that $1 \leq j_1 < j_2 \leq p$, using Equation (2.16). Throughout this work, we require that the conditional Kendall's tau satisfies basic structural regularities. To that end, the following assumption is maintained globally throughout this thesis and is implicitly assumed to hold in all subsequent lemmas, propositions, and theorems.

Assumption. *For any $1 \leq j_1 < j_2 \leq p$, the mapping $\mathbf{z} \mapsto \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ is a measurable function on the compact domain $\mathcal{Z}$.*

Our aim is to analyze the variation of these conditional Kendall's tau curves using *functional principal component analysis* (FPCA). Following the framework detailed in Section 2.9, we begin by defining the *true sample mean function* $\mu_p(\mathbf{z})$ and the *true sample covariance function* $\Gamma_p(\mathbf{z}, \mathbf{z}')$ across the pairs of variables:

$$\mu_p(\mathbf{z}) := \frac{1}{\binom{p}{2}} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} \quad (3.1)$$

$$\Gamma_p(\mathbf{z}, \mathbf{z}') := \frac{1}{\binom{p}{2} - 1} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})) (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}'} - \mu_p(\mathbf{z}')). \quad (3.2)$$

To apply the Karhunen-Loève expansion (Theorem 2.9.2), we require Γ_p to define a Hilbert-Schmidt integral operator. Since Kendall's tau is inherently bounded between -1 and 1 , and the domain $\mathcal{Z}$ is compact, we directly obtain the following result.

Lemma 3.1.1. *We have $\tau_{j_1, j_2 | \mathbf{Z}=\cdot} \in L^2(\mathcal{Z})$ and the covariance kernel satisfies*

$$\iint_{\mathcal{Z} \times \mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}')^2 d\mathbf{z} d\mathbf{z}' < \infty.$$

By construction, Γ_p is a symmetric and positive semi-definite kernel. While symmetry is immediate from the definition, positive semi-definiteness can be verified by noting that for any finite collection of points $\{\mathbf{z}_1, \dots, \mathbf{z}_m\} \subset \mathcal{Z}$ and real coefficients $c_1, \dots, c_m$, we have

$$\begin{aligned} \sum_{a=1}^m \sum_{b=1}^m c_a c_b \Gamma_p(\mathbf{z}_a, \mathbf{z}_b) &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \sum_{a=1}^m \sum_{b=1}^m c_a c_b g_{j_1, j_2}(\mathbf{z}_a) g_{j_1, j_2}(\mathbf{z}_b) \\ &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \left(\sum_{a=1}^m c_a g_{j_1, j_2}(\mathbf{z}_a) \right) \left(\sum_{b=1}^m c_b g_{j_1, j_2}(\mathbf{z}_b) \right) \\ &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \left(\sum_{a=1}^m c_a g_{j_1, j_2}(\mathbf{z}_a) \right)^2 \\ &\geq 0, \end{aligned}$$

where $g_{j_1, j_2}(\mathbf{z}) := \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})$. It follows from Definition 2.6.15 that Γ_p is a valid positive semi-definite kernel.

Let $T_{\Gamma_p} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ be the linear integral operator associated with the covariance kernel Γ_p , defined for any $f \in L^2(\mathcal{Z})$ as

$$(T_{\Gamma_p} f)(\mathbf{z}) = \int_{\mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'. \quad (3.3)$$

By applying Lemmas 2.6.14 and 2.6.16, it follows that T_{Γ_p} is a compact, self-adjoint, and positive semi-definite Hilbert-Schmidt integral operator. As a consequence, Theorem 2.6.17 establishes the spectral decomposition of Γ_p

$$\Gamma_p(\mathbf{z}, \mathbf{z}') = \sum_{i=1}^{\infty} \lambda_i \xi_i(\mathbf{z}) \xi_i(\mathbf{z}')$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the ordered eigenvalues, and $\{\xi_i\}_{i=1}^\infty \subset L^2(\mathcal{Z})$ is the corresponding set of orthonormal eigenfunctions. This series converges in the $L^2(\mathcal{Z} \times \mathcal{Z})$ sense. These eigenfunctions form an orthonormal basis of $L^2(\mathcal{Z})$, and we refer to them as the *functional principal components*. Consequently, every centred conditional Kendall's tau function admits the expansion

$$\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} = \mu_p(\mathbf{z}) + \sum_{i=1}^{\infty} \zeta_{j_1, j_2, i} \xi_i(\mathbf{z}),$$

with ζ the *functional principal component scores*

$$\zeta_{j_1, j_2, i} = \langle \tau_{j_1, j_2 | \mathbf{Z}=\cdot} - \mu_p, \xi_i \rangle_{L^2} = \int_{\mathcal{Z}} (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})) \xi_i(\mathbf{z}) d\mathbf{z}. \quad (3.4)$$

Remark 4. The functional principal components ξ_k possess a classical optimization interpretation: they sequentially maximize the empirical variance of the conditional dependence curves across the $\binom{p}{2}$ pairs.

To see this explicitly, let $\phi \in L^2(\mathcal{Z})$ be an arbitrary directional function normalized such that $\|\phi\|_{L^2} = 1$. We can project each centred conditional Kendall's tau function, $g_{j_1, j_2}(\mathbf{z}) := \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})$, onto ϕ via the standard inner product to obtain a projection score for each pair:

$$\zeta_{j_1, j_2}(\phi) = \langle g_{j_1, j_2}, \phi \rangle_{L^2} = \int_{\mathcal{Z}} g_{j_1, j_2}(\mathbf{z}) \phi(\mathbf{z}) d\mathbf{z}.$$

By construction, the sample mean of these scores across all pairs is zero, since $\sum_{1 \leq j_1 < j_2 \leq p} (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})) = 0$. Consequently, the sample variance of these projection scores across the pairs is given by

$$\text{Var}(\phi) := \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \langle g_{j_1, j_2}, \phi \rangle_{L^2}^2. \quad (3.5)$$

We can establish a direct link between this variance and the covariance operator T_{Γ_p} . By substituting the definition of the kernel Γ_p into the operator equation, we observe that for any $f \in L^2(\mathcal{Z})$:

$$\begin{aligned} (T_{\Gamma_p} f)(\mathbf{z}) &= \int_{\mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}' \\ &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} g_{j_1, j_2}(\mathbf{z}) \int_{\mathcal{Z}} g_{j_1, j_2}(\mathbf{z}') f(\mathbf{z}') d\mathbf{z}' \\ &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \langle g_{j_1, j_2}, f \rangle_{L^2} g_{j_1, j_2}(\mathbf{z}). \end{aligned}$$

Taking the inner product of $T_{\Gamma_p} \phi$ with ϕ yields the quadratic form:

$$\langle T_{\Gamma_p} \phi, \phi \rangle_{L^2} = \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} \langle g_{j_1, j_2}, \phi \rangle_{L^2}^2 = \text{Var}(\phi).$$

By Theorem 2.6.34, the largest eigenvalue λ_1 and its corresponding eigenfunction ξ_1 solve the maximization problem:

$$\lambda_1 = \max_{\|\phi\|_{L^2}=1} \langle T_{\Gamma_p} \phi, \phi \rangle_{L^2},$$

and inductively, the subsequent components solve the constrained optimization problems:

$$\lambda_k = \max_{\substack{\|\phi\|_{L^2}=1 \\ \phi \perp \xi_1, \dots, \xi_{k-1}}} \langle T_{\Gamma_p} \phi, \phi \rangle_{L^2}.$$

Therefore, each eigenfunction ξ_k is precisely the direction that maximizes the remaining sample variance across the pairs, and the associated eigenvalue λ_k quantifies the magnitude of this variance.

3

3.2. FPCA FOR THE ESTIMATED CONDITIONAL KENDALL'S TAU

The practical challenge with the framework outlined above is that the true conditional Kendall's tau curves are latent and cannot be directly observed. Instead, we must estimate them from the data using Equation (2.16). Consequently, we define the *estimated sample mean function* $\hat{\mu}_p(\mathbf{z})$ and the *estimated sample covariance function* $\hat{\Gamma}_p(\mathbf{z}, \mathbf{z}')$ across the pairs of variables analogously:

$$\hat{\mu}_p(\mathbf{z}) := \frac{1}{\binom{p}{2}} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p \hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}, \quad (3.6)$$

$$\hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') := \frac{1}{\binom{p}{2} - 1} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p (\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \hat{\mu}_p(\mathbf{z})) (\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}'} - \hat{\mu}_p(\mathbf{z}')). \quad (3.7)$$

In classical functional data analysis, the covariance kernel $\Gamma(\mathbf{z}, \mathbf{z}') = \text{Cov}(X(\mathbf{z}), X(\mathbf{z}'))$ is defined in terms of an underlying random process and induces a deterministic Hilbert-Schmidt operator. Similarly, the population kernel Γ_p defined in Equation (3.2) is non-random because it is constructed from the true conditional Kendall's taus.

In contrast, the empirical kernel defined in (3.7) is inherently data-driven and random. More precisely, because $\hat{\Gamma}_p$ is constructed from estimated curves $\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$, it introduces an additional layer of statistical variation compared to a classical empirical covariance operator. In the classical setting, randomness enters solely through averaging observed independent realizations of a random element in a Hilbert space. Here, we do not observe functional realizations, instead, the quantities being averaged are themselves non-parametric estimators of latent dependence curves. The resulting sample randomness reflects both the underlying structural differences across the $\binom{p}{2}$ pairs and the pointwise estimation errors of the CKT estimators. Consequently, $\hat{\Gamma}_p$ cannot be framed as a standard empirical covariance operator of i.i.d. functional data.

To ensure that this data-driven framework preserves the spectral properties of the true operator, we require $\hat{\Gamma}_p$ to be a Hilbert-Schmidt operator. Since the estimators for Kendall's tau are bounded, we directly obtain the following result.

Corollary 3.2.1. *The estimated covariance kernel satisfies*

$$\iint_{\mathcal{Z} \times \mathcal{Z}} \hat{\Gamma}_p(\mathbf{z}, \mathbf{z}')^2 d\mathbf{z} d\mathbf{z}' < \infty, \text{ almost surely.}$$

Let $T_{\hat{\Gamma}_p} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ be the empirical linear integral operator associated with the estimated kernel $\hat{\Gamma}_p$. We define it for any $f \in L^2(\mathcal{Z})$ as

$$(T_{\hat{\Gamma}_p} f)(\mathbf{z}) = \int_{\mathcal{Z}} \hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'. \quad (3.8)$$

Since $\hat{\Gamma}_p \in L^2(\mathcal{Z} \times \mathcal{Z})$, $T_{\hat{\Gamma}_p}$ is a Hilbert-Schmidt operator and therefore compact. Moreover, because $\hat{\Gamma}_p$ is symmetric and positive semi-definite, $T_{\hat{\Gamma}_p}$ is a compact, self-adjoint, positive semidefinite operator. Theorem 2.6.17 now implies

$$\hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') = \sum_{i=1}^{\infty} \hat{\lambda}_i \hat{\xi}_i(\mathbf{z}) \hat{\xi}_i(\mathbf{z}'),$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$ are the sample eigenvalues corresponding to the orthonormal empirical eigenfunctions $\{\hat{\xi}_i\}_{i=1}^{\infty} \subset L^2(\mathcal{Z})$. The eigenfunctions satisfy the operator eigenvalue problem

$$(T_{\hat{\Gamma}_p} \hat{\xi}_i)(\mathbf{z}) = \hat{\lambda}_i \hat{\xi}_i(\mathbf{z}) \quad \text{for all } i \geq 1.$$

Accordingly, since $\{\hat{\xi}_i\}_{i=1}^{\infty}$ forms an orthonormal basis of $L^2(\mathcal{Z})$, each estimated conditional Kendall's tau function admits the expansion

$$\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} = \hat{\mu}_p(\mathbf{z}) + \sum_{i=1}^{\infty} \hat{\zeta}_{j_1, j_2, i} \hat{\xi}_i(\mathbf{z}),$$

where the estimated principal component scores $\hat{\zeta}_{j_1, j_2, i}$ are computed via:

$$\hat{\zeta}_{j_1, j_2, i} = \langle \hat{\tau}_{j_1, j_2 | \mathbf{Z}=\cdot} - \hat{\mu}_p, \hat{\xi}_i \rangle_{L^2(\mathcal{Z})} = \int_{\mathcal{Z}} (\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \hat{\mu}_p(\mathbf{z})) \hat{\xi}_i(\mathbf{z}) d\mathbf{z}.$$

The primary objective of the subsequent chapters of this thesis is to construct the asymptotic framework necessary to prove that these empirical estimators, consistently converge to their true population counterparts as the sample size $n \rightarrow \infty$. But first, we will discuss the simplifying assumption in this context as introduced in Section 2.3.

3.3. THE SIMPLIFYING ASSUMPTION AND ITS SPECTRAL CHARACTERIZATION

A common framework in multivariate conditional dependence modelling is the so-called *simplifying assumption*, which posits that the conditional dependence structure does not vary with the conditioning variables. In our framework, this translates to assuming that the conditional Kendall's tau curves are constant across the entire support of the covariate. Specifically, if the simplifying assumption is valid for pair (j_1, j_2) , this means $\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} = \tau_{j_1, j_2, \text{simp}}$ for any $\mathbf{z} \in \mathcal{Z}$.

If the simplifying assumption holds for any pair (j_1, j_2) , the covariance kernel $\Gamma_p(\mathbf{z}, \mathbf{z}')$ becomes invariant to both arguments $\mathbf{z}$ and $\mathbf{z}'$. For any $(\mathbf{z}, \mathbf{z}') \in \mathcal{Z} \times \mathcal{Z}$,

$$\begin{aligned} \Gamma_p(\mathbf{z}, \mathbf{z}') &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \mu_p(\mathbf{z})) (\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}'} - \mu_p(\mathbf{z}')) \\ &= \frac{1}{\binom{p}{2} - 1} \sum_{1 \leq j_1 < j_2 \leq p} (\tau_{j_1, j_2, \text{simp}} - \mu_{p, \text{simp}})^2 \\ &:= \Gamma_{p, \text{simp}}. \end{aligned}$$

This degenerated structure of the covariance kernel has implications for the eigenvalues and eigenfunctions of the associated integral operator T_{Γ_p} . We formalize this result in the following lemma.

Lemma 3.3.1. *If the simplifying assumption holds for any pair (j_1, j_2) with $1 \leq j_1 < j_2 \leq p$, the integral operator T_{Γ_p} is a rank-one operator whose spectrum is characterized as follows:*

1. *The leading eigenvalue is given by $\lambda_1 = \Gamma_{p, \text{simp}} \cdot \text{Leb}(\mathcal{Z})$, and its corresponding normalized eigenfunction is the constant function $\xi_1(\mathbf{z}) = (\text{Leb}(\mathcal{Z}))^{-1/2}$.*
2. *All subsequent eigenvalues are identically zero, i.e., $\lambda_k = 0$ for all $k \geq 2$. The corresponding eigenspace is the infinite-dimensional subspace of $L^2(\mathcal{Z})$ consisting of all zero-mean functions*

$$\left\{ f \in L^2(\mathcal{Z}) : \int_{\mathcal{Z}} f(\mathbf{z}) d\mathbf{z} = 0 \right\}.$$

Proof. Let $f \in L^2(\mathcal{Z})$. Under the simplifying assumption, acting on T_{Γ_p} reduces to

$$(T_{\Gamma_p} f)(\mathbf{z}) = \int_{\mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}' = \Gamma_{p, \text{simp}} \int_{\mathcal{Z}} f(\mathbf{z}) d\mathbf{z}.$$

It can easily be checked that the constant function is an eigenfunction with eigenvalue $\lambda_1 = \Gamma_{p, \text{simp}} \cdot \text{Leb}(\mathcal{Z})$. Indeed

$$(T_{\Gamma_p} c)(\mathbf{z}) = \Gamma_{p, \text{simp}} \int_{\mathcal{Z}} c d\mathbf{z} = \Gamma_{p, \text{simp}} \cdot \text{Leb}(\mathcal{Z}) \cdot c.$$

Thus, constant functions are eigenfunctions with eigenvalue $\lambda_1 = \Gamma_{p, \text{simp}} \text{Leb}(\mathcal{Z})$. Normalizing via $\|\xi_1\|_{L^2} = 1$ yields $\xi_1(\mathbf{z}) = (\text{Leb}(\mathcal{Z}))^{-1/2}$.

Conversely, if $f \in \{f \in L^2(\mathcal{Z}) : \int_{\mathcal{Z}} f(\mathbf{z}) d\mathbf{z} = 0\}$, then $(T_{\Gamma_p} f)(\mathbf{z}) = 0$. Since this set is the orthogonal complement of the space of constant functions in $L^2(\mathcal{Z})$, this accounts for the remainder of the spectrum. It then follows that $\lambda_k = 0$ for all $k \geq 2$. $\square$

Lemma 3.3.1 provides a powerful framework for evaluating the validity of the simplifying assumption across all pairs of variables. Rather than assessing each pair individually, the simplifying assumption can be characterized directly through the eigenvalues of the covariance operator. Specifically, we observe that the simplifying assumption holds if

$$\lambda_2 = 0.$$

Because the eigenvalues are sorted by construction in descending order, the condition $\lambda_2 = 0$ automatically implies that $\lambda_k = 0$ for all $k \geq 2$. Conversely, if the simplifying assumption is violated for at least one pair of variables, the kernel $\Gamma_p(\mathbf{z}, \mathbf{z}')$ will no longer be constant, and consequently, λ_2 must be strictly positive.

Remark 5. *In practise, when the true curves are unknown and replaced by their estimators, the sample eigenvalue $\hat{\lambda}_2$ will rarely be exactly zero due to sampling and estimation error. Therefore, we treat this relationship more as an empirical rule-of-thumb. As demonstrated in Chapter 7, the size of the violation of the simplifying assumption can be understood by looking at the value of $\hat{\lambda}_2$ or more specifically, at the proportion of the residual variance.*

4

FPCA ON THE ESTIMATED CONDITIONAL KENDALL'S TAU

In this chapter, we construct the mathematical framework necessary to establish the asymptotic consistency of the functional estimators introduced in Chapter 3. Throughout this analysis, we operate in the idealized functional setting, assuming that the estimated conditional Kendall's tau curves $\hat{\tau}_{j_1, j_2 | \mathbf{Z} = \cdot}$ are available over the entire compact domain $\mathcal{Z}$. This allows us to focus entirely on the statistical variations driven by the sample size $n \rightarrow \infty$, independent of any specific discretization or numerical approximation schemes. This later framework will in fact be analysed in Chapter 5. We begin by establishing several foundational results in operator theory that will serve as essential tools throughout the remainder of this thesis.

4.1. AUXILIARY RESULTS IN OPERATOR THEORY

We derive three lemmas that will be central to establishing consistency results. To this end, recall that the resolvent operator is defined as $R(\lambda) = R(\lambda, A) := (A - \lambda I)^{-1}$, see also Definition 2.6.18.

Lemma 4.1.1. *Let H be a Hilbert space and let $A : H \rightarrow H$ be a (deterministic) bounded self-adjoint linear operator. For $\lambda \in \rho(A)$ (Definition 2.6.18) we have,*

$$\|R(\lambda)\|_{op} = \frac{1}{\text{dist}(\lambda, \sigma(A))}.$$

Proof. We start by noting that for any bounded normal operator T we have that the operator norm and the spectral radius are equal (p.60 of [21]), i.e.

$$\|T\|_{op} = \sup_{\lambda \in \sigma(T)} |\lambda|.$$

The resolvent operator of A is defined as $R(\lambda) = (A - \lambda I)^{-1}$. By construction, $R(\lambda)$ is bounded, and since A is self-adjoint (and thus normal), $R(\lambda)$ is also normal. In addition, the spectral mapping theorem ensures that the spectrum of $R(\lambda)$ satisfies

$$\sigma(R(\lambda)) = \frac{1}{\sigma(A) - \lambda} = \left\{ \frac{1}{\mu - \lambda} \mid \mu \in \sigma(A) \right\}.$$

Hence

$$\|R(\lambda)\|_{op} = \sup_{\zeta \in \sigma(R(\lambda))} |\zeta| = \sup_{\mu \in \sigma(A)} \frac{1}{|\mu - \lambda|} = \frac{1}{\text{dist}(\lambda, \sigma(A))}$$

which proves the result. $\square$

4

Having established how the norm of the resolvent behaves relative to the geometry of the spectrum, we now examine how contour integrals of the resolvent can isolate specific spectral spaces. The next lemma demonstrates that the Riesz projection operator (Definition 2.6.21) associated with a simple eigenvalue directly recovers its corresponding eigenspace.

Lemma 4.1.2. *Let H be a Hilbert space and let $A : H \rightarrow H$ be a (deterministic) closed linear operator with a simple eigenvalue λ_0 . Consider the Riesz projection operator P_{λ_0}*

$$P_{\lambda_0} = -\frac{1}{2\pi i} \int_C (A - zI)^{-1} dz$$

with C a simple closed contour separating λ_0 from the rest of the spectrum of A . Then

$$\text{Ran}(P_{\lambda_0}) = E_{\lambda_0}$$

with E_{λ_0} the eigenspace of λ_0 .

Proof. We will first prove the inclusion $E_{\lambda_0} \subseteq \text{Ran}(P_{\lambda_0})$. To do so, take $x \in E_{\lambda_0}$ as the eigenvector corresponding to λ_0 , meaning $Ax = \lambda_0 x$. It holds that then

$$(A - zI)^{-1} x = \frac{1}{\lambda_0 - z} x.$$

We can next obtain

$$\begin{aligned} P_{\lambda_0} x &= -\frac{1}{2\pi i} \int_C (A - zI)^{-1} x dz \\ &= -\frac{1}{2\pi i} \int_C \frac{1}{\lambda_0 - z} dz \cdot x \\ &= x. \end{aligned}$$

In the last equality, we used the Cauchy Residue Theorem (p. 233 of [40]) to evaluate

$$\int_C \frac{1}{\lambda_0 - z} dz = -2\pi i.$$

So $x \in \text{Ran}(P_{\lambda_0})$ and thus $E_{\lambda_0} \subseteq \text{Ran}(P_{\lambda_0})$.

Now take $P_{\lambda_0}x = y \in \text{Ran}(P_{\lambda_0})$. To show $y \in E_{\lambda_0}$, it is sufficient to show that $(A - \lambda_0 I)P_{\lambda_0}x = 0$. Using Lemma 2.6.22 (ii) we can write

$$\begin{aligned} (A - \lambda_0 I)P_{\lambda_0} &= P_{\lambda_0}(A - \lambda_0 I) = -\frac{1}{2\pi i} \int_C (A - \lambda_0 I)(A - zI)^{-1} dz \\ &= -\frac{1}{2\pi i} \int_C (I - (\lambda_0 - z)(A - zI)^{-1}) dz \\ &= -\frac{1}{2\pi i} \int_C I dz + \frac{1}{2\pi i} \int_C (\lambda_0 - z)(A - zI)^{-1} dz. \end{aligned} \quad (4.1)$$

By the Cauchy-Goursat Theorem (p. 149 of [40]), it is clear that $\int_C I dz = 0$ since the identity map is analytic¹ everywhere. Recall that the map $z \mapsto (A - zI)^{-1}$ is analytic on $\rho(A)$ and therefore also inside and on the contour C , except possibly at $z = \lambda_0$. However, because λ_0 is a simple eigenvalue, writing out the Laurent series reveals that the map $z \mapsto (\lambda_0 - z)(A - zI)^{-1}$ has a removable singularity (no pole) at $z = \lambda_0$ and is therefore analytic inside and on the contour C . Applying the Cauchy-Goursat Theorem once more, proves

$$\int_C (\lambda_0 - z)(A - zI)^{-1} dz = 0.$$

and hence our operator equation simplifies to $(A - \lambda_0 I)P_{\lambda_0} = 0$. Consequently, $(A - \lambda_0 I)P_{\lambda_0}x = 0$, meaning $y \in E_{\lambda_0}$. This shows $\text{Ran}(P_{\lambda_0}) \subseteq E_{\lambda_0}$, completing the proof that $E_{\lambda_0} = \text{Ran}(P_{\lambda_0})$. $\square$

We have now established the key ingredients needed to formulate an operator-theoretic version of the Davis–Kahan theorem (Theorem 2.6.37). In particular, the following lemma shows how perturbations of compact self-adjoint operators can be translated into quantitative bounds on the distance between their associated eigenvectors.

Lemma 4.1.3. *Let $T : H \rightarrow H$ and $\widehat{T} : H \rightarrow H$ be two (deterministic) compact self-adjoint linear operators on a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ with their ordered eigenvalues $\{\lambda_i\}_{i \geq 1}$ and $\{\widehat{\lambda}_i\}_{i \geq 1}$ corresponding to the eigenvectors $\{\xi_i(\cdot)\}_{i \geq 1}$ and $\{\widehat{\xi}_i(\cdot)\}_{i \geq 1}$. Let $i \geq 1$ such that λ_i and $\widehat{\lambda}_i$ are simple eigenvalues w.r.t. the operators T and $\widehat{T}$ respectively. Define*

$$\delta := \min_{\mu \in \{\sigma(T) \setminus \lambda_i\}} \text{dist}(\mu, \lambda_i),$$

and assume that $\|T - \widehat{T}\|_{op} < \delta/4$. Then there exists a sequence of signs $s_i \in \{\pm 1\}$ such that

$$\|\widehat{\xi}_i - s_i \xi_i\|_H \leq \frac{4\|\widehat{T} - T\|_{op}}{\delta - 2\|\widehat{T} - T\|_{op}}.$$

¹A function f of the complex variable z is called analytic on the open set U if it is complex differentiable everywhere in U .

Proof. Pick a simple closed contour C_{λ_i} in the complex plane such that it encloses λ_i from the rest of the spectrum of T . Specifically, take a ball of radius $r = \delta/2$ around λ_i . Next, introduce the Riesz projection operators (see Definition 2.6.21) of T and $\hat{T}$

$$P_T := -\frac{1}{2\pi i} \int_{C_{\lambda_i}} R(x, T) dx = -\frac{1}{2\pi i} \int_{C_{\lambda_i}} (T - xI)^{-1} dx$$

$$P_{\hat{T}} := -\frac{1}{2\pi i} \int_{C_{\lambda_i}} R(x, \hat{T}) dx = -\frac{1}{2\pi i} \int_{C_{\lambda_i}} (\hat{T} - xI)^{-1} dx$$

with R being the resolvent operator from Definition 2.6.18. If we apply Lemma 2.6.20 (ii), we can write the difference between the Riesz operators as

$$P_T - P_{\hat{T}} = -\frac{1}{2\pi i} \int_{C_{\lambda_i}} (R(x, T) - R(x, \hat{T})) dx = -\frac{1}{2\pi i} \int_{C_{\lambda_i}} R(x, T)(\hat{T} - T)R(x, \hat{T}) dx.$$

By using the triangle inequality and the consistency of the norms we obtain,

$$\begin{aligned} \|P_T - P_{\hat{T}}\|_{op} &\leq \frac{1}{2\pi} \int_{C_{\lambda_i}} \|R(x, T)\|_{op} \|\hat{T} - T\|_{op} \|R(x, \hat{T})\|_{op} |dx| \\ &\leq \frac{|C_{\lambda_i}|}{2\pi} \sup_{x \in C_{\lambda_i}} \|R(x, T)\|_{op} \|\hat{T} - T\|_{op} \sup_{x \in C_{\lambda_i}} \|R(x, \hat{T})\|_{op}. \end{aligned} \quad (4.2)$$

To bound $\sup_{x \in C_{\lambda_i}} \|R(x, T)\|_{op}$, we apply Lemma 4.1.1. Figure 4.1 schematically illustrates the setting.

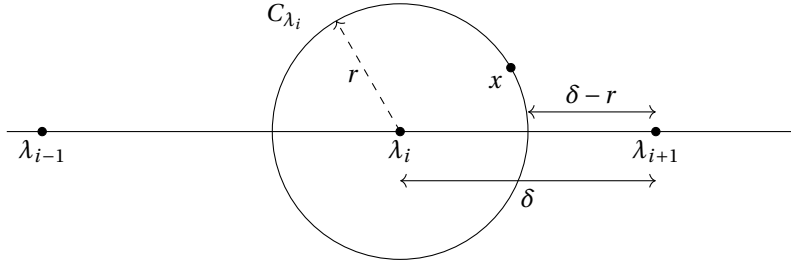


Figure 4.1: Illustration of the curve C_{λ_i} .

Recall that by construction, every point $x \in C_{\lambda_i}$ satisfies $\text{dist}(x, \lambda_i) = r$. For any other spectral element $\mu \in \sigma(T) \setminus \{\lambda_i\}$, the minimal spectral gap δ and the reverse triangle inequality imply:

$$|x - \mu| \geq |\lambda_i - \mu| - |x - \lambda_i| \geq \delta - r.$$

Since $r = \delta/2$, it follows that $\text{dist}(x, \sigma(T)) \geq \delta/2$ for all $x \in C_{\lambda_i}$. Since T is a compact, self-adjoint operator, Lemma 4.1.1 directly yields the resolvent bound

$$\sup_{x \in C_{\lambda_i}} \|R(x, T)\|_{op} = \sup_{x \in C_{\lambda_i}} \frac{1}{\text{dist}(x, \sigma(T))} \leq \frac{1}{\delta/2}. \quad (4.3)$$

Next, we derive a uniform bound for $\sup_{x \in C_{\lambda_i}} \|R(x, \hat{T})\|_{op}$. Figure 4.2 schematically illustrates this setting. The diagram depicts additional shaded disks of radius $t = \|T - \hat{T}\|_{op}$ centred at each unperturbed eigenvalue λ_j . Since $\hat{T}$ is self-adjoint, Weyl's perturbation theorem guarantees that each perturbed eigenvalue $\hat{\lambda}_j \in \sigma(\hat{T})$ lies within these disks, meaning $|\hat{\lambda}_j - \lambda_j| \leq \|T - \hat{T}\|_{op}$ for all j .

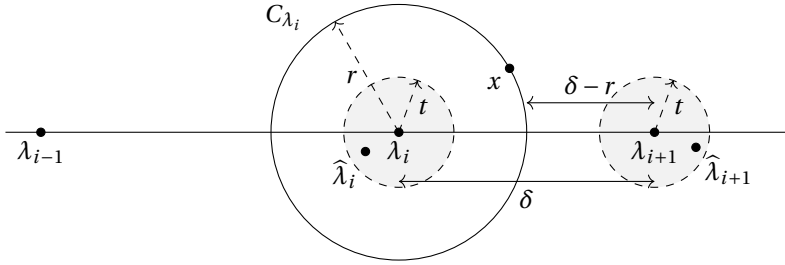


Figure 4.2: Illustration of the curve C_{λ_i} . Where t denotes the distance $\|T - \hat{T}\|_{op}$. The shaded area represents the area where the eigenvalue $\hat{\lambda}$ could be located.

To bound the distance from any point $x \in C_{\lambda_i}$ to the perturbed spectrum $\sigma(\hat{T})$, let $\hat{\mu} \in \sigma(\hat{T})$ be an arbitrary eigenvalue of $\hat{T}$, and let $\mu \in \sigma(T)$ be its corresponding unperturbed counterpart. Applying the reverse triangle inequality yields

$$|x - \hat{\mu}| \geq |x - \mu| - |\mu - \hat{\mu}| \geq |x - \mu| - \|T - \hat{T}\|_{op}.$$

From our previous analysis, we know that every point $x \in C_{\lambda_i}$ maintains a distance of at least $\delta - r$ from all unperturbed spectral elements $\mu \in \sigma(T)$. Substituting this uniform lower bound into the inequality gives

$$|x - \hat{\mu}| \geq \delta - r - \|T - \hat{T}\|_{op}$$

Since $r = \delta/2$ and since this holds uniformly for all $\hat{\mu} \in \sigma(\hat{T})$ and all $x \in C_{\lambda_i}$,

$$\text{dist}(x, \sigma(\hat{T})) \geq \frac{\delta}{2} - \|T - \hat{T}\|_{op}$$

Assuming that the perturbation is small enough such that $\|T - \hat{T}\|_{op} < \delta/2$, we can apply Lemma 4.1.1 to achieve the desired uniform upper bound

$$\sup_{x \in C_{\lambda_i}} \|R(x, \hat{T})\|_{op} = \sup_{x \in C_{\lambda_i}} \frac{1}{\text{dist}(x, \sigma(\hat{T}))} \leq \frac{1}{\frac{\delta}{2} - \|T - \hat{T}\|_{op}}. \quad (4.4)$$

Note that

$$\delta/2 - \|T - \hat{T}\|_{op} > \delta/2 - \delta/4 = \delta/4 > 0,$$

so Equation (4.4) is well defined. Combining Equations (4.2), (4.3) and (4.4) gives the

following inequality

$$\begin{aligned}
 \|P_T - P_{\hat{T}}\|_{op} &\leq \frac{|C_{\lambda_i}|}{2\pi} \frac{\|\hat{T} - T\|_{op}}{(\delta/2)(\delta/2 - \|\hat{T} - T\|_{op})} \\
 &= \frac{(\delta/2)\|\hat{T} - T\|_{op}}{(\delta/2)(\delta/2 - \|\hat{T} - T\|_{op})} \\
 &= \frac{2\|\hat{T} - T\|_{op}}{\delta - 2\|\hat{T} - T\|_{op}}.
 \end{aligned}$$

Next, we try to find an upper bound for the eigenvectors in terms of $\|P_T - P_{\hat{T}}\|_{op}$.

Lemma 2.6.22(iv) establishes that P_T is a projection, and Lemma 4.1.2 identifies its range as the eigenspace E_{λ_i} . Since λ_i is a simple eigenvalue, E_{λ_i} is one-dimensional, showing that P_T is a rank-one projection.

By Weyl's perturbation theorem, the perturbed eigenvalue $\hat{\lambda}_i$ lies within a ball of radius $\|T - \hat{T}\|_{op}$ centred at λ_i . Given the assumption $\|T - \hat{T}\|_{op} < \delta/4$, this perturbed eigenvalue is guaranteed to lie within the contour C_{λ_i} (where $r = \delta/2$). Furthermore, no other eigenvalue of $\hat{T}$ can enter the contour, as the distance between any other $\hat{\lambda}_j$ and $\hat{\lambda}_i$ satisfies

$$\text{dist}(\hat{\lambda}_i, \{\hat{\lambda}_{i-1}, \hat{\lambda}_{i+1}\}) \geq \delta - 2\|T - \hat{T}\|_{op} > \delta/2 = r$$

As illustrated in Figure 4.2, this confirms that exactly one eigenvalue of $\hat{T}$ lies inside C_{λ_i} . Consequently, Lemma 2.6.22(iv) and Lemma 4.1.2 imply that $P_{\hat{T}}$ is also a rank-one projection onto the one-dimensional eigenspace $E_{\hat{\lambda}_i}$.

Since T and $\hat{T}$ are self-adjoint, Lemma 2.6.22(iii) ensures that their respective Riesz projections P_T and $P_{\hat{T}}$ are also self-adjoint. In a Hilbert space, every self-adjoint projection is an orthogonal projection² (see p. 55 of [21]). Thus, P_T and $P_{\hat{T}}$ are orthogonal projections onto the one-dimensional spaces $E_{\lambda_i} = \text{span}\{\xi_i\}$ and $E_{\hat{\lambda}_i} = \text{span}\{\hat{\xi}_i\}$, respectively, for normalized vectors $\xi_i, \hat{\xi}_i \in H$. We may therefore write

$$\begin{aligned}
 P_T(x) &= \langle x, \xi_i \rangle_H \xi_i \\
 P_{\hat{T}}(x) &= \langle x, \hat{\xi}_i \rangle_H \hat{\xi}_i
 \end{aligned}$$

Now choose a sequence of signs $s_i \in \{-1, 1\}$ such that $\langle \hat{\xi}_i, s_i \xi_i \rangle_H \geq 0$. Since P_T and $P_{\hat{T}}$ are bounded normal operators, we will again use the fact that

$$\|P_T - P_{\hat{T}}\|_{op} = \sup_{\lambda \in \sigma(P_T - P_{\hat{T}})} |\lambda|.$$

²Although the Riesz projections are defined via complex contour integrals, the analysis remains in the real Hilbert space. Since the operators T and $\hat{T}$ are self-adjoint, their spectra are real. Therefore, any projections remains real as well. In particular, the Riesz projections P_T and $P_{\hat{T}}$ will map the real Hilbert space into itself.

To find the eigenvalues, let us build an orthonormal basis for $\{\widehat{\xi}_i, s_i \xi_i\}$. Define

$$\begin{aligned} v &= s_i \xi_i \\ w &= \frac{\widehat{\xi}_i - \langle \widehat{\xi}_i, s_i \xi_i \rangle s_i \xi_i}{\|\widehat{\xi}_i - \langle \widehat{\xi}_i, s_i \xi_i \rangle s_i \xi_i\|_H} = \frac{\widehat{\xi}_i - \cos(\theta) s_i \xi_i}{\sin(\theta)}, \end{aligned}$$

where $\theta := \arccos \langle \widehat{\xi}_i, s_i \xi_i \rangle_H$. Now $\{v, w\}$ is an orthonormal basis. Note that

$$\begin{aligned} \xi_i &= s_i v \\ \widehat{\xi}_i &= \cos(\theta) v + \sin(\theta) w \end{aligned}$$

and so rewriting the projection operators P_T and $P_{\widehat{T}}$ into this new basis gives

$$\begin{aligned} [P_T]_{\{v, w\}} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \\ [P_{\widehat{T}}]_{\{v, w\}} &= \begin{pmatrix} \cos^2(\theta) & \cos(\theta) \sin(\theta) \\ \cos(\theta) \sin(\theta) & \sin^2(\theta) \end{pmatrix}. \end{aligned}$$

It can be checked that the eigenvalues of $P_T - P_{\widehat{T}}$ are $\pm \sin(\theta)$ and it follows that

$$\|P_T - P_{\widehat{T}}\|_{op} = \sin(\theta).$$

To see that $\sin(\theta)$ cannot be negative, recall that $\theta := \arccos \langle \widehat{\xi}_i, s_i \xi_i \rangle_{L^2}$ and therefore $\theta \in \text{Ran}(\arccos) = [0, \pi]$. We have chosen s_i such that $\langle \widehat{\xi}_i, s_i \xi_i \rangle_{L^2} \geq 0$ which means that $\theta \in [0, \frac{\pi}{2}]$. On this domain we have that $\sin(\theta) \geq 0$.

Using standard trigonometry we can now also write

$$\begin{aligned} \|\widehat{\xi}_i - s_i \xi_i\|_H &= \langle \widehat{\xi}_i - s_i \xi_i, \widehat{\xi}_i - s_i \xi_i \rangle_H \\ &= \|\widehat{\xi}_i\|_H^2 + \|s_i \xi_i\|_H^2 - 2 \langle \widehat{\xi}_i, s_i \xi_i \rangle_H \\ &= 2 - 2 \cos(\theta) \\ &= 2 \sin(\theta/2) \\ &\leq 2 \sin(\theta). \end{aligned}$$

The last inequality is valid for $0 \leq \theta \leq \pi/2$ which is satisfied in our setting. We now combine it with the previous result and we get

$$\|\widehat{\xi}_i - s_i \xi_i\|_H \leq 2 \sin(\theta) \leq 2 \|P_T - P_{\widehat{T}}\| \leq \frac{4 \|\widehat{T} - T\|_{op}}{\delta - 2 \|\widehat{T} - T\|_{op}}.$$

which proves the result. $\square$

Remark 6. Although Lemma 4.1.3 is formulated for deterministic operators, it extends naturally to random operators via a pathwise argument. Specifically, for any fixed $\omega \in \Omega$, the realizations $T(\omega)$ and $\widehat{T}(\omega)$ are deterministic bounded self-adjoint operators on H . Thus, the bounds established in Lemma 4.1.3 hold almost surely for all $\omega \in \Omega$.

4.2. MODES OF CONVERGENCE FOR RANDOM OPERATORS

Section 2.6.3 reviewed convergence modes for functions and deterministic operators. Because this work concerns random (data-driven) operators, those definitions are insufficient. This section establishes the necessary modes of convergence for random operators.

Definition 4.2.1 (Convergence in operator norm in probability). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T_n, T : \Omega \times X \rightarrow Y$ be random operators between normed spaces X and Y . We say T_n converges to T in operator norm in probability if $\forall \epsilon > 0$*

$$\mathbb{P}(\|T_n - T\|_{op} > \epsilon) \xrightarrow{n \rightarrow \infty} 0.$$

In that case we write

$$\|T_n - T\|_{op} \xrightarrow{prob} 0.$$

Definition 4.2.2 (Convergence in operator norm almost surely). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T_n, T : \Omega \times X \rightarrow Y$ be random operators between normed spaces X and Y . We say T_n converges to T in operator norm almost surely if*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \|T_n - T\|_{op} = 0\right) = 1.$$

In that case we write

$$\|T_n - T\|_{op} \xrightarrow{a.s.} 0.$$

Definition 4.2.3 (Pointwise almost sure convergence). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T_n, T : \Omega \times X \rightarrow Y$ be random operators between normed spaces X and Y . We say that T_n converges to T pointwise almost surely if for every $f \in X$,*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \|T_n f - T f\|_Y = 0\right) = 1.$$

We write

$$T_n \xrightarrow{a.s.} T \quad \text{pointwise.}$$

Definition 4.2.4 (Convergence pointwise in probability). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $T_n, T : \Omega \times X \rightarrow Y$ be random operators between normed spaces X and Y . We say that T_n converges to T pointwise in probability if for every $f \in X$ and $\forall \epsilon > 0$,*

$$\mathbb{P}\left(\|T_n f - T f\|_Y > \epsilon\right) \rightarrow 0.$$

We write

$$T_n \xrightarrow{prob} T \quad \text{pointwise.}$$

4.3. CONSISTENCY RESULTS

Hereafter, we will study the asymptotic behaviour of the operators and its spectral properties. In our study, we need the usual regularity conditions on the kernel.

Assumption 1. *We assume that*

- (a) *The kernel K is bounded, i.e. $\|K\|_\infty < \infty$.*
- (b) *$K(\mathbf{u}) = K(-\mathbf{u})$ for every $\mathbf{u} \in \mathbb{R}^d$*
- (c) *$\int K(\mathbf{u}) d\mathbf{u} = 1, \int |K(\mathbf{u})| d\mathbf{u} < \infty$*
- (d) *This kernel is of order α for some integer $\alpha > 1$, i.e. $\forall j = 1, \dots, \alpha - 1$ and every indices $i_1, \dots, i_j$ in $\{1, \dots, d\}$, $\int K(\mathbf{u}) u_{i_1} \cdots u_{i_j} d\mathbf{u} = 0, \int K(\mathbf{u})^2 u_{i_1} \cdots u_{i_j} d\mathbf{u} = 0$.*
- (e) *$\mathbb{E}(K_h(\mathbf{Z} - \mathbf{z})) > 0$ for every $\mathbf{z} \in \mathcal{Z}$ and $h > 0$.*
- (f) *$\lim K(\mathbf{x})|\mathbf{x}|^d = 0$ when $\mathbf{x} \rightarrow \infty$*

Assumption 2. *Suppose the functions $f_{\mathbf{z}}$ and $\tau_{j_1, j_2 | \mathbf{z} = \mathbf{z}}$ are continuous on $\mathcal{Z}$*

Let $T_{\Gamma_p} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ be the linear integral operator with kernel Γ_p . As in Equation (3.3), for any $f \in L^2(\mathcal{Z})$,

$$(T_{\Gamma_p} f)(\mathbf{z}) = \int_{\mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'.$$

We estimate T_{Γ_p} by the empirical operator $T_{\hat{\Gamma}_p} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$, defined analogously from the estimated kernel $\hat{\Gamma}_p$

$$(T_{\hat{\Gamma}_p} f)(\mathbf{z}) = \int_{\mathcal{Z}} \hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'.$$

To study the convergence of $T_{\hat{\Gamma}_p}$ toward T_{Γ_p} , we first establish that the correspondence between a kernel and its associated integral operator is continuous.

Lemma 4.3.1. *The mapping*

$$T : \left(\begin{array}{cc} (L^2(\mathcal{Z} \times \mathcal{Z}), \|\cdot\|_{L^2(\mathcal{Z} \times \mathcal{Z})}) & \rightarrow (L(L^2(\mathcal{Z})), \|\cdot\|_{op}) \\ \Gamma & \mapsto T_\Gamma \end{array} \right)$$

is 1-Lipschitz and therefore continuous. It is a linear operator with operator norm smaller or equal to 1.

Proof. Let Γ_1 and Γ_2 in $L^2(\mathcal{Z} \times \mathcal{Z})$, let $f \in L^2(\mathcal{Z})$.

$$(T_{\Gamma_1} f - T_{\Gamma_2} f)(\mathbf{z}) = \int_{\mathcal{Z}} (\Gamma_1 - \Gamma_2)(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'.$$

By Cauchy-Schwartz

$$|(T_{\Gamma_1} f)(\mathbf{z}) - (T_{\Gamma_2} f)(\mathbf{z})| \leq \|\Gamma_1(\mathbf{z}, \cdot) - \Gamma_2(\mathbf{z}, \cdot)\|_{L^2(\mathcal{Z})} \|f\|_{L^2(\mathcal{Z})}.$$

Now squaring and integrating over $\mathcal{Z}$

$$\|T_{\Gamma_1} f - T_{\Gamma_2} f\|_{L^2(\mathcal{Z})}^2 \leq \|\Gamma_1 - \Gamma_2\|_{L^2(\mathcal{Z} \times \mathcal{Z})}^2 \|f\|_{L^2(\mathcal{Z})}^2.$$

Taking the supremum over all f with $\|f\|_{L^2(\mathcal{Z})} \leq 1$, we obtain

$$\|T_{\Gamma_1} - T_{\Gamma_2}\|_{op} \leq \|\Gamma_1 - \Gamma_2\|_{L^2(\mathcal{Z} \times \mathcal{Z})},$$

as desired. $\square$

4.3.1. CONVERGENCE OF THE OPERATORS

Under the previous assumptions, we can now prove that the estimated operator converges in probability to the true operator.

Theorem 4.3.2 (Convergence in operator norm). *Let T_{Γ_p} and $T_{\hat{\Gamma}_p}$ the covariance integral operators as defined above. Then under Assumptions 1 and 2, if $nh^d \rightarrow +\infty$ and $h \rightarrow 0$*

$$\|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{HS} \xrightarrow[n \rightarrow \infty]{prob} 0$$

and

$$\|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{op} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Proof. Lemma 2.4.2 shows that $\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ converges pointwise in probability to $\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ for any $\mathbf{z} \in \mathcal{Z}$ as $n \rightarrow \infty$. Define $\Delta_{j_1, j_2}(\mathbf{z}) := \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} - \hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ and note that $|\Delta_{j_1, j_2}| \leq 2$. Recall the fact that since $\Delta_{j_1, j_2}(\mathbf{z}) \xrightarrow[n \rightarrow \infty]{prob} 0$ pointwise in $\mathbf{z}$ and Δ_{j_1, j_2} is bounded, we have that

$$\mathbb{E}(\Delta_{j_1, j_2}^2(\mathbf{z})) \xrightarrow[n \rightarrow \infty]{} 0.$$

Now since $\mathbb{E}(\Delta_{j_1, j_2}^2) \leq 4$ the dominated convergence theorem yields

$$\int_{\mathcal{Z}} \mathbb{E}(\Delta_{j_1, j_2}^2(\mathbf{z})) d\mathbf{z} \xrightarrow[n \rightarrow \infty]{} 0.$$

If we now define

$$S_{j_1, j_2} := \int_{\mathcal{Z}} \Delta_{j_1, j_2}^2(\mathbf{z}) d\mathbf{z}$$

then by Fubini

$$\mathbb{E}(S_{j_1, j_2}) = \int_{\mathcal{Z}} \mathbb{E}(\Delta_{j_1, j_2}^2(\mathbf{z})) d\mathbf{z} \xrightarrow[n \rightarrow \infty]{} 0.$$

Markov's inequality now gives for any $\epsilon > 0$

$$\mathbb{P}(|S_{j_1, j_2}| > \epsilon) \leq \frac{1}{\epsilon} \mathbb{E}(S_{j_1, j_2}) \xrightarrow[n \rightarrow \infty]{} 0.$$

This shows that $S_{j_1, j_2} \xrightarrow[n \rightarrow \infty]{prob} 0$ and hence $\|\tau_{j_1, j_2 | \mathbf{Z}=\cdot} - \hat{\tau}_{j_1, j_2 | \mathbf{Z}=\cdot}\|_{L^2(\mathcal{Z})} \xrightarrow[n \rightarrow \infty]{prob} 0$. It then follows by construction that also $\|\hat{\mu} - \mu\|_{L^2(\mathcal{Z})} \xrightarrow[n \rightarrow \infty]{prob} 0$.

Define $\delta_{j_1, j_2}(\mathbf{z}) := \Delta_{j_1, j_2}(\mathbf{z}) - (\hat{\mu}(\mathbf{z}) - \mu(\mathbf{z}))$. Now using the triangle inequality and Cauchy-

Schwartz, we obtain

$$\begin{aligned}
\|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{HS} &= \left(\iint_{\mathcal{Z} \times \mathcal{Z}} \left(\Gamma_p(\mathbf{z}, \mathbf{z}') - \hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') \right)^2 d\mathbf{z} d\mathbf{z}' \right)^{1/2} \\
&= \left\| \Gamma_p(\mathbf{z}, \mathbf{z}') - \hat{\Gamma}_p(\mathbf{z}, \mathbf{z}') \right\|_{L^2(\mathcal{Z} \times \mathcal{Z})} \\
&\leq \frac{1}{\binom{p}{2} - 1} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p \left\| \delta_{j_1, j_2}(\mathbf{z}) (\tau_{j_1, j_2}(\mathbf{z}') - \mu(\mathbf{z}')) + \delta_{j_1, j_2}(\mathbf{z}') (\tau_{j_1, j_2}(\mathbf{z}) - \mu(\mathbf{z})) \right. \\
&\quad \left. + \delta_{j_1, j_2}(\mathbf{z}) \delta_{j_1, j_2}(\mathbf{z}') \right\|_{L^2(\mathcal{Z} \times \mathcal{Z})} \\
&\leq \frac{1}{\binom{p}{2} - 1} \sum_{j_1=1}^p \sum_{j_2=j_1+1}^p \left(2 \left\| \delta_{j_1, j_2} \right\|_{L^2(\mathcal{Z})} \left\| \tau_{j_1, j_2}(\mathbf{z}') - \mu(\mathbf{z}') \right\|_{L^2(\mathcal{Z})} + \left\| \delta_{j_1, j_2} \right\|_{L^2(\mathcal{Z})}^2 \right).
\end{aligned}$$

Since $\left\| \delta_{j_1, j_2} \right\|_{L^2(\mathcal{Z})} \xrightarrow[n \rightarrow \infty]{prob} 0$, $\hat{\tau}_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ and $\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ are bounded, we obtain

$$\|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{HS} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Using Lemma 2.6.13, we conclude

$$\|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{op} \leq \|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{HS} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

□

4.3.2. CONVERGENCE OF THE EIGENVALUES AND EIGENFUNCTIONS

Building upon the convergence of the covariance operator $T_{\hat{\Gamma}_p}$ in operator norm, we now examine the implications for its spectral components. We use this convergence property to establish that the empirical eigenvalues $\hat{\lambda}_i$ are consistent estimators of the true eigenvalues λ_i as $n \rightarrow \infty$.

Theorem 4.3.3 (Consistency of the eigenvalues). *Let T_{Γ_p} and $T_{\hat{\Gamma}_p}$ the covariance integral operators as defined above with their associated eigenvalues $\{\lambda_i\}_{i \geq 1}$ and $\{\hat{\lambda}_i\}_{i \geq 1}$ corresponding to the eigenfunctions $\{\xi_i(\cdot)\}_{i \geq 1}$ and $\{\hat{\xi}_i(\cdot)\}_{i \geq 1}$ respectively. If Assumptions 1 and 2 hold and $nh^d \rightarrow +\infty$ as $h \rightarrow 0$, then for every fixed $i \geq 1$,*

$$\hat{\lambda}_i \xrightarrow[n \rightarrow \infty]{prob} \lambda_i \quad \text{as } n \rightarrow \infty.$$

Proof. By construction, T_{Γ_p} and $T_{\hat{\Gamma}_p}$ are self-adjoint, positive semi-definite and compact operators. Therefore there exists a set of at most countable nonnegative ordered eigenvalues. First note by Theorem 4.3.2 that

$$\|T_{\hat{\Gamma}_p} - T_{\Gamma_p}\|_{op} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Recall that by Lemma 2.6.36

$$|\lambda_i(T_{\Gamma_p}) - \lambda_i(T_{\hat{\Gamma}_p})| \leq \|T_{\Gamma_p} - T_{\hat{\Gamma}_p}\|_{op} \quad \forall i = 1, 2, \dots$$

Hence the eigenvalue functional $\lambda_i(\cdot)$ is 1-Lipschitz with respect to the operator norm. Therefore by applying the continuous mapping theorem

$$\widehat{\lambda}_i = \lambda_i(T_{\widehat{\Gamma}_p}) \xrightarrow[n \rightarrow \infty]{prob} \lambda_i(T_{\Gamma_p}) = \lambda_i$$

which proves the result. $\square$

Similarly we can show that asymptotically, the empirical eigenfunctions $\widehat{\xi}_i$ are consistent estimators of the true eigenfunctions ξ_1 up to a sequence of signs.

Theorem 4.3.4 (Consistency of the eigenfunctions). *Let T_{Γ_p} and $T_{\widehat{\Gamma}_p}$ the covariance integral operators as defined above with their associated eigenvalues $\{\lambda_i\}_{i \geq 1}$ and $\{\widehat{\lambda}_i\}_{i \geq 1}$ corresponding to the eigenfunctions $\{\xi_i(\cdot)\}_{i \geq 1}$ and $\{\widehat{\xi}_i(\cdot)\}_{i \geq 1}$.*

Let $i \geq 1$ such that λ_i and $\widehat{\lambda}_i$ are simple, if Assumptions 1 and 2 hold and if $nh^d \rightarrow +\infty$ as $h \rightarrow 0$, then there exists a sequence of signs $s_i \in \{\pm 1\}$ such that,

$$\|\widehat{\xi}_i - s_i \xi_i\|_{L^2(\mathcal{Z})} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Proof. Let $i \geq 0$ such that λ_i is simple for the operator T_{Γ_p} and define the positive eigen-gap

$$\delta := \min_{\mu \in \{\sigma(T) \setminus \lambda_i\}} \text{dist}(\mu, \lambda_i) > 0.$$

Theorem 4.3.2 shows that

$$\|T_{\Gamma_p} - T_{\widehat{\Gamma}_p}\|_{op} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

We can now directly apply Lemma 4.1.3 with $H = L^2(\mathcal{Z})$, which gives

$$\|\widehat{\xi}_i - s_i \xi_i\|_{L^2(\mathcal{Z})} \leq \frac{4\|T_{\Gamma_p} - T_{\widehat{\Gamma}_p}\|_{op}}{\delta - 2\|T_{\Gamma_p} - T_{\widehat{\Gamma}_p}\|_{op}} \xrightarrow[n \rightarrow \infty]{prob} 0,$$

which proves the result. $\square$

Remark 7. Theorem 4.3.4 requires the target eigenvalues to be simple. In practice, this is not a restrictive assumption; covariance operators of continuous stochastic objects rarely have eigenvalues with a multiplicity greater than one. In cases where identical eigenvalues do occur, they typically correspond to $\lambda = 0$, representing directions of zero variance that carry no structural information and are thus not of interest.

4.4. CONVERGENCE RATE

In this section, we establish the rate of convergence for the estimated covariance operator under the Hilbert-Schmidt norm. This result serves as a foundational theoretical guarantee for the performance of our estimator. Proving this theorem requires extensive technical machinery and additional material about U -statistics. The derivation of this material and the subsequent proof of Theorem 4.4.1 are deferred to Chapter 6.

Theorem 4.4.1. *Under Assumptions 1, 2, 8, 9, 10, 11 and 12 in the bandwidth regime where $h_n \rightarrow 0$ and $nh_n^{2d}/|\log h| \rightarrow +\infty$,*

$$\|T_{\hat{\Gamma}_p} - T_{\Gamma_p}\|_{HS}^2 = \iint \left(\hat{\Gamma}(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') \right)^2 d\mathbf{z} d\mathbf{z}' = \mathcal{O}_P\left(h^4 + \frac{1}{nh^d}\right).$$

Theorem 4.4.1 characterizes the classic bias-variance trade-off inherent to nonparametric kernel smoothing. The first term, $\mathcal{O}(h^4)$, reflects the squared smoothing bias. Conversely, the second term, $\mathcal{O}((nh^d)^{-1})$, represents the variance of the estimator, which scales inversely with the sample size.

Consequently, selecting a bandwidth h that decays too rapidly minimizes the smoothing bias but inflates the variance, leading to an erratic estimator. Conversely, an oversized bandwidth suppresses variance at the cost of a severe oversmoothing bias. Balancing these competing forces yields an optimal theoretical bandwidth rate of $h_n \asymp n^{-1/(4+d)}$, provided that this rate is compatible with the constraints. Specifically, this optimal rate satisfies the regularizing condition $nh_n^{2d}/|\log h| \rightarrow +\infty$ if and only if the dimension d satisfies $d < 4$. In settings where this holds, the estimator achieves the optimal Hilbert-Schmidt convergence rate of $\mathcal{O}_P\left(n^{-4/(4+d)}\right)$.

The first consequence of Theorem 4.4.1 is that we can establish a convergence rate in operator norm. This follows naturally from Lemma 2.6.13.

Corollary 4.4.2. *Under the assumptions of Theorem 4.4.1, the estimated operator $T_{\hat{\Gamma}_p}$ converges in operator norm to its T_{Γ_p} at the following rate*

$$\|T_{\hat{\Gamma}_p} - T_{\Gamma_p}\|_{op} = \mathcal{O}_P\left(h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

Now this global rate can be seamlessly mapped to the individual components of the spectral decomposition. Specifically, applying Lemma 2.6.36 immediately yields the convergence rate for the estimated eigenvalues.

Corollary 4.4.3. *Let $\hat{\lambda}_i$ and λ_i denote the i -th sorted eigenvalues of $T_{\hat{\Gamma}_p}$ and T_{Γ_p} , respectively. Then, under the assumptions of Theorem 4.4.1*

$$|\hat{\lambda}_i - \lambda_i| = \mathcal{O}_P\left(h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

Finally, we turn our attention to the stability of the eigenvectors. Provided that the target eigenvalue λ_i is isolated and distinct from the rest of the spectrum, we can control the alignment of the estimated eigenvectors. Applying Lemma 4.1.3 yields the identical optimal convergence rate for the sample eigenfunctions up to a sequence of signs $s_i \in \{-1, 1\}$.

Corollary 4.4.4. *Let $\hat{\xi}_i$ and ξ_i be the i -th eigenfunctions of $T_{\hat{\Gamma}_p}$ and T_{Γ_p} such that the corresponding eigenvalues are simple. Then, under the assumptions of Theorem 4.4.1, there exists a sequence of signs $s_i \in \{-1, 1\}$ such that*

$$\|\hat{\xi}_i - s_i \xi_i\|_{L^2(\mathcal{Z})} = \mathcal{O}_P\left(h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

5

PCA ON THE ESTIMATED CONDITIONAL KENDALL'S TAU

In the previous chapter, we established a mathematical framework for the spectral analysis of covariance integral operators under the idealized assumption that conditional Kendall's tau curves are available across the entire continuous domain $\mathcal{Z}$. In practice, however, these curves are typically observed only at a finite number of discrete grid points. This chapter relaxes the assumption of continuous observation and develops the framework necessary to analyse the discretized version of these operators.

We begin by discretizing the variable $\mathbf{z}$ over a fixed grid of points $\{\mathbf{z}_1, \dots, \mathbf{z}_N\} \subset \mathcal{Z}$. By evaluating the kernels of the covariance operators Γ_p and $\hat{\Gamma}_p$ on this grid, we construct the discrete counterparts, denoted by the matrices $\Gamma^{(N)}$ and $\hat{\Gamma}^{(N)}$, respectively:

$$\Gamma^{(N)} = (\Gamma_p(\mathbf{z}_i, \mathbf{z}_j))_{i,j=1}^N \quad \text{and} \quad \hat{\Gamma}^{(N)} = (\hat{\Gamma}_p(\mathbf{z}_i, \mathbf{z}_j))_{i,j=1}^N.$$

These discretized kernels can thus be represented as matrices in $\mathbb{R}^{N \times N}$. Next, we fix some weights $\{w_j^{(N)}\}_{j=1}^N$ which satisfy the conditions

- i) $\forall j = 1, \dots, N, w_j^{(N)} > 0$ and $\sum_{j=1}^N w_j^{(N)} = 1$,
- ii) $\forall N > 0, N \sum_{j=1}^N (w_j^{(N)})^2 < +\infty$.

These weights may for instance arise from a quadrature rule, or simply be uniform weights given by $w_j^{(N)} = \frac{1}{N}$.

We define the induced operators $T_{\Gamma^{(N)}} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ and $T_{\hat{\Gamma}^{(N)}} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ by

$$(T_{\Gamma^{(N)}} \mathbf{x})_i := \sum_{j=1}^N w_j^{(N)} \Gamma_{ij}^{(N)} x_j \quad \text{and} \quad (T_{\hat{\Gamma}^{(N)}} \mathbf{x})_i := \sum_{j=1}^N w_j^{(N)} \hat{\Gamma}_{ij}^{(N)} x_j, \quad (5.1)$$

for $\mathbf{x} \in \mathbb{R}^N$ and $i = 1, \dots, N$. We can now define the k -th-eigenvalue and eigenvector of these operators by

$$T_{\Gamma^{(N)}} \boldsymbol{\xi}_k^{(N)} = \lambda_k^{(N)} \boldsymbol{\xi}_k^{(N)} \quad \text{and} \quad T_{\hat{\Gamma}^{(N)}} \hat{\boldsymbol{\xi}}_k^{(N)} = \hat{\lambda}_k^{(N)} \hat{\boldsymbol{\xi}}_k^{(N)}, \quad \forall k = 1, \dots, N. \quad (5.2)$$

It is often convenient to treat the operators in matrix form. Let $W := \text{diag}(w_1^{(N)}, \dots, w_N^{(N)})$, the diagonal matrix of weights. Then the eigenproblem can be written as

$$\Gamma^{(N)} W \boldsymbol{\xi}_k^{(N)} = \lambda_k^{(N)} \boldsymbol{\xi}_k^{(N)} \quad \text{and} \quad \hat{\Gamma}^{(N)} W \hat{\boldsymbol{\xi}}_k^{(N)} = \hat{\lambda}_k^{(N)} \hat{\boldsymbol{\xi}}_k^{(N)}, \quad \forall k = 1, \dots, N. \quad (5.3)$$

5.1. THEORETICAL FRAMEWORK

Bridging the gap between the continuous Hilbert space and its discrete approximations requires a systematic decomposition of the underlying domain. We begin by partitioning $\mathcal{Z}$ into a finite collection of cells, which allows us to define the functional spaces necessary for our discretization scheme.

Definition 5.1.1. Let $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ be a grid on $\mathcal{Z}$ and define $\{C_j^{(N)}\}_{j=1}^N$ as a measurable partition of $\mathcal{Z}$ such that

$$i) \quad \mathbf{z}_j \in C_j^{(N)}$$

$$ii) \quad \text{Leb}(C_j^{(N)}) > 0$$

for each $j \in \{1, \dots, N\}$ and $\text{Leb}(\cdot)$ is the Lebesgue measure on $\mathcal{Z}$.

To compare the discretized operators with the continuous covariance operator T_{Γ_p} , we associate the matrix $\Gamma^{(N)}$ with a finite-rank operator on the continuous space $L^2(\mathcal{Z})$. For that, we first introduce the space of piecewise constant functions.

Definition 5.1.2. Define

$$\text{Piece}_N(\mathcal{Z}) = \left\{ f \in L^2(\mathcal{Z}) : f|_{C_i^{(N)}} \text{ is constant for each } i \right\}. \quad (5.4)$$

The following lemma characterizes $\text{Piece}_N(\mathcal{Z})$ as a finite-dimensional subspace and provides an explicit orthonormal basis, which is useful for deriving the matrix representations of our operators.

Lemma 5.1.3. Let $\{C_1^{(N)}, \dots, C_N^{(N)}\}$ be a partition of $\mathcal{Z}$ as in Definition 5.1.1, and define the functions

$$e_j^{(N)}(\mathbf{z}) = \frac{1}{\sqrt{\text{Leb}(C_j^{(N)})}} \mathbb{1}_{C_j^{(N)}}(\mathbf{z}), \quad j = 1, \dots, N.$$

Then $\{e_j^{(N)}\}_{j=1}^N$ is an orthonormal basis for $\text{Piece}_N(\mathcal{Z})$ and $\dim(\text{Piece}_N(\mathcal{Z})) = N$.

Proof. To prove $\{e_j^{(N)}\}_{j=1}^N$ is a basis we shall first prove the spanning property and then prove linear independence. We view $\text{Piece}_N(\mathcal{Z})$ as a subspace of $L^2(\mathcal{Z})$, equipped with the inner product inherited from $L^2(\mathcal{Z})$. First, take any $f \in \text{Piece}_N(\mathcal{Z})$. Then by definition, there exist $a_1, \dots, a_N$ such that for each $i = 1, \dots, N$

$$f(\mathbf{z}) = a_i, \quad \text{for } \mathbf{z} \in C_i^{(N)}.$$

We can therefore write

$$f = \sum_{i=1}^N a_i \mathbb{1}_{C_i^{(N)}} = \sum_{i=1}^N a_i \sqrt{\text{Leb}(C_i^{(N)})} e_i^{(N)}$$

This means

$$\text{Piece}_N(\mathcal{Z}) = \text{Span}\{e_1^{(N)}, \dots, e_N^{(N)}\}$$

as desired. To prove linear independence suppose

$$\sum_{i=1}^N c_i e_i^{(N)}(\mathbf{z}) = 0, \quad \text{in } L^2(\mathcal{Z}).$$

For any fixed k integrate over $C_k^{(N)}$

$$\int_{C_k^{(N)}} \sum_{i=1}^N c_i e_i^{(N)}(\mathbf{z}) d\mathbf{z} = 0.$$

Since the cells are disjoint, only the $i = k$ case remains, giving

$$c_k \int_{C_k^{(N)}} e_k^{(N)}(\mathbf{z}) d\mathbf{z} = 0.$$

As $\text{Leb}(C_k^{(N)}) > 0$, this implies that $c_k = 0$. Since k was arbitrary, all coefficients vanish, and the family is linearly independent. This proves that the set $\{e_j^{(N)}\}_{j=1}^N$ is a basis for $\text{Piece}_N(\mathcal{Z})$ and we naturally obtain $\dim(\text{Piece}_N(\mathcal{Z})) = N$. To show this set is also orthonormal, note that

$$\langle e_j^{(N)}, e_i^{(N)} \rangle_{L^2} = \frac{1}{\sqrt{\text{Leb}(C_j^{(N)})\text{Leb}(C_i^{(N)})}} \int_{\mathcal{Z}} \mathbb{1}_{C_j^{(N)}} \mathbb{1}_{C_i^{(N)}} d\mathbf{z} = \delta_{ij}(\mathbf{z})$$

where δ_{ij} denotes the Kronecker delta function. As a consequence we obtain that $\{e_j^{(N)}\}_{j=1}^N$ is an orthonormal set in $L^2(\mathcal{Z})$ and hence $\{e_j^{(N)}\}_{j=1}^N$ is an orthonormal basis of $\text{Piece}_N(\mathcal{Z})$. $\square$

We can now associate the matrix $\Gamma^{(N)}$ with a finite-rank operator acting on this new continuous space $\text{Piece}_N(\mathcal{Z})$. Since $\text{Piece}_N(\mathcal{Z}) \subset L^2(\mathcal{Z})$, this space can be used to compare with operators acting upon $L^2(\mathcal{Z})$.

Definition 5.1.4. Let $\{C_1^{(N)}, \dots, C_N^{(N)}\}$ a measurable partition of the domain $\mathcal{Z}$ as in Definition 5.1.1. Let $\tilde{T}_{\Gamma^{(N)}} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ be the linear operator that maps any function $f \in L^2(\mathcal{Z})$ to the piecewise constant function $\tilde{T}_{\Gamma^{(N)}} f$ defined by

$$\tilde{T}_{\Gamma^{(N)}} f := \begin{pmatrix} \mathcal{Z} = \bigsqcup_{i=1}^N C_i^{(N)} & \rightarrow & \mathbb{R} \\ \mathbf{z} \in C_i^{(N)} & \mapsto & \sum_{j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \bar{f}_j \end{pmatrix} \in \text{Piece}_N(\mathcal{Z})$$

where $\bar{f}_j$ denotes the cell average of f ,

$$\bar{f}_j := \frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{x}) d\mathbf{x},$$

and Leb is the Lebesgue measure on $\mathcal{Z}$.

5

With this definition, $\tilde{T}_{\Gamma^{(N)}}$ produces piecewise-constant functions that are adapted to the given partition. By Lemma 5.1.3, we have $\dim(\text{Piece}_N(\mathcal{Z})) = N$. Since the range of $\tilde{T}_{\Gamma^{(N)}}$ is contained in $\text{Piece}_N(\mathcal{Z})$, it follows that $\tilde{T}_{\Gamma^{(N)}}$ is a finite-rank operator with $\text{rank}(\tilde{T}_{\Gamma^{(N)}}) \leq N$. It can also be shown that under some conditions $\tilde{T}_{\Gamma^{(N)}}$ is a Hilbert-Schmidt integral operator.

Lemma 5.1.5. Let $N > 0$. If

$$\sum_{i,j=1}^N \frac{\text{Leb}(C_i^{(N)})}{\text{Leb}(C_j^{(N)})} w_j^2 < +\infty. \quad (5.5)$$

then the linear operator $\tilde{T}_{\Gamma^{(N)}}$ from Definition 5.1.4 is a Hilbert-Schmidt integral operator on both $L^2(\mathcal{Z})$ and $\text{Piece}_N(\mathcal{Z})$ with kernel $\tilde{k}$ defined for $\mathbf{x} \in C_i^{(N)}$ and $\mathbf{y} \in C_j^{(N)}$ as

$$\tilde{k}(\mathbf{x}, \mathbf{y}) = \frac{1}{\text{Leb}(C_j^{(N)})} w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j). \quad (5.6)$$

Proof. Let $\mathbf{x} \in C_i^{(N)}$ and $f \in L^2(\mathcal{Z})$, then by definition

$$\begin{aligned} (\tilde{T}_{\Gamma^{(N)}} f)(\mathbf{x}) &= \sum_{j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{y}) d\mathbf{y} \\ &= \sum_{j=1}^N \int_{C_j^{(N)}} \frac{1}{\text{Leb}(C_j^{(N)})} w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) f(\mathbf{y}) d\mathbf{y} \\ &= \int_{\mathcal{Z}} \tilde{k}(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Therefore $\tilde{k}$ is indeed the kernel of $\tilde{T}_{\Gamma^{(N)}}$. To see that this operator is Hilbert-Schmidt,

note

$$\begin{aligned}
 \|\tilde{T}_{\Gamma^{(N)}}\|_{HS}^2 &= \iint_{\mathcal{Z} \times \mathcal{Z}} |\tilde{k}(\mathbf{x}, \mathbf{y})|^2 d\mathbf{x} d\mathbf{y} \\
 &= \sum_{i=1}^N \sum_{j=1}^N \int_{C_i^{(N)}} \int_{C_j^{(N)}} \frac{1}{\text{Leb}(C_j^{(N)})^2} w_j^2 \Gamma_p(\mathbf{z}_i, \mathbf{z}_j)^2 d\mathbf{x} d\mathbf{y} \\
 &\leq 16 \sum_{i,j} \frac{\text{Leb}(C_i^{(N)})}{\text{Leb}(C_j^{(N)})} w_j^2 < \infty,
 \end{aligned}$$

where we used the fact that $\Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \leq 4$. This proves the claim $\square$

Remark 8. It is straightforward to verify that Equation (5.5) holds under mild regularity assumptions on the partition. Suppose that the cells have comparable sizes, in the sense that there exist constants $a_1, a_2 > 0$ such that

$$a_1 \leq \frac{\text{Leb}(C_i^{(N)})}{\text{Leb}(C_j^{(N)})} \leq a_2, \quad \text{for all } i, j = 1, \dots, N,$$

and assume uniform weights, $w_j^{(N)} = \frac{1}{N}$. Then

$$\sum_{i,j=1}^N \frac{\text{Leb}(C_i^{(N)})}{\text{Leb}(C_j^{(N)})} w_j^2 \leq CN \sum_{j=1}^N w_j^2 = C < +\infty.$$

It is useful to express Definition 5.1.4 in a form that makes its structure more explicit. To this end, we introduce two linear operators that connect $L^2(\mathcal{Z})$, $\mathbb{R}^N$, and $\text{Piece}_N(\mathcal{Z})$. First, introduce the linear projection operator $P_N := \text{proj}_{\mathbb{R}^N} : L^2(\mathcal{Z}) \rightarrow \mathbb{R}^N$ defined by

$$(P_N f)_j := \frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{z}) d\mathbf{z} = \bar{f}_j. \quad (5.7)$$

Next, define the linear operator $\Phi_N : \mathbb{R}^N \rightarrow \text{Piece}_N(\mathcal{Z})$ for any $g \in \mathbb{R}^N$ as

$$(\Phi_N g)(\mathbf{z}) := \sum_{i=1}^N g_i \mathbb{1}_{C_i^{(N)}}(\mathbf{z}). \quad (5.8)$$

Recall from Equations (5.2) and (5.3) that $T_{\Gamma^{(N)}}$ can also be viewed as an operator $T_{\Gamma^{(N)}} : \mathbb{R}^N \rightarrow \mathbb{R}^N$ given by

$$T_{\Gamma^{(N)}} g := \Gamma^{(N)} W g, \quad g \in \mathbb{R}^N.$$

Consequently, the operator $\tilde{T}_{\Gamma^{(N)}}$ admits the important decomposition

$$\tilde{T}_{\Gamma^{(N)}} = \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N. \quad (5.9)$$

This identity establishes a formal bridge between the continuous operator on $L^2(\mathcal{Z})$ and its discrete counterpart on $\mathbb{R}^N$.

We conclude this section by introducing the projection operator onto $\text{Piece}_N(\mathcal{Z})$, which provides a concrete link to the theoretical framework developed above.

Definition 5.1.6. For any $N > 0$, the projection operator $\text{proj}_{\text{Piece}_N(\mathcal{Z})} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ is defined as $\text{proj}_{\text{Piece}_N(\mathcal{Z})} = \Phi_N \circ P_N$ with P_N and Φ_N as in Equations (5.7) and (5.8) respectively.

For any $f \in L^2(\mathcal{Z})$, this projection can be written explicitly as

$$(\text{proj}_{\text{Piece}_N(\mathcal{Z})} f)(\mathbf{z}) = \sum_{i=1}^n \tilde{f}_i \mathbb{1}_{C_i^{(N)}}(\mathbf{z})$$

where $\tilde{f}_i = \frac{1}{\text{Leb}(C_i^{(N)})} \int_{C_i^{(N)}} f(\mathbf{z}) d\mathbf{z}$. Equivalently, in terms of the orthonormal basis $\{e_i^{(N)}\}_{i=1}^N$,

$$(\text{proj}_{\text{Piece}_N(\mathcal{Z})} f)(\mathbf{z}) = \sum_{i=1}^N \langle f, e_i^{(N)} \rangle_{L^2(\mathcal{Z})} e_i^{(N)}(\mathbf{z}). \quad (5.10)$$

This representation shows that $\text{proj}_{\text{Piece}_N(\mathcal{Z})}$ is the orthogonal projection onto $\text{Piece}_N(\mathcal{Z})$ in $L^2(\mathcal{Z})$.

To view this projection as an operator on the full space $L^2(\mathcal{Z})$, we extend its codomain using the *canonical injection*. Since $\text{Piece}_N(\mathcal{Z}) \subset L^2(\mathcal{Z})$, define the canonical injection operator map $i_{\text{Piece}_N(\mathcal{Z})} : \text{Piece}_N(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ by $i_{\text{Piece}_N(\mathcal{Z})}(g) = g$ for all $g \in \text{Piece}_N(\mathcal{Z})$. We then define the operator $\Pi_N : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ by

$$\Pi_N = i_{\text{Piece}_N(\mathcal{Z})} \circ \text{proj}_{\text{Piece}_N(\mathcal{Z})}. \quad (5.11)$$

5.2. PROPERTIES OF $\tilde{T}_{\Gamma^{(N)}}$

To understand the theoretical behaviour of $\tilde{T}_{\Gamma^{(N)}}$, we first establish its fundamental algebraic properties. While the operator is generally compact, its self-adjointness depends on the relationship between weights and partition cells.

Assumption 3. The weights $w_j^{(N)}$ and partition cells $C_j^{(N)}$ satisfy $w_j^{(N)} = C \cdot \text{Leb}(C_j^{(N)})$ for some constant C , for all $1 \leq j \leq N$.

Since the weights sum up to 1, this assumption enforces $w_j^{(N)} = \frac{1}{|\mathcal{Z}|} \text{Leb}(C_j^{(N)})$

Lemma 5.2.1. The operator $\tilde{T}_{\Gamma^{(N)}} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ is compact and linear. Furthermore, $\tilde{T}_{\Gamma^{(N)}}$ is self-adjoint if and only if Assumption 3 holds.

Proof. Linearity is clear by construction. For compactness, note that $\tilde{T}_{\Gamma^{(N)}}$ has its image in $\text{Piece}_N(\mathcal{Z})$ which is finite-dimensional by Lemma 5.1.3. Remember that any linear operator with finite-dimensional range is finite-rank and every finite-rank operator on a Hilbert space is compact.

It remains to characterize when $\tilde{T}_{\Gamma^{(N)}}$ is self-adjoint. For $f, g \in L^2(\mathcal{Z})$, we compute

$$\begin{aligned}\langle \tilde{T}_{\Gamma^{(N)}} f, g \rangle_{L^2(\mathcal{Z})} &= \sum_{i=1}^N \int_{C_i^{(N)}} \left(\sum_{j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \tilde{f}_j \right) g(\mathbf{z}) d\mathbf{z} \\ &= \sum_{i,j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \tilde{f}_j \tilde{g}_i \text{Leb}(C_i^{(N)}), \\ \langle f, \tilde{T}_{\Gamma^{(N)}} g \rangle_{L^2(\mathcal{Z})} &= \sum_{i=1}^N \int_{C_i^{(N)}} \left(\sum_{j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \tilde{g}_j \right) f(\mathbf{z}) d\mathbf{z} \\ &= \sum_{i,j=1}^N w_j^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \tilde{g}_j \tilde{f}_i \text{Leb}(C_i^{(N)}).\end{aligned}$$

By swapping indices i and j in the second expression and using the symmetry of Γ_p , we obtain

$$\langle f, \tilde{T}_{\Gamma^{(N)}} g \rangle_{L^2(\mathcal{Z})} = \sum_{i,j=1}^N w_i^{(N)} \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \tilde{f}_j \tilde{g}_i \text{Leb}(C_j^{(N)}).$$

Hence $\tilde{T}_{\Gamma^{(N)}}$ is self-adjoint if and only if, for all i, j ,

$$w_j^{(N)} \text{Leb}(C_i^{(N)}) = w_i^{(N)} \text{Leb}(C_j^{(N)}).$$

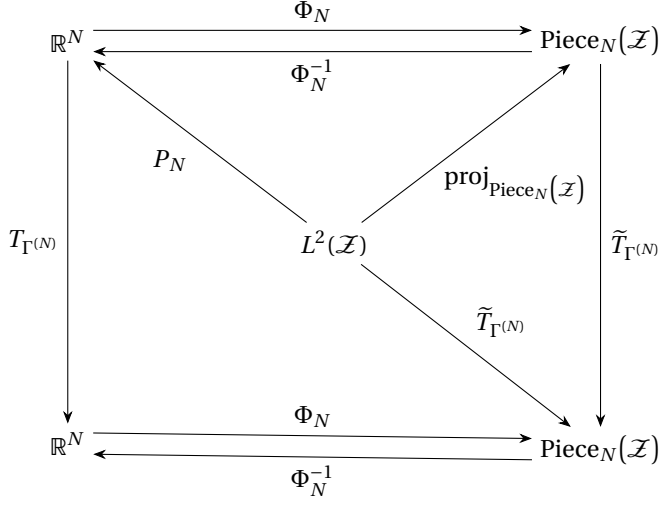
This holds if and only if

$$\frac{w_i^{(N)}}{\text{Leb}(C_i^{(N)})} = C, \quad \text{for all } i,$$

for some constant C . This means $w_i^{(N)} = C \cdot \text{Leb}(C_i^{(N)})$. This proves that $\tilde{T}_{\Gamma^{(N)}}$ is self-adjoint if and only if Assumption 3 holds. $\square$

To bridge the finite-dimensional space $\mathbb{R}^N$ with the function space $\text{Piece}_N(\mathcal{Z}) \subset L^2(\mathcal{Z})$, we establish a formal correspondence that links the operators $T_{\Gamma^{(N)}}$ and $\tilde{T}_{\Gamma^{(N)}}$. This relationship is formalized via a commutative diagram, demonstrating the equivalence of the two spaces and enabling a direct spectral comparison between $T_{\Gamma^{(N)}}$ and $\tilde{T}_{\Gamma^{(N)}}$.

Lemma 5.2.2. *We have the following commutative graph:*



Proof. We shall first prove that on $\mathbb{R}^N$ we have $\Phi_N \circ T_{\Gamma^{(N)}} = \tilde{T}_{\Gamma^{(N)}} \circ \Phi_N$. For that, take $\mathbf{x} \in \mathbb{R}^N$. Then by Equation (5.9)

$$(\tilde{T}_{\Gamma^{(N)}} \circ \Phi_N)\mathbf{x} = \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N \circ (\Phi_N \mathbf{x}).$$

Note that for any function in $\mathbb{R}^N$, $P_N \circ \Phi_N = \text{Id}_{\mathbb{R}^N}$. Therefore

$$(\tilde{T}_{\Gamma^{(N)}} \circ \Phi_N)\mathbf{x} = \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N \circ (\Phi_N \mathbf{x}) = (\Phi_N \circ T_{\Gamma^{(N)}})\mathbf{x}.$$

Next, we shall prove that on $\text{Piece}_N(\mathcal{Z})$, $\Phi_N^{-1} \circ \tilde{T}_{\Gamma^{(N)}} = T_{\Gamma^{(N)}} \circ \Phi_N^{-1}$. Note that on $\text{Piece}_N(\mathcal{Z})$, $\Phi_N^{-1} = P_N$. Thus using Equation (5.9)

$$P_N \circ \tilde{T}_{\Gamma^{(N)}} = P_N \circ \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N = \text{Id}_{\mathbb{R}^N} \circ T_{\Gamma^{(N)}} \circ P_N = T_{\Gamma^{(N)}} \circ P_N$$

so they commute.

We continue by proving that on $L^2(\mathcal{Z})$, we have $\tilde{T}_{\Gamma^{(N)}} = \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N$. Note

$$\Phi_N \circ T_{\Gamma^{(N)}} \circ P_N = \tilde{T}_{\Gamma^{(N)}} \circ \Phi_N \circ P_N = \tilde{T}_{\Gamma^{(N)}} \circ \text{Id}_{\text{Piece}_N(\mathcal{Z})} = \tilde{T}_{\Gamma^{(N)}}.$$

We end with proving $P_N = \Phi_N^{-1} \circ \text{proj}_{\text{Piece}_N(\mathcal{Z})}$. Let $f \in L^2(\mathcal{Z})$. Then by Definition 5.1.6, $\text{proj}_{\text{Piece}_N(\mathcal{Z})} = \Phi_N \circ P_N$. Then

$$\Phi_N^{-1} \circ \text{proj}_{\text{Piece}_N(\mathcal{Z})} = \Phi_N^{-1} \circ \Phi_N \circ P_N = P_N$$

as desired. Thus the full diagram commutes. $\square$

This commutative diagram clarifies the relationship between the finite-dimensional space $\mathbb{R}^N$ and the subspace $\text{Piece}_N(\mathcal{Z})$. Although these spaces are defined in different ways,

one being a discrete vector space and the other a continuous functional space, the diagram shows that they can be put in one-to-one correspondence via the map Φ_N and are thus identical in their algebraic structure. Specifically this means, that the action of $\tilde{T}_{\Gamma(N)}$ on $\text{Piece}_N(\mathcal{Z})$ is a direct translation of the action of $T_{\Gamma(N)}$ on $\mathbb{R}^N$. This observation will allow us to transfer structural and spectral properties between the two settings. We formalize this correspondence in the following theorem. For this, we use the notation $A \simeq B$ to denote that two normed vector spaces A and B are isomorphic, meaning that there exists a continuous bijection between them with a continuous inverse.

Theorem 5.2.3. *For every $N > 0$, we have $\mathbb{R}^N \simeq \text{Piece}_N(\mathcal{Z})$.*

Proof. To show that $\mathbb{R}^N$ and $\text{Piece}_N(\mathcal{Z})$ are isomorphic, we will prove that $\Phi_N : \mathbb{R}^N \rightarrow \text{Piece}_N(\mathcal{Z})$ is an isomorphism. From Lemma 5.2.2 we obtain

$$\Phi_N \circ \Phi_N^{-1} = \text{Id}_{\mathbb{R}^N}$$

which proves that Φ_N is injective. Since we have proven in Lemma 5.1.3 that

$$\dim(\mathbb{R}^N) = \dim(\text{Piece}_N(\mathcal{Z})) = N < \infty,$$

Φ_N is also surjective and hence a bijection. Linearity of Φ_N proves the result, since linear functions on finite-dimensional spaces are always continuous. $\square$

This theorem proves that the mapping Φ_N is in an isomorphism, establishing that the spaces $\mathbb{R}^N$ and $\text{Piece}_N(\mathcal{Z})$ possess an identical mathematical structure. Beyond the direct equivalence, the commutative diagram also clarifies how the operator $\tilde{T}_{\Gamma(N)}$ interacts with the rest of the function space. The following lemma shows that $\text{Piece}_N(\mathcal{Z})$ is the effective support of the operator's action on $L^2(\mathcal{Z})$.

Corollary 5.2.4. *We have that*

$$\ker(P_N) = \text{Piece}_N(\mathcal{Z})^\perp$$

and then for any $f \in \text{Piece}_N(\mathcal{Z})^\perp$,

$$\tilde{T}_{\Gamma(N)} f = 0.$$

Proof. Φ_N is an isomorphism by Theorem 5.2.3, so we obtain that

$$\ker(P_N) = \ker(\Phi_N^{-1} \circ \text{proj}_{\text{Piece}_N(\mathcal{Z})}) = \ker(\text{proj}_{\text{Piece}_N(\mathcal{Z})}) = \text{Piece}_N(\mathcal{Z})^\perp$$

by definition. Now take any $f \in \text{Piece}_N(\mathcal{Z})^\perp$. By Lemma 5.2.2

$$\begin{aligned} \tilde{T}_{\Gamma(N)} f &= (\Phi_N \circ T_{\Gamma(N)} \circ P_N)(f) \\ &= (\Phi_N \circ T_{\Gamma(N)}) \circ (P_N f) \\ &= 0 \end{aligned}$$

since $\text{Piece}_N(\mathcal{Z})^\perp = \ker(P_N)$. $\square$

These structural properties now provide the necessary foundation to establish a rigorous connection between the discrete and continuous eigenvalue problems. The previously obtained equivalence between $\mathbb{R}^N$ and $\text{Piece}_N(\mathcal{Z})$ essentially allow us to interchange the spectral properties of $T_{\Gamma^{(N)}}$ and $\tilde{T}_{\Gamma^{(N)}}$.

Theorem 5.2.5.

- i) $(\lambda, \mathbf{v})$ is an eigenvalue-eigenvector pair of $T_{\Gamma^{(N)}}$ if and only if $(\lambda, \Phi_N \mathbf{v})$ is an eigenvalue-eigenvector pair of $\tilde{T}_{\Gamma^{(N)}}|_{\text{Piece}_N(\mathcal{Z})}$.
- ii) Every $\xi \in \text{Piece}_N(\mathcal{Z})^\perp$ is an eigenvectors of $\tilde{T}_{\Gamma^{(N)}}$ with eigenvalue 0, which has infinite multiplicity.
- iii) Let $(\lambda_j, \mathbf{v}_j)$, $1 \leq j \leq N$, be eigenpairs of $T_{\Gamma^{(N)}}$, where the eigenvalues are ordered such that $|\lambda_1| \geq \dots \geq |\lambda_N|$. Then the eigenvalues of $\tilde{T}_{\Gamma^{(N)}}$ can be ordered as $|\mu_1| \geq |\mu_2| \geq \dots$, such that for every $1 \leq j \leq N$, $\mu_j = \lambda_j$, and $\Phi_N \mathbf{v}_j$ is an eigenvector of $\tilde{T}_{\Gamma^{(N)}}$ corresponding to μ_j .

Proof.

i) Suppose that $(\lambda, \mathbf{v})$ is an eigenvalue-eigenvector pair of $T_{\Gamma^{(N)}}$ with $\mathbf{v} \neq 0$. It holds that

$$T_{\Gamma^{(N)}} \mathbf{v} = \lambda \mathbf{v}.$$

Now by Lemma 5.2.2

$$\tilde{T}_{\Gamma^{(N)}}(\Phi_N \mathbf{v}) = (\tilde{T}_{\Gamma^{(N)}} \circ \Phi_N) \mathbf{v} = (\Phi_N \circ T_{\Gamma^{(N)}} \circ P_N \circ \Phi_N) \mathbf{v} = \Phi_N(T_{\Gamma^{(N)}} \mathbf{v}) = \lambda(\Phi_N \mathbf{v}).$$

Hence $(\lambda, \Phi_N \mathbf{v})$ is an eigenvalue-eigenvector pair of $\tilde{T}_{\Gamma^{(N)}}$. Now suppose that on $\text{Piece}_N(\mathcal{Z})$, $(\lambda, \Phi_N \mathbf{v})$ is an eigenvalue-eigenvector pair. Now by applying Lemma 5.2.2 we can easily show that

$$\begin{aligned} \lambda \Phi_N \mathbf{v} &= \tilde{T}_{\Gamma^{(N)}}(\Phi_N \mathbf{v}) \\ \lambda(P_N \circ \Phi_N) \mathbf{v} &= (P_N \circ \tilde{T}_{\Gamma^{(N)}} \circ \Phi_N) \mathbf{v} \\ \lambda \mathbf{v} &= T_{\Gamma^{(N)}} \mathbf{v} \end{aligned}$$

since $P_N \circ \Phi_N = \text{Id}_{\mathbb{R}^N}$. This proves (i).

ii) The fact $\xi \in \text{Piece}_N(\mathcal{Z})^\perp$ is an eigenvectors of $\tilde{T}_{\Gamma^{(N)}}$ with eigenvalue 0 follows naturally from Lemma 5.2.4. So see it has infinite multiplicity, recall by Lemma 5.1.3 that

$$\dim(\text{Piece}_N(\mathcal{Z})) = N,$$

and $L^2(\mathcal{Z})$ is infinite-dimensional so

$$\dim(\text{Piece}_N(\mathcal{Z})^\perp) = \infty.$$

iii) Recall by part (i) that for each $1 \leq j \leq N$,

$$T_{\Gamma^{(N)}} \mathbf{v}_j = \lambda_j \mathbf{v}_j \iff \tilde{T}_{\Gamma^{(N)}} (\Phi_N \mathbf{v}_j) = \lambda_j (\Phi_N \mathbf{v}_j).$$

In addition, part (ii) shows that 0 is an eigenvalue of $\tilde{T}_{\Gamma^{(N)}}$ with infinite multiplicity. Since $\dim(\text{Piece}_N(\mathcal{Z})) = N$, the operator $\tilde{T}_{\Gamma^{(N)}}$ has at most N nonzero eigenvalues, and these are precisely coincides with $\{\lambda_1, \dots, \lambda_N\}$, counting multiplicities. Ordering the eigenvalues (μ_j) of $\tilde{T}_{\Gamma^{(N)}}$ by decreasing magnitude, the first N eigenvalues must therefore be exactly the nonzero ones. It follows that

$$\mu_j = \lambda_j, \quad 1 \leq j \leq N.$$

Moreover, for each such j , $\Phi_N \mathbf{v}_j$ is an eigenvector of $\tilde{T}_{\Gamma^{(N)}}$ corresponding to μ_j which proves (iii). $\square$

Remark 9. Recall that we have defined the projection operator $P_N^{\text{avg}} : L^2(\mathcal{Z}) \rightarrow \mathbb{R}^N$ in Equation (5.7) as

$$(P_N^{\text{avg}} f)_j = \frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{z}) d\mathbf{z}.$$

Alternatively, we could also consider the projection operator $P_N^{\text{points}} : L^2(\mathcal{Z}) \rightarrow \mathbb{R}^N$ defined by

$$(P_N^{\text{points}} f)_j := f(\mathbf{z}_j), \quad (5.12)$$

which gives rise to the operator

$$\tilde{T}_{\Gamma^{(N)}}^{\text{points}} := \Phi_N \circ T_{\Gamma^{(N)}} \circ P_N^{\text{points}}.$$

However, working with this operators creates problems as point evaluation is not a continuous linear functional on $L^2(\mathcal{Z})$. In fact, since elements of $L^2(\mathcal{Z})$ are equivalence classes defined up to sets of measure zero, point evaluation is not even well-defined. Consequently, the operator P_N^{points} is not bounded and therefore not continuous as a map $L^2(\mathcal{Z}) \rightarrow \mathbb{R}^N$. Because of this lack of continuity, replacing P_N^{avg} by P_N^{points} breaks the commutative diagram of Lemma 5.2.2, and the results derived from it do not hold in full generality.

In practical computations, however, the exact cell averages

$$\frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{z}), d\mathbf{z}$$

are typically not evaluated analytically. Instead, one naturally works with point evaluations, which leads to the apparent use of P_N^{points} in numerical implementations.

This discrepancy is resolved by noting that, in computational settings, the functions we consider are already discretized on the chosen grid. In particular, they are piecewise constant on a partition $\{C_j^{(N)}\}_{j=1}^N$, i.e., they belong to $\text{Piece}_N(\mathcal{Z})$ for some $N > 0$. For $f \in \text{Piece}_N(\mathcal{Z})$ and evaluation points $\mathbf{z}_j \in C_j^{(N)}$, we have

$$P_N^{\text{avg}} f = P_N^{\text{points}} f.$$

Hence the two operators coincide on $\text{Piece}_N(\mathcal{Z})$, which is what we use in practice.

5.3. CONVERGENCE OF DISCRETIZED OPERATORS

In this section, we establish the convergence properties of the discretized operators. We begin by introducing several technical lemmas regarding the projection operator Π_N and its relationship to the finite-rank operator $\tilde{T}_{\Gamma^{(N)}}$. These results serve as the foundation for proving our main theorem on the consistency of the finite-rank operator $\tilde{T}_{\Gamma^{(N)}}$ with respect to T_{Γ_p} as the partition refinement increases.

Recall that for a set A , $\text{diam}(A)$ is called the diameter of a set and is defined by

$$\text{diam}(A) = \sup_{x, y \in A} \rho(x, y)$$

for the given metric ρ .

Assumption 4. *The partition cells satisfy*

$$\delta_N = \max_j \text{diam}(C_j^{(N)}) \rightarrow 0$$

as $N \rightarrow \infty$.

Assumption 4 is a mild regularity condition satisfied by most standard partition schemes. For example, if $\mathcal{Z}$ is bounded and partitioned into cells of comparable size (such as uniform grids), the diameter of each cell necessarily decreases as the partition is refined.

Lemma 5.3.1. *Let $N > 0$, and let Π_N the orthogonal projection operator as in Equation (5.11). If Assumption 4 holds, then for every $f \in L^2(\mathcal{Z})$,*

$$\Pi_N f \xrightarrow{L^2(\mathcal{Z})} f$$

as $N \rightarrow \infty$. Additionally, if f is continuously differentiable ($f \in C^1(\mathcal{Z})$) with bounded derivatives, the convergence exhibits the rate

$$\|\Pi_N f - f\|_{L^2}^2 = \mathcal{O}(\delta_N^2).$$

Proof. For any $f \in L^2(\mathcal{Z})$, we decompose the error as

$$\|f - \Pi_N f\|_{L^2(\mathcal{Z})} \leq \|f - g\|_{L^2(\mathcal{Z})} + \|g - \Pi_N g\|_{L^2(\mathcal{Z})} + \|\Pi_N g - \Pi_N f\|_{L^2(\mathcal{Z})}.$$

Recall that C_c^∞ is dense in $L^2(\mathcal{Z})$ where

$$C_c^\infty := \{f \in C^\infty(\mathcal{Z}) \mid f \text{ is compactly supported}\}.$$

If we let $\epsilon > 0$, there exists a function $g \in C_c^\infty$ such that

$$\|f - g\|_{L^2(\mathcal{Z})} < \epsilon/3.$$

Next,

$$\|\Pi_N g - \Pi_N f\|_{L^2(\mathcal{Z})} = \|\Pi_N(g - f)\|_{L^2(\mathcal{Z})} \leq \|g - f\|_{L^2(\mathcal{Z})} < \epsilon/3$$

since Π_N is a orthogonal projection operator. For the middle term we can write

$$\begin{aligned}
 \|g - \Pi_N g\|_{L^2(\mathcal{Z})}^2 &= \int_{\mathcal{Z}} (g(\mathbf{z}) - \Pi_N g(\mathbf{z}))^2 d\mathbf{z} \\
 &= \sum_{i=1}^N \int_{C_i^{(N)}} (g(\mathbf{z}) - \bar{g}_i)^2 d\mathbf{z} \\
 &\leq \sum_{i=1}^N \int_{C_i^{(N)}} \sup_{\mathbf{y} \in C_i^{(N)}} |g(\mathbf{z}) - g(\mathbf{y})|^2 d\mathbf{z} \\
 &\leq \sum_{i=1}^N \text{Leb}(C_i^{(N)}) \sup \left\{ |g(\mathbf{z}) - g(\mathbf{y})| : \|\mathbf{z} - \mathbf{y}\| < \text{diam}(C_i^{(N)}) \right\}^2.
 \end{aligned}$$

Since g is uniformly continuous on the compact set $C_i^{(N)}$, we get

$$\sup \left\{ |g(\mathbf{z}) - g(\mathbf{y})| : \|\mathbf{z} - \mathbf{y}\| < \text{diam}(C_i^{(N)}) \right\} \longrightarrow 0$$

because $\delta_N \rightarrow 0$ when $N \rightarrow \infty$ by Assumption 4. Therefore

$$\|g - \Pi_N g\|_{L^2(\mathcal{Z})} < \epsilon/3$$

for sufficiently large N . Combining the bounds gives

$$\|f - \Pi_N f\|_{L^2(\mathcal{Z})} \leq \|f - g\|_{L^2(\mathcal{Z})} + \|g - \Pi_N g\|_{L^2(\mathcal{Z})} + \|\Pi_N g - \Pi_N f\|_{L^2(\mathcal{Z})} < \epsilon$$

and therefore

$$\Pi_N f \xrightarrow{N \rightarrow \infty} f$$

in $L^2(\mathcal{Z})$.

Now, suppose additionally that $f \in C^1(\mathcal{Z})$ with bounded derivatives such that $\sup_{\mathbf{z} \in \mathcal{Z}} \|\nabla f(\mathbf{z})\| \leq M < \infty$. We can directly evaluate $\|f - \Pi_N f\|_{L^2(\mathcal{Z})}^2$ without the density approximation by substituting f for g into the expression above

$$\|f - \Pi_N f\|_{L^2(\mathcal{Z})}^2 \leq \sum_{i=1}^N \text{Leb}(C_i^{(N)}) \sup \left\{ |f(\mathbf{z}) - f(\mathbf{y})| : \|\mathbf{z} - \mathbf{y}\| < \text{diam}(C_i^{(N)}) \right\}^2.$$

By the Mean Value Theorem, for any $\mathbf{z}, \mathbf{y} \in C_i^{(N)}$, we have $|f(\mathbf{z}) - f(\mathbf{y})| \leq M\|\mathbf{z} - \mathbf{y}\| \leq M\delta_N$. Squaring this bound and substituting it back into our summation yields:

$$\begin{aligned}
 \|f - \Pi_N f\|_{L^2(\mathcal{Z})}^2 &\leq \sum_{i=1}^N \text{Leb}(C_i^{(N)}) (M\delta_N)^2 \\
 &= M^2 \delta_N^2 \sum_{i=1}^N \text{Leb}(C_i^{(N)}) = M^2 \text{Leb}(\mathcal{Z}) \delta_N^2.
 \end{aligned}$$

Since $M^2 \text{Leb}(\mathcal{Z})$ is a constant independent of N , this explicitly establishes that

$$\|\Pi_N f - f\|_{L^2}^2 = \mathcal{O}(\delta_N^2).$$

□

While Lemma 5.3.1 establishes the convergence of the projection operator, the approximation of the underlying integral operator requires additional control over the discretization error. To this end, we introduce assumptions on the quadrature rule and the regularity of the partition.

Assumption 5. Suppose that the numerical quadrature rule satisfies

$$\epsilon_N := \sup_{j=1,\dots,N} \sup_{\mathbf{y} \in \mathcal{Z}} \left| \int_{C_j^{(N)}} \Gamma_p(\mathbf{x}, \mathbf{y}) d\mathbf{x} - w_j^{(N)} \Gamma_p(\mathbf{z}_j, \mathbf{y}) \right| = \mathcal{O}\left(\frac{1}{N^2}\right).$$

Assumption 6. The partition cells satisfy, for some constants $0 < a_1 \leq a_2 < \infty$,

$$\frac{a_1}{N} \leq \text{Leb}(C_j^{(N)}) \leq \frac{a_2}{N}, \quad j = 1, \dots, N, \quad \forall N > 0$$

and there exists a constant $C_w > 0$ such that for all N and j ,

$$0 < w_j^{(N)} \leq C_w \text{Leb}(C_j^{(N)}).$$

Remark 10. Assumption 5 is satisfied, for example, if the cells form a regular partition with $\text{diam}(C_j^{(N)}) \leq C_d/N$, the weights match the cell volumes $w_j^{(N)} = \text{Leb}(C_j^{(N)}) \leq a/N$, and the kernel Γ_p is continuously differentiable in its first argument with a uniformly bounded gradient

$$M := \sup_{\mathbf{x}, \mathbf{y} \in \mathcal{Z}} \|\nabla_{\mathbf{x}} \Gamma_p(\mathbf{x}, \mathbf{y})\| < \infty.$$

Indeed, for any fixed $\mathbf{y} \in \mathcal{Z}$ and any cell j , the Mean Value Theorem yields

$$\begin{aligned} \left| \int_{C_j^{(N)}} \Gamma_p(\mathbf{x}, \mathbf{y}) d\mathbf{x} - w_j^{(N)} \Gamma_p(\mathbf{z}_j, \mathbf{y}) \right| &\leq \int_{C_j^{(N)}} \|\nabla_{\mathbf{x}} \Gamma_p(\xi_{\mathbf{x}}, \mathbf{y})\| \|\mathbf{x} - \mathbf{z}_j\| d\mathbf{x} \\ &\leq M \cdot \text{diam}(C_j^{(N)}) \text{Leb}(C_j^{(N)}) \\ &\leq M \left(\frac{C_d}{N}\right) \left(\frac{a}{N}\right) = \mathcal{O}\left(\frac{1}{N^2}\right). \end{aligned}$$

Because this upper bound depends only on global constants (M, C_d, a_2) and is completely independent of the choice of j and $\mathbf{y}$, taking the supremum yields $\epsilon_N = \mathcal{O}(1/N^2)$, as required.

These assumptions ensure that the discrete approximation remains a valid representation of the continuous integral. Assumption 5 controls the local discretization error; it implies that as the cells shrink, the function value at a single representative point $\mathbf{z}_j$ becomes an increasingly accurate approximation of the local integral. Assumption 6 imposes a regularity condition on the geometry of the partition, preventing any single cell from becoming disproportionately large or small. It can be seen as a weaker Assumption than 3. Together, these conditions ensure that no single cell dominates the approximation and that the quadrature rule behaves like a uniform Riemann sum.

Lemma 5.3.2. For $N > 0$, let Π_N the orthogonal projection operator as in Equation 5.11 and let $\tilde{T}_{\Gamma^{(N)}} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ as in Definition 5.1.4. Then under Assumptions 2, 4, 5 and 6, we have

$$\|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} \xrightarrow{N \rightarrow \infty} 0.$$

where $\|\cdot\|_{op}$ denotes the operator norm on $L^2(\mathcal{Z})$. Moreover, if Γ_p has uniformly bounded derivatives on $\mathcal{Z} \times \mathcal{Z}$ the convergence exhibits the rate

$$\|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} = \mathcal{O}\left(\frac{1}{N} + \delta_N\right).$$

Proof. Under Assumption 6 we have that

$$\sum_{i,j}^N \frac{\text{Leb}(C_i^{(N)})}{\text{Leb}(C_j^{(N)})} w_j^2 \leq \sum_{i,j}^N \frac{a_2}{N} \frac{N}{a_1} w_j^2 = \frac{a_2}{a_1} N \sum_{j=1}^N w_j^2 < +\infty.$$

Hence by Lemma 5.1.5, $\tilde{T}_{\Gamma^{(N)}}$ is Hilbert-Schmidt operator with a kernel defined for $\mathbf{x} \in C_i^{(N)}$ and $\mathbf{y} \in C_j^{(N)}$ as

$$\tilde{k}(\mathbf{x}, \mathbf{y}) = \frac{1}{\text{Leb}(C_j^{(N)})} w_j \Gamma_p(\mathbf{z}_i, \mathbf{z}_j).$$

For any $f \in L^2(\mathcal{Z})$ and $\mathbf{z} \in C_i^{(N)}$ we can write

$$(\Pi_N T_{\Gamma_p} \Pi_N f)(\mathbf{z}) = \sum_{j=1}^N \frac{1}{\text{Leb}(C_j^{(N)})} \tilde{f}_j \iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}$$

where

$$\tilde{f}_j = \frac{1}{\text{Leb}(C_j^{(N)})} \int_{C_j^{(N)}} f(\mathbf{u}) d\mathbf{u}$$

is the cell average of f over $C_j^{(N)}$. It then follows that we can rewrite $\Pi_N T_{\Gamma_p} \Pi_N$ as an integral operator with kernel

$$k(\mathbf{x}, \mathbf{y}) = \frac{1}{\text{Leb}(C_i^{(N)}) \text{Leb}(C_j^{(N)})} \iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v}$$

for $\mathbf{x} \in C_i^{(N)}$ and $\mathbf{y} \in C_j^{(N)}$. Indeed, it can be checked that $\mathbf{z} \in C_i^{(N)}$ and $f \in L^2(\mathcal{Z})$

$$(\Pi_N \circ T_{\Gamma_p} \circ \Pi_N) f = \sum_{j=1}^N \frac{1}{\text{Leb}(C_j^{(N)})} \tilde{f}_j \iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} = \int_{\mathcal{Z}} k(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) d\mathbf{y}.$$

To see that this integral operator is in fact Hilbert-Schmidt, we compute

$$\begin{aligned}
 \|(\Pi_N \circ T_{\Gamma_p} \circ \Pi_N) f\|_{HS}^2 &= \iint_{\mathcal{Z} \times \mathcal{Z}} |k(\mathbf{x}, \mathbf{y})|^2 d\mathbf{x} d\mathbf{y} \\
 &= \sum_{i,j}^N \iint_{C_i^{(N)} \times C_j^{(N)}} \frac{1}{\text{Leb}(C_i^{(N)})^2 \text{Leb}(C_j^{(N)})^2} \left(\iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} \right)^2 d\mathbf{x} d\mathbf{y} \\
 &= \sum_{i,j}^N \frac{1}{\text{Leb}(C_i^{(N)}) \text{Leb}(C_j^{(N)})} \left(\iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} \right)^2 \\
 &\leq 16 \text{Leb}(\mathcal{Z})^2 < +\infty,
 \end{aligned}$$

using that $\|\Gamma_p\|_\infty \leq 4$. We can see, since the operators are bounded that

$$\begin{aligned}
 \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - \tilde{T}_{\Gamma^{(N)}}\|_{op}^2 &\leq \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - \tilde{T}_{\Gamma^{(N)}}\|_{HS}^2 \\
 &= \iint_{\mathcal{Z} \times \mathcal{Z}} |k(\mathbf{x}, \mathbf{y}) - \tilde{k}(\mathbf{x}, \mathbf{y})|^2 d\mathbf{x} d\mathbf{y} \\
 &= \sum_{i=1}^N \sum_{j=1}^N \iint_{C_i^{(N)} \times C_j^{(N)}} |k(\mathbf{x}, \mathbf{y}) - \tilde{k}(\mathbf{x}, \mathbf{y})|^2 d\mathbf{x} d\mathbf{y} \\
 &= \sum_{i=1}^N \sum_{j=1}^N \text{Leb}(C_i^{(N)}) \text{Leb}(C_j^{(N)}) |d_{ij}^{(N)}|^2
 \end{aligned} \tag{5.13}$$

where

$$d_{ij}^{(N)} := \frac{1}{\text{Leb}(C_i^{(N)}) \text{Leb}(C_j^{(N)})} \iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} - \frac{1}{\text{Leb}(C_j^{(N)})} w_j \Gamma_p(\mathbf{z}_i, \mathbf{z}_j).$$

We start by looking at the value $\text{Leb}(C_j^{(N)}) d_{ij}$ and rewrite it as

$$\begin{aligned}
 &\frac{1}{\text{Leb}(C_i^{(N)})} \iint_{C_i^{(N)} \times C_j^{(N)}} \Gamma_p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} - w_j \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \\
 &= \frac{1}{\text{Leb}(C_i^{(N)})} \int_{C_i^{(N)}} \underbrace{\left(\int_{C_j^{(N)}} \Gamma_p(\mathbf{x}, \mathbf{y}) d\mathbf{y} - w_j \Gamma_p(\mathbf{x}, \mathbf{z}_j) \right)}_{A(\mathbf{x})} + \underbrace{w_j (\Gamma_p(\mathbf{x}, \mathbf{z}_j) - \Gamma_p(\mathbf{z}_i, \mathbf{z}_j))}_{B(\mathbf{x})} d\mathbf{x}.
 \end{aligned}$$

For $A(\mathbf{x})$, we directly apply the quadrature rule error from Assumption 5 to find that for any $\mathbf{x} \in C_i^{(N)}$:

$$|A(\mathbf{x})| \leq C_1 \epsilon_N = \mathcal{O}\left(\frac{1}{N^2}\right).$$

For $B(\mathbf{x})$, uniform continuity of Γ_p (Assumption 2) gives

$$\begin{aligned}
 |B(\mathbf{x})| &= w_j^{(N)} \left| \Gamma_p(\mathbf{x}, \mathbf{z}_j) - \Gamma_p(\mathbf{z}_i, \mathbf{z}_j) \right| \leq w_j^{(N)} \sup \left\{ |\Gamma_p(\mathbf{x}, \mathbf{z}_j) - \Gamma_p(\mathbf{y}, \mathbf{z}_j)| : \|\mathbf{x} - \mathbf{y}\| < \text{diam}(C_i^{(N)}) \right\} \\
 &=: w_j^{(N)} \gamma(\delta_N).
 \end{aligned}$$

Hence

$$\left| \text{Leb}(C_j^{(N)}) d_{ij}^{(N)} \right| \leq C_1 \epsilon_N + w_j \gamma(\delta_N)$$

and so

$$\left| d_{ij}^{(N)} \right| \leq \frac{C_1 \epsilon_N}{\text{Leb}(C_j^{(N)})} + \frac{w_j^{(N)}}{\text{Leb}(C_j^{(N)})} \gamma(\delta_N).$$

Now applying Assumption 6 we have

$$\frac{\epsilon_N}{\text{Leb}(C_j^{(N)})} \leq \frac{\epsilon_N N}{a_1} \quad \text{and} \quad \frac{w_j^{(N)}}{\text{Leb}(C_j^{(N)})} \gamma(\delta_N) \leq C_w \gamma(\delta_N).$$

Hence combining this gives the uniform bound

$$\sup_{i,j} \left| d_{ij}^{(N)} \right| \leq C' (\epsilon_N N + \gamma(\delta_N)).$$

Now since $\epsilon_N N \rightarrow 0$ and $\gamma(\delta_N) \rightarrow 0$ as $N \rightarrow \infty$ (by Assumption 4) we obtain $\sup_{i,j} \left| d_{ij}^{(N)} \right| \rightarrow 0$. Plugging this into Equation (5.13) yields

$$\left\| \Pi_N \circ T_{\Gamma_p} \circ \Pi_N - \tilde{T}_{\Gamma^{(N)}} \right\|_{op}^2 \leq \text{Leb}(\mathcal{Z})^2 \left(\sup_{i,j} \left| d_{ij}^{(N)} \right| \right)^2 \xrightarrow{N \rightarrow \infty} 0.$$

If we additionally assume that Γ_p has uniformly bounded derivatives, the Mean Value Theorem show $\gamma(\delta_N) \lesssim \delta_N$. Therefore

$$\begin{aligned} \left\| \Pi_N \circ T_{\Gamma_p} \circ \Pi_N - \tilde{T}_{\Gamma^{(N)}} \right\|_{op}^2 &\leq \text{Leb}(\mathcal{Z})^2 \left(\sup_{i,j} \left| d_{ij}^{(N)} \right| \right)^2 \\ &\lesssim \left(\frac{1}{N^2} N + \gamma(\delta_N) \right)^2 \\ &\lesssim \left(\frac{1}{N} + \delta_N \right)^2, \end{aligned}$$

which completes the proof. $\square$

While Lemma 5.3.2 establishes that the finite-rank operator $\tilde{T}_{\Gamma^{(N)}}$ is a consistent approximation of the projected integral operator $\Pi_N \circ T_{\Gamma_p} \circ \Pi_N$, it remains to show that the discretized operator as a whole converges to the original operator T_{Γ_p} . By combining this result with Lemma 5.3.1, we obtain the following main consistency result.

Theorem 5.3.3 (Consistency of the operators). *Let $\tilde{T}_{\Gamma^{(N)}} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ as in Definition 5.1.4. Then under Assumptions 2, 4, 5 and 6,*

$$\left\| \tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p} \right\|_{op} \xrightarrow{N \rightarrow \infty} 0$$

where $\|\cdot\|_{op}$ denotes the operator norm on $L^2(\mathcal{Z})$.

Proof. Recall from Lemma 5.3.1 that the operator $\Pi_N : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ converges to the identity

$$\Pi_N f \xrightarrow{N \rightarrow \infty} f$$

strongly or in the pointwise sense, i.e. for every fixed $f \in L^2(\mathcal{Z})$. Therefore, since T_{Γ_p} is compact, by Lemma 2.6.38 we obtain

$$\|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} \xrightarrow{N \rightarrow \infty} 0.$$

Also Lemma 5.3.2 gives that

$$\|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} \xrightarrow{N \rightarrow \infty} 0.$$

By using the triangle inequality we can conclude

$$\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} \leq \|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} + \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} \xrightarrow{N \rightarrow \infty} 0.$$

□

5.4. CONSISTENCY OF THE EIGENVALUES

Having established the convergence of the operators, we now examine the consistency of their eigenvalues. In this classical PCA setting, the error arises from two distinct sources, the stochastic error from finite sampling and the error from discretizing the continuous domain. We first address the stochastic component, showing that for a fixed partition, the empirical eigenvalues converge to their discretized counterparts as the sample size grows.

Theorem 5.4.1 (Consistency of the eigenvalues). *Let $T_{\Gamma^{(N)}}$ be the deterministic discretized operator with eigenvalues $\{\lambda_i^{(N)}\}_{i=1}^N$ and let $\tilde{T}_{\hat{\Gamma}^{(N)}}$ be the corresponding random operator with eigenvalues $\{\hat{\lambda}_i^{(n,N)}\}_{i=1}^N$. If Assumption 1 and 2 hold and $nh^d \rightarrow +\infty$ as $h \rightarrow 0$, then for every fixed $1 \leq i \leq N$,*

$$\hat{\lambda}_i^{(n,N)} \xrightarrow[n \rightarrow \infty]{prob} \lambda_i^{(N)}.$$

Proof. First recall that the operators $T_{\Gamma^{(N)}}$ and $\tilde{T}_{\hat{\Gamma}^{(N)}}$ can also be written as matrices. From Equation (5.3) it follows that it is sufficient to consider the eigenvalues of $\Gamma^{(N)}W$ and $\hat{\Gamma}^{(N)}W$. However, we will instead look at the eigenproblem

$$\begin{aligned} (W^{1/2} \Gamma^{(N)} W^{1/2})(W^{1/2} \xi_k^{(N)}) &= \lambda_k^{(N)} (W^{1/2} \xi_k^{(N)}) \\ (W^{1/2} \hat{\Gamma}^{(N)} W^{1/2})(W^{1/2} \hat{\xi}_k^{(N)}) &= \hat{\lambda}_k^{(n,N)} (W^{1/2} \hat{\xi}_k^{(N)}), \quad \forall k = 1, \dots, N. \end{aligned}$$

It can be seen that the matrix $W^{1/2} \Gamma^{(N)} W^{1/2}$ is similar to $\Gamma^{(N)}W$, meaning their eigenvalues are exactly the same.

Lemma 2.4.2 shows that $\hat{\tau}_{j_1, j_2 | \mathbf{Z} = \mathbf{z}_j}$ converges pointwise in probability to $\tau_{j_1, j_2 | \mathbf{Z} = \mathbf{z}_j}$ for each $j = 1, \dots, N$. As a result also $\hat{\Gamma}^{(N)}$ converges to $\Gamma^{(N)}$ pointwise in probability since it consists of a finite sum of converging elements. So by definition

$$\|W^{1/2} \hat{\Gamma}^{(N)} W^{1/2} - W^{1/2} \Gamma^{(N)} W^{1/2}\|_F \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Since all norms are equivalent in finite dimensions we obtain that also

$$\|W^{1/2} \hat{\Gamma}^{(N)} W^{1/2} - W^{1/2} \Gamma^{(N)} W^{1/2}\|_2 \xrightarrow[n \rightarrow \infty]{prob} 0$$

where $\|\cdot\|_2$ denotes the spectral norm. Since $W^{1/2} \Gamma^{(N)} W^{1/2}$ and $W^{1/2} \hat{\Gamma}^{(N)} W^{1/2}$ are symmetric matrices, Lemma 2.6.35 then shows

$$|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_2 \quad \forall i = 1, 2, \dots, N.$$

Hence the eigenvalue functional $\lambda_i(\cdot)$ is 1-Lipschitz with respect to the spectral norm. Therefore by applying the continuous mapping theorem

$$\hat{\lambda}_i^{(n, N)} = \lambda_i(\hat{\Gamma}^{(N)} W) = \lambda_i(W^{1/2} \hat{\Gamma}^{(N)} W^{1/2}) \xrightarrow[n \rightarrow \infty]{prob} \lambda_i(W^{1/2} \Gamma^{(N)} W^{1/2}) = \lambda_i(\Gamma^{(N)} W) = \lambda_i^{(N)}$$

which proves the result. $\square$

While the previous result ensures that we can recover the eigenvalues of the discretized system, it does not account for the error introduced by the partition itself. The following theorem completes the consistency analysis by showing that the eigenvalues of the discretized operator converge to the true eigenvalues of the continuous integral operator as the number of grid points increases.

Theorem 5.4.2 (Consistency of the discretization). *Let $T_{\Gamma^{(N)}}$ be the deterministic discretized operator with eigenvalues $\{\lambda_i^{(N)}\}_{i=1}^N$, and let T_{Γ_p} be the continuous integral operator with eigenvalues $\{\lambda_i\}_{i \in \mathbb{N}}$. Then if Assumptions 2, 3, 4, 5 and 6 hold, for any fixed $i \geq 1$,*

$$\lambda_i^{(N)} \xrightarrow[N \rightarrow \infty]{} \lambda_i.$$

Note that for every fixed $i \geq 1$, $\lambda_i^{(N)}$ is defined for every $N \geq i$, so that there is no definition problem in the large N regime.

Proof. By Theorem 5.2.5, the operators $T_{\Gamma^{(N)}}$ and $\tilde{T}_{\Gamma^{(N)}}$ share the same nonzero eigenvalues. In particular,

$$\lambda_i^{(N)} = \lambda_i(T_{\Gamma^{(N)}}) = \lambda_i(\tilde{T}_{\Gamma^{(N)}}).$$

Moreover, by Theorem 5.3.3,

$$\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Since both operators are self-adjoint and compact by Lemma 5.2.1, Lemma 2.6.36 yields

$$|\lambda_i(\tilde{T}_{\Gamma^{(N)}}) - \lambda_i(T_{\Gamma_p})| \leq \|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op}.$$

Combining the above, we obtain

$$\lambda_i^{(N)} = \lambda_i(\tilde{T}_{\Gamma^{(N)}}) \xrightarrow[N \rightarrow \infty]{} \lambda_i(T_{\Gamma_p}) = \lambda_i,$$

which proves the claim. $\square$

The previous two theorems can now be combined into the following corollary on the full consistency of the eigenvalues.

Corollary 5.4.3 (Full consistency). *Under the assumptions of Theorems 5.4.1 and 5.4.2, for every fixed $i \geq 1$,*

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \hat{\lambda}_i^{(n,N)} = \lambda_i,$$

5.5. CONSISTENCY OF THE EIGENVECTORS

After having established the convergence of the eigenvalues we continue by examining the consistency of the eigenvectors of the operators. Again, the error in this setting arises from the stochastic error from finite sampling and the error which comes with discretizing the domain. We will start by examining the influence of the stochastic error, showing that the eigenvectors converge up to a sequence of signs, as the sample size increases.

Theorem 5.5.1 (Consistency of the eigenvectors). *Let $T_{\Gamma^{(N)}}$ be the deterministic discretized operator and let $T_{\hat{\Gamma}^{(N)}}$ be the corresponding random operator from Equation (5.1). Denote $\{\lambda_i^{(N)}\}_{i=1}^N, \{\hat{\lambda}_i^{(n,N)}\}_{i=1}^N$ the eigenvalues of $T_{\Gamma^{(N)}}$ and $T_{\hat{\Gamma}^{(N)}}$ corresponding to the eigenvectors $\{\xi_i^{(N)}(\cdot)\}_{i=1}^N$ and $\{\hat{\xi}_i^{(n,N)}(\cdot)\}_{i=1}^N$ respectively as defined in Equation (5.2). If Assumption 1 and 2 hold and $nh^d \rightarrow +\infty$ and $h \rightarrow 0$, then for any $i = 1, \dots, N$ such that $\lambda_i^{(N)}$ and $\hat{\lambda}_i^{(n,N)}$ are simple, then there exists $(s_i^{(n,N,\omega)} \in \{-1, 1\})_{n \geq 1, \omega \in \Omega}$, such that*

$$\|\hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \xi_i^{(N)}\|_2 \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Proof. First recall that the operators $T_{\Gamma^{(N)}}$ and $T_{\hat{\Gamma}^{(N)}}$ can also be written as matrices. From Equation (5.3) it follows that it is sufficient to consider the eigenvectors of $\Gamma^{(N)}W$ and $\hat{\Gamma}^{(N)}W$. However, we will instead be looking at the eigenproblem

$$\begin{aligned} (W^{1/2} \Gamma^{(N)} W^{1/2})(W^{1/2} \xi_k^{(N)}) &= \lambda_k^{(N)} (W^{1/2} \xi_k^{(N)}) \\ (W^{1/2} \hat{\Gamma}^{(N)} W^{1/2})(W^{1/2} \hat{\xi}_k^{(N)}) &= \hat{\lambda}_k^{(N)} (W^{1/2} \hat{\xi}_k^{(N)}), \quad \forall k = 1, \dots, N. \end{aligned}$$

Denote $S^{(N)} = W^{1/2} \Gamma^{(N)} W^{1/2}$ and $\mathbf{v}_k = W^{1/2} \xi_k^{(N)}$ then the eigenproblem rewrites as

$$S^{(N)} \mathbf{v}_k^{(N)} = \lambda_k^{(N)} \mathbf{v}_k \quad \text{and} \quad \hat{S}^{(N)} \hat{\mathbf{v}}_k^{(n,N)} = \hat{\lambda}_k^{(N)} \hat{\mathbf{v}}_k.$$

Note that $S^{(N)}$ is similar to $\Gamma^{(N)}W$, meaning their eigenvalues are exactly the same.

Lemma 2.4.2 shows that under the given conditions $\hat{\tau}_{j_1, j_2 | \mathbf{Z} = \mathbf{z}_j}$ converges pointwise in probability to $\tau_{j_1, j_2 | \mathbf{Z} = \mathbf{z}_j}$ for each grid point $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$. Because $\Gamma^{(N)}$ and $\hat{\Gamma}^{(N)}$ are a finite

linear combinations these also converge elementwise in probability. Hence by definition of the Frobenius norm

$$\|\widehat{S}^{(N)} - S^{(N)}\|_F \xrightarrow[n \rightarrow \infty]{prob} 0.$$

Now by the equivalence of norm in finite dimensions we also obtain

$$\|\widehat{S}^{(N)} - S^{(N)}\|_2 \xrightarrow[n \rightarrow \infty]{prob} 0,$$

where $\|\cdot\|_2$ denotes the spectral norm. Since $\lambda_i^{(N)}$ is simple

$$\min\left(\lambda_{i-1}^{(N)} - \lambda_i^{(N)}, \lambda_i^{(N)} - \lambda_{i+1}^{(N)}\right) > 0,$$

where $\lambda_0^{(N)} = \infty$ and $\lambda_{N+1}^{(N)} = -\infty$. Since $S^{(N)}$ and $\widehat{S}^{(N)}$ are symmetric matrices, they have orthonormal eigenvectors. Choose $s_i^{(n,N,\omega)} \in \{\pm 1\}$ such that $\langle \widehat{\mathbf{v}}_i^{(n,N)}, s_i^{(n,N,\omega)} \mathbf{v}_i^N \rangle_{\mathbb{R}^N} \geq 0$. Since, the matrices are symmetric, we can apply Theorem 2.6.37 to obtain

$$\|\widehat{\mathbf{v}}_i^{(n,N)} - s_i^{(n,N,\omega)} \mathbf{v}_i^{(N)}\|_2 \leq \frac{2^{3/2} \|\widehat{S}^{(N)} - S^{(N)}\|_2}{\min\left(\lambda_{i-1}^{(N)} - \lambda_i^{(N)}, \lambda_i^{(N)} - \lambda_{i+1}^{(N)}\right)} \xrightarrow[n \rightarrow \infty]{prob} 0.$$

We can now map them back to the eigenvectors using $\boldsymbol{\xi}_i^{(N)} = W^{-1/2} \mathbf{v}_i^{(N)}$. We obtain

$$\|\widehat{\boldsymbol{\xi}}_i^{(n,N)} - s_i^{(n,N,\omega)} \boldsymbol{\xi}_i^{(N)}\|_2 \leq \|W^{-1/2}\|_2 \|\widehat{\mathbf{v}}_i^{(n,N)} - s_i^{(n,N,\omega)} \mathbf{v}_i^{(N)}\|_2 \xrightarrow[n \rightarrow \infty]{prob} 0.$$

This completes the proof. $\square$

The previous lemma establishes that the random eigenvectors converge to their deterministic counterparts in $\mathbb{R}^N$, effectively showing that the stochastic error vanishes asymptotically. To complete the argument, it remains to control the discretization error. To problem that arises is that the eigenvectors of T_{Γ_p} and $T_{\Gamma^{(N)}}$ reside in different spaces, $L^2(\mathcal{Z})$ and $\mathbb{R}^N$. To facilitate a direct comparison, we lift the discrete eigenvectors into $L^2(\mathcal{Z})$ via the isomorphism $\Phi_N : \mathbb{R}^N \rightarrow \text{Piece}_N(\mathcal{Z}) \subset L^2(\mathcal{Z})$. This allows us to compare objects in a common space.

Theorem 5.5.2 (Consistency of the discretization). *Let $T_{\Gamma^{(N)}}$ be the deterministic discretized operator with eigenvectors $\{\boldsymbol{\xi}_i^{(N)}\}_{i=1}^N$ corresponding to the eigenvalues $\{\lambda_i^{(N)}\}_{i=1}^N$. Let T_{Γ_p} be the continuous integral operator with eigenvectors $\{\xi_i\}_{i \in \mathbb{N}}$ corresponding to eigenvalues $\{\lambda_i\}_{i \in \mathbb{N}}$. Let $\Phi_N : \mathbb{R}^N \rightarrow \text{Piece}_N(\mathcal{Z})$ as in Equation (5.8) be the isomorphism between $\mathbb{R}^N$ and $\text{Piece}_N(\mathcal{Z})$. If Assumptions 2, 3, 4, 5 and 6 hold, then there exists a sequence of signs $s_i^{(N)} \in \{\pm 1\}$ such that for any fixed $i \geq 1$ where $\lambda_i^{(N)}$ and λ_i are simple*

$$\|\Phi_N \boldsymbol{\xi}_i^{(N)} - s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})} \xrightarrow[N \rightarrow \infty]{} 0.$$

Proof. Let $(\lambda_i^{(N)}, \boldsymbol{\xi}_i^{(N)})$ be an i 'th eigenpair of $T_{\Gamma^{(N)}}$. Then by Theorem 5.2.5, the eigenvalues of $\widetilde{T}_{\Gamma^{(N)}}$ can also be ordered as $|\mu_1| \geq |\mu_2| \geq \dots$. In particular, $\mu_i = \lambda_i^{(N)}$ and $\Phi_N \boldsymbol{\xi}_j^{(N)}$

is an eigenvector of $\tilde{T}_{\Gamma^{(N)}}$ corresponding to μ_i . Also note that since the eigenvalue $\lambda_i^{(N)}$ of $T_{\Gamma^{(N)}}$ is simple, also the eigenvalue μ_i of $\tilde{T}_{\Gamma^{(N)}}$ is simple (by Theorem 5.2.5). We view $\tilde{T}_{\Gamma^{(N)}}$ as an operator on $L^2(\mathcal{Z})$ by identifying $\text{Piece}_N(\mathcal{Z}) \subset L^2(\mathcal{Z})$, so that $\tilde{T}_{\Gamma^{(N)}} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$. By construction, T_{Γ_p} is compact and self-adjoint, and under Assumption 3, the operator $\tilde{T}_{\Gamma^{(N)}}$ is also compact and self-adjoint (see Lemma 5.2.1). Furthermore, by Theorem 5.3.3,

$$\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} \xrightarrow{N \rightarrow \infty} 0.$$

Let $\delta := \min_{\mu \in \sigma(T_{\Gamma_p}) \setminus \{\lambda_i\}} |\mu - \lambda_i| > 0$, which is strictly positive since λ_i is simple. Then for all sufficiently large N , we have

$$\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} < \delta/4.$$

We may therefore apply Lemma 4.1.3 to the operators T_{Γ_p} and $\tilde{T}_{\Gamma^{(N)}}$. It follows that there exists a sequence of signs $s_i^{(N)} \in \{\pm 1\}$ such that

$$\|\Phi_N \hat{\xi}_i^{(N)} - s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})} \leq \frac{4\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op}}{\delta - 2\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op}} \xrightarrow{N \rightarrow \infty} 0,$$

which concludes the proof. $\square$

Note that $\tilde{s}_i^{(n,N,\omega)} := s_i^{(n,N,\omega)} s_i^{(N)}$ belongs to $\{\pm 1\}$ and satisfies

$$\begin{aligned} & \|\Phi_N \hat{\xi}_i^{(n,N)} - \tilde{s}_i^{(n,N,\omega)} \xi_i\|_{L^2(\mathcal{Z})} \\ &= \|\Phi_N \hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \Phi_N \xi_i^{(N)} + s_i^{(n,N,\omega)} \Phi_N \xi_i^{(N)} - s_i^{(n,N,\omega)} s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})} \\ &\leq \|\Phi_N \hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \Phi_N \xi_i^{(N)}\|_{L^2(\mathcal{Z})} + \|s_i^{(n,N,\omega)} \Phi_N \xi_i^{(N)} - s_i^{(n,N,\omega)} s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})} \\ &\leq \|\Phi_N \hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \Phi_N \xi_i^{(N)}\|_{L^2(\mathcal{Z})} + \|\Phi_N \xi_i^{(N)} - s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})} \\ &= \|\hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \xi_i^{(N)}\|_2 + \|\Phi_N \xi_i^{(N)} - s_i^{(N)} \xi_i\|_{L^2(\mathcal{Z})}, \end{aligned} \quad (5.14)$$

such that the previous two theorems can now be combined into the following corollary on the full consistency of the eigenvectors.

Corollary 5.5.3 (Full eigenvector consistency). *Under the assumptions of both Theorem 5.5.1 and Theorem 5.5.2, we have*

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \min_{s \in \{-1, 1\}} \|\Phi_N \hat{\xi}_i^{(n,N)} - s \xi_i\|_{L^2(\mathcal{Z})} = 0,$$

or, equivalently, there exists $\tilde{s}_i^{(n,N,\omega)} \in \{\pm 1\}$ such that

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \|\Phi_N \hat{\xi}_i^{(n,N)} - \tilde{s}_i^{(n,N,\omega)} \xi_i\|_{L^2(\mathcal{Z})} = 0,$$

where the limits are to be understood in the sense of convergence in probability.

5.6. CONVERGENCE RATE

In this section, we establish the convergence rate for the discretized covariance operator. These results provide theoretical guarantees for our scheme, characterizing the exact rate at which the numerical approximation converges to the true infinite-dimensional operator. By assuming stronger regularity conditions on the kernel, we derive an explicit convergence rate in the operator norm. Particularly, we need the partial derivatives of Γ_p to be uniformly bounded on $\mathcal{Z} \times \mathcal{Z}$. We can show that this holds under some conditions.

Lemma 5.6.1. *Let $f_{\mathbf{X},\mathbf{Z}}$ denote their joint density function. Assume that:*

1. $f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ is continuously differentiable with respect to $\mathbf{z}$.
2. The marginal density of $\mathbf{Z}$ is strictly bounded away from zero on $\mathcal{Z}$, such that $f_{\mathbf{Z}}(\mathbf{z}) \geq c_0 > 0$ for any $\mathbf{z} \in \mathcal{Z}$.
3. For $k \in \{0, 1\}$ we have $\sup_{\mathbf{z} \in \mathcal{Z}} \|\nabla_{\mathbf{z}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})\| \leq M_k(\mathbf{x})$ with $\int M_k(\mathbf{x}) d\mathbf{x} < +\infty$.

Then, $\Gamma_p(\mathbf{z}, \mathbf{z}')$ is Lipschitz continuous on $\mathcal{Z} \times \mathcal{Z}$.

Proof. We start by proving that $\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}}$ is Lipschitz continuous for any pair (j_1, j_2) and then it will follow that also Γ_p is Lipschitz. Recall by Definition 2.2.5 that

$$\tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} = \mathbb{E} \left[\text{sign} \left((X_{j_1} - X'_{j_1})(X_{j_2} - X'_{j_2}) \right) \middle| \mathbf{Z} = \mathbf{Z}' = \mathbf{z} \right].$$

Let $\mathbf{x} = (x_1, x_2)^\top$ and $\mathbf{x}' = (x'_1, x'_2)^\top$ denote realizations of two independent and identically distributed copies of $(X_{j_1}, X_{j_2})^\top \in \mathbb{R}^2$ given $\mathbf{Z} = \mathbf{z}$. We define the kernel function $H: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \{-1, 0, 1\}$ as

$$H(\mathbf{x}, \mathbf{x}') = \text{sign}((x_1 - x'_1)(x_2 - x'_2)).$$

Note that $H(\mathbf{x}, \mathbf{x}')$ is bounded, satisfying $|H(\mathbf{x}, \mathbf{x}')| \leq 1$ everywhere on its domain. By expanding the conditional expectations using the conditional density we obtain

$$\begin{aligned} \tau_{j_1, j_2 | \mathbf{Z}=\mathbf{z}} &= \iint_{\mathbb{R}^2 \times \mathbb{R}^2} \frac{H(\mathbf{x}, \mathbf{x}') f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}', \mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})^2} d\mathbf{x} d\mathbf{x}' \\ &= \frac{\iint_{\mathbb{R}^2 \times \mathbb{R}^2} H(\mathbf{x}, \mathbf{x}') f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}', \mathbf{z}) d\mathbf{x} d\mathbf{x}'}{\left(\int_{\mathbb{R}^2} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) d\mathbf{x} \right)^2} \\ &=: \frac{N(\mathbf{z})}{D(\mathbf{z})}. \end{aligned}$$

Let ∂_{z_i} denote the partial derivative with respect to the i -th component of $\mathbf{z}$ for $i \in \{1, \dots, d\}$. By assumption $f_{\mathbf{X},\mathbf{Z}} \in C^3(\mathcal{Z})$, its partial derivatives exist and are continuous. Hence, the partial derivative of the integrand of $N(\mathbf{z})$ is given by

$$\Psi_i(\mathbf{x}, \mathbf{x}', \mathbf{z}) = H(\mathbf{x}, \mathbf{x}') \left[\partial_{z_i} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}', \mathbf{z}) + f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}) \partial_{z_i} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}', \mathbf{z}) \right].$$

Applying the envelope conditions from Assumption (3), where $M_0(\mathbf{x})$ dominates the density and $M_1(\mathbf{x})$ dominates its first partial derivatives, we can bound the magnitude of Ψ_i uniformly for all $\mathbf{z} \in \mathcal{Z}$

$$|\Psi_i(\mathbf{x}, \mathbf{x}', \mathbf{z})| \leq M_1(\mathbf{x})M_0(\mathbf{x}') + M_0(\mathbf{x})M_1(\mathbf{x}') = \tilde{M}(\mathbf{x}, \mathbf{x}').$$

Because M_0 and M_1 are integrable on $\mathbb{R}^2$, we have

$$\iint_{\mathbb{R}^2 \times \mathbb{R}^2} \tilde{M}(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' < +\infty.$$

Since the partial derivatives of the integrand are uniformly dominated by the integrable function $G(\mathbf{x}, \mathbf{x}')$, $N(\mathbf{z})$ is differentiable and

$$\nabla_{\mathbf{z}} N(\mathbf{z}) = \iint_{\mathbb{R}^2 \times \mathbb{R}^2} H(\mathbf{x}, \mathbf{x}') \left[\partial_{z_i} f_{\mathbf{x}, \mathbf{z}}(\mathbf{x}, \mathbf{z}) f_{\mathbf{x}, \mathbf{z}}(\mathbf{x}', \mathbf{z}) + f_{\mathbf{x}, \mathbf{z}}(\mathbf{x}, \mathbf{z}) \partial_{z_i} f_{\mathbf{x}, \mathbf{z}}(\mathbf{x}', \mathbf{z}) \right] d\mathbf{x} d\mathbf{x}'.$$

Furthermore, the Dominated Convergence Theorem ensures that $\nabla_{\mathbf{z}} N(\mathbf{z})$ is continuous, so $N(\mathbf{z}) \in C^1(\mathcal{Z})$. By a similar argument we can show that also $D(\mathbf{z}) \in C^2(\mathcal{Z})$.

By Assumption (2), $f_{\mathbf{z}}(\mathbf{z}) \geq c_0 > 0$ meaning $D(\mathbf{z})$ never vanishes. The quotient rule means that $\tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}}$ is continuously differentiable on $\mathcal{Z}$ with

$$\nabla_{\mathbf{z}} \tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}} = \frac{\nabla_{\mathbf{z}} N(\mathbf{z}) D(\mathbf{z}) + N(\mathbf{z}) \nabla_{\mathbf{z}} D(\mathbf{z})}{(D(\mathbf{z}))^2}.$$

Since $\mathcal{Z}$ is compact and $\nabla_{\mathbf{z}} \tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}}$ is continuous, the Extreme Value Theorem states that the finite maximum value is achieved on $\mathcal{Z}$. We define this uniform bound as

$$L := \sup_{\mathbf{z} \in \mathcal{Z}} \|\nabla \tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}}\| < \infty.$$

By again applying the Mean Value Theorem we obtain

$$|\tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}} - \tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}'}| \leq \|\nabla_{\mathbf{z}} \tau_{j_1, j_2 | \mathbf{z}=\xi}\| \|\mathbf{z} - \mathbf{z}'\| \leq L \|\mathbf{z} - \mathbf{z}'\|.$$

Therefore, $\tau_{j_1, j_2 | \mathbf{z}=\mathbf{z}}$ is Lipschitz continuous on $\mathcal{Z}$ with constant L . Since Γ_p is a finite linear combination of τ 's it follows that also Γ_p is Lipschitz continuous on $\mathcal{Z} \times \mathcal{Z}$. This completes the proof. $\square$

Lemma 5.6.1 shows that under the given conditions, Γ_p is Lipschitz continuous. The Mean Value Theorem that yields that its partial derivatives are uniformly bounded on $\mathcal{Z} \times \mathcal{Z}$. For convenience we write the conditions of Lemma 5.6.1 as an assumption.

Assumption 7. $f_{\mathbf{x}, \mathbf{z}}$ is continuously differentiable on $\mathcal{Z}$ and for $k \in \{0, 1\}$ we have $\sup_{\mathbf{z} \in \mathcal{Z}} \|\nabla_{\mathbf{z}}^k f_{\mathbf{x}, \mathbf{z}}(\mathbf{x}, \mathbf{z})\| \leq M_k(\mathbf{x})$ with $\int M_k(\mathbf{x}) d\mathbf{x} < +\infty$.

Assumption 8. Suppose there exist two positive constants $f_{\mathbf{z}, \min}$ and $f_{\mathbf{z}, \max}$ such that for any $\mathbf{z} \in \mathcal{Z}$, $f_{\mathbf{z}, \min} \leq f_{\mathbf{z}}(\mathbf{z}) \leq f_{\mathbf{z}, \max}$.

Remark 11. Note that Assumption 7 is for example satisfied when the support of $\mathbf{X}$, $\mathcal{X}$ is compact. Indeed a supremum on a compact domain $\mathcal{Z}$ is always attained on $\mathcal{Z}$. Also Assumption 8 is most often satisfied with the commonly used continuous distributions.

Theorem 5.6.2. Let $\tilde{T}_{\Gamma(N)} : L^2(\mathcal{Z}) \rightarrow \text{Piece}_N(\mathcal{Z})$ as in Definition 5.1.4 and $T_{\Gamma_p} : L^2(\mathcal{Z}) \rightarrow L^2(\mathcal{Z})$ as in Equation 3.3. Then under Assumptions 2, 3, 4, 5, 6, 7 and 8 we obtain

$$\|\tilde{T}_{\Gamma(N)} - T_{\Gamma_p}\|_{op} = \mathcal{O}\left(\frac{1}{N^{1/d}}\right).$$

Proof. By applying the triangle inequality, we decompose the operator difference into

$$\begin{aligned} \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} &= \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p} \circ \Pi_N + T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} \\ &\leq \|(\Pi_N - I) \circ T_{\Gamma_p} \circ \Pi_N\|_{op} + \|T_{\Gamma_p} \circ (\Pi_N - I)\|_{op}. \end{aligned}$$

Since Π_N is an orthogonal projection, its operator norm is bounded by 1. Therefore, the first term satisfies

$$\|(\Pi_N - I) \circ T_{\Gamma_p} \circ \Pi_N\|_{op} \leq \|(\Pi_N - I) \circ T_{\Gamma_p}\|_{op}.$$

For any $f \in L^2(\mathcal{Z})$, let $g = T_{\Gamma_p} f$. By the definition

$$g(\mathbf{z}) = \int_{\mathcal{Z}} \Gamma_p(\mathbf{z}, \mathbf{z}') f(\mathbf{z}') d\mathbf{z}'.$$

Differentiating with respect to $\mathbf{z}$ and applying the Cauchy-Schwarz inequality yields

$$\|\nabla g(\mathbf{z})\| \leq \int_{\mathcal{Z}} \|\nabla_{\mathbf{z}} \Gamma_p(\mathbf{z}, \mathbf{z}')\| |f(\mathbf{z}')| d\mathbf{z}' \leq \left(\int_{\mathcal{Z}} \|\nabla_{\mathbf{z}} \Gamma_p(\mathbf{z}, \mathbf{z}')\|^2 d\mathbf{z}' \right)^{1/2} \|f\|_{L^2(\mathcal{Z})}.$$

By Assumptions 7 and 8, Lemma 5.6.1 yields that the partial derivatives of the kernel Γ_p are uniformly bounded since Γ_p is Lipschitz. This implies that also the gradient of g is uniformly bounded by some constant M . Applying Lemma 5.3.1 directly to g gives

$$\|(\Pi_N - I) \circ T_{\Gamma_p} f\|_{L^2(\mathcal{Z})} = \|\Pi_N g - g\|_{L^2(\mathcal{Z})} = \mathcal{O}(\delta_N).$$

Taking the supremum over all $\|f\|_{L^2} = 1$, gives that $\|(\Pi_N - I) \circ T_{\Gamma_p}\|_{op} = \mathcal{O}(\delta_N)$.

Since T_{Γ_p} and Π_N are self-adjoint, the second term yields a similar rate

$$\|T_{\Gamma_p} \circ (\Pi_N - I)\|_{op} = \|(\Pi_N - I) \circ T_{\Gamma_p}\|_{op} = \mathcal{O}(\delta_N).$$

Combining both bounds

$$\|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} \leq \|(\Pi_N - I) \circ T_{\Gamma_p} \circ \Pi_N\|_{op} + \|T_{\Gamma_p} \circ (\Pi_N - I)\|_{op} = \mathcal{O}(\delta_N). \quad (5.15)$$

Lemma 5.3.2 has shown that

$$\|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} = \mathcal{O}\left(\frac{1}{N} + \delta_N\right). \quad (5.16)$$

Now combining Equations (5.15) and (5.16) and applying the triangle inequality, yields

$$\begin{aligned} \|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} &\leq \|\tilde{T}_{\Gamma^{(N)}} - \Pi_N \circ T_{\Gamma_p} \circ \Pi_N\|_{op} + \|\Pi_N \circ T_{\Gamma_p} \circ \Pi_N - T_{\Gamma_p}\|_{op} \\ &= \mathcal{O}\left(\frac{1}{N} + \delta_N\right). \end{aligned}$$

By Assumption 6, the cell volumes scale as $\mathcal{O}(1/N)$, meaning their diameters scale as $\delta_N = \mathcal{O}(1/N^{1/d})$. For any spatial dimension $d \geq 1$, we have $\frac{1}{N} \leq \frac{1}{N^{1/d}}$, allowing us to absorb the $1/N$ term and conclude

$$\|\tilde{T}_{\Gamma^{(N)}} - T_{\Gamma_p}\|_{op} = \mathcal{O}\left(\frac{1}{N^{1/d}}\right),$$

which completes the proof. $\square$

Theorem 5.6.2 isolates the pure approximation error introduced by projecting the continuous operator onto a finite grid of size N . The rate $\mathcal{O}(1/N^{1/d})$ highlights the classical curse of dimensionality. In a one-dimensional setting, the error decays linearly with the number of grid points. However, as the dimension d of the covariate space grows, the volume of the domain scales exponentially, meaning a much denser grid is required to maintain the same level of spatial resolution.

With this discretization rate established, we now transition to the practical setting in which we estimate the covariance kernel by $\hat{\Gamma}^{(N)}$. The following theorem combines the numerical and statistical error, showing how the total operator error breaks down when evaluating the empirically estimated operator $\tilde{T}_{\hat{\Gamma}^{(N)}}$.

Theorem 5.6.3. *Under Assumptions 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^{2d}/|\log h| \rightarrow +\infty$,*

$$\|T_{\Gamma_p} - \tilde{T}_{\hat{\Gamma}^{(N)}}\|_{op} = \mathcal{O}_P\left(\frac{1}{N^{1/d}} + h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

Proof. We can combine Theorem 5.6.2 and Theorem 4.4.1 to obtain the result

$$\begin{aligned} \|T_{\Gamma_p} - \tilde{T}_{\hat{\Gamma}^{(N)}}\|_{op} &\leq \|T_{\Gamma_p} - \tilde{T}_{\Gamma^{(N)}}\|_{op} + \|\tilde{T}_{\Gamma^{(N)}} - \tilde{T}_{\hat{\Gamma}^{(N)}}\|_{op} \\ &= \mathcal{O}_P\left(\frac{1}{N^{1/d}} + h^2 + \frac{1}{\sqrt{nh^d}}\right). \end{aligned}$$

$\square$

The explicit rate derived in Theorem 5.6.3 characterizes structural asymmetry in how the numerical resolution parameter N and the statistical smoothing parameters h_n and n govern the overall operator error. Note that the discretization error term, $\mathcal{O}(1/N^{1/d})$, acts as a stand-alone bottleneck decoupled from the stochastic components of the model.

The choice of N does not induce a bias-variance trade-off. An arbitrarily dense grid ($N \rightarrow \infty$) reduces the discretization error toward zero without inflating the stochastic variance or introducing systematic estimation bias. Consequently, from a theoretical standpoint, maximizing the grid resolution is always beneficial. The only constraint on N is computational complexity, as a larger number of partition cells greatly increase the complexity of the problem especially when the dimension d is high.

Conversely, the bandwidth parameter $h = h_n$ is subject to the classical bias-variance trade-off as already discussed in Section 4.4. Choosing a bandwidth that is too small minimizes the localized smoothing bias, $\mathcal{O}(h_n^2)$, but simultaneously causes the stochastic variance term $\mathcal{O}((nh_n^d)^{-1/2})$ to diverge. On the other hand, choosing h_n too large, stabilizes the variance but inflates the overall bias resulting in over-smoothing.

Recall that in Section 4.4 we computed that the optimal bandwidth has a scaling of $h_n \asymp n^{-\frac{1}{4+d}}$ if $d < 4$. In that case

$$\|T_{\Gamma_p} - \tilde{T}_{\tilde{\Gamma}(N)}\|_{op} = \mathcal{O}_P\left(\frac{1}{N^{1/d}} + n^{-\frac{2}{4+d}}\right). \quad (5.17)$$

To ensure that the numerical discretization scheme does not act as the primary asymptotic bottleneck of the estimator, the grid resolution must scale at a rate that exceeds the decay rate of the statistical sampling error. Specifically

$$\frac{1}{N^{1/d}} \lesssim n^{-\frac{2}{4+d}} \implies N \gtrsim n^{\frac{2d}{4+d}}.$$

This condition shows that in low dimensional cases ($d < 4$), N only needs to grow sub-linear with respect to the sample size n . However, as the dimension increases, the required grid density must scale progressively faster.

A consequence of Theorem 5.6.3 is that we can establish convergence rates for the components of the spectral decomposition. Specifically, applying Lemma 2.6.36 immediately yields the convergence rate for the estimated eigenvalues.

Corollary 5.6.4. *Let $\hat{\lambda}_i^{(n,N)}$ and λ_i denote the i -th sorted eigenvalues of $\tilde{T}_{\tilde{\Gamma}(N)}$ and T_{Γ_p} respectively. Then under the assumptions of Theorem 5.6.3,*

$$|\hat{\lambda}_i^{(n,N)} - \lambda_i| = \mathcal{O}_P\left(\frac{1}{N^{1/d}} + h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

Recall that in Corollary 5.4.3 we also obtain the convergence of the eigenvalues. Specifically we established that

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} \hat{\lambda}_i^{(n,N)} = \lambda_i.$$

Corollary 5.6.4 strengthens Corollary 5.4.3 by establishing an explicit convergence rate. As a consequence, it immediately yields consistency under joint asymptotics $(n, N) \rightarrow \infty$, whereas the earlier result only considered the iterated limit.

We end this chapter with the stability of the eigenvectors. Provided that the eigenvalues are isolated from the rest of the spectrum, we can control the discretized estimated eigenvectors. Combining the decomposition of Equation (5.14) with Lemma 4.1.3 yields the following corollary.

Corollary 5.6.5. *Let $\hat{\xi}_i^{(n,N)}$ and ξ_i be the i -th eigenvectors of $\tilde{T}_{\hat{\Gamma}^{(N)}}$ and T_{Γ_p} respectively such that the corresponding eigenvalues are simple. Then, under the assumptions of Theorem 5.6.3 there exists a sequence of signs $s_i^{(n,N,\omega)} \in \{-1, 1\}$ such that*

$$\|\Phi_N \hat{\xi}_i^{(n,N)} - s_i^{(n,N,\omega)} \xi_i\|_{L^2(\mathcal{Z})} = \mathcal{O}_P\left(\frac{1}{N^{1/d}} + h^2 + \frac{1}{\sqrt{nh^d}}\right).$$

6

ON CONDITIONAL U -STATISTICS

In this chapter, we further explore the theoretical properties of conditional U -statistics. We begin with establishing a generalization on the probability bound from Lemma 2.7.4 that serves as a key technical tool for our subsequent results.

Let $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ be an i.i.d. sample from a distribution F . We define the U -statistic of order k , denoted by U_n , as in Definition 2.7.1,

$$U_n = U(\mathbf{X}_1, \dots, \mathbf{X}_n) = |\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}),$$

where $\mathcal{J}_{k,n}$ denotes the set of injective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$ and $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ a measurable function.

Lemma 6.0.1. *Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ be a measurable function such that $a \leq g(\mathbf{x}_1, \dots, \mathbf{x}_k) \leq b$ and let U_n a U -statistic of order k . Set $\theta = \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]$ and $\sigma^2 = \text{Var}(g(\mathbf{X}_1, \dots, \mathbf{X}_k))$. Then for $k, m < n$*

$$\begin{aligned} & \mathbb{P}\left(U_n - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] \geq t \mid X_1, \dots, X_m\right) \\ & \leq \exp\left(-\frac{(n-m)\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)^2}{2k\left(\sigma^2 + 1/3(b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)\right)}\right) \\ & + \sum_{s=1}^{k-1} \exp\left(-\frac{(n-m)(t_s^*)^2}{2(k-s)\left(\tilde{\sigma}^2 + (1/3)(b - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)])\right)t_s^*}\right), \end{aligned}$$

with

$$\begin{aligned} t_s^* &= \frac{t}{2(k-1)} \frac{|\mathcal{J}_{k,n}| \binom{k}{s}^{-1}}{|\mathcal{J}_{k-s,n-m}| |\mathcal{J}_{s,m}|} - \frac{1}{2(k-1)} \frac{\binom{k}{s}^{-1}}{|\mathcal{J}_{k-s,n-m}| |\mathcal{J}_{s,m}|} \\ & \quad \times \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)]. \end{aligned}$$

and where for $s = 0, \dots, k$,

$$\begin{aligned} I_{k,n}(s) &= \left\{ \sigma \in \mathcal{I}_{k,n} : |\{\sigma(1), \dots, \sigma(k)\} \cap \{1, \dots, m\}| = s \right\} \\ \tilde{g}(\mathbf{x}_{s+1}, \dots, \mathbf{x}_k) &= \frac{1}{|\mathcal{I}_{s,m}|} \sum_{\tilde{\sigma} \in \mathcal{I}_{s,m}} g(\mathbf{X}_{\tilde{\sigma}(1)}, \dots, \mathbf{X}_{\tilde{\sigma}(s)}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_k) \\ \tilde{\sigma}^2 &= \text{Var}\left(\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)\right). \end{aligned}$$

It then follows that, if n is sufficiently large and m, k are fixed, that there exists a constant $C_{g,k,m}$ such that

$$\mathbb{P}\left(U_n - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] \geq t \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \lesssim \exp\left(-\frac{C_{g,k} n t^2}{\sigma^2 + (b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])t}\right).$$

Proof. Assume w.l.o.g. that g is symmetric and consider a U -statistic of order k

$$U_n = \frac{1}{|\mathcal{I}_{k,n}|} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}),$$

where $\mathcal{I}_{k,n}$ denotes the set of injective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$. Define the set

$$I_{k,n}(s) = \left\{ \sigma \in \mathcal{I}_{k,n} : |\{\sigma(1), \dots, \sigma(k)\} \cap \{1, \dots, m\}| = s \right\}$$

such that we can rewrite U_n as

$$U_n = \frac{1}{|\mathcal{I}_{k,n}|} \sum_{s=0}^k \sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}).$$

Note that the sum over $I_{k,n}(0)$ consists of purely random terms, $I_{k,n}(k)$ of fully deterministic terms and $I_{k,n}(s)$ for $1 \leq s \leq k-1$ have both random and deterministic parts.

We can now apply the following decomposition

$$\begin{aligned} U_n &= \frac{1}{|\mathcal{I}_{k,n}|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) + \frac{1}{|\mathcal{I}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \\ &\quad + \frac{1}{|\mathcal{I}_{k,n}|} \sum_{s=1}^{k-1} \sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}). \end{aligned}$$

Now, for a fixed s and for any $\sigma \in I_{k,n}(s)$, since g is symmetric, the value of $g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})$ depends only on the set of indices assigned to $\{1, \dots, m\}$, not on their ordering within the arguments. Therefore,

$$\sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) = \binom{k}{s} \sum_{\tilde{\sigma}: [1, s] \rightarrow [1, m]} \sum_{\sigma: [s+1, k] \rightarrow [m+1, n]} g(\mathbf{X}_{\tilde{\sigma}(1)}, \dots, \mathbf{X}_{\tilde{\sigma}(s)}, \mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)}),$$

where we sum over all injective functions from $\{1, \dots, s\}$ to $\{1, \dots, m\}$ and from $\{s+1, \dots, k\}$ to $\{m+1, \dots, n\}$ respectively. If we define

$$\tilde{g}(\mathbf{x}_{s+1}, \dots, \mathbf{x}_k) = \frac{1}{|\mathcal{I}_{s,m}|} \sum_{\tilde{\sigma}: [1, s] \rightarrow [1, m]} g(\mathbf{X}_{\tilde{\sigma}(1)}, \dots, \mathbf{X}_{\tilde{\sigma}(s)}, \mathbf{x}_{s+1}, \dots, \mathbf{x}_k)$$

then conditional on $\mathbf{X}_1, \dots, \mathbf{X}_m$, $\tilde{g}$ is a deterministic kernel function in $[a, b]$. Recall that $\mathcal{J}_{s,m}$ is again the set of injective functions from $\{1, \dots, s\}$ to $\{1, \dots, m\}$. Therefore, after the right normalization,

$$\frac{1}{|\mathcal{J}_{k-s, n-m}|} \sum_{\sigma: [s+1, k] \rightarrow [m+1, n]} \tilde{g}(\mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)}),$$

is a U -statistic of order $k-s$. Now we write

$$\begin{aligned} & \mathbb{P}\left(U_n - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] \geq t \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &= \mathbb{P}\left(\frac{1}{|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) + \frac{1}{|\mathcal{J}_{k,n}|} \sum_{s=1}^{k-1} \sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]\right. \\ & \quad \left. \geq t - \frac{1}{|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &\leq \mathbb{P}\left(\frac{1}{|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]\right. \\ & \quad \left. \geq t/2 - \frac{1}{2|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &+ \mathbb{P}\left(\frac{1}{|\mathcal{J}_{k,n}|} \sum_{s=1}^{k-1} \sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \geq t/2 - \frac{1}{2|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right). \end{aligned}$$

We will now look at these two expressions separately.

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]\right. \\ & \quad \left. \geq t/2 - \frac{1}{2|\mathcal{J}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &= \mathbb{P}\left(\frac{1}{|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \frac{|\mathcal{J}_{k,n}|}{|I_{k,n}(0)|} \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]\right. \\ & \quad \left. \geq \frac{t}{2} \frac{|\mathcal{J}_{k,n}|}{|I_{k,n}(0)|} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &\leq \mathbb{P}\left(\frac{1}{|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(0)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]\right. \\ & \quad \left. \geq \frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \mid \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\ &\leq \exp\left(-\frac{(n-m)\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)^2}{2k\left(\sigma^2 + 1/3(b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)\right)}\right). \end{aligned}$$

We used Lemma 2.7.4 in the last inequality. Next,

$$\begin{aligned}
& \mathbb{P}\left(\frac{1}{|\mathcal{I}_{k,n}|} \sum_{s=1}^{k-1} \sum_{\sigma \in I_{k,n}(s)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \geq t/2 - \frac{1}{2|\mathcal{I}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) | \mathbf{X}_1, \dots, \mathbf{X}_m\right) \\
&= \mathbb{P}\left(\frac{1}{|\mathcal{I}_{k,n}|} \sum_{s=1}^{k-1} \binom{k}{s} \sum_{\tilde{\sigma}: [1,s] \rightarrow [1,m]} \sum_{\sigma: [s+1,k] \rightarrow [m+1,n]} g(\mathbf{X}_{\tilde{\sigma}(1)}, \dots, \mathbf{X}_{\tilde{\sigma}(s)}, \mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)})\right) \\
&\quad \geq t/2 - \frac{1}{2|\mathcal{I}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) | \mathbf{X}_1, \dots, \mathbf{X}_m \\
&= \mathbb{P}\left(\sum_{s=1}^{k-1} \frac{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|}{|\mathcal{I}_{k,n}|} \binom{k}{s} \frac{1}{|\mathcal{I}_{k-s,n-m}|} \sum_{\sigma: [s+1,k] \rightarrow [m+1,n]} \tilde{g}(\mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)})\right) \\
&\quad \geq t/2 - \frac{1}{2|\mathcal{I}_{k,n}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) | \mathbf{X}_1, \dots, \mathbf{X}_m \\
&\leq \sum_{s=1}^{k-1} \mathbb{P}\left(\frac{1}{|\mathcal{I}_{k-s,n-m}|} \sum_{\sigma: [s+1,k] \rightarrow [m+1,n]} \tilde{g}(\mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)})\right) \\
&\quad \geq \frac{t}{2(k-1)} \frac{|\mathcal{I}_{k,n}| \binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} - \frac{1}{2(k-1)} \frac{\binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) | \mathbf{X}_1, \dots, \mathbf{X}_m \\
&= \sum_{s=1}^{k-1} \mathbb{P}\left(\frac{1}{|\mathcal{I}_{k-s,n-m}|} \sum_{\sigma: [s+1,k] \rightarrow [m+1,n]} \tilde{g}(\mathbf{X}_{\sigma(s+1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)]\right) \\
&\quad \geq \frac{t}{2(k-1)} \frac{|\mathcal{I}_{k,n}| \binom{k}{j}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} - \frac{1}{2(k-1)} \frac{\binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \\
&\quad \quad - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)] | \mathbf{X}_1, \dots, \mathbf{X}_m \\
&\leq \sum_{s=1}^{k-1} \exp\left(-\frac{(n-m)(t_s^*)^2}{2(k-s)(\tilde{\sigma}^2 + (1/3)(b - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)])t_s^*)}\right)
\end{aligned}$$

where

$$\begin{aligned}
t_s^* &= \frac{t}{2(k-1)} \frac{|\mathcal{I}_{k,n}| \binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} - \frac{1}{2(k-1)} \frac{\binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \\
&\quad \times \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)].
\end{aligned}$$

If we now combine the two results we obtain

$$\begin{aligned}
& \mathbb{P}\left(U_n - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] \geq t | X_1, \dots, X_m\right) \\
&\leq \exp\left(-\frac{(n-m)\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)^2}{2k\left(\sigma^2 + 1/3(b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])\left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})\right)\right)}\right)
\end{aligned}$$

$$+ \sum_{s=1}^{k-1} \exp \left(- \frac{(n-m)(t_s^*)^2}{2(k-s)(\tilde{\sigma}^2 + (1/3)(b - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)])t_s^*)} \right).$$

When looking at the order we find that for fixed m and k and for n sufficiently large,

$$\left(\frac{t}{2} - \frac{1}{|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \right)^2 \sim C_{k,g} t^2$$

such that

$$\begin{aligned} & \exp \left(- \frac{(n-m) \left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \right)^2}{2k \left(\sigma^2 + 1/3(b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)]) \left(\frac{t}{2} - \frac{1}{2|I_{k,n}(0)|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \right) \right)} \right) \\ & \lesssim \exp \left(- \frac{C_{k,g} n t^2}{\sigma^2 + (b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])t} \right) \end{aligned}$$

Also for any $1 \leq s \leq k-1$

$$\frac{|\mathcal{I}_{k,n}| \binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \sim C_{k,s,m} n^s.$$

and

$$\frac{\binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \lesssim \frac{1}{n^{k-s}} |I_{n,k}(k)| b \sim \frac{b}{n^{k-s}}.$$

This shows that

$$\begin{aligned} (t_s^*)^2 &= \left(\frac{t}{2(k-1)} \frac{|\mathcal{I}_{k,n}| \binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} - \frac{1}{2(k-1)} \frac{\binom{k}{s}^{-1}}{|\mathcal{I}_{k-s,n-m}| |\mathcal{I}_{s,m}|} \right. \\ & \quad \left. \sum_{\sigma \in I_{k,n}(k)} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)] \right)^2 \sim C_{k,s,m} n^{2s} t^2 \end{aligned}$$

and so for each $1 \leq s \leq k-1$ each summand behaves as

$$\exp \left(- C_{k,s,m} \frac{3n^{2s+1} t^2}{3\tilde{\sigma} + (b - \mathbb{E}[\tilde{g}(\mathbf{X}_{s+1}, \dots, \mathbf{X}_k)]) n^s t} \right).$$

Note this sum is always being dominated by

$$\exp \left(- \frac{C_{k,g} n t^2}{\sigma^2 + (b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])t} \right).$$

Hence we conclude that

$$\mathbb{P}(U_n - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)] \geq t | \mathbf{X}_1, \dots, \mathbf{X}_m) \lesssim \exp \left(- \frac{C_{k,g} n t^2}{\sigma^2 + (b - \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k)])t} \right).$$

This proves the claim. $\square$

6.1. BIAS OF A CONDITIONAL U -STATISTIC

Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ be a measurable function and let $(\mathbf{X}_1, \mathbf{Z}_1), (\mathbf{X}_2, \mathbf{Z}_2), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$ be i.i.d. observations in $\mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $F_{\mathbf{X}, \mathbf{Z}}$. We consider the conditional functional

$$\theta(\mathbf{z}_1, \dots, \mathbf{z}_k) = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]$$

for fixed $(\mathbf{z}_1, \dots, \mathbf{z}_k) \in (\mathbb{R}^d)^k$. We are using random design, allowing the variables $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ to be random. Using Definition 2.7.5, we define the conditional U -statistic estimator of $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ as

$$U_n(\mathbf{z}_1, \dots, \mathbf{z}_k) = \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}.$$

We start by computing the bias of U_n for which we need an additional assumption to control the bias.

Assumption 9. Suppose that $f_{\mathbf{X}, \mathbf{Z}}$ is 3-times continuously differentiable on $\mathcal{Z}$ and for every sequence $\alpha_1, \dots, \alpha_k \in \{0, 1, 2, 3\}$ and all $\mathbf{z} \in \mathcal{Z}$:

$$I_1(\mathbf{z}) = \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \prod_{j=1}^k |K(\mathbf{u}_j)| \prod_{i=1}^k \sum_{k_1, \dots, k_{\alpha_i}=1}^d |u_{i, k_1} \cdots u_{i, k_{\alpha_i}}| \sup_{t^* \in [0, 1]} \left| \frac{\partial^{\alpha_i} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z} + t^* \mathbf{u}_i h)}{\partial \mathbf{z}_{k_1} \cdots \partial \mathbf{z}_{k_{\alpha_i}}} \right| d\mathbf{x}_i d\mathbf{u}_i < \infty.$$

Theorem 6.1.1. Define the event

$$A_{n,h} := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

Then under Assumptions 1, 2, 8 and 9 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^d \rightarrow \infty$ such that $\frac{1}{nh^d} = o(h^2)$

$$\begin{aligned} \mathbb{E}[U_n(\mathbf{z}_1, \dots, \mathbf{z}_k) \mid A_{n,h}] &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right] \\ &\quad + h^2 \left(\frac{A_{2,g} - A_{2,1} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]}{A_{0,1}} \right) \\ &\quad + \mathcal{O} \left(h^3 + \frac{1}{nh^d} \right) \end{aligned}$$

where

$$A_{0,1} = \prod_{j=1}^k f_Z(\mathbf{z}_j)$$

$$A_{2,g} = \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X},Z}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j$$

$$A_{2,1} = \int \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X},Z}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j.$$

$A_{3,g}$ and $A_{3,1}$ are as in Equation (6.9). $A_2(\mathbf{x}, \mathbf{z}, \mathbf{u})$ is as in Equation (6.8).

The exact forms of the remaining constants are established within the subsequent proof.

Proof. We want to compute

$$\mathbb{E} \left[\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})} \middle| A_{n,h} \right].$$

The inside of the expectation can be rewritten as

$$\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}$$

$$= \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \frac{1}{1 + \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}}.$$

If we now apply a Taylor-Lagrange formula with explicit remainder to $x \mapsto (1+x)^{-1}$ we get the existence of some $|x^*| \leq |x|$ such that

$$\frac{1}{1+x} = 1 - \frac{x}{(1+x^*)^2}.$$

If we apply this, we obtain

$$\frac{1}{1 + \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}}$$

$$= 1 - \frac{1}{(1+\Delta)^2} \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}$$

where

$$|\Delta| \leq \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \right|.$$

We can now write

$$\begin{aligned}
 & \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})} \\
 &= \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \\
 &\times \left(1 - \frac{1}{(1+\Delta)^2} \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \right) \\
 &=: f_1 - f_2
 \end{aligned}$$

with

$$\begin{aligned}
 f_1 &= \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \\
 f_2 &= \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]^2} \\
 &\quad \times \left(\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \frac{1}{(1+\Delta)^2}.
 \end{aligned}$$

The following Lemma is proven in Section 6.1.1.

Lemma 6.1.2. *There exists a constant $C^* \in [\pm h^2(A_{2,1} + hA_{3,1})/A_{0,1}]$ such that*

$$\begin{aligned}
 & \mathbb{E} \left[\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \middle| A_{n,h} \right] \\
 &= \frac{A_{0,g}}{A_{0,1}} + h^2 \left(\frac{A_{2,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{2,1} \right) + h^3 \left(\frac{A_{3,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{3,1} \right) \\
 &\quad + h^4 \left(\frac{A_{0,g}}{A_{0,1}^3} A_{2,1}^2 - \frac{A_{2,g}}{A_{0,1}^2} A_{2,1} \right) + h^5 \left(\frac{2A_{0,g}}{A_{0,1}^3} A_{2,1} A_{3,1} - \frac{A_{2,g}}{A_{0,1}^2} A_{3,1} + \frac{A_{2,1} A_{3,g}}{A_{0,1}^2} \right) \\
 &\quad + h^6 \left(\frac{A_{0,g}}{A_{0,1}^3} A_{3,1}^2 - \frac{A_{3,g}}{A_{0,1}^2} A_{3,1} + \frac{A_{2,1}^2 A_{2,g}}{A_{0,1}^2 (1+C^*)^3} \right) + h^7 \left(\frac{A_{2,1}^2 A_{3,g} + 2A_{2,1} A_{2,g} A_{3,1}}{A_{0,1}^3 (1+C^*)^3} \right) \\
 &\quad + h^8 \frac{A_{3,1}^2 A_{2,g}}{A_{0,1}^3 (1+C^*)^3} + h^9 \frac{A_{3,1}^2 A_{3,g}}{A_{0,1}^3 (1+C^*)^3} + \mathcal{O} \left(\exp \left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k} \right) \right).
 \end{aligned}$$

Lemma 6.1.2 than shows that

$$\mathbb{E}[f_1 | A_{n,h}] = \frac{A_{0,g}}{A_{0,1}} + h^2 \left(\frac{A_{2,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{2,1} \right) + \mathcal{O}(h^3) \quad (6.1)$$

since the exponential decays much faster than any power of h . Note that we can write

$$\begin{aligned} \frac{A_{0,g}}{A_{0,1}} &= \frac{1}{\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}[f_1 | A_{n,h}] &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right] \\ &\quad + h^2 \left(\frac{A_{2,g} - A_{2,1} \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k]}{A_{0,1}} \right) + \mathcal{O}(h^3) \end{aligned}$$

The following Lemma is proven in Section 6.1.3.

Lemma 6.1.3. *In the same setting as Lemma 6.1.2,*

$$\begin{aligned} &\mathbb{E} \left[\frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E} \left[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]^2} \right. \\ &\quad \times \left. \left(\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \right] \Big| A_{n,h} \Big] \\ &= \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} \frac{|I_{k,n}(1)|}{|\mathfrak{J}_{k,n}|} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right. \right. \\ &\quad + \frac{h^2}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &\quad + \frac{h^3}{h^{kd}} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \Big) \\ &\quad \times \left(\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j \neq i}^k K(\mathbf{u}_j) \sum_{l \neq i}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{r \neq l} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_r, \mathbf{z}_r) \right) d\mathbf{x}_j d\mathbf{u}_j \right. \\ &\quad \left. \left. + h^3 \int \prod_{j \neq i}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \right) + \mathcal{O} \left(\frac{1}{(nh^d)^2} \right) \right] \\ &- \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} \frac{|I_{k,n}(1)|}{|\mathfrak{J}_{k,n}|} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(A_{0,g} A_{0,1} + h^2 (A_{2,1} A_{0,g} + A_{2,g} A_{0,1}) + \mathcal{O}(h^3) \right). \end{aligned}$$

Where we define

$$I_{k,n}(r) = \left\{ \sigma \in \mathfrak{J}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\}$$

with

$$|I_{k,n}(r)| = k! \binom{k}{r} \binom{n-k}{k-r}.$$

Note that by Lemma 6.1.3

$$\frac{|I_{n,k}(1)|}{|\mathfrak{I}_{k,n}|} = \frac{k \binom{n-k}{k-1}}{\binom{n}{k}} = \frac{k^2 [(n-k)!]^2}{n! (n-2k+1)!} = \mathcal{O}\left(\frac{1}{n}\right).$$

Also recall that under $A_{n,h}$ we have the bound $\left|\frac{1}{(1+\Delta)^2}\right| \leq 4$. We then obtain under Assumption 9

$$\begin{aligned} & \left| \mathbb{E}[f_2 | A_{n,h}] \right| \\ & \lesssim \left| \frac{1}{(1+\Delta)^2} \right| \frac{|I_{k,n}(1)|}{|\mathfrak{I}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \prod_{j=1}^k |f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j)| d\mathbf{x}_j + o\left(\frac{1}{nh^d}\right) \\ & = \mathcal{O}\left(\frac{1}{nh^d}\right). \end{aligned} \tag{6.2}$$

If we now combine Equation (6.1) and Equation (6.2) we obtain

$$\begin{aligned} & \mathbb{E} \left[\frac{\sum_{\sigma \in \mathfrak{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})} \middle| A_{n,h} \right] \\ & = \mathbb{E}[f_1 | A_{n,h}] - \mathbb{E}[f_2 | A_{n,h}] \\ & = \frac{A_{0,g}}{A_{0,1}} + h^2 \left(\frac{A_{2,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{2,1} \right) + \mathcal{O}\left(h^3 + \frac{1}{nh^d}\right) \\ & = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right] \\ & \quad + h^2 \left(\frac{A_{2,g} - A_{2,1} \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_k) | \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k]}{A_{0,1}} \right) \\ & \quad + \mathcal{O}\left(h^3 + \frac{1}{nh^d}\right) \end{aligned}$$

as desired. $\square$

Remark 12. In Lemma 6.1.1, a remainder term of order $\mathcal{O}\left(\frac{1}{nh^d}\right)$ appears, for which the explicit constants are not tracked in the proof. For certain choices of the bandwidth h , this term may dominate the bias term of order $\mathcal{O}(h^2)$, in which case knowledge of the corresponding constants becomes important. However, under the common bandwidth choice $h \asymp n^{-1/(4+d)}$, which typically minimizes the mean squared error, we have $\frac{1}{nh^d} = o(h^2)$, so that the $\mathcal{O}\left(\frac{1}{nh^d}\right)$ term is asymptotically negligible compared to the $\mathcal{O}(h^2)$ term.

If a different bandwidth regime is considered, the leading constant in the $\frac{1}{nh^d}$ term can be made explicit by refining the proof of Lemma 6.1.1, in particular by carrying the expansion one order further and keeping precise control of the remainder terms.

6.1.1.1. PROOF OF LEMMA 6.1.2

Start by writing

$$\begin{aligned} & \mathbb{E} \left[\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \middle| A_{n,h} \right] \\ &= \frac{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \middle| A_{n,h} \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \end{aligned} \quad (6.3)$$

where we recall that

$$A_{n,h} := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

Lemma 6.1.4. *Under the same setting as Lemma 6.1.2 we have that*

$$\begin{aligned} & \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \middle| A_{n,h} \right] \\ &= A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} + \mathcal{O} \left(\exp \left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k} \right) \right). \end{aligned}$$

where

$$\begin{aligned} A_{0,g} &:= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \\ A_{2,g} &:= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ A_{3,g} &:= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \end{aligned}$$

The following result will be proven in Section 6.1.2. Note that we can rewrite Equation (6.3) to apply Lemma 6.1.4

$$\begin{aligned} & \frac{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \middle| A_{n,h} \right]}{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \\ &= \frac{A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} + \mathcal{O} \left(\exp \left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k} \right) \right)}{A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1}} \end{aligned} \quad (6.4)$$

where we used Lemma 6.1.4 with function $g = 1$ to obtain the denominator. We have,

$$\begin{aligned} A_{0,1} &= \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) \\ A_{2,1} &= \int \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ A_{3,1} &= \int \prod_{j=1}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j. \end{aligned} \quad (6.5)$$

To simplify this expression, we can rewrite Equation (6.4) in the form

$$\frac{x + \epsilon}{y + \eta} = \frac{x}{y} \left(1 + \frac{\epsilon}{x} \right) \frac{1}{1 + \frac{\eta}{y}}.$$

If we now apply the Taylor-Lagrange formula with explicit remainder to $x \mapsto (1 + x)^{-1}$ we get the existence of some $|x^*| \leq |x|$ such that

$$\frac{1}{1 + x} = 1 - x + \frac{x^2}{(1 + x^*)^3}.$$

This means that

$$\begin{aligned} \frac{x + \epsilon}{y + \eta} &= \frac{x}{y} \left(1 + \frac{\epsilon}{x} \right) \frac{1}{1 + \frac{\eta}{y}} = \frac{x}{y} \left(1 + \frac{\epsilon}{x} \right) \left(1 - \frac{\eta}{y} + \frac{\eta^2}{y^2(1 + x^*)^3} \right) \\ &= \frac{x}{y} + \frac{\epsilon}{y} - \frac{x\eta}{y^2} - \frac{\epsilon\eta}{y^2} + \frac{x}{y} \frac{\eta^2}{y^2(1 + x^*)^3} + \frac{\epsilon}{y} \frac{\eta^2}{y^2(1 + x^*)^3}, \end{aligned}$$

for some $x^* \in [\pm\eta/y]$. If we apply this expression on Equation (6.4), we obtain that there exists $C^* \in [\pm\eta/y] = [\pm h^2(A_{2,1} + hA_{3,1})/A_{0,1}]$ such that

$$\begin{aligned} &\mathbb{E} \left[\frac{\sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \middle| A_{n,h} \right] \\ &= \frac{A_{0,g}}{A_{0,1}} + \frac{h^2 A_{2,g} + h^3 A_{3,g}}{A_{0,1}} - \frac{A_{0,g}}{(A_{0,1})^2} (h^2 A_{2,1} + h^3 A_{3,1}) \\ &\quad + \frac{A_{0,g} (h^2 A_{2,1} + h^3 A_{3,1})^2}{(A_{0,1})^3 (1 + C^*)^3} - \frac{(h^2 A_{2,g} + h^3 A_{3,g}) (h^2 A_{2,1} + h^3 A_{3,1})}{(A_{0,1})^2} \\ &\quad + \frac{(h^2 A_{2,g} + h^3 A_{3,g}) (h^2 A_{2,1} + h^3 A_{3,1})^2}{(A_{0,1})^3 (1 + C^*)^3} + \mathcal{O} \left(\exp \left(-C_g \frac{(n-k)h^{2dk}}{\|K\|_{\infty}^{2k}} f_{\min}^{2k} \right) \right) \\ &= \frac{A_{0,g}}{A_{0,1}} + h^2 \left(\frac{A_{2,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{2,1} \right) + h^3 \left(\frac{A_{3,g}}{A_{0,1}} - \frac{A_{0,g}}{A_{0,1}^2} A_{3,1} \right) \\ &\quad + h^4 \left(\frac{A_{0,g}}{A_{0,1}^3} A_{2,1}^2 - \frac{A_{2,g} A_{2,1}}{A_{0,1}^2} \right) + h^5 \left(\frac{2A_{0,g}}{A_{0,1}^3} A_{2,1} A_{3,1} - \frac{A_{2,g} A_{3,1} + A_{2,1} A_{3,g}}{A_{0,1}^2} \right) \end{aligned}$$

$$\begin{aligned}
& + h^6 \left(\frac{A_{0,g} A_{3,1}^2}{A_{0,1}^3} - \frac{A_{3,g} A_{3,1}}{A_{0,1}^2} + \frac{A_{2,1}^2 A_{2,g}}{A_{0,1}^3 (1 + C^*)^3} \right) + h^7 \left(\frac{A_{2,1}^2 A_{3,g} + 2 A_{2,1} A_{2,g} A_{3,1}}{A_{0,1}^3 (1 + C^*)^3} \right) \\
& + h^8 \frac{A_{3,1}^2 A_{2,g}}{A_{0,1}^3 (1 + C^*)^3} + h^9 \frac{A_{3,1}^2 A_{3,g}}{A_{0,1}^3 (1 + C^*)^3} + \mathcal{O} \left(\exp \left(-n h^{kd} \frac{C f_{\mathbf{Z}, \min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k} \right) \right)
\end{aligned}$$

which proves Lemma 6.1.2.

6.1.2. PROOF OF LEMMA 6.1.4

First write

$$\begin{aligned}
& \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \middle| A_{n,h} \right] \\
& = \frac{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}} \right]}{\mathbb{P}(A_{n,h})}. \tag{6.6}
\end{aligned}$$

Where

$$\begin{aligned}
& \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}} \right] \\
& = \underbrace{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}_{(A)} \\
& - \underbrace{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}^c} \right]}_{(B)}.
\end{aligned}$$

EXPANSION OF (A)

We can rewrite (A) as

$$\begin{aligned}
& \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\
& = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}'_j - \mathbf{Z}_j) \right] \\
& = \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K_h(\mathbf{z}'_j - \mathbf{z}_j) \prod_{i=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mathbf{x}_i d\mathbf{z}_i \\
& = \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, h\mathbf{u}_j + \mathbf{z}'_j) d\mathbf{x}_j d\mathbf{u}_j. \tag{6.7}
\end{aligned}$$

Set

$$\phi(t) = f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z} + t\mathbf{u}h), \text{ for } t \in [0, 1].$$

This function has at least the same regularity as $f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ so it is differentiable by Assumption 9. By the Taylor-Lagrange formula with explicit remainder, we get the existence of some $t^* \in [0, 1]$ such that

$$f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z} + \mathbf{u}h) = \sum_{l=0}^3 h^l A_l$$

where

$$\begin{aligned} A_0(\mathbf{x}, \mathbf{z}, \mathbf{u}) &:= f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}), \\ A_1(\mathbf{x}, \mathbf{z}, \mathbf{u}) &:= \mathbf{u} \cdot \nabla_{\mathbf{z}} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}), \\ A_2(\mathbf{x}, \mathbf{z}, \mathbf{u}) &:= \frac{1}{2} \sum_{k_1=1}^d \sum_{k_2=1}^d u_{k_1} u_{k_2} \partial_{z_{k_1}, z_{k_2}} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z}), \\ A_3(\mathbf{x}, \mathbf{z}, \mathbf{u}) &:= \frac{1}{6} \sum_{k_1=1}^d \sum_{k_2=1}^d \sum_{k_3=1}^d u_{k_1} u_{k_2} u_{k_3} \partial_{z_{k_1}, z_{k_2}, z_{k_3}} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z} + t^* \mathbf{u}h). \end{aligned} \quad (6.8)$$

Then

$$\begin{aligned} \prod_{j=1}^k K(\mathbf{u}_j) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, h\mathbf{u}_j + \mathbf{z}'_j) &= \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=0}^3 h^l A_l(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) \\ &= \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=\{0,2,3\}}^3 h^l A_l(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) \end{aligned}$$

where we have used that $\int K(\mathbf{u}) \mathbf{u} d\mathbf{u} = 0$, from Assumption 1. Equation (6.7) becomes

$$\begin{aligned} &\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) \left(A_0(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) + h^2 A_2(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) + h^3 A_3(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}'_j) d\mathbf{x}_j \\ &\quad + h^2 \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}'_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}'_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &\quad + h^3 \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}'_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \\ &=: A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g}. \end{aligned} \quad (6.9)$$

UPPER BOUND FOR (B)

We now aim to find an upper bound for (B) by using the tower property.

$$\left| \mathbb{E} \left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}^c} \right] \right|$$

$$\begin{aligned}
&= \left| \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}'_j - \mathbf{Z}_j) \mathbb{E}[\mathbb{1}_{A_{n,h}^c} | \mathbf{X}_1, \dots, \mathbf{X}_k, \mathbf{Z}_1, \dots, \mathbf{Z}_k] \right] \right| \\
&\leq \mathbb{E} \left[|g(\mathbf{X}_1, \dots, \mathbf{X}_k)| \prod_{j=1}^k |K_h(\mathbf{z}'_j - \mathbf{Z}_j)| \mathbb{P}(A_{n,h}^c | \mathbf{X}_1, \dots, \mathbf{X}_k, \mathbf{Z}_1, \dots, \mathbf{Z}_k) \right] \\
&= \mathbb{E} \left[|g(\mathbf{X}_1, \dots, \mathbf{X}_k)| \prod_{j=1}^k |K_h(\mathbf{z}'_j - \mathbf{Z}_j)| \mathbb{P}(A_{n,h}^c | \mathbf{Z}_1, \dots, \mathbf{Z}_k) \right] \\
&= \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \mathbb{P}(A_{n,h}^c | \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k) \prod_{j=1}^k |K_h(\mathbf{z}'_j - \mathbf{z}_j)| f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j d\mathbf{z}_j. \quad (6.10)
\end{aligned}$$

Now in order to compute the conditional probability we rewrite the event $A_{n,h}^c$ as

$$\begin{aligned}
A_{n,h}^c &= \left\{ \left| \frac{|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| > \frac{1}{2} \right\} \\
&= \left\{ \left| |\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right| > \frac{1}{2} \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right\}.
\end{aligned}$$

Now define

$$f(\mathbf{x}_1, \dots, \mathbf{x}_k) := \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{x}_j)$$

such that

$$\begin{aligned}
&\mathbb{P}(A_{n,h}^c | \mathbf{Z}_1, \dots, \mathbf{Z}_k) \\
&= \mathbb{P} \left(\left| |\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} f(\mathbf{Z}_{\sigma(1)}, \dots, \mathbf{Z}_{\sigma(k)}) - \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)] \right| > \frac{1}{2} \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)] \right).
\end{aligned}$$

We can now apply Lemma 6.0.1 with $t = \frac{1}{2} \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]$, $|f(\mathbf{x}_1, \dots, \mathbf{x}_k)| \leq h^{-kd} \|K\|_\infty^k$ and the fact that $\text{Var}(f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)) \sim \frac{1}{h^{kd}} \|K\|_\infty^{2k} f_{\mathbf{Z}, \max}^k$ to obtain

$$\begin{aligned}
&\mathbb{P}(A_{n,h}^c | \mathbf{Z}_1, \dots, \mathbf{Z}_k) \\
&\lesssim \exp \left(- \frac{C_{k,g} n \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]^2}{h^{-kd} \|K\|_\infty^{2k} f_{\mathbf{Z}, \max}^k + (h^{-kd} \|K\|_\infty^k - \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]) \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]} \right).
\end{aligned}$$

We can plug this into Equation (6.10) to obtain

$$\begin{aligned}
&\left| \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}^c} \right] \right| \\
&\lesssim \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \cdot \exp \left(- \frac{C_{k,g} n \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]^2}{h^{-kd} \|K\|_\infty^{2k} f_{\mathbf{Z}, \max}^k + (h^{-kd} \|K\|_\infty^k - \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]) \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]} \right) \\
&\quad \times \prod_{j=1}^k |K_h(\mathbf{z}'_j - \mathbf{z}_j)| f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j d\mathbf{z}_j
\end{aligned}$$

$$= \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \exp \left(- \frac{C_{k,g} n \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]^2}{h^{-kd} \|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k + (h^{-kd} \|K\|_{\infty}^k - \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]) \mathbb{E}[f(\mathbf{Z}_1, \dots, \mathbf{Z}_k)]} \right) \\ \times \prod_{j=1}^k |K(\mathbf{u}_j)| f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, h\mathbf{u}_j + \mathbf{z}_j) d\mathbf{x}_j d\mathbf{u}_j.$$

We can now use the result we found earlier

$$\mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] = \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j=1}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ + h^3 \int \prod_{j=1}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j$$

and so obtain

$$\left| \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}^c} \right] \right| \\ \lesssim \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \exp \left(- \frac{C_{k,g} n \left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) \right)^2}{h^{-kd} \|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k + h^{-kd} \|K\|_{\infty}^k \left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) \right)} \right) \\ \times \prod_{j=1}^k |K(\mathbf{u}_j)| f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, h\mathbf{u}_j + \mathbf{z}_j) d\mathbf{x}_j d\mathbf{u}_j \\ \leq \exp \left(- \frac{C_{k,g} n f_{\mathbf{Z}, \min}^{2k}}{h^{-kd} \|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k} \right) \int |g(\mathbf{x}_1, \dots, \mathbf{x}_k)| \prod_{j=1}^k |K(\mathbf{u}_j)| f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, h\mathbf{u}_j + \mathbf{z}_j) d\mathbf{x}_j d\mathbf{u}_j \\ = \mathcal{O} \left(\exp \left(- nh^{kd} \frac{C f_{\mathbf{Z}, \min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k} \right) \right).$$

Therefore, (B) is negligible compared to (A) as n increases and so

$$\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}} \right] \\ = \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ - \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \mathbb{1}_{A_{n,h}^c} \right] \\ = A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} + \mathcal{O} \left(\exp \left(- nh^{kd} \frac{C f_{\mathbf{Z}, \min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z}, \max}^k} \right) \right).$$

By applying a Taylor expansion we find that Equation (6.6) becomes

$$\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) | A_{n,h} \right]$$

$$\begin{aligned}
&= \frac{A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} + \mathcal{O}\left(\exp\left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k}\right)\right)}{\mathbb{P}(A_{n,h})} \\
&= A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} + \mathcal{O}\left(\exp\left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k}\right)\right),
\end{aligned}$$

where we used Lemma 2.7.4 to obtain

$$1 - \mathbb{P}(A_{n,h}) = \mathcal{O}\left(\exp\left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k}\right)\right).$$

This proves Lemma 6.1.4. For future use, we will combine this result into a corollary.

Corollary 6.1.5. *Define*

$$A_{n,h} = \left\{ \left| \frac{|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}\left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]}{\mathbb{E}\left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]} \right| < \frac{1}{2} \right\}$$

then

$$1 - \mathbb{P}(A_{n,h}) = \mathcal{O}\left(\exp\left(-nh^{kd} \frac{C f_{\mathbf{Z},\min}^{2k}}{\|K\|_{\infty}^{2k} f_{\mathbf{Z},\max}^k}\right)\right).$$

From this proof, the following useful generalization can also be made

Corollary 6.1.6. *Terms of the form*

$$\mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k)^m \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j)^m \mathbb{1}_{A_{n,h}^c}\right]$$

converge to 0 exponentially fast for any m and so conditioning with respect to $A_{n,h}$ can be ignored.

6.1.3. PROOF OF LEMMA 6.1.3

Start by noting that due to Corollary 6.1.6 we can just consider

$$\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}\left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]^2}\right. \\
&\quad \times \left.\left(\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}\left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]\right)\right] \\
&= \frac{1}{\mathbb{E}\left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j)\right]^2} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j)\right]
\end{aligned}$$

$$\begin{aligned}
& \times \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \\
& = \frac{1}{\mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2} |\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right. \\
& \quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \right].
\end{aligned}$$

To split the sum $\sum_{\sigma \in \mathcal{I}_{k,n}}$ into different cases, we define the set

$$I_{k,n}(r) := \left\{ \sigma \in \mathcal{I}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\},$$

such that we can decompose $\sum_{\sigma \in \mathcal{I}_{k,n}} = \sum_{\sigma \in I_{k,n}(0)} + \sum_{\sigma \in I_{k,n}(k)} + \sum_{r=1}^{k-1} \sum_{\sigma \in I_{k,n}(r)}$. The following lemma can be proven.

Lemma 6.1.7. *We have that*

$$|I_{n,k}(r)| = k! \binom{k}{r} \binom{n-k}{k-r}.$$

Proof. Recall that

$$I_{n,k}(r) = \left\{ \sigma \in \mathcal{I}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\}.$$

is the set of injective maps $\sigma : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ such that the size of the intersection of the domain and the image is exactly r . To determine $|I_{n,k}(r)|$, we count the number of ways to construct such a map σ through the following steps:

1. We must choose exactly r elements from the set $\{1, \dots, k\}$ to be included in the image of σ . There are $\binom{k}{r}$ ways to choose these values.
2. The remaining $k - r$ values in the image must be chosen from the set $\{k+1, \dots, n\}$. Since this set contains $n - k$ elements, there are $\binom{n-k}{k-r}$ ways to choose them.
3. At this stage, we have selected a set of k distinct values to serve as the image of σ . To define the injective map σ , we must assign these k values to the k elements of the domain $\{1, \dots, k\}$. There are $k!$ ways to order these values.

This yields that

$$|I_{k,n}(r)| = k! \binom{k}{r} \binom{n-k}{k-r}$$

as desired.

To ensure this partition accounts for all possible injective maps, we sum over all possible values of r :

$$\sum_{r=0}^k |I_{n,k}(r)| = k! \sum_{r=0}^k \binom{k}{r} \binom{n-k}{k-r} = k! \binom{n}{k} = |\mathcal{I}_{k,n}|$$

The summation $\sum_{r=0}^k \binom{k}{r} \binom{n-k}{k-r} = \binom{n}{k}$ follows from Vandermonde's Identity, confirming that the expression is correct [41]. $\square$

Lemma 6.1.7 shows that the dominator cases occur when $r = 0$ and $r = 1$. We will hence focus on these cases.

i) Suppose $\sigma \in I_{k,n}(0)$. Then by independence

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \right] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right] \\ &= 0. \end{aligned}$$

ii) Let $\sigma \in I_{k,n}(1)$ and suppose index $\sigma(i)$ overlaps with index m .

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \right] \quad (6.11) \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &\quad - \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]. \end{aligned}$$

By using Equation (6.9) and Equation (6.5) we find

$$\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] = A_{0,g} A_{0,1} + h^2 (A_{2,1} A_{0,g} + A_{2,g} A_{0,1}) + \mathcal{O}(h^3).$$

We can also compute

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \prod_{\substack{r=1 \\ r \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(r)}) K_h(\mathbf{z}_m - \mathbf{Z}_m) K_h(\mathbf{z}_i - \mathbf{Z}_m) \right]. \end{aligned}$$

Since K is compact,

$$K_h(\mathbf{z}_m - \mathbf{Z}_m) K_h(\mathbf{z}_i - \mathbf{Z}_m) = 0$$

whenever $\mathbf{z}_m \neq \mathbf{z}_i$ and n sufficiently large. So the expectation is only non-zero if $\mathbf{z}_m = \mathbf{z}_i$. In that case,

$$\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \prod_{\substack{r=1 \\ r \neq i}}^k K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) K_h(\mathbf{z}_m - \mathbf{Z}_m)^2 \right]$$

$$= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) K_h(\mathbf{z}_m - \mathbf{Z}_m)^2 \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \mathbb{E} \left[\prod_{\substack{r=1 \\ r \neq i}}^k K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) \right].$$

We have previously shown that

$$\begin{aligned} \mathbb{E} \left[\prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] &= \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j \neq i}^k K(\mathbf{u}_j) \sum_{l \neq i}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{r \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_r, \mathbf{z}_r) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &\quad + h^3 \int \prod_{j \neq i}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j. \end{aligned}$$

Now also

$$\begin{aligned} &\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) K_h(\mathbf{z}_m - \mathbf{Z}_m)^2 \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \\ &= h^{-d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j + h\mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j. \end{aligned}$$

If we now apply a Taylor expansion on $f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}, \mathbf{z} + t\mathbf{h}\mathbf{u})$, we can use Equation (6.9) and Assumption 1 to show that the previous expression equals

$$\begin{aligned} &\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \\ &+ \frac{h^2}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &+ \frac{h^3}{h^{kd}} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j. \end{aligned}$$

Equation (6.11) now becomes

$$\begin{aligned} &\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right. \\ &\quad + \frac{h^2}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ &\quad + \frac{h^3}{h^{kd}} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \Big) \\ &\quad \times \left(\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j \neq i}^k K(\mathbf{u}_j) \sum_{l \neq i}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{r \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_r, \mathbf{z}_r) \right) d\mathbf{x}_j d\mathbf{u}_j \right. \\ &\quad \left. + h^3 \int \prod_{j \neq i}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \right) - A_{0,g} A_{0,1} + h^2 (A_{2,1} A_{0,g} + A_{2,g} A_{0,1}). \end{aligned}$$

All the other cases are at least of order $\mathcal{O}\left(\frac{1}{h^{2d}}\right)$, so we will not consider them.

This shows that

$$\begin{aligned}
 & \mathbb{E} \left[\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j)]^2} \right. \\
 & \quad \times \left(\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \Big| A_{n,h} \Big] \\
 &= \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} |\mathcal{J}_{k,n}|^{-1} \sum_{r=0}^k \sum_{\sigma \in I_{k,n}(r)} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right. \\
 & \quad \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \Big] \\
 &= \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} \frac{|I_{k,n}(1)|}{|\mathcal{J}_{k,n}|} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right. \right. \\
 & \quad + \frac{h^2}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\
 & \quad + \frac{h^3}{h^{kd}} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \Big) \\
 & \quad \times \left(\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j \neq i}^k K(\mathbf{u}_j) \sum_{l \neq i}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{r \neq l} f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_r, \mathbf{z}_r) \right) d\mathbf{x}_j d\mathbf{u}_j \right. \\
 & \quad \left. \left. + h^3 \int \prod_{j \neq i}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \right) + \mathcal{O}\left(\frac{1}{(nh^d)^2}\right) \right] - A_{0,g} A_{0,1} + h^2 (A_{2,1} A_{0,g} + A_{2,g} A_{0,1}) + \mathcal{O}(h^3) \Big]
 \end{aligned}$$

or similarly

$$\begin{aligned}
 & \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} \frac{|I_{k,n}(1)|}{|\mathcal{J}_{k,n}|} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right. \right. \\
 & \quad + \frac{h^2}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) \sum_{l=1}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\
 & \quad + \frac{h^3}{h^{kd}} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) K(\mathbf{u}_m)^2 \prod_{j \neq m}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \Big) \\
 & \quad \times \left(\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 \int \prod_{j \neq i}^k K(\mathbf{u}_j) \sum_{l \neq i}^k \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{r \neq l} f_{\mathbf{x},\mathbf{Z}}(\mathbf{x}_r, \mathbf{z}_r) \right) d\mathbf{x}_j d\mathbf{u}_j \right. \\
 & \quad \left. \left. + h^3 \int \prod_{j \neq i}^k K(\mathbf{u}_j) I_3(\mathbf{x}_j, \mathbf{z}_j, \mathbf{u}_j) d\mathbf{x}_j d\mathbf{u}_j \right) + \mathcal{O}\left(\frac{1}{(nh^d)^2}\right) \right] \\
 & - \frac{1}{(A_{0,1} + h^2 A_{2,1} + h^3 A_{3,1})^2} \frac{|I_{k,n}(1)|}{|\mathcal{J}_{k,n}|} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(A_{0,g} A_{0,1} + h^2 (A_{2,1} A_{0,g} + A_{2,g} A_{0,1}) + \mathcal{O}(h^3) \right).
 \end{aligned}$$

6.2. VARIANCE OF A CONDITIONAL U -STATISTIC

Let $g : (\mathbb{R}^p)^k \rightarrow \mathbb{R}$ some measurable function and let $(\mathbf{X}_1, \mathbf{Z}_1), (\mathbf{X}_2, \mathbf{Z}_2), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$ a sequence of i.i.d. observations in $\mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $F_{\mathbf{X}, \mathbf{Z}}$. Using Definition 2.7.5 we again define the conditional U -statistic estimator of

$$\theta(\mathbf{z}_1, \dots, \mathbf{z}_k) = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]$$

as

$$U_n(\mathbf{z}_1, \dots, \mathbf{z}_k) = \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}.$$

This section aims to find an expression for the variance of a conditional U -statistic of order k . We need one additional assumption to ensure the integrals are bounded.

Assumption 10. Suppose that $f_{\mathbf{X}, \mathbf{Z}}$ is 3-times continuously differentiable on $\mathcal{Z}$. For every $p \in \{0, 1, \dots, k\}$, every valid pair of index mappings (ϕ_1, ϕ_2) spanning $2k - p$ unique indices, every sequence of derivative orders $\alpha_1, \dots, \alpha_k \in \{0, 1, 2, 3\}$, and all $\mathbf{z} \in \mathcal{Z}$:

$$I_2(\mathbf{z}) = \int \left| g(\mathbf{x}_{\phi_1(1)}, \dots, \mathbf{x}_{\phi_1(k)}) g(\mathbf{x}_{\phi_2(1)}, \dots, \mathbf{x}_{\phi_2(k)}) \right| \prod_{v=1}^k |K(\mathbf{u}_v)|^{\alpha_v} \\ \prod_{v=1}^k \left(\sum_{k_1, \dots, k_{\alpha_v}=1}^d |u_{v,k_1} \cdots u_{v,k_{\alpha_v}}| \sup_{t^* \in [0,1]} \left| \frac{\partial^{\alpha_v} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_v, \mathbf{z} + t^* \mathbf{u}_v h)}{\partial \mathbf{z}_{k_1} \cdots \partial \mathbf{z}_{k_{\alpha_v}}} \right| \right) \prod_{v=1}^k d\mathbf{x}_v d\mathbf{u}_v < \infty.$$

Theorem 6.2.1. Define the event

$$A_{n,h} := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

Then under Assumptions 1, 2, 8, 9 and 10 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^d \rightarrow \infty$, we have

$$\text{Var} \left(U_n(\mathbf{z}_1, \dots, \mathbf{z}_k) \mid A_{n,h} \right) \\ \sim \frac{1}{nh^d} \sum_{i=1}^k \sum_{m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_i)} \left(\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_m, \dots, \mathbf{X}_{2k}) \mid \mathbf{Z}_j = \mathbf{z}_j \right. \right. \\ \left. \left. \text{for all } j, \mathbf{Z}_{k+j} = \mathbf{z}_j \text{ for } j \neq i \right] - \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 \right) + o \left(\frac{1}{nh^d} \right).$$

Proof. We start by taking a similar approach as in Theorem 6.1.1 and rewrite

$$\frac{\sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}$$

$$= \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \frac{1}{1 + \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}}.$$

If we now again apply a Taylor-Lagrange formula with explicit remainder to $x \mapsto (1+x)^{-1}$ we get the existence of some $|x^*| \leq |x|$ such that

$$\frac{1}{1+x} = 1 - x + \frac{x^2}{(1+x^*)^3}.$$

We can now write

$$\begin{aligned} & \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})} = \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \\ & \times \left(1 - \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \right) \\ & + \frac{1}{(1+\Delta)^3} \left(\frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \right)^2 \\ & =: f_1 - f_2 + \frac{1}{(1+\Delta)^2} f_3 \end{aligned}$$

with

$$\begin{aligned} f_1 &= \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \\ f_2 &= \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]^2} \\ & \quad \times \left(\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right) \\ f_3 &= \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]^3} \\ & \quad \times \left(\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \right)^2 \end{aligned}$$

and

$$|\Delta| \leq \left| \frac{\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]}{\mathbb{E}[\sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})]} \right|.$$

Note that when conditioning over $A_{n,h}$, we have that $|\Delta| \leq \frac{1}{2}$. The following lemma is proven in Section 6.2.1.

Lemma 6.2.2. Let $I_\sigma(1) = \left\{ \sigma \in \mathcal{I}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = 1 \right\}$, then

$$\begin{aligned} & \text{Var} \left(\left| \mathcal{I}_{k,n} \right|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right) \\ &= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathcal{I}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m}_{i'\text{th}}, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\ & \quad \times \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Big) \\ &+ \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\left(\frac{|I_\sigma(0)|}{|\mathcal{I}_{k,n}|} - 1 \right) \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \end{aligned}$$

When n is sufficiently large

$$\begin{aligned} & \text{Var} \left(\left| \mathcal{I}_{k,n} \right|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right) \\ & \sim \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m}_{i'\text{th}}, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\ & \quad \times \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Big) \\ & - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{1}{n} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \end{aligned}$$

Recall from Lemma 6.1.4 that

$$\mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] = \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 A_{2,1} + h^3 A_{3,1}.$$

If we combine this with Lemma 6.2.2 we obtain

$$\begin{aligned} \text{Var}(f_1 | A_{n,h}) &= \frac{\text{Var} \left(\left| \mathcal{I}_{k,n} \right|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \middle| A_{n,h} \right)}{\mathbb{E} \left[\left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right)^2 \right]} \\ &= \frac{1}{\left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 A_{2,1} + h^3 A_{3,1} \right)^2} \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathcal{I}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \right. \right. \\ & \quad \times \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Big) \\ & \quad \left. + \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\left(\frac{|I_\sigma(0)|}{|\mathcal{I}_{k,n}|} - 1 \right) \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \right) \end{aligned}$$

A Taylor expansion allows us to rewrite this as

$$\begin{aligned}
 \text{Var}\left(f_1 \mid A_{n,h}\right) &= \frac{1}{\left(\prod_{j=1}^k f_Z(\mathbf{z}_j)\right)^2} \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathfrak{J}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \right. \right. \\
 &\quad \times \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},Z}(\mathbf{x}_l, \mathbf{z}_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},Z}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Bigg) \\
 &\quad + \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\left(\frac{|I_\sigma(0)|}{|\mathfrak{J}_{k,n}|} - 1 \right) \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},Z}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \\
 &= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathfrak{J}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d f_Z(\mathbf{z}_i)} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_m, \dots, \mathbf{X}_{2k}) \mid \mathbf{Z}_j = \mathbf{z}_j \right. \right. \\
 &\quad \left. \left. \text{for all } j, \mathbf{Z}_{k+j} = \mathbf{z}_j \text{ for } j \neq i \right] + o\left(\frac{1}{nh^d}\right) \right) \\
 &\quad + \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\left(\frac{|I_\sigma(0)|}{|\mathfrak{J}_{k,n}|} - 1 \right) \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right) \\
 &\sim \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{1}{nh^d} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_Z(\mathbf{z}_i)} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_m, \dots, \mathbf{X}_{2k}) \mid \mathbf{Z}_j = \mathbf{z}_j \right. \right. \\
 &\quad \left. \left. \text{for all } j, \mathbf{Z}_{k+j} = \mathbf{z}_j \text{ for } j \neq i \right] + o\left(\frac{1}{nh^d}\right) \right) \\
 &\quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\frac{1}{n} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right). \tag{6.12}
 \end{aligned}$$

We used from Lemma 6.1.7 that

$$\begin{aligned}
 \frac{|I_\sigma(1)|}{|\mathfrak{J}_{k,n}|} &= \frac{k! \binom{k}{1} \binom{n-k}{k-1}}{k! \binom{n}{k}} = \frac{k^2 [(n-k)!]^2}{n! (n-2k+1)!} \sim \frac{k^2}{n} \\
 \frac{|I_\sigma(0)|}{|\mathfrak{J}_{k,n}|} &= \frac{k! \binom{n-k}{k}}{k! \binom{n}{k}} = \frac{[(n-k)!]^2}{n! (n-2k)!} \sim 1 - \frac{1}{n}.
 \end{aligned}$$

This following lemma is proven in Section 6.2.2.

Lemma 6.2.3. Define for $k \leq r \leq 4k$, the set $I_{\eta,\tau,\theta}(r) = \left\{ \eta, \tau, \theta \in \mathfrak{J}_{k,n} : \left| \{1, \dots, k\} \cup \{\theta(1), \dots, \theta(k)\} \cup \{\eta(1), \dots, \eta(k)\} \cup \{\tau(1), \dots, \tau(k)\} \right| = r \right\}$, then

$$\begin{aligned}
 \text{Var}\left(|\mathfrak{J}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
 \left. \times \left(|\mathfrak{J}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}\left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right)
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{nh^d}\right) \right] \\
&- \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} A_{0,g}^2 \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right].
\end{aligned}$$

For n sufficiently large, the variance is similar to

$$\begin{aligned}
&\left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{nh^d}\right) \right] \\
&- \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{1}{n} \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right].
\end{aligned}$$

By combining Lemma 6.1.4 and Lemma 6.2.3 we obtain

$$\begin{aligned}
\text{Var}(f_2 | A_{n,h}) &= \frac{1}{\mathbb{E}[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j)]^4} \text{Var}\left(|\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
&\quad \left. \times \left(|\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}\left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) | A_{n,h} \right) \\
&= \frac{1}{\left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + h^2 A_{2,1} + h^3 A_{3,1} \right)^4} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 \right. \\
&\quad \left. + o\left(\frac{1}{nh^d}\right) \right] - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} A_{0,g}^2 \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right].
\end{aligned}$$

A Taylor series allows us to rewrite this as

$$\begin{aligned}
\text{Var}(f_2 | A_{n,h}) &= \frac{1}{\left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) \right)^4} \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 \right. \\
&\quad \left. + o\left(\frac{1}{nh^d}\right) \right] - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} A_{0,g}^2 \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \\
&= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_i) h^d} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + o\left(\frac{1}{nh^d}\right) \right) \\
&\quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathfrak{I}_{k,n}|^3} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right) \\
&\sim \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{1}{nh^d} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_i)} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + o\left(\frac{1}{nh^d}\right) \right) \\
&\quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\frac{1}{n} \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right) \tag{6.13}
\end{aligned}$$

The following Lemma is proven in Section 6.2.3.

Lemma 6.2.4. *Define the sets*

$$I_{\eta,\tau}(r) = \left\{ \eta, \tau \in \mathfrak{J}_{k,n} : \left| \{1, \dots, k\} \cup \{\eta(1), \dots, \eta(k)\} \cup \{\tau(1), \dots, \tau(k)\} \right| = r \right\}$$

$$I_{\sigma}(r) = \left\{ \sigma \in \mathfrak{J}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\}$$

then

$$\begin{aligned} \text{Cov} & \left(\left| \mathfrak{J}_{k,n} \right|^{-1} \sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}), \left| \mathfrak{J}_{k,n} \right|^{-1} \sum_{\sigma \in \mathfrak{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \right. \\ & \times \left. \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \left(\left| \mathfrak{J}_{k,n} \right|^{-1} \sum_{\sigma \in \mathfrak{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \middle| A_{n,h} \right) \\ & = \left(\frac{2|I_{\sigma,\tau}(3k-1)|}{3|\mathfrak{J}_{k,n}|^2} - \frac{|I_{\sigma}(1)|}{|\mathfrak{J}_{k,n}|} \right) \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 \right. \right. \\ & \quad \left. \left. + o\left(\frac{1}{h^d}\right) \right] - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \right). \end{aligned}$$

For n sufficiently large

$$\begin{aligned} & \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \\ & \quad \left. - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{1}{n} A_{0,g}^2 A_{0,1} + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \right) \end{aligned}$$

6

A Taylor expansion allows us to write

$$\begin{aligned} \text{Cov}(f_1, f_2 \mid A_{n,h}) & = \frac{1}{\left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) \right)^3} \left(\frac{2|I_{\sigma,\tau}(3k-1)|}{3|\mathfrak{J}_{k,n}|^2} - \frac{|I_{\sigma}(1)|}{|\mathfrak{J}_{k,n}|} \right) \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \right. \right. \\ & \quad \times \left. \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{h^d}\right) \right] - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \right) \\ & = \left(\frac{2|I_{\sigma,\tau}(3k-1)|}{3|\mathfrak{J}_{k,n}|^2} - \frac{|I_{\sigma}(1)|}{|\mathfrak{J}_{k,n}|} \right) \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_j) h^d} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 \right. \right. \\ & \quad \left. \left. + o\left(\frac{1}{h^d}\right) \right] - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}(h^2) \right] \right) \\ & \sim \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{1}{nh^d} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_j)} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + o\left(\frac{1}{nh^d}\right) \right) \\ & \quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\frac{1}{n} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right). \end{aligned} \quad (6.14)$$

We omit the full derivation for brevity, however, it can be shown that $\text{Var}(f_3)$, $\text{Cov}(f_1, f_3)$, and $\text{Cov}(f_2, f_3)$ are all of $o(nh^d)^{-1}$. Combining Equations (6.12), (6.13) and (6.14) we

obtain

$$\begin{aligned} \text{Var}(U_n | A_{n,h}) &= \text{Var}(f_1 | A_{n,h}) + \text{Var}(f_2 | A_{n,h}) - 2\text{Cov}(f_1, f_2 | A_{n,h}) + o\left(\frac{1}{nh^d}\right) \\ &\sim \frac{1}{nh^d} \sum_{i=1}^k \sum_{m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z}_i)} \left(\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_m, \dots, \mathbf{X}_{2k}) \mid \mathbf{Z}_j = \mathbf{z}_j \right. \right. \\ &\quad \left. \left. \text{for all } j, \mathbf{Z}_{k+j} = \mathbf{z}_j \text{ for } j \neq i \right] - \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \mid \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k \right]^2 \right) + o\left(\frac{1}{nh^d}\right) \end{aligned}$$

as desired. $\square$

6.2.1. PROOF OF LEMMA 6.2.2

By Corollary 6.1.6 we can ignore the conditioning on $A_{n,h}$, hence

$$\begin{aligned} &\text{Var} \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right) \\ &= |\mathcal{J}_{k,n}|^{-2} \mathbb{E} \left[\sum_{\sigma, \eta \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &\quad - \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]^2 \end{aligned}$$

By Lemma 6.1.4

$$\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] = \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + h^2 A_{2,g} + h^3 A_{3,g}.$$

For the other expression, we will first apply exchangeability to write

$$\begin{aligned} &|\mathcal{J}_{k,n}|^{-2} \mathbb{E} \left[\sum_{\sigma, \eta \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= |\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= \sum_{r=0}^k \sum_{\sigma \in I_{\sigma}(r)} \frac{1}{|\mathcal{J}_{k,n}|} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \end{aligned}$$

where we defined the set

$$I_{\sigma}(r) = \left\{ \sigma \in \mathcal{J}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\}.$$

Lemma 6.1.7 had shown that $|I_{\sigma}(r)| \sim n^{k-r}$, implying that the low values of r are the most dominant.

i) Suppose $\sigma \in I_\sigma(0)$.

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2 \\ &= \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + h^2 A_{2,g} + h^3 A_{3,g} \right)^2 \end{aligned}$$

ii) Suppose $\eta \in I_\sigma(1)$ and assume index m overlaps at position i .

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(i-1)}, \mathbf{X}_m, \mathbf{X}_{\sigma(i+1)}, \dots, \mathbf{X}_{\sigma(k)}) \right. \\ & \quad \times \left. \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) K_h(\mathbf{z}_i - \mathbf{Z}_m) \right] \\ &= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+i-1}, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{k+j}) K_h(\mathbf{z}_i - \mathbf{Z}_m) \\ & \quad \times \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r}. \end{aligned}$$

Now, if K has compact support

$$K\left(\frac{\mathbf{z}_i - \mathbf{Z}_m}{h}\right) K\left(\frac{\mathbf{z}_m - \mathbf{Z}_m}{h}\right) = 0$$

whenever $\mathbf{z}_i \neq \mathbf{z}_m$ and n is sufficiently large. Thus for $\sigma \in I_\sigma(1)$

$$\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] = 0 \quad \text{when } \mathbf{z}_i \neq \mathbf{z}_m.$$

When $\mathbf{z}_i = \mathbf{z}_m$,

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\ &= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+i-1}, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{z}_j - \mathbf{z}'_j) \prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{z}'_{k+j}) K_h(\mathbf{z}_m - \mathbf{z}'_m)^2 \\ & \quad \times \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}'_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}'_r) d\mathbf{x}_l d\mathbf{x}_{k+r} d\mathbf{z}'_l d\mathbf{z}'_{k+r} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{\substack{j=1 \\ j \neq m}}^k K_h(\mathbf{u}_j) \prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{u}_{k+j}) K_h(\mathbf{u}_m)^2 \\
&\quad \times \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l + h\mathbf{u}_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r + h\mathbf{u}_{k+r}) d\mathbf{x}_l d\mathbf{x}_{k+r} d\mathbf{u}_l d\mathbf{u}_{k+r} \\
&= \frac{\int K(\mathbf{u}) d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_m, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} \\
&\quad + o\left(\frac{1}{h^d}\right).
\end{aligned}$$

We applied the usual Taylor expansion. We will omit the derivation of the cases when $r > 1$ since these are negligible compared to the $r = 0$ and $r = 1$ case.

We now obtain that

$$\begin{aligned}
&\sum_{r=0}^k \sum_{\sigma \in I_\sigma(r)} |\mathcal{J}_{k,n}|^{-1} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right] \\
&= \frac{|I_\sigma(0)|}{|\mathcal{J}_{k,n}|} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + h^2 A_{2,g} + h^3 A_{3,g} \right)^2 \\
&+ \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathcal{J}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m, \dots, \mathbf{x}_{2k}}_{i'\text{th}}) \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\
&\quad \times \left. \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \right).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\text{Var} \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right) \\
&= \left(\frac{|I_\sigma(0)|}{|\mathcal{J}_{k,n}|} - 1 \right) \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + h^2 A_{2,g} + h^3 A_{3,g} \right)^2 \\
&+ \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{|I_\sigma(1)|}{|\mathcal{J}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m, \dots, \mathbf{x}_{2k}}_{i'\text{th}}) \prod_{l=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\
&\quad \times \left. \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \right).
\end{aligned}$$

It should be noted that $\frac{|I_\sigma(0)|}{|\mathcal{J}_{k,n}|} = 1 - \mathcal{O}\left(\frac{1}{n}\right)$, hence

$$\text{Var} \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right)$$

$$\begin{aligned}
&= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{|I_{\sigma}(1)|}{|\mathcal{I}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m}_{i'\text{th}}, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\
&\quad \times \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Big) \\
&+ \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left(\frac{|I_{\sigma}(0)|}{|\mathcal{I}_{k,n}|} - 1 \right) \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + \mathcal{O}\left(\frac{h^2}{n}\right)
\end{aligned}$$

When n is sufficiently large, the variance is similar to

$$\begin{aligned}
&\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) g(\mathbf{x}_{k+1}, \dots, \underbrace{\mathbf{x}_m}_{i'\text{th}}, \dots, \mathbf{x}_{2k}) \prod_{l=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_l, \mathbf{z}_l) \right. \\
&\quad \times \prod_{\substack{r=1 \\ r \neq i}}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+r}, \mathbf{z}_r) d\mathbf{x}_l d\mathbf{x}_{k+r} + o\left(\frac{1}{nh^d}\right) \Big).
\end{aligned}$$

This proves the lemma.

6.2.2. PROOF OF LEMMA 6.2.3

Corollary 6.1.6 shows that conditioning on $A_{n,h}$ decays exponentially, hence

$$\begin{aligned}
&\text{Var} \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
&\quad \times \left. \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right) \\
&= \mathbb{E} \left[\left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \right. \\
&\quad \times \left. \left. \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right)^2 \right] \\
&- \mathbb{E} \left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
&\quad \times \left. \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right)^2 \right].
\end{aligned}$$

Recall that we have already computed the second expression in Lemma 6.1.3

$$\begin{aligned}
&\mathbb{E} \left[|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
&\quad \times \left. \left(|\mathcal{I}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right)^2 \right]
\end{aligned}$$

$$= \left(\sum_{r=1}^k \frac{|I_{n,k}(r)|}{|\mathcal{I}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{x}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{nh^d}\right). \quad (6.15)$$

We now focus on the first expression, which can be rewritten using exchangeability as

$$\begin{aligned} & |\mathcal{I}_{k,n}|^{-4} \sum_{\sigma, \eta, \tau, \theta \in \mathcal{I}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \Big] \\ &= |\mathcal{I}_{k,n}|^{-3} \sum_{\eta, \tau, \theta \in \mathcal{I}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_j) \right. \\ & \quad \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \Big] \end{aligned} \quad (6.16)$$

Now define the set for $k \leq r \leq 4k$

$$I_{\eta, \tau, \theta}(r) = \left\{ \eta, \mu, \theta \in \mathcal{I}_{k,n} : \left| \{1, \dots, k\} \cup \{\theta(1), \dots, \theta(k)\} \cup \{\eta(1), \dots, \eta(k)\} \cup \{\tau(1), \dots, \tau(k)\} \right| = r \right\},$$

for the number of different indices. It can be shown that $|I_{\eta, \tau, \theta}(r)| \sim n^{r-k}$ so the cases where $r = 4k$ and $r = 4k - 1$ are the main contributions and other cases are negligible for n large enough. We will analyse these two cases separately.

- i) Let $\eta, \tau, \theta \in I_{\eta, \tau, \theta}(4k)$. This means there is no overlap between any of the indices and by independence we write

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_j) \right. \\ & \quad \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \Big] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) \right] \\ & \quad \times \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right] \cdot \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right] = 0. \end{aligned}$$

- ii) Let $\eta, \tau, \theta \in I_{\eta, \tau, \theta}(4k - 1)$ such that there is one overlapping index. Note that the expectation is again 0 unless the overlap is between $\{\tau(1), \dots, \tau(k)\}$ and $\{\theta(1), \dots, \theta(k)\}$. In that case

$$\mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]$$

$$\begin{aligned}
& \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \\
& = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) \right] \\
& \quad \times \mathbb{E} \left[\left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right].
\end{aligned} \tag{6.17}$$

Assume index $\theta(m)$ overlaps $\tau(i)$,

$$\begin{aligned}
& \mathbb{E} \left[\left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\
& = \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) \right] - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2 \\
& = \mathbb{E} \left[\prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) \prod_{\substack{r=1 \\ r \neq m}}^k K_h(\mathbf{z}_r - \mathbf{Z}_{\theta(r)}) K_h(\mathbf{z}_i - \mathbf{Z}) K_h(\mathbf{z}_m - \mathbf{Z}) \right] - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2 \\
& = \mathbb{E} \left[\prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) \right] \mathbb{E} \left[\prod_{\substack{r=1 \\ r \neq m}}^k K_h(\mathbf{z}_r - \mathbf{Z}_{\theta(r)}) \right] \mathbb{E} \left[K_h(\mathbf{z}_i - \mathbf{Z}) K_h(\mathbf{z}_m - \mathbf{Z}) \right] - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2.
\end{aligned}$$

6

We again have that since K has compact support,

$$K\left(\frac{\mathbf{z}_i - \mathbf{Z}}{h}\right) K\left(\frac{\mathbf{z}_m - \mathbf{Z}}{h}\right) = 0$$

for $\mathbf{z}_i \neq \mathbf{z}_m$ and n large enough. In that case

$$\begin{aligned}
& \mathbb{E} \left[\left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\
& = 0 - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2 = - \left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + \mathcal{O}(h^2) \right)^2.
\end{aligned}$$

If $\mathbf{z}_i = \mathbf{z}_m$

$$\begin{aligned}
& \mathbb{E} \left[\prod_{\substack{j=1 \\ j \neq i}}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) \right] \mathbb{E} \left[\prod_{\substack{r=1 \\ r \neq m}}^k K_h(\mathbf{z}_r - \mathbf{Z}_{\theta(r)}) \right] \mathbb{E} \left[K_h(\mathbf{z}_i - \mathbf{Z})^2 \right] - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right]^2 \\
& = \left(\frac{1}{h^d} \left(\int K(\mathbf{u})^2 d\mathbf{u} \right) f_{\mathbf{Z}}(\mathbf{z}_i) + o\left(\frac{1}{h^d}\right) \right) \left(\prod_{\substack{j=1 \\ j \neq i}}^k f_{\mathbf{Z}}(\mathbf{z}_j) + \mathcal{O}(h^2) \right) \left(\prod_{\substack{r=1 \\ r \neq m}}^k f_{\mathbf{Z}}(\mathbf{z}_r) + \mathcal{O}(h^2) \right) - \left(\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j) + \mathcal{O}(h^2) \right)^2 \\
& = \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{\substack{j=1 \\ j \neq i}}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 - \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) + \mathcal{O}(h^2).
\end{aligned}$$

We can conclude that Equation (6.17) becomes

$$\begin{aligned} & \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + \mathcal{O}(h^2) \right)^2 \left(- \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}(h^2) \right) \\ & + \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{\substack{j=1 \\ j \neq i}}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) \right). \end{aligned}$$

This implies

$$\begin{aligned} & \sum_{\eta, \tau, \theta \in I(4k-1)} |\mathcal{J}_{k,n}|^{-3} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) \right] \\ & \times \mathbb{E} \left[\left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ & = \frac{|I_{\eta, \tau, \theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} \left(A_{0,g} + \mathcal{O}(h^2) \right)^2 \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left(\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) \right) - \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 \right) \\ & = \frac{|I_{\eta, \tau, \theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} A_{0,g}^2 \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) \right] - \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 \right) + \mathcal{O}(h^2). \end{aligned}$$

We scale $|I_{\eta, \tau, \theta}(4k-1)|$ by $1/6$ because the expectation vanishes unless the overlap occurs specifically between τ and θ , one of six equally likely pairing possibilities. We will not consider the other possibilities since it can be shown that they are of at least $\mathcal{O}\left(\frac{1}{(nh^d)^2}\right)$.

This now means that Equation (6.16) can be rewritten as

$$\begin{aligned} & \sum_{r=0}^{4k} \sum_{I_{\eta, \tau, \theta}(r)} \frac{1}{|\mathcal{J}_{k,n}|^3} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) K_h(\mathbf{z}_j - \mathbf{Z}_j) \right. \\ & \times \left. \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\theta(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ & = \frac{|I_{\eta, \tau, \theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \\ & \quad \left. - A_{0,g}^2 \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}(h^2) \right] \right) \end{aligned}$$

If we now combine this result with Equation (6.15) we obtain

$$\begin{aligned} & \text{Var} \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \left. \times \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} \left(\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \\
&\quad \left. - A_{0,g}^2 \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}(h^2) \right] \right) \\
&\quad - \left(\sum_{r=1}^k \frac{|I_{n,k}(r)|}{|\mathcal{J}_{k,n}|} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 \\
&= \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{nh^d}\right) \right] \\
&\quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{|I_{\eta,\tau,\theta}(4k-1)|}{6|\mathcal{J}_{k,n}|^3} A_{0,g}^2 \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right].
\end{aligned}$$

When n is sufficiently big, this expression is similar to

$$\begin{aligned}
&\sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i=\mathbf{z}_m} \left[\frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} A_{0,g}^2 f_{\mathbf{Z}}(\mathbf{z}_i) \prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + o\left(\frac{1}{nh^d}\right) \right] \\
&\quad - \sum_{i,m=1}^k \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{1}{n} A_{0,g}^2 \prod_{j=1}^k f_{\mathbf{Z}}(\mathbf{z}_j)^2 + \mathcal{O}\left(\frac{h^2}{n}\right) \right]
\end{aligned}$$

which proves the claim.

6

6.2.3. PROOF OF LEMMA 6.2.4

Start with recalling that by definition

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

So we start by computing

$$\begin{aligned}
&\mathbb{E} \left[|\mathcal{J}_{k,n}|^{-2} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \sum_{\eta \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\eta(1)}, \dots, \mathbf{X}_{\eta(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\eta(j)}) \right. \\
&\quad \left. \times \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\tau \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\
&= |\mathcal{J}_{k,n}|^{-2} \sum_{\tau, \sigma \in \mathcal{J}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\
&\quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right].
\end{aligned}$$

We will use a similar approach as before and define the set $k \leq r \leq 3k$

$$I_{\sigma,\tau}(r) = \left\{ \sigma, \mu \in \mathcal{J}_{k,n} : \left| \{1, \dots, k\} \cup \{\sigma(1), \dots, \sigma(k)\} \cup \{\tau(1), \dots, \tau(k)\} \right| = r \right\},$$

which captures the number of different indices. It can be shown that $|I_{\sigma,\tau}(r)| \sim n^{r-k}$ so we will focus on the cases when $r = 3k$ and $r = 3k - 1$.

i) Suppose that $\sigma, \tau \in I_{\sigma, \tau}(3k)$ such that there is no overlap between the indices.

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] = 0 \end{aligned}$$

ii) Suppose that $\sigma, \tau \in I_{\sigma, \tau}(3k-1)$ such that there is one overlapping index. Note that the expectation is 0 if the overlap is between $\{1, \dots, k\}$ and $\{\sigma(1), \dots, \sigma(k)\}$. So assume that the overlap is between $\sigma(i)$ and $\tau(m)$. Then by Lemma 6.1.3 and Lemma 6.1.4

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ &= \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ &= \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j + o\left(\frac{1}{h^d}\right) \right] \right. \\ & \quad \left. - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g} A_{0,1} + \mathcal{O}(h^2) \right] \right) \cdot \left(A_{0,g} + h^2 A_{2,g} + h^3 A_{3,g} \right) \\ &= \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \\ & \quad \left. - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \right). \end{aligned}$$

It can be shown that all the other cases are of at least $\mathcal{O}\left(\frac{1}{h^{2d}}\right)$ and are thus negligible compared to the above.

We now obtain that

$$\begin{aligned} & |\mathcal{J}_{k,n}|^{-2} \sum_{\tau, \sigma \in \mathcal{J}_{k,n}} \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \left. \times \left(\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\tau(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ &= \frac{2|I_{\sigma, \tau}(3k-1)|}{3|\mathcal{J}_{k,n}|^2} \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \end{aligned}$$

$$- \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \Bigg). \quad (6.18)$$

The factor $2/3$ is because the expectation vanishes if the overlap is between $\{1, \dots, k\}$ and $\{\sigma(1), \dots, \sigma(k)\}$, one of three equally likely pairing possibilities.

Using Lemma 6.1.3 and 6.1.4 once more, we obtain

$$\begin{aligned} & \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \cdot \mathbb{E} \left[|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right. \\ & \quad \times \left. \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right] \\ &= \frac{|I_\sigma(1)|}{|\mathcal{J}_{k,n}|} \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{h^d}\right) \right] \right. \\ & \quad \left. - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \right). \quad (6.19) \end{aligned}$$

With $I_\sigma(r) = \left\{ \sigma \in \mathcal{J}_{k,n} : \left| \{1, \dots, k\} \cap \{\sigma(1), \dots, \sigma(k)\} \right| = r \right\}$. By combining Equation (6.18) and (6.19) we obtain

$$\begin{aligned} & \text{Cov} \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}), |\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \right. \\ & \quad \times \left. \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \left(|\mathcal{J}_{k,n}|^{-1} \sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_j) \right] \right) \right) \\ &= \left(\frac{2|I_{\sigma,r}(3k-1)|}{3|\mathcal{J}_{k,n}|^2} - \frac{|I_\sigma(1)|}{|\mathcal{J}_{k,n}|} \right) \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{h^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 \right. \right. \\ & \quad \left. \left. + o\left(\frac{1}{h^d}\right) \right] - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[A_{0,g}^2 A_{0,1} + \mathcal{O}(h^2) \right] \right) \end{aligned}$$

and for n sufficiently large

$$\begin{aligned} & \sum_{i,m=1}^k \left(\mathbb{1}_{\mathbf{z}_i = \mathbf{z}_m} \left[\prod_{j \neq i}^k f_{\mathbf{Z}}(\mathbf{z}_j) \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{nh^d} \left(\int g(\mathbf{x}_1, \dots, \mathbf{x}_k) \prod_{j=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) d\mathbf{x}_j \right)^2 + o\left(\frac{1}{nh^d}\right) \right] \right. \\ & \quad \left. - \mathbb{1}_{\mathbf{z}_i \neq \mathbf{z}_m} \left[\frac{1}{n} A_{0,g}^2 A_{0,1} + \mathcal{O}\left(\frac{h^2}{n}\right) \right] \right). \end{aligned}$$

This proves Lemma 6.2.4.

6.3. EXPECTATION AND VARIANCE OF THE CKT ESTIMATOR

In this section, we apply the general theory for conditional U-statistics developed in the previous subsections to our estimator of the conditional Kendall's tau. For that, let $(\mathbf{X}_1, \mathbf{Z}_1), (\mathbf{X}_2, \mathbf{Z}_2), \dots, (\mathbf{X}_n, \mathbf{Z}_n)$ be i.i.d. observations in $\mathbb{R}^p \times \mathbb{R}^d$ with joint distribution $F_{\mathbf{X}, \mathbf{Z}}$, where each $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,p}) \in \mathbb{R}^p$. We focus on the estimator $\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z})$, which can be written as a normalized conditional U-statistic. Specifically, it is defined as $\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) = \frac{\hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z})}{1 - s_n}$ where $s_n = \sum_{j=1}^n w_{j,n}^2(\mathbf{z})$ and $w_{j,n}(\mathbf{z})$ are the Nadaraya–Watson weights introduced in Equation (2.12). The numerator is given by

$$\hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z}) = \sum_{j=1}^n \sum_{i=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) g^{(1)}(\mathbf{X}_{i, (j_1, j_2)}, \mathbf{X}_{j, (j_1, j_2)})$$

where the kernel $g^{(1)}$ is defined in Equation (2.13). This representation makes clear that $\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z})$ inherits the structure of a conditional U-statistic. As a result, the asymptotic results derived earlier, can be applied.

Theorem 6.3.1. *Define the event*

$$A_{n,h} := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{I}_{2,n}} \prod_{j=1}^2 K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{I}_{2,n}} \prod_{j=1}^2 K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{I}_{2,n}} \prod_{j=1}^2 K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

then under Assumptions 1, 2, 8, 9 and 10 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^d \rightarrow \infty$ such that $\frac{1}{nh^d} = o(h^2)$, we have

$$\mathbb{E} \left(\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) | A_{n,h} \right) = \tau_{j_1, j_2}(\mathbf{z}) + h^2 \left(\frac{A_{2,g}}{f_{\mathbf{Z}}(\mathbf{z})^2} - \frac{A_{2,1}}{f_{\mathbf{Z}}(\mathbf{z})^2} \tau_{j_1, j_2}(\mathbf{z}) \right) + \mathcal{O} \left(h^3 + \frac{1}{nh^d} \right)$$

and

$$\text{Var} \left(\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) | A_{n,h} \right) \sim \frac{1}{nh^d} \frac{4 \int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z})} \text{Cov} \left(g^{(1)}(\mathbf{X}_1, \mathbf{X}_2), g^{(1)}(\mathbf{X}_1, \mathbf{X}_3) \middle| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{z} \right) + o \left(\frac{1}{nh^d} \right)$$

where

$$A_{2,g} = \int g(\mathbf{x}_1, \mathbf{x}_2) K(\mathbf{u}_1) K(\mathbf{u}_2) \left(A_2(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_2, \mathbf{z}) + A_2(\mathbf{x}_2, \mathbf{z}_2, \mathbf{u}_2) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_1, \mathbf{z}) \right) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{u}_1 d\mathbf{u}_2$$

$$A_{2,1} = \int K(\mathbf{u}_1) K(\mathbf{u}_2) \left(A_2(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_2, \mathbf{z}) + A_2(\mathbf{x}_2, \mathbf{z}_2, \mathbf{u}_2) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_1, \mathbf{z}) \right) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{u}_1 d\mathbf{u}_2$$

with $A_2(\mathbf{x}, \mathbf{z}, \mathbf{u})$ as in Equation (6.8) and g as in Equation (2.13).

Proof. We start by showing the estimator of the conditional Kendall's tau can be written as a conditional U-statistic of order 2. Indeed, let $g^{(1)}: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ as in Equation (2.13). Where we can choose $g^{(1)}$ without loss of generality. Following Section 2.7.1 we can define the functional as

$$\theta(F_{\mathbf{X}}|_{\mathbf{Z}=\mathbf{z}_1}, F_{\mathbf{X}}|_{\mathbf{Z}=\mathbf{z}_2}) = \mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \mathbf{Z}_2 = \mathbf{z}_2 \right].$$

From Definition 2.7.5 we then obtain that the conditional U -statistic estimator writes as

$$U_n(\mathbf{z}_1, \mathbf{z}_2) = \frac{\sum_{\sigma \in \mathfrak{I}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1)}, \mathbf{X}_{\sigma(2)}) \prod_{j=1}^2 K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathfrak{I}_{2,n}} \prod_{j=1}^2 K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}.$$

In order to compare U_n with the estimator

$$\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) = \frac{\hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z})}{1 - s_n}$$

from Equation (2.16), we start by rewriting

$$\begin{aligned} \hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z}) &= \sum_{j=1}^n \sum_{i=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) g^{(1)}(\mathbf{X}_{i, (j_1, j_2)}, \mathbf{X}_{j, (j_1, j_2)}) \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{K_h(\mathbf{z} - \mathbf{Z}_j) \cdot K_h(\mathbf{z} - \mathbf{Z}_i)}{(\sum_{k=1}^n K_h(\mathbf{z} - \mathbf{Z}_k))^2} g^{(1)}(\mathbf{X}_{i, (j_1, j_2)}, \mathbf{X}_{j, (j_1, j_2)}) \\ &= \frac{\sum_{\sigma \in S_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(2), (j_1, j_2)}) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{\sigma \in S_{2,n}} \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}. \end{aligned} \quad (6.20)$$

We have that $\mathbf{X}_{i, (j_1, j_2)} = (X_{i, j_1}, X_{i, j_2}) \in \mathbb{R}^2$ and $S_{2,n} = \{\sigma = (i, j) : 1 \leq i, j \leq n\}$. Note that Equation (6.20) resembles U_n but is not yet in the same framework since $S_{2,n} \neq \mathfrak{I}_{2,n}$. However, note that we can decompose the numerator as

$$\begin{aligned} &\sum_{\sigma \in S_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(2), (j_1, j_2)}) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)}) \\ &= \sum_{\sigma \in \mathfrak{I}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(2), (j_1, j_2)}) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)}) \\ &\quad + \sum_{\sigma \in \mathfrak{I}_{1,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(1), (j_1, j_2)}) K_h(\mathbf{z} - \mathbf{Z}_{\sigma(1)})^2. \end{aligned}$$

Now recall by the Equation (2.13) that

$$g^{(1)}(\mathbf{x}, \mathbf{x}) = 0.$$

Hence Equation (6.20) becomes

$$\frac{\sum_{\sigma \in \mathfrak{I}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(2), (j_1, j_2)}) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{i,j}^n K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j)}.$$

Now by definition

$$\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) = \frac{1}{1 - \sum_{j=1}^n w_{j,n}^2} \frac{\sum_{\sigma \in \mathfrak{I}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \mathbf{X}_{\sigma(2), (j_1, j_2)}) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{i,j}^n K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j)}$$

$$\begin{aligned}
&= \frac{1}{1 - \frac{\sum_{j=1}^n K_h(\mathbf{z} - \mathbf{Z}_j)^2}{\sum_{i,j}^n K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j)}} \frac{\sum_{\sigma \in \mathcal{J}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1)}, (j_1, j_2), \mathbf{X}_{\sigma(2)}, (j_1, j_2)) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{i,j}^n K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j)} \\
&= \frac{\sum_{\sigma \in \mathcal{J}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1)}, (j_1, j_2), \mathbf{X}_{\sigma(2)}, (j_1, j_2)) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{i,j} K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j) - \sum_{j=1}^n K_h(\mathbf{z} - \mathbf{Z}_j)^2} \\
&= \frac{\sum_{\sigma \in \mathcal{J}_{2,n}} g^{(1)}(\mathbf{X}_{\sigma(1)}, (j_1, j_2), \mathbf{X}_{\sigma(2)}, (j_1, j_2)) \prod_{r=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(r)})}{\sum_{\sigma \in \mathcal{J}_{2,n}} \prod_{j=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(j)})}.
\end{aligned}$$

Note that now $\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z})$ exactly resembles a conditional U -statistic of order 2. Then by Lemma 6.1.1 we obtain

$$\mathbb{E} \left[\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \mid A_{n,h} \right] = \tau_{j_1, j_2}(\mathbf{z}) + h^2 \left(\frac{A_{2,g}}{f_{\mathbf{Z}}(\mathbf{z})^2} - \frac{\tau_{j_1, j_2}(\mathbf{z}) A_{2,1}}{f_{\mathbf{Z}}(\mathbf{z})^2} \right) + \mathcal{O} \left(h^3 + \frac{1}{nh^d} \right),$$

where we used that

$$\mathbb{E} \left[g^{(1)}(\mathbf{X}_1, (j_1, j_2), \mathbf{X}_2, (j_1, j_2)) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z} \right] = \tau_{j_1, j_2}(\mathbf{z}),$$

and

$$A_{2,g} = \int g(\mathbf{x}_1, \mathbf{x}_2) K(\mathbf{u}_1) K(\mathbf{u}_2) \left(A_2(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_2, \mathbf{z}) + A_2(\mathbf{x}_2, \mathbf{z}_2, \mathbf{u}_2) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_1, \mathbf{z}) \right) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{u}_1 d\mathbf{u}_2$$

$$A_{2,1} = \int K(\mathbf{u}_1) K(\mathbf{u}_2) \left(A_2(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_2, \mathbf{z}) + A_2(\mathbf{x}_2, \mathbf{z}_2, \mathbf{u}_2) f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_1, \mathbf{z}) \right) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{u}_1 d\mathbf{u}_2$$

with $A_2(\mathbf{x}, \mathbf{z}, \mathbf{u})$ as in Equation (6.8). Similarly Theorem 6.2.1 can be applied

$$\begin{aligned}
\text{Var}(\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \mid A_{n,h}) &\sim \frac{1}{nh^d} \frac{\int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z})} \left(\left(\mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) g^{(1)}(\mathbf{X}_1, \mathbf{X}_4) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_4 = \mathbf{z} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z} \right]^2 \right) + \left(\mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) g^{(1)}(\mathbf{X}_2, \mathbf{X}_4) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_4 = \mathbf{z} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z} \right]^2 \right) + \left(\mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) g^{(1)}(\mathbf{X}_3, \mathbf{X}_1) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{z} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z} \right]^2 \right) + \left(\mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) g^{(1)}(\mathbf{X}_3, \mathbf{X}_2) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{z} \right] \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left[g^{(1)}(\mathbf{X}_1, \mathbf{X}_2) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z} \right]^2 \right) \right) + o \left(\frac{1}{nh^d} \right) \\
&= \frac{1}{nh^d} \frac{4 \int K(\mathbf{u})^2 d\mathbf{u}}{f_{\mathbf{Z}}(\mathbf{z})} \text{Cov} \left(g^{(1)}(\mathbf{X}_1, \mathbf{X}_2), g^{(1)}(\mathbf{X}_3, \mathbf{X}_1) \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{z} \right) + o \left(\frac{1}{nh^d} \right)
\end{aligned}$$

which proves the result. $\square$

We can now also find an expression for the *mean squared error* of $\hat{\tau}_{j_1, j_2}^{(4)}$ by using

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\theta - \hat{\theta})^2] = \text{Var}(\hat{\theta}) + \text{Bias}(\theta, \hat{\theta})^2.$$

Lemma 6.3.2 (Mean Squared Error of $\hat{\tau}_{j_1, j_2}^{(4)}$). *Under the assumptions of Theorem 6.3.1*

$$\text{MSE}(\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \mid A_{n,h}) = h^4 \left(\frac{A_{2,g}}{f_{\mathbf{Z}}(\mathbf{z})^2} - \frac{A_{2,1}}{f_{\mathbf{Z}}(\mathbf{z})^2} \tau_{j_1, j_2}(\mathbf{z}) \right) + \mathcal{O} \left(h^5 + \frac{1}{nh^d} \right).$$

6.4. MOMENTS OF THE CKT ESTIMATOR

In this section, we extend the results on the CKT estimator by computing its moments.

Theorem 6.4.1. *Define the event*

$$A_{n,h,k}(\mathbf{z}_1, \dots, \mathbf{z}_k) := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

Then under Assumptions 1, 2, 8, 9 and 10 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^d \rightarrow \infty$ such that $\frac{1}{nh^d} = o(h^2)$, we have

$$\begin{aligned} & \mathbb{E} \left[\underbrace{\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(4)}(\mathbf{z})}_p \underbrace{\hat{\tau}_{k_1, k_2}^{(4)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(4)}(\mathbf{z}')}_m \middle| A_{n,h,2(p+m)}(\mathbf{z}_1, \dots, \mathbf{z}_{2(p+m)}) \right] \\ &= \tau_{j_1, j_2}^{(4)}(\mathbf{z}) \cdots \tau_{l_1, l_2}^{(4)}(\mathbf{z}) \tau_{k_1, k_2}^{(4)}(\mathbf{z}') \cdots \tau_{i_1, i_2}^{(4)}(\mathbf{z}') \\ &+ h^2 \left(\frac{A_{2,g} - A_{2,1} \tau_{j_1, j_2}^{(4)}(\mathbf{z}) \cdots \tau_{l_1, l_2}^{(4)}(\mathbf{z}) \tau_{k_1, k_2}^{(4)}(\mathbf{z}') \cdots \tau_{i_1, i_2}^{(4)}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z})^{2p} f_{\mathbf{Z}}(\mathbf{z}')^{2m}} \right) + \mathcal{O} \left(h^3 + \frac{1}{nh^d} \right) \end{aligned}$$

where

$$\begin{aligned} A_{2,g} &= \int g(\mathbf{x}_1, \dots, \mathbf{x}_{2(p+m)}) \prod_{j=1}^{2(p+m)} K(\mathbf{u}_j) \sum_{l=1}^{2(p+m)} \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \\ A_{2,1} &= \int \prod_{j=1}^{2(p+m)} K(\mathbf{u}_j) \sum_{l=1}^{2(p+m)} \left(A_2(\mathbf{x}_l, \mathbf{z}_l, \mathbf{u}_l) \prod_{i \neq l} f_{\mathbf{X}, \mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) \right) d\mathbf{x}_j d\mathbf{u}_j \end{aligned}$$

with $A_2(\mathbf{x}, \mathbf{z}, \mathbf{u})$ as in Equation (6.8), $(\mathbf{z}_1, \dots, \mathbf{z}_{2p}, \mathbf{z}_{2p+1}, \dots, \mathbf{z}_{2(p+m)}) = (\mathbf{z}, \dots, \mathbf{z}, \mathbf{z}', \dots, \mathbf{z}')$ and

$$g(\mathbf{x}_1, \dots, \mathbf{x}_{2(p+m)}) = g^{(1)}(\mathbf{x}_1, \mathbf{x}_2) \cdots g^{(1)}(\mathbf{x}_{2(p+m)-1}, \mathbf{x}_{2(p+m)})$$

from Equation (2.13).

Proof. We aim to show that the estimator can be written as a conditional U -statistic of order $2(p+m)$. To start, let $g : (\mathbb{R}^2)^{2(p+m)} \rightarrow \mathbb{R}$, and follow Section 2.7.1 to define the functional as

$$\theta = \mathbb{E} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_{2(p+m)}) \middle| \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_{2(p+m)} = \mathbf{z}_{2(p+m)} \right].$$

From Definition 2.7.5 we obtain that θ can be estimated by

$$U_n(\mathbf{z}_1, \dots, \mathbf{z}_{2(p+m)}) = \frac{\sum_{\sigma \in \mathcal{I}_{4p,n}} g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(2(p+m))}) \prod_{j=1}^{2(p+m)} K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}{\sum_{\sigma \in \mathcal{I}_{4p,n}} \prod_{j=1}^{2(p+m)} K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})}.$$

To see that

$$\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(4)}(\mathbf{z}) \hat{\tau}_{k_1, k_2}^{(4)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(4)}(\mathbf{z}') = \frac{\hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(1)}(\mathbf{z}) \hat{\tau}_{k_1, k_2}^{(1)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(1)}(\mathbf{z}')}{(1 - s_{n,\mathbf{z}})^p (1 - s_{n,\mathbf{z}'})^m} \quad (6.21)$$

also defines a U -statistic, take $g^{(1)}$ as in Equation (2.13) and write

$$\begin{aligned}
 & \hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(1)}(\mathbf{z}) \hat{\tau}_{k_1, k_2}^{(1)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(1)}(\mathbf{z}') \\
 &= \sum_{m_1, \dots, m_{2(p+m)}=1}^n w_{m_1, n}(\mathbf{z}) \cdots w_{m_{2p}, n}(\mathbf{z}) w_{m_{2p+1}, n}(\mathbf{z}') \cdots w_{m_{2(p+m)}, n}(\mathbf{z}') \\
 & \quad \times g^{(1)}(\mathbf{X}_{m_1, (j_1, j_2)}, \mathbf{X}_{m_2, (j_1, j_2)}) \cdots g^{(1)}(\mathbf{X}_{m_{2(p+m)-1}, (i_1, i_2)}, \mathbf{X}_{m_{2(p+m)}, (i_1, i_2)}) \\
 &= \sum_{m_1, \dots, m_{2(p+m)}=1}^n \frac{\prod_{j=1}^{2p} K_h(\mathbf{z} - \mathbf{Z}_{m_j}) \prod_{j=2p+1}^{2(p+m)} K_h(\mathbf{z}' - \mathbf{Z}_{m_j})}{\left(\sum_{i=1}^n K_h(\mathbf{z} - \mathbf{Z}_i) \right)^{2p} \left(\sum_{i=1}^n K_h(\mathbf{z}' - \mathbf{Z}_i) \right)^{2m}} g^*(\mathbf{X}_{m_1, (j_1, j_2)}, \dots, \mathbf{X}_{m_{2(p+m)}, (i_1, i_2)}) \\
 &= \frac{\sum_{\sigma \in S_{2(p+m), n}} g^*(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \dots, \mathbf{X}_{\sigma(2(p+m)), (i_1, i_2)}) \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})}{\sum_{\sigma \in S_{2(p+m), n}} \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})} \quad (6.22)
 \end{aligned}$$

where $(\mathbf{z}_1, \dots, \mathbf{z}_{2p}, \mathbf{z}_{2p+1}, \dots, \mathbf{z}_{2(p+m)}) = (\mathbf{z}, \dots, \mathbf{z}, \mathbf{z}', \dots, \mathbf{z}')$ and

$$g^*(\mathbf{x}_1, \dots, \mathbf{x}_{4p}) := g^{(1)}(\mathbf{x}_1, \mathbf{x}_2) \cdots g^{(1)}(\mathbf{x}_{2(p+m)-1}, \mathbf{x}_{2(p+m)}).$$

We have that $\mathbf{X}_{i, (j_1, j_2)} = (X_{i, j_1}, X_{i, j_2}) \in \mathbb{R}^2$ and $S_{2(p+m), n} = \{\sigma = (m_1, \dots, m_{2(p+m)}) : 1 \leq m_1, \dots, m_{2(p+m)} \leq n\}$. Note that Equation (6.22) resembles U_n but is not yet in the same framework since $S_{2(p+m), n} \neq \mathcal{I}_{2(p+m), n}$.

In the proof of Theorem 6.3.1 we have seen that

$$(1 - s_{n, \mathbf{z}}) \sum_{i, j} K_h(\mathbf{z} - \mathbf{Z}_i) K_h(\mathbf{z} - \mathbf{Z}_j) = \sum_{\sigma \in \mathcal{I}_{2, n}} \prod_{j=1}^2 K_h(\mathbf{z} - \mathbf{Z}_{\sigma(j)}).$$

Similarly, we can combine Equation (6.21) with the denominator of Equation (6.22) to obtain

$$(1 - s_{n, \mathbf{z}})^p (1 - s_{n, \mathbf{z}'})^m \sum_{\sigma \in S_{2(p+m), n}} \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) = \sum_{\sigma \in \mathcal{I}_{2p, n}} \prod_{j=1}^{2p} K_h(\mathbf{z} - \mathbf{Z}_{\sigma(j)}) \sum_{\eta \in \mathcal{I}_{2m, n}} \prod_{j=1}^{2m} K_h(\mathbf{z}' - \mathbf{Z}_{\eta(j)}).$$

Now this rewrites as

$$\begin{aligned}
 & \sum_{\sigma \in \mathcal{I}_{2p, n}} \prod_{j=1}^{2p} K_h(\mathbf{z} - \mathbf{Z}_{\sigma(j)}) \sum_{\eta \in \mathcal{I}_{2m, n}} \prod_{j=1}^{2m} K_h(\mathbf{z}' - \mathbf{Z}_{\eta(j)}) \\
 &= \sum_{\sigma \in \mathcal{I}_{2(p+m), n}} \prod_{j=1}^{2(p+m)} K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) + \sum_{i=1}^{2 \min(p, m)} \sum_{\sigma \in \mathcal{I}_{4p-i, n}} C_i H_i(\sigma)
 \end{aligned}$$

where $H_i(\sigma)$ consists of a product of $2(p+m)$ kernels where it collects all terms in which exactly i indices coincide between the two index sets. Recall that since K has compact support and $\mathbf{z} \neq \mathbf{z}'$, for n large enough we have

$$K_h(\mathbf{z} - \mathbf{Z}_j) K_h(\mathbf{z}' - \mathbf{Z}_j) = 0.$$

Hence, terms with overlapping indices involve factors of the form $K_h(\mathbf{z} - \mathbf{Z}_j)K_h(\mathbf{z}' - \mathbf{Z}_j)$ will vanish for n large enough. Therefore all overlap terms disappear and so

$$\sum_{\sigma \in \mathcal{I}_{2p,n}} \prod_{j=1}^{2p} K_h(\mathbf{z} - \mathbf{Z}_{\sigma(j)}) \sum_{\sigma \in \mathcal{I}_{2m,n}} \prod_{j=1}^{2m} K_h(\mathbf{z}' - \mathbf{Z}_{\sigma(j)}) = \sum_{\sigma \in \mathcal{I}_{2(p+m),n}} \prod_{j=1}^{2(p+m)} K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}).$$

To show that the numerator of Equation (6.22) can be written as a U-statistic, we decompose this sum according to the number of distinct indices in σ . This yields

$$\begin{aligned} & \sum_{\sigma \in \mathcal{S}_{2(p+m),n}} g^*(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \dots, \mathbf{X}_{\sigma(2(p+m)), (i_1, i_2)}) \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) \\ &= \sum_{\sigma \in \mathcal{I}_{2(p+m),n}} g^*(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \dots, \mathbf{X}_{\sigma(2(p+m)), (i_1, i_2)}) \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) \\ & \quad + \sum_{k=1}^{2(p+m)-1} \sum_{\sigma \in \mathcal{I}_{2(p+m)-k,n}} C_{p,k} g_k^*(\sigma) H_k(\sigma), \end{aligned}$$

where the second sum collects all terms in which at least one index appears more than once. $C_{p,k}$ is a combinatorial coefficient counting the configurations with exactly k index collisions, g_k^* is the function obtained from g^* by identifying arguments corresponding to repeated indices. Specifically, if a configurations maps the $2(p+m)$ positions to $2(p+m) - k$ distinct indices via a surjection $\pi : \{1, \dots, 2(p+m)\} \rightarrow \{1, \dots, 2(p+m) - k\}$ then

$$g_k^*(\mathbf{x}_1, \dots, \mathbf{x}_{2(p+m)-k}) := g^*(\mathbf{x}_{\pi(1)}, \dots, \mathbf{x}_{\pi(2(p+m))}).$$

H_k is the corresponding kernel product

$$H_k(\mathbf{z}_1, \dots, \mathbf{z}_{2(p+m)-k}) := \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{z}_{\pi(r)}).$$

Recall that $g^{(1)}(\mathbf{x}, \mathbf{x}) = 0$. This implies that for any $1 \leq j \leq p+m$, if two arguments within a pair coincide, i.e. $\mathbf{x}_{2j-1} = \mathbf{x}_{2j}$, then $g^* = 0$. Similarly, because $K_h(\mathbf{z} - \mathbf{Z}_j)K_h(\mathbf{z}' - \mathbf{Z}_j) = 0$ for n large enough, we find that $H_k = 0$ whenever there is overlap between the first $2p$ and last $2m$ arguments. With these two conditions it can in fact be shown that for $m = p = 1$,

$$\begin{aligned} & \sum_{\sigma \in \mathcal{S}_{4,n}} g^*(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \dots, \mathbf{X}_{\sigma(4), (k_1, k_2)}) \prod_{r=1}^4 K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}) \\ &= \sum_{\sigma \in \mathcal{I}_{4,n}} g^*(\mathbf{X}_{\sigma(1), (j_1, j_2)}, \dots, \mathbf{X}_{\sigma(4), (k_1, k_2)}) \prod_{r=1}^4 K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)}). \end{aligned}$$

We can rewrite Equation (6.22) and obtain

$$\hat{\tau}_{j_1, j_2}^{(4)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(4)}(\mathbf{z}) \hat{\tau}_{k_1, k_2}^{(4)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(4)}(\mathbf{z}') = \frac{\hat{\tau}_{j_1, j_2}^{(1)}(\mathbf{z}) \cdots \hat{\tau}_{l_1, l_2}^{(1)}(\mathbf{z}) \hat{\tau}_{k_1, k_2}^{(1)}(\mathbf{z}') \cdots \hat{\tau}_{i_1, i_2}^{(1)}(\mathbf{z}')}{(1 - s_{n,\mathbf{z}})^p (1 - s_{n,\mathbf{z}'})^m}$$

$$\begin{aligned}
&= \frac{\sum_{\sigma \in \mathcal{I}_{2(p+m),n}} g^*(\mathbf{X}_{\sigma(1),(j_1,j_2)}, \dots, \mathbf{X}_{\sigma(2(p+m)),(i_1,i_2)}) \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})}{\sum_{\sigma \in \mathcal{I}_{2(p+m),n}} \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})} \\
&+ \sum_{k=1}^{2(p+m)-1} \frac{\sum_{\sigma \in \mathcal{I}_{2(p+m)-k,n}} C_{p,k} g_k^*(\sigma) H_k(\sigma)}{\sum_{\sigma \in \mathcal{I}_{2(p+m),n}} \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})}.
\end{aligned}$$

where $(\mathbf{z}_1, \dots, \mathbf{z}_{2p}, \mathbf{z}_{2p+1}, \dots, \mathbf{z}_{2(p+m)}) = (\mathbf{z}, \dots, \mathbf{z}, \mathbf{z}', \dots, \mathbf{z}')$. It can be shown using the proof of Theorem 6.1.1 that

$$\frac{\sum_{\sigma \in \mathcal{I}_{2(p+m)-k,n}} C_{p,k} g_k^*(\sigma) H_k(\sigma)}{\sum_{\sigma \in \mathcal{I}_{2(p+m),n}} \prod_{r=1}^{2(p+m)} K_h(\mathbf{z}_r - \mathbf{Z}_{\sigma(r)})} = \mathcal{O}\left(\frac{1}{(nh^d)^k}\right).$$

Then we obtain for n sufficiently large that and using Lemma 6.1.1 that

$$\begin{aligned}
&\mathbb{E}\left[\hat{\tau}_{j_1,j_2}^{(4)}(\mathbf{z}) \cdots \hat{\tau}_{l_1,l_2}^{(4)}(\mathbf{z}) \hat{\tau}_{k_1,k_2}^{(4)}(\mathbf{z}') \cdots \hat{\tau}_{i_1,i_2}^{(4)}(\mathbf{z}') \mid A_{n,h,2(p+m)}(\mathbf{z}, \dots, \mathbf{z}')\right] \\
&= \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \cdots \tau_{l_1,l_2}^{(4)}(\mathbf{z}) \tau_{k_1,k_2}^{(4)}(\mathbf{z}') \cdots \tau_{i_1,i_2}^{(4)}(\mathbf{z}') \\
&+ h^2 \left(\frac{A_{2,g} - A_{2,1} \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \cdots \tau_{l_1,l_2}^{(4)}(\mathbf{z}) \tau_{k_1,k_2}^{(4)}(\mathbf{z}') \cdots \tau_{i_1,i_2}^{(4)}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z})^2 p f_{\mathbf{Z}}(\mathbf{z}')^{2m}} \right) + \mathcal{O}\left(h^3 + \frac{1}{nh^d}\right)
\end{aligned}$$

where we used that

$$\mathbb{E}\left[g^*(\mathbf{X}_{1,(j_1,j_2)}, \dots, \mathbf{X}_{2(p+m),(i_1,i_2)}) \mid \mathbf{Z}_{2p+1} = \dots = \mathbf{Z}_{2(p+m)} = \mathbf{z}'\right] = \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \cdots \tau_{l_1,l_2}^{(4)}(\mathbf{z}) \tau_{k_1,k_2}^{(4)}(\mathbf{z}') \cdots \tau_{i_1,i_2}^{(4)}(\mathbf{z}').$$

□

Note that for the special case when $m = p = 1$ and $(j_1, j_2) = (k_1, k_2)$, Theorem 6.4.1 gives

$$\begin{aligned}
&\mathbb{E}\left[\hat{\tau}_{j_1,j_2}^{(4)}(\mathbf{z}) \hat{\tau}_{j_1,j_2}^{(4)}(\mathbf{z}') \mid A_{n,h,4}(\mathbf{z}, \mathbf{z}, \mathbf{z}', \mathbf{z}')\right] = \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \tau_{j_1,j_2}^{(4)}(\mathbf{z}') \\
&+ h^2 \left(\frac{A_{2,g} - A_{2,1} \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \tau_{j_1,j_2}^{(4)}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z})^2 f_{\mathbf{Z}}(\mathbf{z}')^2} \right) + \mathcal{O}\left(h^3 + \frac{1}{nh^d}\right).
\end{aligned}$$

Combining this with Lemma 6.1.1 gives the following corollary.

Corollary 6.4.2. *Define the event*

$$A_{n,h,k}(\mathbf{z}_1, \dots, \mathbf{z}_k) := \left\{ \left| \frac{\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E}\left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]}{\mathbb{E}\left[\sum_{\sigma \in \mathcal{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})\right]} \right| \leq \frac{1}{2} \right\}.$$

Then under Assumptions 1, 2, 8, 9 and 10 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^d \rightarrow \infty$ such that $\frac{1}{nh^d} = o(h^2)$, we have

$$\begin{aligned}
&\text{Cov}\left(\hat{\tau}_{j_1,j_2}^{(4)}(\mathbf{z}, \mathbf{z}), \hat{\tau}_{j_1,j_2}^{(4)}(\mathbf{z}') \mid A_{n,h,4}(\mathbf{z}, \mathbf{z}, \mathbf{z}', \mathbf{z}') \cap A_{n,h,2}(\mathbf{z}, \mathbf{z}) \cap A_{n,h,2}(\mathbf{z}', \mathbf{z}')\right) \\
&= h^2 \left(\frac{A_{2,g} - A_{2,1} \tau_{j_1,j_2}^{(4)}(\mathbf{z}) \tau_{j_1,j_2}^{(4)}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z})^2 f_{\mathbf{Z}}(\mathbf{z}')^2} \right) + \mathcal{O}\left(h^3 + \frac{1}{nh^d}\right).
\end{aligned}$$

6.5. CONVERGENCE OF THE INTEGRAL OPERATORS

In order to prove convergence of the operator $\hat{\Gamma}_p$ we need some additional assumptions on the kernel.

Assumption 11. *We have the following properties for kernel K*

- i) K is piecewise Lipschitz. This means there exist finitely many sets $A_1, \dots, A_M \subset \mathcal{Z}$ such that $\mathcal{Z} = \bigcup_{m=1}^M A_m$, and for every m there exists a constant $L_m > 0$ such that

$$|K(\mathbf{z}) - K(\mathbf{z}')| \leq L_m \|\mathbf{z} - \mathbf{z}'\|, \quad \text{for all } \mathbf{z}, \mathbf{z}' \in A_m.$$

Moreover, the boundaries ∂A_m have Lebesgue measure zero.

- ii) $\inf_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}[K_h(\mathbf{z} - \mathbf{Z})] > 0$ for all sufficiently small h .

Assumption 12. *Suppose the integrals*

$$\int I_1(\mathbf{z}) d\mathbf{z} < \infty \quad \text{and} \quad \int I_2(\mathbf{z}) d\mathbf{z} < \infty$$

with I_1 and I_2 as in Assumption 9 and 10 respectively.

Remark 13. *It can be shown that Assumption 11 (ii) follows from Assumptions 1 and 8. Define*

$$\phi(\mathbf{z}) := \mathbb{E}[K_h(\mathbf{z} - \mathbf{Z})] = \int K_h(\mathbf{z} - \mathbf{Z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} = \int K_h(\mathbf{u}) f_{\mathbf{Z}}(\mathbf{z} + h\mathbf{u}) d\mathbf{u}.$$

To show that ϕ is continuous, let $(\mathbf{z}_n)_{n \geq 1}$ be a sequence such that $\mathbf{z}_n \rightarrow \mathbf{z}$. For every $\mathbf{u}$, define

$$g_n(\mathbf{u}) := K(\mathbf{u}) f_{\mathbf{Z}}(\mathbf{z}_n + h\mathbf{u}), \quad g(\mathbf{u}) := K(\mathbf{u}) f_{\mathbf{Z}}(\mathbf{z} + h\mathbf{u}).$$

Since $f_{\mathbf{Z}}$ is continuous, it follows that for every fixed $\mathbf{u}$,

$$g_n(\mathbf{u}) \rightarrow g(\mathbf{u}) \quad \text{as } n \rightarrow \infty.$$

Now by Assumptions 1(a) and 8 we can bound the sequence

$$|g_n(\mathbf{u})| = |K_h(\mathbf{u}) f_{\mathbf{Z}}(\mathbf{z}_n + h\mathbf{u})| \leq f_{\mathbf{Z}, \max} \|K\|_{\infty}.$$

Because K is integrable and the domain is compact, the function $f_{\mathbf{Z}, \max} \|K\|_{\infty}$ is integrable as well. Hence, the dominated convergence theorem yields

$$\phi(\mathbf{z}_n) \rightarrow \phi(\mathbf{z}),$$

proving that ϕ is continuous on $\mathcal{Z}$. Since $\mathcal{Z}$ is compact, the continuous function ϕ attains its minimum on $\mathcal{Z}$. Therefore, there exists some $\mathbf{z}' \in \mathcal{Z}$ such that $\inf_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}[K_h(\mathbf{z} - \mathbf{Z})] = \phi(\mathbf{z}')$. Assumption 1(e) then implies that $\inf_{\mathbf{z} \in \mathcal{Z}} \mathbb{E}[K_h(\mathbf{z} - \mathbf{Z})] > 0$.

We start by proving an extension of Corollary 6.1.5.

Lemma 6.5.1. *Define the set*

$$A_{n,h,k}^{\text{unif}} = \left\{ \sup_{\mathbf{z} \in \mathcal{Z}} \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right| \leq \frac{1}{2} \right\}.$$

Under Assumptions 1, 2, 8, 9, 10, 11 and 12 and in the bandwidth regime where $h_n \rightarrow 0$ and $nh^{kd}/|\log h| \rightarrow +\infty$, then $\mathbb{P}(A_{n,h,k}^{\text{unif}}) \xrightarrow{n \rightarrow \infty} 1$.

Proof. Start by writing

$$R_n(\mathbf{z}) = \left| \frac{\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) - \mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]}{\mathbb{E} \left[\sum_{\sigma \in \mathcal{J}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}) \right]} \right|.$$

By compactness of $\mathcal{Z}$, for every $\delta > 0$, there exists a finite set $\{\mathbf{z}_1, \dots, \mathbf{z}_{N_\delta}\} \subset \mathcal{Z}$ such that every $\mathbf{z} \in \mathcal{Z}$ is contained in a ball of radius δ around some grid point. Specifically, $\mathcal{Z} \subset \bigcup_{m=1}^{N_\delta} B(\mathbf{z}_m, \delta)$. Now for any $\mathbf{z}$ pick m such that $\|\mathbf{z} - \mathbf{z}_m\| \leq \delta$. Then $R_n(\mathbf{z}) \leq R_n(\mathbf{z}_m) + |R_n(\mathbf{z}) - R_n(\mathbf{z}_m)|$ and also

$$\sup_{\mathbf{z} \in \mathcal{Z}} R_n(\mathbf{z}) \leq \max_{1 \leq m \leq N_\delta} R_n(\mathbf{z}_m) + \sup_{\|\mathbf{z} - \mathbf{z}'\| \leq \delta} |R_n(\mathbf{z}) - R_n(\mathbf{z}')|.$$

Now since there are only finitely many grid points we use Corollary 6.1.5 to control the first term

$$\mathbb{P} \left(\max_{1 \leq m \leq N_\delta} R_n(\mathbf{z}_m) > \epsilon \right) \leq \sum_{m=1}^{N_\delta} \mathbb{P} \left(R_n(\mathbf{z}_m) > \epsilon \right) \leq CN_\delta e^{-cnh^{kd}} \xrightarrow{n \rightarrow \infty} 0,$$

for any fixed m . Hence $\max_{1 \leq m \leq N_\delta} R_n(\mathbf{z}_m) \xrightarrow[n \rightarrow \infty]{\text{prob}} 0$. Next we need to control the second term. By Assumption 11, K is piecewise Lipschitz on a compact domain and so, there exists a finite collection of regions on which K is Lipschitz continuous. Let D denote the union of the corresponding boundaries. Now suppose $\|\mathbf{z} - \mathbf{z}'\| \leq \delta$. If both $\frac{\mathbf{z} - \mathbf{Z}}{h}$ and $\frac{\mathbf{z}' - \mathbf{Z}}{h}$ belong to the same Lipschitz region of K , then

$$|K_h(\mathbf{z} - \mathbf{Z}) - K_h(\mathbf{z}' - \mathbf{Z})| \leq h^{-(d+1)} L_K \|\mathbf{z} - \mathbf{z}'\|.$$

Otherwise, the segment joining $\frac{\mathbf{z} - \mathbf{Z}}{h}$ and $\frac{\mathbf{z}' - \mathbf{Z}}{h}$ intersects the boundary set D . We then bound it by

$$|K_h(\mathbf{z} - \mathbf{Z}) - K_h(\mathbf{z}' - \mathbf{Z})| \leq 2h^{-d} \|K\|_\infty.$$

It then follows that for any $\mathbf{z}$

$$|K_h(\mathbf{z} - \mathbf{Z}) - K_h(\mathbf{z}' - \mathbf{Z})| \leq h^{-(d+1)} L_K \|\mathbf{z} - \mathbf{z}'\| \mathbb{1}_{E_\delta^c(\mathbf{z}, \mathbf{z}')} + 2h^{-d} \|K\|_\infty \mathbb{1}_{E_\delta(\mathbf{z}, \mathbf{z}')}. \quad (6.23)$$

where

$$E_\delta(\mathbf{z}, \mathbf{z}') = \left\{ \mathbf{z}, \mathbf{z}' \in \mathcal{Z} : \text{the segment joining } \frac{\mathbf{z} - \mathbf{Z}}{h} \text{ and } \frac{\mathbf{z}' - \mathbf{Z}}{h} \text{ intersects } D \right\}.$$

If $\mathbf{z}, \mathbf{z}' \in E_\delta$ and $\|\mathbf{z} - \mathbf{z}'\| \leq \delta$ then

$$E_\delta(\mathbf{z}, \mathbf{z}') \subset \left\{ \text{dist}\left(\frac{\mathbf{z} - \mathbf{Z}}{h}, D\right) \leq \frac{\delta}{h} \right\}.$$

Now introduce the ϵ -neighbourhood $D^\epsilon = \{x : \text{dist}(x, D) \leq \epsilon\}$. Then the previous inclusion becomes

$$E_\delta(\mathbf{z}, \mathbf{z}') \subset \left\{ \frac{\mathbf{z} - \mathbf{Z}}{h} \in D^{\delta/h} \right\} = \left\{ \mathbf{Z} \in \mathbf{z} - hD^{\delta/h} \right\}.$$

By Assumption 11, D has measure 0 so $|D^\epsilon| \rightarrow 0$ as $\epsilon \rightarrow 0$. We can use this to bound

$$\mathbb{P}(E_\delta(\mathbf{z}, \mathbf{z}')) \leq \mathbb{P}(\mathbf{Z} \in \mathbf{z} - hD^{\delta/h}) = \int_{\mathbf{z} - hD^{\delta/h}} f_{\mathbf{Z}}(\mathbf{u}) d\mathbf{u} \leq f_{\mathbf{Z}, \max} |\mathbf{z} - hD^{\delta/h}| = f_{\mathbf{Z}, \max} h^d |D^{\delta/h}|.$$

In particular this means that if we choose $\delta = \delta_n = o(h)$, we have $|D^{\delta/h}| \rightarrow 0$ and hence

$$\sup_{\|\mathbf{z} - \mathbf{z}'\| \leq \delta} \mathbb{P}(E_\delta(\mathbf{z}, \mathbf{z}')) \rightarrow 0.$$

So the boundary-crossing contribution in Equation (6.23) are asymptotically negligible in probability. We continue bounding $|R_n(\mathbf{z}) - R_n(\mathbf{z}')|$ by using the inequality

$$|a_1 \dots a_k - b_1 \dots b_k| \leq \sum_{i=1}^k |a_i - b_i| \left(\prod_{j=1}^{i-1} |b_j| \right) \left(\prod_{j=i+1}^k |a_j| \right).$$

If we combine this with Equation (6.23) we can bound $S_n(\mathbf{z}) = \sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})$. This yields

$$|S_n(\mathbf{z}) - S_n(\mathbf{z}')| \lesssim k n^k \left(\frac{\|K\|_\infty}{h^d} \right)^{k-1} \frac{\|\mathbf{z} - \mathbf{z}'\|}{h^{d+1}} \lesssim k n^k \frac{\|\mathbf{z} - \mathbf{z}'\|}{h^{dk+1}}$$

and similarly

$$|\mathbb{E}[S_n(\mathbf{z})] - \mathbb{E}[S_n(\mathbf{z}')]| \lesssim k n^k \left(\frac{\|K\|_\infty}{h^d} \right)^{k-1} \frac{\|\mathbf{z} - \mathbf{z}'\|}{h^{d+1}} \lesssim k n^k \frac{\|\mathbf{z} - \mathbf{z}'\|}{h^{dk+1}}.$$

Now using that Assumption 11 we find that $\mathbb{E}[\sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{j=1}^k K_h(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)})] \gtrsim n^k$. Define $A_n(\mathbf{z}) = S_n(\mathbf{z}) - \mathbb{E}[S_n(\mathbf{z})]$ and $B_n(\mathbf{z}) = \mathbb{E}[S_n(\mathbf{z})]$. Using

$$\left| \frac{A_n(\mathbf{z})}{B_n(\mathbf{z})} - \frac{A_n(\mathbf{z}')}{B_n(\mathbf{z}')} \right| \leq \frac{|A_n(\mathbf{z}) - A_n(\mathbf{z}')|}{|B_n(\mathbf{z})|} + \frac{|A_n(\mathbf{z}')| |B_n(\mathbf{z}) - B_n(\mathbf{z}')|}{|B_n(\mathbf{z}) B_n(\mathbf{z}')|}$$

yields

$$|R_n(\mathbf{z}) - R_n(\mathbf{z}')| \lesssim \frac{\|\mathbf{z} - \mathbf{z}'\|}{h^{dk+1}}.$$

Therefore

$$\sup_{\|\mathbf{z} - \mathbf{z}'\| \leq \delta} |R_n(\mathbf{z}) - R_n(\mathbf{z}')| \lesssim \frac{\delta}{h^{dk+1}}.$$

Choose $\delta = \delta_n = h^{d h + 2}$ and then $\frac{\delta_n}{h^{d k + 1}} = h_n \rightarrow 0$. To ensure that also $\max_{1 \leq m \leq N_\delta} R_n(\mathbf{z}_m)$ converges to 0, we want $N_\delta e^{-c n h^{k d}} \rightarrow 0$. Recall that

$$N_\delta = \left(\frac{\text{diam}(\mathcal{Z})}{\delta} \right)^d \sim h^{-(k + \frac{2}{d})}.$$

Thus

$$h^{-(k + \frac{2}{d})} e^{-n h^{k d}} = \exp \left(-n h^{k d} - \left(k + \frac{2}{d} \right) \log h \right).$$

For this expression to go to 0, we need that $n h^{k d} / |\log h| \rightarrow +\infty$. As a result,

$$\sup_{\mathbf{z} \in \mathcal{Z}} R_n(\mathbf{z}) \leq \max_{1 \leq m \leq N_\delta} R_n(\mathbf{z}_m) + \sup_{\|\mathbf{z} - \mathbf{z}'\| \leq \delta} |R_n(\mathbf{z}) - R_n(\mathbf{z}')| \xrightarrow[n \rightarrow \infty]{\text{prob}} 0.$$

This concludes that

$$\mathbb{P} \left(A_{n,h,k}^{\text{unif}} \right) = \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{Z}} R_n(\mathbf{z}) \leq \frac{1}{2} \right) \xrightarrow[n \rightarrow \infty]{} 1.$$

□

6

6.5.1. PROOF OF THEOREM 4.4.1

To prove the desired result we will use Lemma 2.5.2 and show

$$\mathbb{E} \left[\iint \left(\widehat{\Gamma}(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') \right)^2 d\mathbf{z} d\mathbf{z}' \mid A_{n,h}^{\text{unif}} \right] = \mathcal{O} \left(h^4 + \frac{1}{n h^d} \right) \quad \text{and} \quad \mathbb{P} \left(A_{n,h}^{\text{unif}} \right) \xrightarrow[n \rightarrow \infty]{} 1.$$

Start with writing $\Delta_{j_1, j_2}(\mathbf{z}) = \widehat{\tau}_{j_1, j_2}(\mathbf{z}) - \tau_{j_1, j_2}(\mathbf{z})$ for any pair (j_1, j_2) . Therefore, we can write

$$\widehat{\tau}_{j_1, j_2}(\mathbf{z}) \widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z}) \tau_{k_1, k_2}(\mathbf{z}') = \Delta_{j_1, j_2}(\mathbf{z}) \Delta_{k_1, k_2}(\mathbf{z}') + \Delta_{j_1, j_2}(\mathbf{z}) \tau_{k_1, k_2}(\mathbf{z}') + \Delta_{k_1, k_2}(\mathbf{z}') \tau_{j_1, j_2}(\mathbf{z}).$$

Since the τ are bounded we have

$$\left| \widehat{\tau}_{j_1, j_2}(\mathbf{z}) \widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z}) \tau_{k_1, k_2}(\mathbf{z}') \right| \leq \left| \Delta_{j_1, j_2}(\mathbf{z}) \right| + \left| \Delta_{k_1, k_2}(\mathbf{z}') \right| + \left| \Delta_{j_1, j_2}(\mathbf{z}) \Delta_{k_1, k_2}(\mathbf{z}') \right|$$

and

$$\left(\widehat{\tau}_{j_1, j_2}(\mathbf{z}) \widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z}) \tau_{k_1, k_2}(\mathbf{z}') \right)^2 \leq 3 \left(\Delta_{j_1, j_2}(\mathbf{z})^2 + \Delta_{k_1, k_2}(\mathbf{z}')^2 + \Delta_{j_1, j_2}(\mathbf{z})^2 \Delta_{k_1, k_2}(\mathbf{z}')^2 \right).$$

Now taking the expectation and applying Lemma 6.3.2 we find

$$\mathbb{E} \left[\Delta_{j_1, j_2}(\mathbf{z})^2 \mid A_{n,h,2}^{\text{unif}} \right] = \mathcal{O} \left(h^4 + \frac{1}{n h^d} \right).$$

Also if we use

$$\Delta_{j_1, j_2}(\mathbf{z})^2 \Delta_{k_1, k_2}(\mathbf{z}')^2 \leq \frac{1}{2} \left(\Delta_{j_1, j_2}(\mathbf{z})^4 + \Delta_{k_1, k_2}(\mathbf{z}')^4 \right) \leq C \left(\Delta_{j_1, j_2}(\mathbf{z})^2 + \Delta_{k_1, k_2}(\mathbf{z}')^2 \right),$$

we find $\mathbb{E}[\Delta_{j_1, j_2}(\mathbf{z})^2 \Delta_{k_1, k_2}(\mathbf{z}')^2 \mid A_{n, h, 2}^{\text{unif}}] = \mathcal{O}\left(h^4 + \frac{1}{nh^d}\right)$. Thus

$$\mathbb{E}\left[\left(\widehat{\tau}_{j_1, j_2}(\mathbf{z})\widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z})\tau_{k_1, k_2}(\mathbf{z}')\right)^2 \mid A_{n, h}^{\text{unif}}\right] = \mathcal{O}\left(h^4 + \frac{1}{nh^d}\right).$$

Next, recall

$$\begin{aligned} & \widehat{\Gamma}_p(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') \\ &= \frac{1}{\left(\binom{p}{2} - 1\right)^2} \sum_{j_1 < j_2}^p \left((\widehat{\tau}_{j_1, j_2}(\mathbf{z}) - \widehat{\mu}(\mathbf{z}))(\widehat{\tau}_{j_1, j_2}(\mathbf{z}') - \widehat{\mu}(\mathbf{z}')) - (\tau_{j_1, j_2}(\mathbf{z}) - \mu(\mathbf{z}))(\tau_{j_1, j_2}(\mathbf{z}') - \mu(\mathbf{z}')) \right) \\ &= \frac{1}{\left(\binom{p}{2} - 1\right)^2} \sum_{j_1 < j_2}^p \left((\widehat{\tau}_{j_1, j_2}(\mathbf{z})\widehat{\tau}_{j_1, j_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z})\tau_{j_1, j_2}(\mathbf{z}')) \right. \\ &\quad - \frac{1}{\binom{p}{2}} (\widehat{\tau}_{j_1, j_2}(\mathbf{z}) \sum_{k_1 < k_2}^p \widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z}) \sum_{k_1 < k_2}^p \tau_{k_1, k_2}(\mathbf{z}')) \\ &\quad - \frac{1}{\binom{p}{2}} (\widehat{\tau}_{j_1, j_2}(\mathbf{z}') \sum_{k_1 < k_2}^p \widehat{\tau}_{k_1, k_2}(\mathbf{z}) - \tau_{j_1, j_2}(\mathbf{z}') \sum_{k_1 < k_2}^p \tau_{k_1, k_2}(\mathbf{z})) \\ &\quad \left. + \frac{1}{\binom{p}{2}^2} \left(\sum_{k_1 < k_2}^p \sum_{l_1 < l_2}^p \widehat{\tau}_{k_1, k_2}(\mathbf{z})\widehat{\tau}_{l_1, l_2}(\mathbf{z}') - \sum_{k_1 < k_2}^p \sum_{l_1 < l_2}^p \tau_{k_1, k_2}(\mathbf{z})\tau_{l_1, l_2}(\mathbf{z}') \right) \right). \end{aligned}$$

This expression is simply a finite linear combination of $\widehat{\tau}_{j_1, j_2}(\mathbf{z})\widehat{\tau}_{k_1, k_2}(\mathbf{z}') - \tau_{j_1, j_2}(\mathbf{z})\tau_{k_1, k_2}(\mathbf{z}')$ for any combination of pairs (j_1, j_2) and (k_1, k_2) . So we can write this expression as

$$\widehat{\Gamma}_p(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') = \sum_{r=1}^N c_r \left(\widehat{\tau}_{a_r}(\mathbf{z})\widehat{\tau}_{b_r}(\mathbf{z}') - \tau_{a_r}(\mathbf{z})\tau_{b_r}(\mathbf{z}') \right),$$

where $N < \infty$ and only depends on p . Now by Cauchy-Schwartz, we get the existence of a constant $C > 0$ such that

$$\left(\widehat{\Gamma}_p(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') \right)^2 \leq C \sum_{r=1}^N \left(\widehat{\tau}_{a_r}(\mathbf{z})\widehat{\tau}_{b_r}(\mathbf{z}') - \tau_{a_r}(\mathbf{z})\tau_{b_r}(\mathbf{z}') \right)^2.$$

This means

$$\begin{aligned} \mathbb{E}\left[\iint \left(\widehat{\Gamma}_p(\mathbf{z}, \mathbf{z}') - \Gamma_p(\mathbf{z}, \mathbf{z}') \right)^2 d\mathbf{z}d\mathbf{z}' \mid A_{n, h, 2}^{\text{unif}}\right] &\leq C \sum_{r=1}^N \iint \mathbb{E}\left[\left(\widehat{\tau}_{a_r}(\mathbf{z})\widehat{\tau}_{b_r}(\mathbf{z}') - \tau_{a_r}(\mathbf{z})\tau_{b_r}(\mathbf{z}') \right)^2 \mid A_{n, h, 2}^{\text{unif}}\right] d\mathbf{z}d\mathbf{z}' \\ &= \mathcal{O}\left(h^4 + \frac{1}{nh^d}\right). \end{aligned}$$

We used Assumption 12 to show that the integrals are finite. Now by Lemma 6.5.1 we have that $\mathbb{P}\left(A_{n, h, 2}^{\text{unif}}\right) \xrightarrow{n \rightarrow \infty} 1$ and the result follows from Lemma 2.5.2.

7

APPLICATION TO FINANCIAL DATA

In this final chapter, we apply the theoretical framework developed in the preceding chapters to a selection of returns of 14 financial assets. Following the previous framework, we estimate the conditional Kendall's tau over all 1092 different asset triplets. Since these conditional dependencies may vary across the support of the conditioning variable, the CKT estimators are functional objects. We aim to extract the dominant modes of variation among this collection of curves, however it becomes highly impractical to analyse these curves individually. Hence, we will apply the theory developed in Chapter 4 and 5 to apply FPCA and find the relevant structural phenomena present among all the CKT curves.

As is typical with financial asset returns, the raw data exhibits strong dependencies over time, violating the standard independent and identically distributed (i.i.d.) assumptions required in our framework. Consequently, we employ a filtering procedure as a preprocessing step to isolate the underlying independent innovations before computing the principal components. But before doing this, we start by introducing the dataset.

7.1. DESCRIPTION AND VISUALIZATION OF THE DATASET

Our empirical analysis focuses on evaluating how various finance markets relate and depend on one other. The raw dataset comprises daily closing prices for 14 distinct financial assets spanning from January 2007 to September 2025, sourced via Yahoo Finance. This collection consists of different global equity indices, energy commodities, and currency markets, as detailed in Table 7.1. For notational convenience we collect their names in the vector $\text{NAMES} = (\text{"DAX"}, \text{"AEX"}, \dots)$. For every time $t \in \mathbb{T}$, where $\mathbb{T}$ denotes the set of all trading days between January 2007 and September 2025, and every asset $i \in \text{NAMES}$, we denote by $X_{t,i}$ the return on day t of asset i .

Table 7.1: Overview of the asset classes, market indices, abbreviation and geographical regions included in the dataset.

Type	Name	Abbreviation	Location
European Markets	DAX Index	DAX	Germany
	AEX Index	AEX	the Netherlands
	CAC 40 Index	FCHI	France
American Markets	Dow Jones Index	DJI	United States
	NASDAQ Composite	IXIC	United States
	S&P 500	SP500	United States
Asian Markets	Moscow Exchange Market	MOEX	Russia
	Nikkei 225	N225	Japan
	Shanghai Stock Exchange	SHC	China
Currency Markets	US Treasury Yield	FVX	United States
	Euro to US Dollar Exchange	EURUSD/X	-
	Bitcoin Market	BTC	-
Oil Markets	Brent Crude Oil	Brent	North Sea
	West Texas Intermediate	WTI	United States

Before going into the analysis of this dataset we start by visualizing some of the market behaviours. We will consider the AEX Index as a representative example. Figure 7.1 illustrates the historical evolution of the AEX Index over the time period T . The trajectory clearly captures major global shocks which have occurred. Notably, the severe downfall associated with the 2008 global financial crisis is prominently visible, followed by a prolonged recovery period. Similarly, the sharp drop in early 2020 reflects the onset of the COVID-19 pandemic, succeeded by an rapid recovery of the market and the stagnation of the market in the recent years.

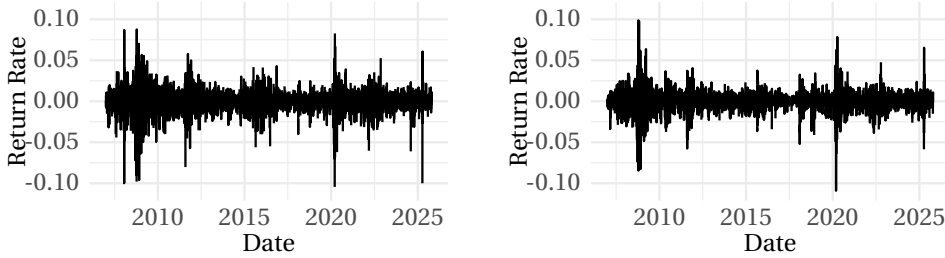


Figure 7.1: Daily closing prices of the AEX Index from January 2007 to September 2025.

The figure shows that the daily closing prices of a market is not an i.i.d. sample and it is in fact not even stationary. To make the sample more stationary, we instead consider the return rates on each day. A return rate measures the relative change in price between two consecutive trading periods. Specifically, if P_t denotes the asset price at time $t \in \mathbb{T}$, we compute the returns R_t as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

For example, if the price of an asset increases by 2%, the corresponding return rate equals 0.02. Figure 7.2 displays the return rates of the AEX and the Dow Jones Index.



(a) Return rate of the AEX over the period January 2007 till September 2025

(b) Return rate of the Dow Jones Index over the period January 2007 till September 2025

Figure 7.2: Return rate of the AEX and Dow Jones Index as a function of time.

Figure 7.2 shows that periods of high volatility tend to occur during times of financial stress, such as the 2008 financial crisis, while calmer market periods, for example around 2017, exhibit substantially lower volatility. We can see that this volatility clustering occurs for both markets, though it seems a bit more extreme in the AEX market.

7.2. ARMA-GARCH FILTERING OF RETURNS

As established in Chapter 3, the validity of our estimators hinges on the assumption that the observations are independent and identically distributed. Figure 7.2 show that the return rates are now mean-stationary but still have non-constant variance. They exhibit phenomena as volatility clustering and serial correlation. To resolve this, we model each asset price series using an univariate ARMA-GARCH model, as detailed in Section 2.10. Consequently, we extract the estimated standardized residuals, denoted by $\{\hat{\eta}_t^{\text{DAX}}, \hat{\eta}_t^{\text{AEX}}, \dots\}_{t \in \mathbb{T}}$. These quantities serve as feasible empirical counterparts to the true, unobservable standardized innovations $\{\eta_t^{\text{DAX}}, \eta_t^{\text{AEX}}, \dots\}_{t \in \mathbb{T}}$. The standardized residuals can be reliably modelled as an (approximately) i.i.d. process with a mean of zero and unit variance, making them suitable for non-parametric conditional Kendall's tau estimation.

Our primary objective is to estimate the conditional dependence between the standardized innovations of two assets, conditioned on a third. For any asset triplet $(j_1, j_2, k) \in$

NAMES, the true conditional Kendall's tau of the underlying standardized innovations is defined as

$$\tau_{j_1, j_2 | \eta^k = z}^\eta := \tau_{\eta_t^{j_1}, \eta_t^{j_2} | \eta^k = z},$$

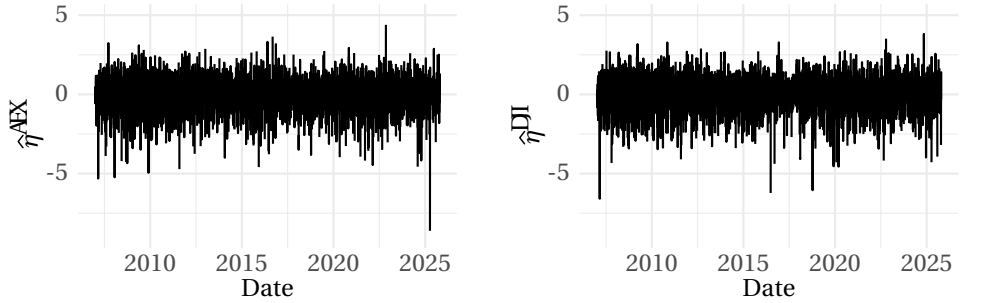
and assumed not to be depending on t . If the true standardized innovations were observable, we would estimate the CKT by $\hat{\tau}_{j_1, j_2 | \eta^k = z}^\eta$. However, we do not directly observe the η , instead we work with the standardized residuals $\hat{\eta}$. Therefore, we will work with the estimator $\hat{\tau}_{j_1, j_2 | \eta^k = z}^{\hat{\eta}}$.

Remark 14. Note that the ARMA-GARCH model adds an additional layer of randomness to the problem. Indeed, the total error in our functional estimates decomposes now into two distinct operational layers

$$\hat{\tau}_{j_1, j_2 | \eta^k = z}^{\hat{\eta}} - \tau_{j_1, j_2 | \eta^k = z}^\eta = \underbrace{\left(\hat{\tau}_{j_1, j_2 | \eta^k = z}^{\hat{\eta}} - \hat{\tau}_{j_1, j_2 | \eta^k = z}^\eta \right)}_{\text{Filtering Error}} + \underbrace{\left(\hat{\tau}_{j_1, j_2 | \eta^k = z}^\eta - \tau_{j_1, j_2 | \eta^k = z}^\eta \right)}_{\text{Sampling Error}}. \quad (7.1)$$

The first component represents the variation introduced by replacing η with $\hat{\eta}$, while the second component represents the sampling error in kernel estimation.

Figure 7.3 shows the standardized residuals $\hat{\eta}_t^{\text{AEX}}$ and $\hat{\eta}_t^{\text{DJI}}$. If we compare this figure with Figure 7.2, it shows that the filtering process has successfully normalized the variance profile over time. The structural clusters of high volatility have been removed, leaving behind a sequence that closely resembles i.i.d. data.



(a) Standardized residuals of the AEX Index over the period January 2007 till September 2025

(b) Standardized residuals of the Dow Jones Index over the period January 2007 till September 2025

Figure 7.3: Standardized residuals of the AEX and Dow Jones Index.

7.3. UNCONDITIONAL DEPENDENCE

It is also of interest to study the unconditional dependence structure between the different financial markets. We begin by considering the standardized residuals of the Amsterdam market (AEX) and the Dow Jones Index. Figure 7.4 displays the corresponding scatterplot.

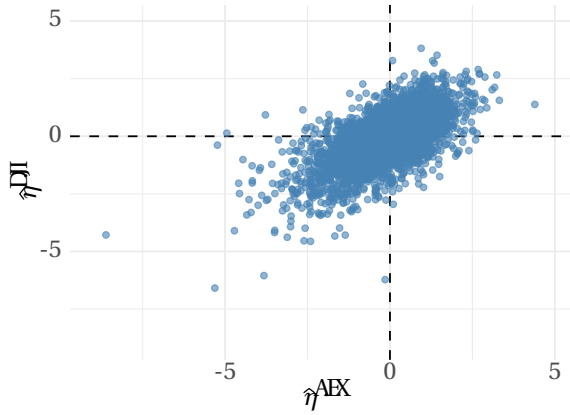


Figure 7.4: Scatterplot of the standardized residuals of the Dow Jones Index, $\hat{\eta}^{DJI}$, and the AEX Index, $\hat{\eta}^{AEX}$.

Figure 7.4 reveals a clear positive dependence between the two markets. The scatterplot also contains several notable outliers, illustrating the advantages of using Kendall's tau as a measure of dependence. Whereas a linear regression model can be heavily influenced by outliers, Kendall's tau remains relatively robust because it is based on the ranking of observations rather than their exact numerical values. Our objective is to understand the dependence structure across the entire set of markets. To this end, we next examine the unconditional pairwise dependencies among all considered assets. Figure 7.5 presents a heatmap of the pairwise unconditional Kendall's tau values for the 14 standardized residual series.

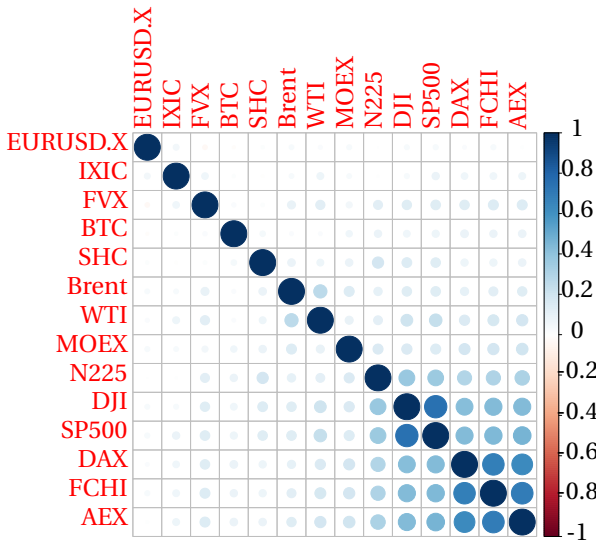


Figure 7.5: Heat map of the unconditional Kendall's tau given two assets.

The unconditional dependence structure reveals distinct clustering. As expected, the European equity markets, the AEX, DAX, and CAC 40 (FCHI), exhibit highly positive dependencies ($\tau > 0.6$) with each other, reflecting the regional integration. In contrast, markets such as the NASDAQ Composite (IXIC) and the Shanghai Stock Exchange (SHC) display weaker dependence with the European indices, reflecting more localized and independent market dynamics.

As shown in Figure 7.5, the strongest unconditional dependencies occur among the assets in the bottom-right quadrant of the heatmap (from the AEX to the Nikkei 225). These strong relationships suggest that the corresponding asset pairs may also exhibit pronounced variation in their conditional Kendall's tau curves. Although the main analysis considers all 14 assets, Appendix A presents a complementary analysis restricted to this highly dependent subset.

7.4. FPCA CONDITIONAL TO THE AMSTERDAM MARKET

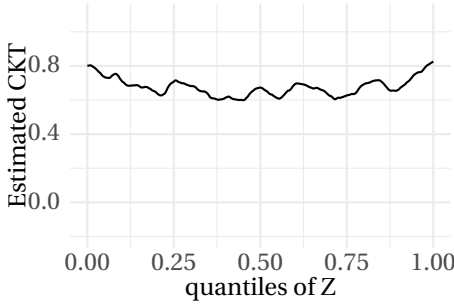
While Figure 7.5 provides insights into the dependence structure, it does not capture any conditional dependence dynamics. We therefore transition to the conditional Kendall's tau framework. The initial correlation plot shows that the AEX market has a high unconditional correlation with many different assets. Therefore, we expect the CKT curves conditioned on the AEX to exhibit interesting structural patterns. In this section, we analyse the set of CKT curves conditioned on the AEX and identify their dominant structural modes. In Chapter 5 we have developed a mathematical framework for such setting which we can hence apply.

7

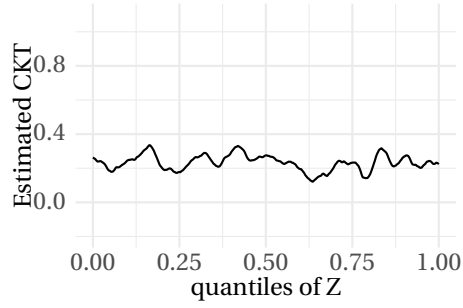
Following the methods of Chapter 5, we evaluate the conditional Kendall's tau on a finite grid $\{z_1, \dots, z_N\} \subset \mathcal{Z}_{\text{AEX}}$. Specifically, we employ a quantile grid to ensure an equitable discretization. Figure 7.6 illustrates four examples of these estimated conditional Kendall's tau curves, denoted as $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}} = z}^{\hat{\eta}}$.

Figures 7.6b and 7.6c show that their CKT curve is relative flat. The two oil markets (Brent and West Texas Intermediate) remain bound near $\tau \approx 0.3$, showing minimal sensitivity to the value of the AEX Index. The same is true for the Nikkei Index and the Shanghai Stock Exchange. This might suggest that the simplifying assumption (see Section 2.3 and 3.3) is valid for these markets.

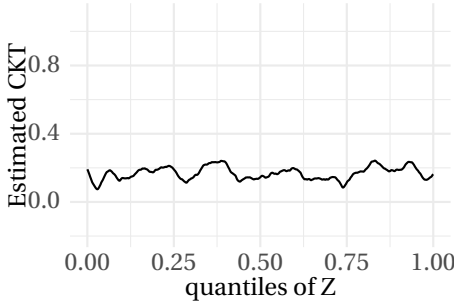
This cannot be said for the Dow Jones/S&P 500 pair and the CAC 40/Moscow pair, where the trajectories exhibit slightly parabolic shapes. Specifically during periods of extreme AEX values, both positive and negative, the conditional Kendall's tau values are higher than during periods of moderate market performance. This indicates that during times of extreme prosperity or systemic crisis, these markets depend more heavily on one another than during calmer, baseline periods. Interestingly, during normal periods, the conditional Kendall's tau between the French and Russian markets hovers around 0, indicating a lack of baseline correlation. However, during extreme market periods, their τ increases. This suggests that the two markets develop a form of tail dependence during times of distress. This phenomenon where financial crisis are rapidly spread between markets is known as the *contagion effect*.



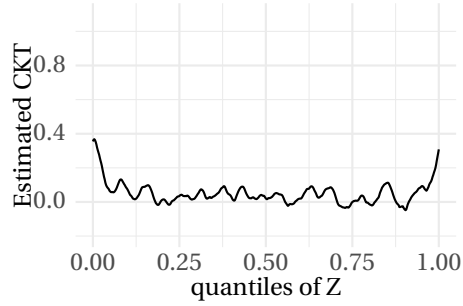
(a) The estimated conditional Kendall's tau curve for the Dow Jones Index and the S&P 500, conditioned on the AEX Index.



(b) The estimated conditional Kendall's tau curve for the Brent Crude Oil and the West Texas Intermediate, conditioned on the AEX Index.



(c) The estimated conditional Kendall's tau curve for the Nikkei Index and the Shanghai Stock Exchange, conditioned on the AEX Index.



(d) The estimated conditional Kendall's tau curve for the CAC 40 and the Moscow Exchange Market, conditioned on the AEX Index.

Figure 7.6: Estimated conditional Kendall's tau curves, $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}} = z}^{\eta}$.

Our goal is now to identify the dominant modes of variation among the different Kendall's tau curves. However, with $\binom{14}{2} = 91$ distinct asset pairs, analysing each curve individually becomes impractical. We therefore apply FPCA to these random curves in order to extract their main patterns of variation.

Although the theoretical framework of FPCA is formulated in an infinite-dimensional functional setting, practical implementation relies on discretised observations over a finite grid. Consequently, FPCA is carried out numerically through PCA applied to the discretised functions. The functional perspective is then recovered as the grid resolution increases, corresponding to the limiting regime where the number of discretisation points tends to infinity. The theoretical basis of applying PCA is laid out in Chapter 5.

Figure 7.7 shows the estimates of the first four principal components $\hat{\xi}_i$. Recall, that each estimated CKT curve, can be expressed as a linear combination of the components where each function captures a distinct mode of variation within the collection of curves. To understand the relative importance of these modes, we can look at their associated eigenvalues, $\hat{\lambda}_i$. Figure 7.8 outlines the proportion of total variance explained by the first six components.

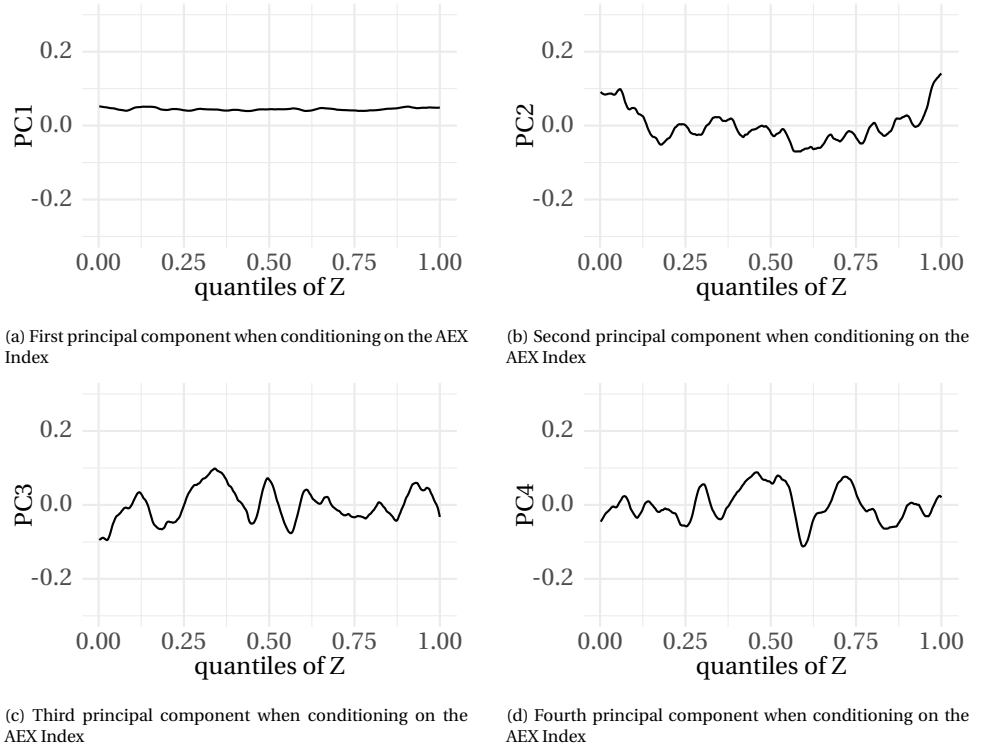


Figure 7.7: The first four principal component when conditioning on $\hat{\eta}^{\text{AEX}}$.

Noticeably, the first principal component is overwhelmingly dominant, accounting for nearly 80% of the total variation, whereas higher-order components sharply drop to less than 3% each. This stark contrast in variance explanation is visually reflected in the shapes of the estimated principal components displayed in Figure 7.7. The dominant first component ($\hat{\xi}_1$) in Figure 7.7a, shows a uniform vertical shift. This represents a baseline correlation, independent of the specific level of the AEX Index. In contrast, the second principal component ($\hat{\xi}_2$) in Figure 7.7b reveals a distinct parabolic trajectory. This confirms that a secondary source of variation captures stronger dependence during extreme market conditions. These patterns align precisely with the empirical behaviour observed in Figure 7.6. The fact that the parabolic shapes seen in Figures 7.6a and 7.6c emerge as a principal component confirms that these dynamics represent genuine underlying phenomena among the curves.

Beyond the second component, however, it is more difficult to interpret the components. As seen in Figure 7.7, the third and fourth components appear highly erratic and noisy. While it is likely that the components indeed represent noise, it is theoretically also possible that these higher-order components possess underlying structure but are simply obscured by estimation difficulties. Indeed, recall from Section 5.5 that the error of the eigenvectors is inversely proportional to its eigenvalue gap, $\min(\lambda_{i-1} - \lambda_i, \lambda_i - \lambda_{i+1})$.

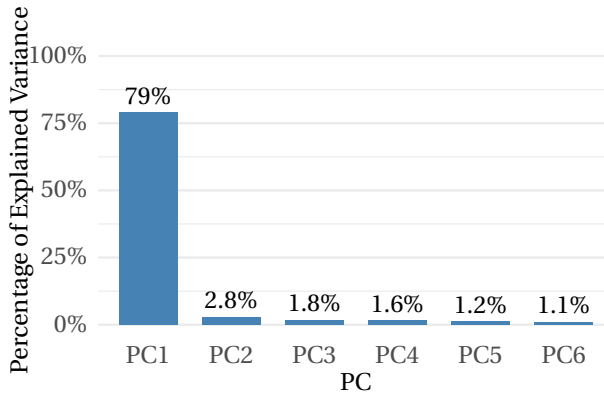


Figure 7.8: Histogram showing the percentage of the variance explain by the first six principal component when conditioning on the AEX Index.

Because the higher-order components each capture less than 2% of the variance, their eigenvalues are tightly clustered together near zero, resulting in a very small eigenvalue gap and thus a high estimation error. Nevertheless, it is more probable that these component simply capture noise.

As established in Section 3.3, a strictly valid simplifying assumption implies that all higher-order eigenvalues asymptotically vanish (i.e., $\lambda_i = 0$ for all $i \geq 2$). Under this assumption, the first principal component (PC1) would account for 100% of the total variation, leaving 0% for the remaining components. However, the empirical variance spectrum in Figure 7.8 paints a different picture: PC1 explains roughly 80% of the variation, while the subsequent components still capture minor shares (under 3% each).

The important question is whether these remaining eigenvalues represent true underlying structures or are simply non-zero because of estimation error. Since our current framework lacks a formal hypothesis test for $\lambda_i = 0$ ($i \geq 1$), we rely on rules-of-thumb like the *elbow rule* [10]. The method states to truncate the number of components at the point where the variance histogram flattens out. As shown in Figure 7.8, the variance drops by 1% from PC2 to PC3, but shifts by only 0.5% for the subsequent steps. This distinct elbow suggests that PC2 retains meaningful information and should be included. Furthermore, our analysis of Figure 7.7b demonstrates that this second component is also interpretable and aligns with known economic phenomena, specifically the contagion effect. Taken together, this strongly suggest that the simplifying assumption does not hold across the entire collection of curves. This conclusion is entirely coherent with the economic intuition that the conditional dependence between financial markets are dynamic rather than static.

7.5. FPCA ON ALL ASSET TRIPLETS

In the previous section we have considered the case where we only conditioned on the AEX Index. However, we can of course also condition on other assets and thereby extending the analysis to more curves. Even though the theory established in Chapters 3 till Chapter 5 does not directly apply when conditioning on multiple covariates, it is certainly possible to extend it accordingly. Hence, we still consider that analysis here.

In our system of 14 assets $\{DAX, AEX, \dots\}$, we consider all combinations, choosing 2 conditioned assets and 1 conditioning asset among NAMES. This combinatorial arrangement yields $\binom{14}{2} \times (14 - 2) = 1092$ unique triplets. Following Chapter 5, we evaluate each conditional Kendall's tau curve over a localized grid of N points, $\{z_{k,1}, \dots, z_{k,N}\} \subset \mathcal{Z}_k$ for each $k \in \text{NAMES}$. A quantile grid was selected to capture higher resolution in high-probability density zones and to ensure more robustness to extreme outliers. The complete set of estimated conditional Kendall's tau curves for all triplets can therefore be organized into a large-scale matrix $\text{AllCKT} \in \mathbb{R}^{N \times 1092}$, where each curve correspond to one column, giving the following structure:

$$\text{AllCKT} = \begin{bmatrix} \hat{\tau}_{AEX,DAX|\eta^{FCHI}=z_{FCHI,1}}^{\eta} & \hat{\tau}_{AEX,DAX|\eta^{DJI}=z_{DJI,1}}^{\eta} & \cdots & \hat{\tau}_{BTC,Brent|\eta^{WTI}=z_{WTI,1}}^{\eta} \\ \hat{\tau}_{AEX,DAX|\eta^{FCHI}=z_{FCHI,2}}^{\eta} & \hat{\tau}_{AEX,DAX|\eta^{DJI}=z_{DJI,2}}^{\eta} & \cdots & \hat{\tau}_{BTC,Brent|\eta^{WTI}=z_{WTI,2}}^{\eta} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\tau}_{AEX,DAX|\eta^{FCHI}=z_{FCHI,N}}^{\eta} & \hat{\tau}_{AEX,DAX|\eta^{DJI}=z_{DJI,N}}^{\eta} & \cdots & \hat{\tau}_{BTC,Brent|\eta^{WTI}=z_{WTI,N}}^{\eta} \end{bmatrix}. \quad (7.2)$$

So each column represents on CKT curve discretized over the grid. Figure 7.9 shows all conditional Kendall's tau curves from AllCKT.

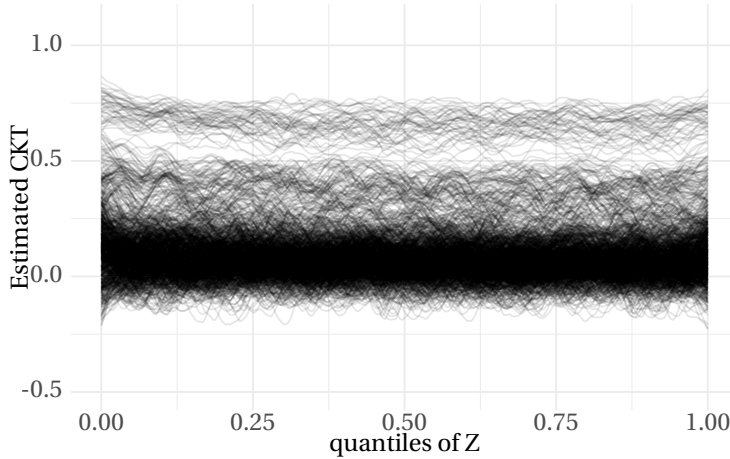


Figure 7.9: All estimates of the conditional Kendall's tau for all combinations of assets. The x -axis is the quantiles of the associated covariate variable.

When looking at Figure 7.9 we can distinguish patterns across the set of curves. We can note that some curve experience more fluctuation near the boundaries of the interval, suggesting some tail dependency for these markets. We can also clearly note that most of the markets experience relative weak correlation ($\tau \approx 0.1$) but there is also a group of assets having a relative big correlation with curves fluctuating between values 0.6 and 0.75.

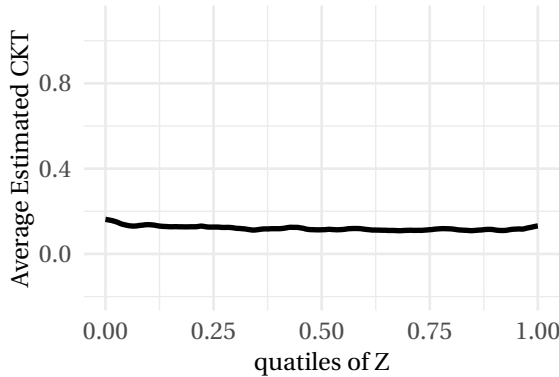


Figure 7.10: Average curve of all the estimated conditional Kendall's tau trajectories.

Figure 7.10 shows the average curve of the conditional Kendall's tau. As expected, the curve is a constant over the entire quantile domain but we can also notice some small parabolic shapes near the boundaries. To better understand the modes describing all the curves, we can compute the estimates for the principal components and look at their respective importance.

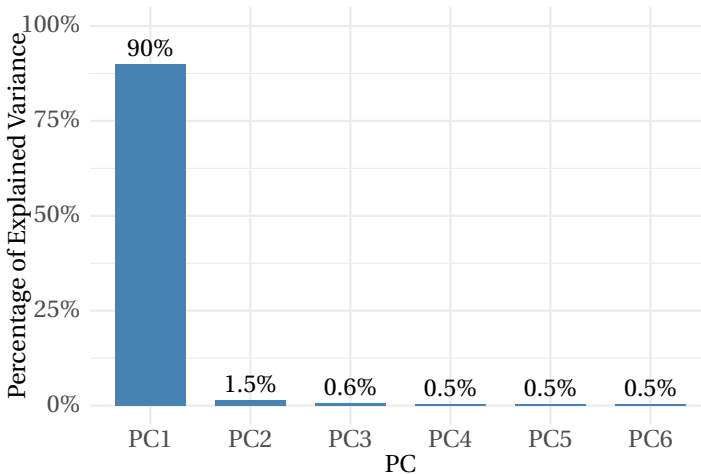
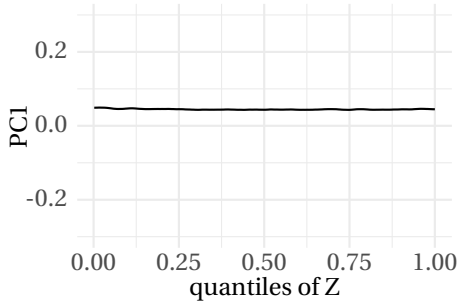
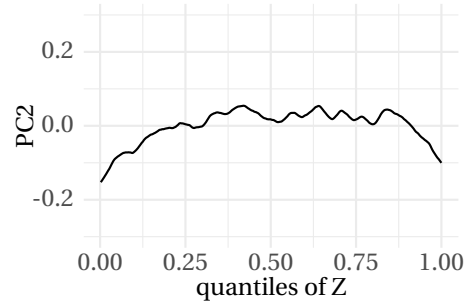


Figure 7.11: Histogram showing the percentage of the variance explain by the first six principal component.

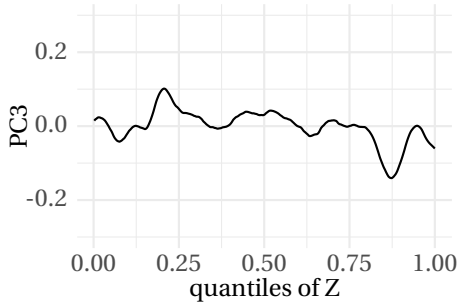
Figure 7.11 shows the percentage of variance explained by each of the estimated principal components from Figure 7.12. As already observed in the single-market analysis, the first principal component ($\hat{\xi}_1$) represents a uniform vertical shift, capturing a baseline correlation between markets. Again, the component is responsible for the vast majority of the variance of the data. The second principal component ($\hat{\xi}_2$) maintains its characteristic parabolic shape, modelling the contagion effect during periods of extreme market prices. This curve quantifies how asset correlations deviate sharply during severe market periods relative to periods of intermediate calmness.



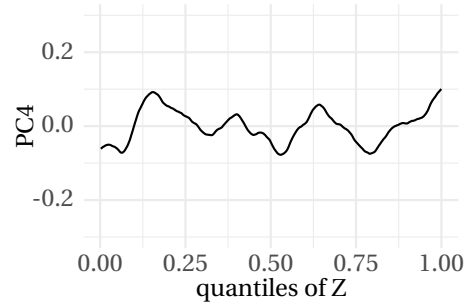
(a) First principal component when considering the entire dataset.



(b) Second principal component when considering the entire dataset.



(c) Third principal component when considering the entire dataset.



(d) Fourth principal component when considering the entire dataset.

Figure 7.12: The first four principal component of the entire dataset.

From the third component onward, structural interpretation becomes more difficult. One could argue that the third eigenfunction ($\hat{\xi}_3$) captures variation during moderately positive or negative market states. In that way, $\hat{\xi}_3$ might be seen as a "near-crisis" component, accounting for systemic variations that occur immediately preceding extreme tail events. In contrast, the fourth component ($\hat{\xi}_4$) is entirely structureless and probably captures only residual noise.

7.5.1. DEFLATION OF THE FIRST PRINCIPAL COMPONENT

Figure 7.12 strongly indicates that the first principal component, which accounts for the vast majority of the variance, is a constant curve representing the average conditional Kendall's tau. We can exploit this structural knowledge to improve the estimation of the remaining principal components. Since the true version of PC1 is known to be flat, we can normalize each CKT curve by subtracting its own average value, rather than deflating the system using an empirically estimated PC1. Differently, this corresponds to projecting the functional data directly onto the orthogonal complement of the space spanned by the true PC1. As a result, the curves no longer possess a vertical shift, and the dominant constant curve, is no longer one of the principal components. This way allows us to better examine the remaining modes of variation since some noise has been removed.

Figure 7.13 shows the variance decomposition after normalizing each CKT curve¹. It illustrates that the proportion of total variance explained by PC3 is nearly identical to that of the subsequent higher-order components, suggesting that also PC3 lacks genuine structural meaning and also behaves as noise.

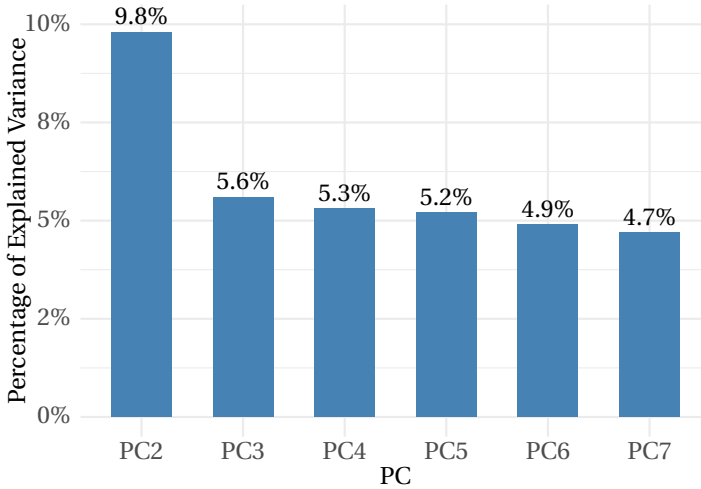


Figure 7.13: Histogram showing the relative percentage of variance explained by the principal components after removing the dominant first mode.

Additionally, Figure 7.13 shows a sharp drop occurring between PC2 and PC3, after which the variance explained stabilizes into a flat plateau across all subsequent modes. By the elbow method, this transition strongly indicates that PC3 marks the threshold where signal gives way to noise. This justifies the idea that only PC1 and PC2 describe genuine patterns in the collection of CKT curves.

To gain a better understanding of PC2's importance in characterizing the curves, we evaluate the functional PCA scores associated with each individual CKT curve. Since each

¹Strictly speaking, PC2 in Figure 7.13 is the first principal component of the transformed dataset, not the second. Still we retain the label PC2 to facilitate direct comparison with Figures 7.11 and 7.12.

CKT curve can be expressed as a linear combination of the principal components, the PCA scores define the respective coefficients.

Figure 7.14 shows the histogram of the scores from PC2 in Figure 7.12b. Note that a positive score represents a more concave CKT curve while a negative value means the curve is more convex shaped. As illustrated, the scores exhibit an approximately symmetric distribution centered close to zero. The number of markets that exhibit a drop in dependence during extreme states, is approximately equal to the number of markets that experience higher correlation during extreme periods.

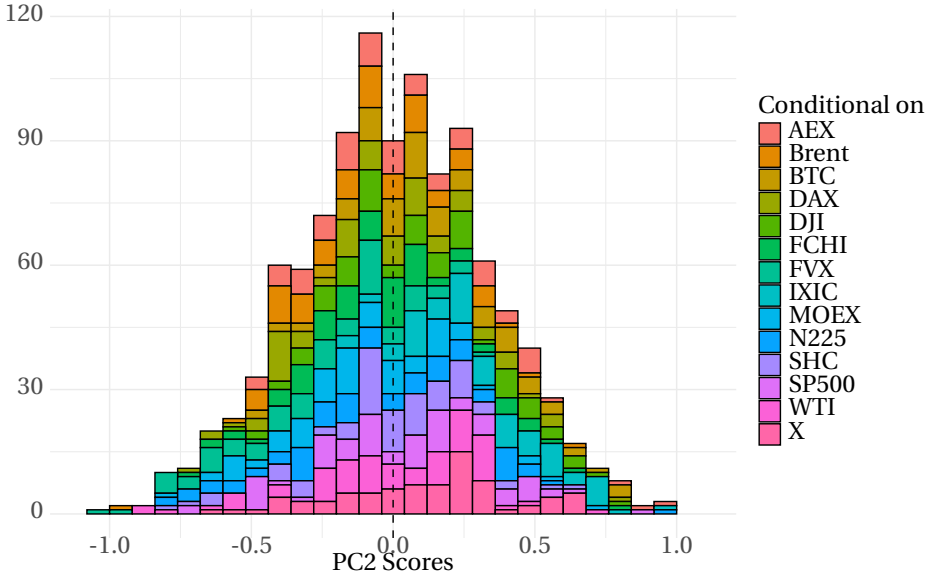


Figure 7.14: Histogram of the scores of the second principal component.

Concluding, these empirical observations provide compelling evidence against the simplifying assumption being valid for A11CKT. As demonstrated by both Figures 7.12 and 7.13, the second principal component captures a distinct, non-trivial structural pattern within the CKT curves. In fact, the variation across all CKT curves can be successfully summarized by just two functional degrees of freedom: a constant baseline level representing the average conditional dependence, and a parabolic component that contrasts extreme market states with more typical periods, capturing the contagion effects. More specifically, assuming the conditioning variable is mapped to the unit interval $z \in [0, 1]$, each individual CKT curve can be approximately characterized by

$$\tau_{1,2|Z=z} = a + b \left(z - \frac{1}{2} \right)^2 \quad (7.3)$$

where a and b are some coefficients representing the baseline dependence level and the sensitivity to market extremes, respectively.

7.6. SMOOTHED FPCA

We will end this chapter by noting that the principal components in Figure 7.12 exhibit some noisy which potentially obscures their underlying structural interpretation. This suggests that some form of smoothing may be appropriate when estimating the principal components. During the process of PCA we can smooth at three different instances: before applying PCA, during the application of PCA or after applying PCA. This section analyses the three moments of smoothing and compares their differences. Specifically, we consider, pre-smoothing the estimated curves, regularized FPCA (detailed in Section 2.9.2) and post-smoothing the estimated functional PCs.

7.6.1. PRE-SMOOTHING THE ESTIMATED CURVES

Our first regularizing strategy operates by smoothing the conditional dependency curves prior to executing the spectral decomposition. This approach fundamentally relies on substituting the standard pointwise estimator with a structurally smooth alternative. Specifically, Derumigny and Fermanian [42] proposed a alternative estimator, $\hat{\tau}^{2\text{-step}}$, rooted in the framework of *Kendall's Regression*. For the sake of generality, we do the derivation using an i.i.d. sample $\{(X_1^{(i)}, X_2^{(i)}, \mathbf{Z}_i)\}_{i=1}^n$.

We begin by introducing an increasing, bijective, and continuously differentiable function $\Lambda : [-1, 1] \rightarrow \mathbb{R}$ that maps the conditional Kendall's tau to the real line. Following a well-established functional approximation strategy, we assume that the transformed curve $\Lambda(\tau_{X_1, X_2 | \mathbf{Z}=\mathbf{z}})$ can be expressed as a linear combination of m linearly independent basis functions $\{\phi_1, \phi_2, \dots, \phi_m\}$

$$\Lambda(\tau_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}) = \sum_{i=1}^m \phi_i(\mathbf{z}) \beta_i = \Phi(\mathbf{z})^\top \boldsymbol{\beta} \quad (7.4)$$

for any $\mathbf{z} \in \mathcal{Z}$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^\top \in \mathbb{R}^m$ represents the vector of expansion coefficients and $\Phi(\mathbf{z}) = (\phi_1(\mathbf{z}), \dots, \phi_m(\mathbf{z}))^\top$ defines the vector of basis functions.

As always, we start by estimating $\tau_{X_1, X_2 | \mathbf{Z}=\mathbf{z}}$ at the grid points $\mathbf{z}_1, \dots, \mathbf{z}_N \in \mathcal{Z}$ using the estimator from Equation (2.16). Given these point-wise estimates, $\hat{\boldsymbol{\beta}}$ is obtained as the minimizer of the penalized least-squares criterion. To enforce smoothness, we introduce a specific roughness penalty²,

$$\text{RP}(f) := \int \left(\frac{d^2 f}{dx^2} \right)^2 dx. \quad (7.5)$$

Applying this penalty yields the following minimization problem

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^m}{\text{argmin}} \left[\frac{1}{N} \sum_{j=1}^N \left(\Lambda(\hat{\tau}_{X_1, X_2 | \mathbf{Z}=\mathbf{z}_j}) - \Phi(\mathbf{z}_j)^\top \boldsymbol{\beta} \right)^2 + \eta \cdot \text{RP}(\Lambda(\tau)) \right] \quad (7.6)$$

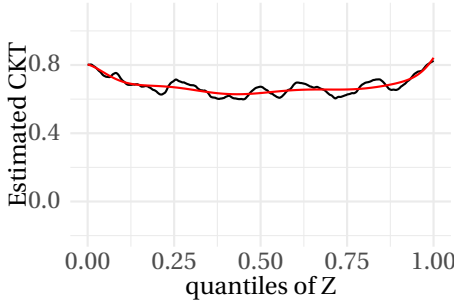
²The second derivative of a function is called its curvature. Calculating the total curvature over an interval is a way to quantify 'roughness'.

where $\eta \geq 0$ is a tuning parameter typically found using cross-validation. Ultimately, the regularized *two-step estimator* of the conditional Kendall's tau curve is defined as

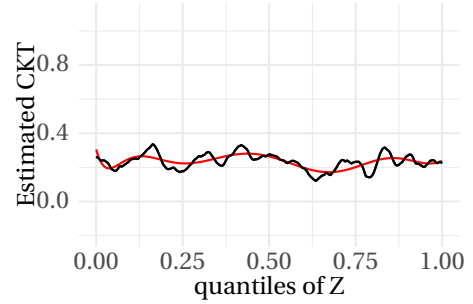
$$\hat{\tau}_{X_1, X_2 | Z=z}^{2\text{-step}} = \Lambda^{-1}(\Phi(\mathbf{z})\hat{\boldsymbol{\beta}}). \quad (7.7)$$

Properties of this estimator can be found in [42]. Because the approach relies on a penalized least-squares framework, it effectively smooths the curves. As a result, the estimated conditional Kendall's tau curves exhibit substantially less noise. When FPCA is subsequently performed on this pre-smoothed data, the resulting principal components display reduced roughness and improved interpretability compared to those obtained from the raw data.

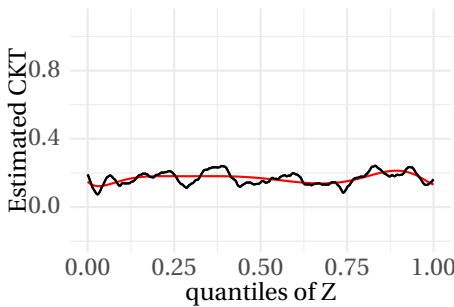
To demonstrate this practical benefit, we apply the two-step estimation framework to our empirical financial dataset. Figure 7.15 shows four examples of the estimated conditional Kendall's tau curves, $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}}=z}$ alongside the regressed conditional Kendall's tau estimator $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}}=z}^{\hat{\eta}, 2\text{-step}}$.



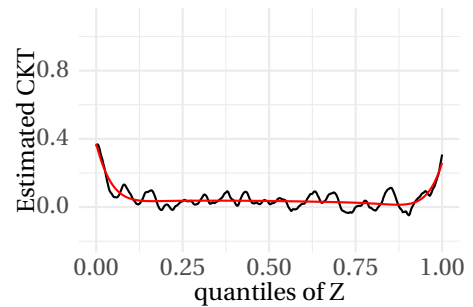
(a) The regressed (red) and non-regressed (black) estimated conditional Kendall's tau curve for the Dow Jones Index and the S&P 500, conditioned on the AEX Index.



(b) The regressed (red) and non-regressed (black) estimated conditional Kendall's tau curve for the Brent Crude Oil and the West Texas Intermediate, conditioned on the AEX Index.



(c) The regressed (red) and non-regressed (black) estimated conditional Kendall's tau curve for the Nikkei Index and the Shanghai Stock Exchange, conditioned on the AEX Index.



(d) The regressed (red) and non-regressed (black) estimated conditional Kendall's tau curve for the CAC 40 and the Moscow Exchange Market, conditioned on the AEX Index.

Figure 7.15: The regressed (red) and non-regressed (black) estimated conditional Kendall's tau curves, $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}}=z}^{\hat{\eta}}$ and $\hat{\tau}_{j_1, j_2 | \eta^{\text{AEX}}=z}^{\hat{\eta}, 2\text{-step}}$.

As expected, the red regressed conditional Kendall's tau curves in Figure 7.15 exhibit significantly less noise and are substantially more regular than before. As a result, the resulting principal components will also be more smoothing which makes interpreting them easier.

7.6.2. SMOOTHING EMBEDDED IN FPCA

Regularized FPCA is a smoothing approach in which the smoothing step is integrated in the FPCA procedure itself. Rather than estimating the principal component functions ξ solely to maximize the empirical variance, we also include a roughness penalty to ensure regularity. The goal is to find orthogonal components that simultaneously maximize variance explained and maintain geometric smoothness, formulated as:

$$\max_{\|\xi\|^2=1} \frac{\text{Var}\left(\int \xi(t) X_i(t) dt\right)}{\|\xi\|^2 + \eta \cdot \text{RP}(\xi)}.$$

Here, $\eta \geq 0$ serves as the smoothing tuning parameter. The roughness penalty is defined as $\text{RP}(\xi) = \int (\xi''(t))^2 dt$, which penalizes the total curvature of the eigenfunction.

A benefit of this embedded approach is that it looks at all the data simultaneously. Instead of smoothing out each curve one by one, it uses the shared patterns of the entire dataset to guide the smoothing. This ensures that roughness is only penalized in directions that do not contribute to the major structural modes.

The derivation is detailed in Section 2.9.2 is based on the assumption that the functional data curves $X_i(t)$ are directly observed. However, in our setting the target functions are estimates. We do not yet have a rigorous theoretical framework for applying regularized FPCA on estimated functional data but we can assume it to work accordingly.

7.6.3. POST-SMOOTHING OF THE PRINCIPAL COMPONENTS

The third regularization approach moves the smoothing towards the end by smoothing the raw functional principal components (PCs) after they have been extracted from unregularized PCA.

Let $(\tilde{\xi}_1(t), \dots, \tilde{\xi}_M(t))$ denote the first M principal components obtained by applying standard, unregularized FPCA to the raw estimated conditional Kendall's tau curves. Our objective is to filter out high-frequency sampling noise from each raw trajectory $\tilde{\xi}_j(t)$ to obtain an optimized, smooth functional counterpart, denoted as $\xi_j(t)$. To achieve this, we assume that each smooth eigenfunction $\xi_j(t)$ for $j = 1, \dots, M$ can be represented as linear combination of K linearly independent basis functions $\{\psi_1(t), \dots, \psi_K(t)\}$:

$$\xi_j(t) = \sum_{i=1}^m \beta_{i,j} \psi_i(t) = \Psi(t)^\top \boldsymbol{\beta}_j$$

where $\Psi(t) = (\psi_1(t), \dots, \psi_K(t))^\top$ represents the vector of basis functions, and $\boldsymbol{\beta}_j = (\beta_{1,j}, \dots, \beta_{K,j})^\top \in \mathbb{R}^K$ is the corresponding vector of coefficients. Just as in pre-smoothing we estimate $\boldsymbol{\beta}$ using least-squares in combination with a roughness penalty.

Let the raw FPCs be evaluated across a finite grid of N discretization points, $\{t_1, \dots, t_N\} \subset \mathcal{Z}$. We then formulate the optimization problem for each component j as

$$\hat{\boldsymbol{\beta}}_j = \operatorname{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^m} \left[\frac{1}{N} \sum_{i=1}^N (\tilde{\xi}_j(t_i) - \Psi(t_i)^\top \boldsymbol{\beta}_j)^2 + \eta_j \cdot \operatorname{RP}(\xi_j) \right], \quad (7.8)$$

where $\operatorname{RP}(\cdot)$ is as in Equation (7.5). Importantly, this formulation introduces a distinct smoothing parameter $\eta_j \geq 0$ for each individual principal component. This allows for flexible regularization, as lower-order principal components (which capture dominant signals) may require less smoothing than higher-order components (which often capture more volatile sampling noise). Each parameter η_j can be isolated and optimized independently via cross-validation techniques. Once the optimal coefficients $\hat{\boldsymbol{\beta}}_j$ are determined, the post-smoothed FPC is constructed as

$$\xi_j^{\text{smooth}} = \Psi(t)^\top \hat{\boldsymbol{\beta}}_j.$$

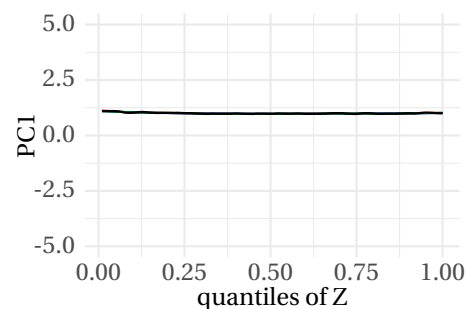
7.6.4. EMPIRICAL COMPARISON OF THE THREE SMOOTHING APPROACHES

In this section, we examine the empirical performance of the three distinct smoothing approaches, pre-smoothing, post-smoothing, and regularized FPCA, applied the ALLCKT. Figure 7.16 illustrates the unregularized empirical principal components alongside their three smoothed counterparts.

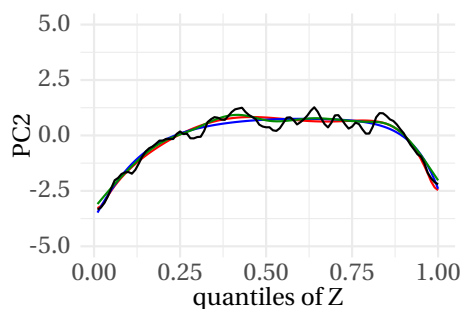
Clearly, all three regularizing approaches succeed in substantially reducing the noise, generating smooth trajectories. For the first two principal components, the curves produced by the different smoothing methods are nearly indistinguishable. This close alignment indicates that the dominant modes of variation capture highly robust structural signals that are invariant to the specific choice of regularization.

Differences between the smoothing methods begin to manifest in the third component and become pronounced in the fourth component. As previously noted, the first two components likely represent genuine economic variations in the conditional dependence structure, whereas the third and fourth component displays stochastic noise rather than a stable structural mode. This interpretation is supported by the divergence of the regularized curves in Figure 7.16d where the different smoothing techniques smooth the noise in different ways.

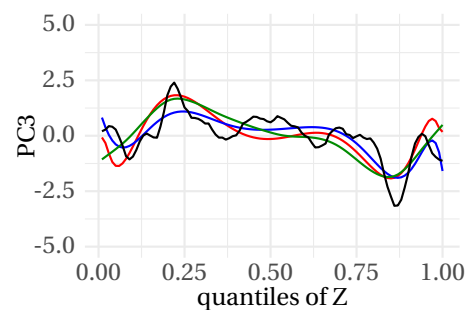
The framework established in Chapter 4 and 5 are not directly compatible with these regularizations steps. Moreover, the existing framework for these smoothing techniques is often based on the assumption of direct observable functional data. Therefore, our ability to mathematically evaluate the trade-offs between the three different smoothing approaches is limited. Extending the current framework to formally integrate these smoothing approaches could be an important step for future work. This could provide useful insights into optimizing the performance of FPCA.



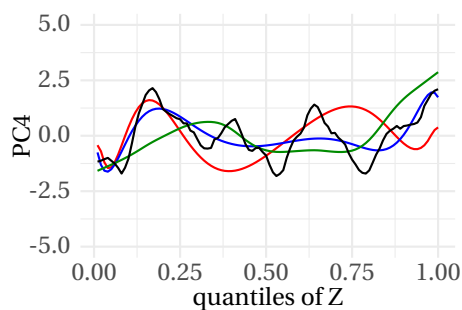
(a) First principal component alongside the three smoothed versions.



(b) Second principal component alongside the three smoothed versions.



(c) Third principal component alongside the three smoothed versions.



(d) Fourth principal component alongside the three smoothed versions.

Figure 7.16: The first four functional principal components extracted from the complete dataset. The black trajectories represent the unregularized baseline, while red denotes the pre-smoothing, blue the post-smoothing, and green the regularized FPCA frameworks.

8

CONCLUSION

In this thesis, we have established a new mathematical framework for recovering low-dimensional modes of variation within a collection of estimated curves. While the existing Functional Principal Component Analysis (FPCA) approach assumes that the functional objects are directly observable, this work extends the methodology to accommodate estimated functional objects, specifically, curves of estimated conditional Kendall's tau (CKT).

By shifting the focus to estimated curves, this framework introduces an additional layer of statistical variation that complicates standard asymptotic theory. The observed functional data points are no longer exact, rather, they are non-parametric estimators evaluated on a sample. Hence, the resulting sample randomness reflects both the underlying differences across the sample and the pointwise estimation error of the CKT estimators.

To resolve this, we proposed an estimator for the covariance operator upon which spectral decomposition yields the principal components. The resulting eigenfunctions capture the dominant modes of variation across the CKT trajectories, while the corresponding eigenvalues quantify the proportion of variation explained by each component. A core contribution of this research is the derivation of explicit convergence rates for the spectral properties of the estimated covariance operator. By extending the work on conditional U -statistics, we demonstrated that in the idealized setting where curves are observed continuously over the entire domain, the spectral estimators achieve a convergence rate of order $\mathcal{O}(h^2 + (nh^d)^{-1/2})$ where n is the sample size, h is the smoothing bandwidth, and d is the dimension of the covariate space. We successfully extended this result to the more realistic scenario where curves are observed on a finite number of grid points N , yielding an asymptotic error bound of $\mathcal{O}(N^{-1/d} + h^2 + (nh^d)^{-1/2})$. This theoretical result uncovers a critical structural insight regarding the estimation mechanics. The grid discretization level N acts as an isolated, standalone bottleneck, increasing the number of grid points always improves the error bound. Conversely, selecting the smoothing bandwidth h governs a classic bias-variance trade-off, where minimizing the bias term ($\mathcal{O}(h^2)$) inevitably inflates the variance component ($\mathcal{O}((nh^d)^{-1/2})$).

Besides the theoretical derivations, we demonstrated the practical use through an empirical study on 14 global financial markets. After applying ARMA-GARCH filtering to create approximately i.i.d. data, we estimated the CKT curves of the 1092 unique asset triplets. Applying FPCA to these estimated CKT curves successfully isolated the different dependence structures and their relative importance in the dataset. It was found that all CKT curves could be approximately summarized by only two modes of variation: one constant baseline curve, representing the average conditional Kendall's tau, and a parabolic component that contrasts extreme market periods with more typical periods capturing the contagion effect. In other words, all these CKT curves approximately live in a two-dimensional space, having only two degrees of freedom.

Lastly, this framework offers immediate value for evaluating the validity of the widely used *simplifying assumption* in copula modelling, which assumes that conditional dependence structures remain invariant across covariate values. Under this assumption, the CKT curves are perfectly flat, implying that the first eigenvalue accounts for the entirety of the system's variance while all subsequent eigenvalues are zero. Our framework serves as an empirical test to check whether the simplifying assumptions holds true. Seeing dominant non-zero eigenvalues for any of the higher components, indicate that the assumption is violated.

We have proposed a first mathematical framework for FPCA on estimated curves. To continue studying these topics, several extensions of the current theory can be done as future work:

- Our theoretical derivations rely partly on the assumption that the eigenvalues of the underlying covariance operator are simple. While in practice, overlapping eigenvalues are often near zero, generalizing the consistency proofs to accommodate eigenvalues with multiplicity greater than one would broaden the framework's generality.
- Currently, the framework provides a rule-of-thumb baseline for examining the simplifying assumption. Future work should focus on establishing the asymptotic distributions of the estimated eigenvalues and eigenfunctions. Deriving these limiting laws will enable the construction of formal statistical hypothesis tests to rigorously accept or reject the simplifying assumption.
- The current asymptotic theory assumes i.i.d. data. Given that financial applications naturally exhibit serial dependence, extending the framework to also include time series generalizes the use of the model. This would also formally validate the ARMA-GARCH filtering.
- Finally, recent literature highlights multiple ways for optimizing the smoothness of CKT curves. One can pre-smooth the initial curves via Kendall's regression [42], post-smooth the estimated principal components using a roughness penalty approach, or integrate a roughness penalty directly into the optimization step via regularized FPCA [10]. A systematic mathematical comparison of these three approaches would provide valuable insight in how to most efficiently apply FPCA.

BIBLIOGRAPHY

- [1] Eike C Brechmann, Claudia Czado, and Kjersti Aas. “Truncated regular vines in high dimensions with application to financial data”. In: *Canadian journal of statistics* 40.1 (2012), pp. 68–85.
- [2] Abe Sklar. “Fonctions de répartition à n dimensions et leurs marges”. In: *Publications de l’Institut de Statistique de l’Université de Paris* 8 (1959), pp. 229–231.
- [3] Maurice G Kendall. “A new measure of rank correlation”. In: *Biometrika* 30.1-2 (1938), pp. 81–93.
- [4] Claudia Czado. “Analyzing dependent data with vine copulas”. In: *Lecture Notes in Statistics, Springer* 222 (2019).
- [5] Thomas Nagler. “Simplified vine copula models: state of science and affairs”. In: *Risk Sciences* 1 (2025), p. 100022.
- [6] Irene Gijbels, Marek Omelka, and Noel Veraverbeke. “Nonparametric testing for no covariate effects in conditional copulas”. In: *Statistics* 51.3 (2017), pp. 475–509.
- [7] Alexis Derumigny and Jean-David Fermanian. “About tests of the “simplifying” assumption for conditional copulas”. In: *arXiv preprint arXiv:1612.07349* (2016).
- [8] Malte S Kurz and Fabian Spanhel. “Testing the simplifying assumption in high-dimensional vine copulas”. In: *Electronic Journal of Statistics* 16.2 (2022), pp. 5226–5276.
- [9] Alexis Derumigny. “Measures of non-simplifyingness for conditional copulas and vines”. In: *arXiv preprint arXiv:2504.07704* (2025).
- [10] James O Ramsay and Bernard W Silverman. *Functional data analysis*. Springer, 2005.
- [11] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. “Functional data analysis for sparse longitudinal data”. In: *Journal of the American statistical association* 100.470 (2005), pp. 577–590.
- [12] Debangana Dey et al. “Functional Principal Component Analysis for Continuous Non-Gaussian, Truncated, and Discrete Functional Data”. In: *Statistics in Medicine* 43.28 (2024), pp. 5431–5445.
- [13] Abe Sklar. “Random variables, joint distribution functions, and copulas”. In: *Kybernetika* 9.6 (1973), pp. 449–460.
- [14] Roger B Nelsen. *An introduction to copulas*. Springer, 2006.
- [15] Marius Hofert et al. *Elements of copula modeling with R*. Springer, 2018.
- [16] Andrew J Patton. “Modelling time-varying exchange rate dependence using the conditional copula”. In: (2001).

- [17] Alexis Derumigny and Jean-David Fermanian. “On kernel-based estimation of conditional Kendall’s tau: finite-distance bounds and asymptotic behavior”. In: *arXiv preprint arXiv:1810.06234* (2018).
- [18] Rutger van der Spek and Alexis Derumigny. “Fast estimation of Kendall’s Tau and conditional Kendall’s Tau matrices under structural assumptions”. In: *Dependence Modeling* 13.1 (2025), p. 20250012.
- [19] Alexis Derumigny and Johannes Schmidt-Hieber. “On lower bounds for the bias-variance trade-off”. In: *The Annals of Statistics* 51.4 (2023), pp. 1510–1533.
- [20] Aad W Van der Vaart. *Asymptotic statistics*. Vol. 3. Cambridge university press, 2000.
- [21] Tosio Kato. *Perturbation theory for linear operators*. Vol. 132. Springer Science & Business Media, 2013.
- [22] Giordano Giambartolomei. “The Karhunen-Loeve Theorem”. In: (2015).
- [23] Jiří Blank, Pavel Exner, and Miloslav Havlíček. *Hilbert space operators in quantum physics*. Springer, 2008.
- [24] Michael Renardy and Robert C Rogers. *An introduction to partial differential equations*. Springer, 2004.
- [25] Steven Roman. *Advanced linear algebra*. Springer, 2005.
- [26] Tailen Hsing and Randall L Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Vol. 997. Wiley Online Library, 2015.
- [27] N Dunford and JT Schwartz. *Linear Operators. Part 2. Spectral Theory*. 1971.
- [28] Leslie Hogben. *Handbook of linear algebra*. CRC press, 2006.
- [29] Yi Yu, Tengyao Wang, and Richard J Samworth. “A useful variant of the Davis–Kahan theorem for statisticians”. In: *Biometrika* 102.2 (2015), pp. 315–323.
- [30] Harold Widom. “Asymptotic behavior of block Toeplitz matrices and determinants. II”. In: *Advances in Mathematics* 21.1 (1976), pp. 1–29.
- [31] Wassily Hoeffding. “A Non-Parametric Test of Independence”. In: *The Annals of Mathematical Statistics* 19.4 (1948), pp. 546–557.
- [32] Alexis Derumigny. “Estimation of a regular conditional functional by conditional U-statistic regression”. In: *Statistica Neerlandica* 79.1 (2025), e12350.
- [33] Robert J Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 2009.
- [34] Winfried Stute. “Conditional U-statistics”. In: *The Annals of Probability* (1991), pp. 812–825.
- [35] Ian Jolliffe. “Principal component analysis”. In: *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.
- [36] Theodore Wilbur Anderson et al. *An introduction to multivariate statistical analysis*. Vol. 2. Wiley New York, 1958.

- [37] AW Davis. "Asymptotic Theory for Principal Component Analysis: Non-Normal Case 1". In: *Australian Journal of Statistics* 19.3 (1977), pp. 206–212.
- [38] Hans F Weinberger. *Variational methods for eigenvalue approximation*. SIAM, 1974.
- [39] Christian Francq and Jean-Michel Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.
- [40] James Ward Brown and Ruel V Churchill. *Complex variables and applications*. McGraw-Hill, 2009.
- [41] Ronald L Graham. *Concrete mathematics: a foundation for computer science*. Pearson Education India, 1994.
- [42] Alexis Derumigny and Jean-David Fermanian. "On Kendall's regression". In: *Journal of Multivariate Analysis* 178 (2020), p. 104610. ISSN: 0047-259X.

A

SUBSET ANALYSIS OF HIGHLY DEPENDENT MARKETS

Figure 7.5 demonstrated that many asset pairs in the empirical dataset are unconditionally uncorrelated. Conversely, the specific subset of markets $\{\text{AEX}, \text{DAX}, \text{FCHI}, \text{SP500}, \text{DJI}, \text{N225}\}$ exhibits strong unconditional dependency, suggesting that their conditional Kendall's tau (CKT) curves may display more pronounced dynamic variation. Consequently, we restrict this supplementary analysis to the combinations within these six markets. Figure A.1 displays the resulting estimated CKT curves.

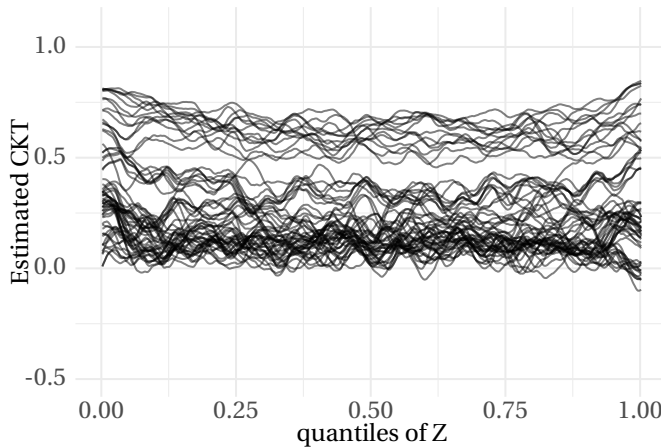


Figure A.1: The estimates of the conditional Kendall's tau for all combinations of assets based on the sample $\{\text{AEX}, \text{DAX}, \text{FCHI}, \text{SP500}, \text{DJI}, \text{N225}\}$. The x -axis is the quantiles of the associated covariate variable.

To extract the distinct functional profiles among these CKT curves, we apply Functional Principal Component Analysis (FPCA). The resulting eigenfunctions in Figure A.2 closely

mirror the components identified in the full-sample analysis (Figure 7.9). This alignment is expected, as isolating this subset effectively removes asset pairs with negligible conditional dependence that otherwise yield flat CKT curves. The variance decomposition histogram in Figure A.3 confirms that the first principal component (FPC1) remains highly dominant, explaining nearly 96% of the total functional variance. This is even higher than in the full-dataset case. Since we have reduced the sample size from 1092 to 61 curves, the primary common mode of variation becomes structurally more prominent.

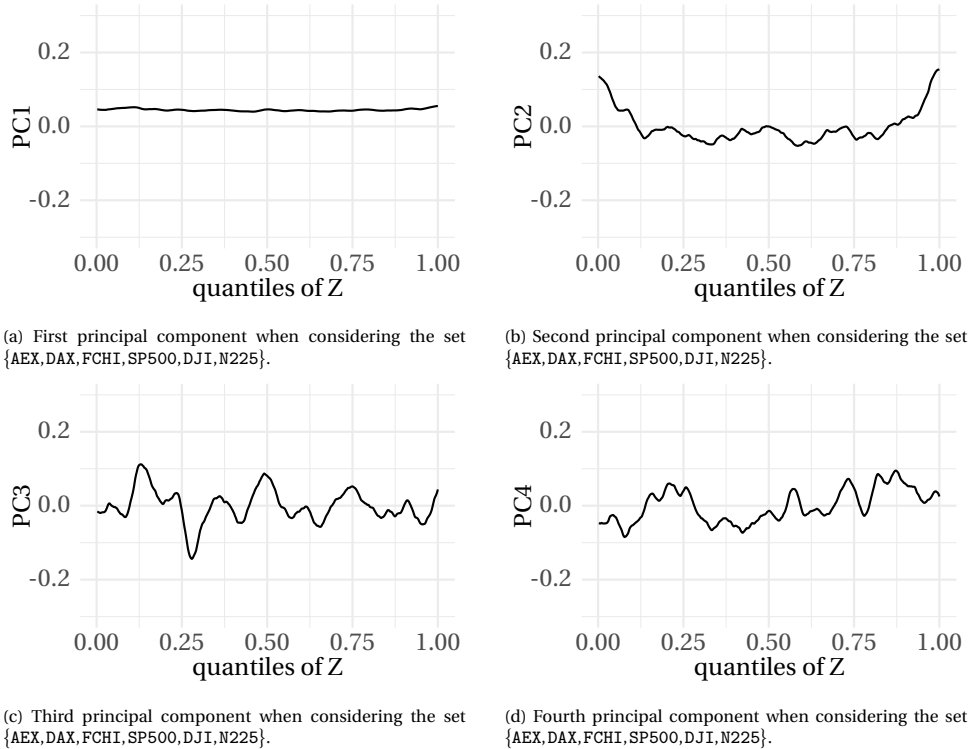


Figure A.2: The first four principal component of the set {AEX, DAX, FCHI, SP500, DJI, N225}.

To better visualize the remaining modes of variation, we normalize all the curves by subtracting the average value. In this way, the curves no longer possess vertical shifts, and PC1 is eliminated. This allows us to get more stable estimates. Figure A.4 shows the resulting variance decomposition.

This normalized view reveals that the second principal component captures an important share of the remaining variation, contributing nearly 30% of the residual variance. The subsequent principal components explain significantly less variation and primarily capture noise.

In summary, restricting the dataset to highly dependent markets does not reveal new structural dynamics compared to the global analysis. However, because the number of

functional observations is drastically reduced, the estimated components may be subject to higher sampling noise. This means we should cautiously interpret the economic meaning of higher-order components.

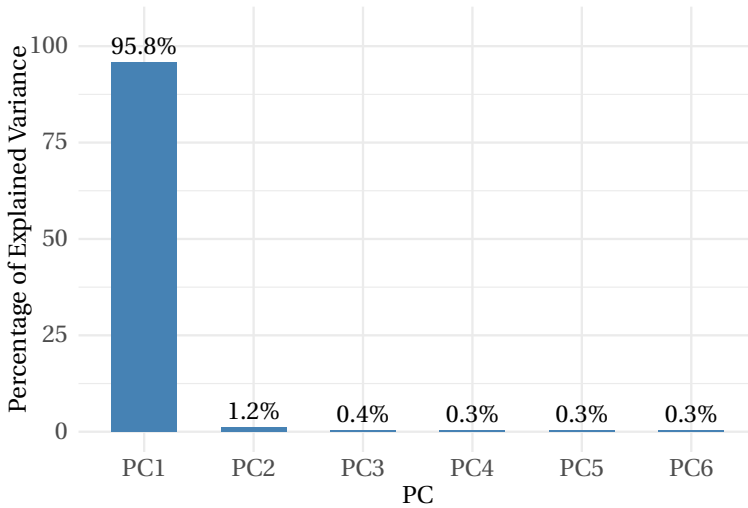


Figure A.3: Histogram showing the relative percentage of variance explained by the principal components on the sample {AEX, DAX, FCHI, SP500, DJI, N225}.

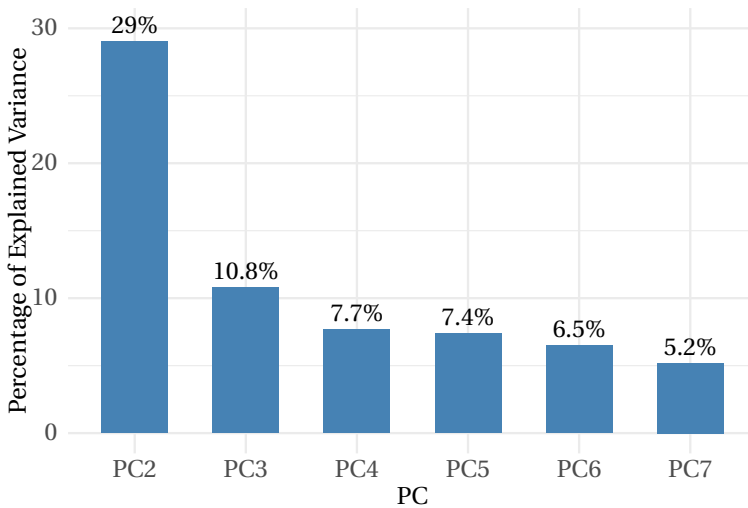


Figure A.4: Histogram showing the relative percentage of variance explained by the principal components after normalizing each curve by subtracting the average value on the sample {AEX, DAX, FCHI, SP500, DJI, N225}.