Robust Statistical Methods and Advanced Computational Techniques for High-Dimensional Data Analysis

by

Tianyi Zhang

(Under the Direction of Yuan Ke)

Abstract

In this dissertation, we explore the problems of high-dimensional feature screening and sampling techniques using Quantum Walk. In the first project, we introduce a novel feature screening methodology that is robust to the underlying distributions of the data, making it well-suited for high-dimensional heterogeneous data. This method is built upon a dependence measure induced by Wasserstein distance, and Gaussianization of the data. We analyze its non-asymptotic properties. We also establish sure screening and rank consistency properties for the proposed screening method upon mild signal strength conditions. Simulation studies demonstrate that our approach outperforms classical feature screening methods in highly nonlinear and heterogeneous cases. In the second project, we propose a model-free feature screening procedure tailored to high-dimensional quantile regressions. We introduce a novel dependence measure to quantify quantile dependence using Copula theory and corresponding non-parametric Kernel estimator. We derive the optimal bandwidth selection for the estimator, and analyze asymptotic properties of the estimator. We also prove sure screening and rank consistency properties for this screening method upon mild signal strength conditions. Additionally, we propose a data-driven threshold selection method for the screening procedure, which effectively controls false discoveries. The feature screening and FDR control performance of our proposals is validated through simulations. In the third project, we apply our feature screening methods to the U.S. 2020 economic data to identify variables related to the GDP growth rate from 2019 to 2020. Using the selected variables and downstream statistical analysis, we explore strategies for maintaining economic stability during major crises, such as the COVID-19 pandemic. In the final project, we investigate the problem of sampling using the 2-state Quantum Walk on the line. We overview the 2-state Quantum Walk on the line and highlight the limitations of this sampling method. We propose a novel approach that combines the strengths of the 2-state Quantum Walk and Kernel smoothing techniques. Experiments indicate that our proposal outperforms traditional Quantum Walk in terms of both density estimation and sampling efficacy.

INDEX WORDS: feature screening, high-dimensional data, heterogeneous data, Wasserstein distance, quantile regression, FDR control, Quantum Walk, Kernel Smoothing

Robust Statistical Methods and Advanced Computational Techniques for High-Dimensional Data Analysis

by

Tianyi Zhang

B.S., Nankai University, 2018

M.S., University of Minnesota Duluth, 2020

A Dissertation Submitted to the Graduate Faculty of the

University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2025

Robust Statistical Methods and Advanced Computational Techniques for High-Dimensional Data Analysis

by

Tianyi Zhang

Major Professor: Yuan Ke

Committee: Shuyang Bai

Jaxk Reeves

Ruizhi Zhang

Electronic Version Approved:

Ron Walcott

Vice Provost for Graduate Education and Dean of the Graduate School

The University of Georgia

May 2025

Acknowledgments

I would like to express my heartfelt gratitude to my advisor, Professor Yuan Ke. Though I may not be the most gifted student, his unwavering support, guidance, and encouragement throughout my research and studies have been invaluable. His mentorship has made a significant impact on my academic journey. I am deeply grateful to have had him as my advisor.

I also want to extend my thanks to the faculty members in the Department of Statistics at the University of Georgia. Whether it was through their teachings in the classroom or their insightful discussions on various projects, they have all contributed immensely to my growth as a researcher. I am especially thankful to the staff at the Statistical Consulting Center, whose help and guidance have enriched my journey.

My deepest thanks go to my wife, Li Di. Her constant support and understanding have been my anchor throughout this journey. Having her by my side has made the challenges of academic life much easier to bear. Our time together, especially as we built our lives and shared countless moments of happiness, has been a source of great joy and strength for me.

Finally, I am forever grateful to my precious daughter, Chlorissa, who was born in the final year of my Ph.D. Although her arrival added to my responsibilities, it also filled my life with an immense sense of purpose. She has been a constant source of motivation, both in my personal life and in my research.

Contents

Acknowledgments **iv**

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Challenges in High-Dimensional Data Analysis 1

1.2 Literature Review for Feature Screening 2

1.3 Literature Review for Quantum Sampling 5

1.4 Organization of this dissertation 6

1.5 Notation 7

2 Distribution-Robust High-dimensional Feature Screening **9**

2.1 Wasserstein Dependence 10

2.2 Wasserstein Dependence via Guassianization 12

2.3 Distribution Robust Feature Screening 15

2.4 Simulation Study 19

2.5 Proofs for Chapter 2 23

3 Model-free Feature Screening and False Discovery Control for High-dimensional Quantile Regressions 35

3.1 Quantile Dependence and Conditional Copula Density 36

3.2 Estimation of Copula Density 38

3.3 Estimation of Conditional Copula 41

3.4 Estimation of Copula Quantile Dependence 44

3.5 Model-free Feature Screening for Quantile Regressions 45

3.6 False Discovery Control 49

3.7 Simulation Study 54

3.8 Important Parameters 60

3.9 Proof of Key Theorems in Chapter 3 62

4 Applications to the U.S. 2020 Economic Data 83

4.1 Feature Screening with DR-WD$_r$-SIS 84

4.2 Feature Screening with QC-FDR 86

5 Consistent Sampling With Smoothed Quantum Walk 90

5.1 Sampling with 2-state Quantum Walk on the Line 91

5.2 Kernel Smoothing 97

5.3 Kernel Smoothed Quantum Sampling 98

5.4 Experiments 100

Appendices **105**

A.1 Additional Lemmas for Chapter 3 105

A.2 Simulations for Bandwidth Selection in Chapter 3 157

A.3 Bias Adjustification Term in Chapter 3 160

A.4 Variable Names of the U.S. 2020 Economic Data 163

Bibliography **177**

List of Figures

4.1 Top 12 and bottom 12 states in GDP growth rate from 2019 to 2020. 86
4.2 Scatter plots and fitted local linear quantile regressions between GDP growth rate and selected variables at quantile level $\alpha = 0.25, 0.75$. 89
5.1 Quantum Walk distributions and target distributions. 96
5.2 SQS densities. 101
5.3 TSQS densities. 102
5.4 Histograms and distributions of p-values of KS tests on QW samples and SQS samples. 104
A.5 Bandwidth selection for sample Copula Quantile Dependence. 159
A.6 Scatter plots, fitted lines and confidence intervals of linear regressions between $\log\left(\delta_n^*(\alpha)\right)$ and $\log(n)$. 161

List of Tables

2.1 The quantiles of minimum model size for Experiment 1 over 200 replications. 19

2.2 The quantiles of minimum model size for Experiment 2 over 200 replications. 21

2.3 The quantiles of minimum model size for Experiment 3 over 200 replications. 22

3.1 The median and IQR of minimum model size over 200 replications for Experiment 1.a. 54

3.2 The median and IQR of minimum model size over 200 replications for Experiment 1.b. 55

3.3 The median and IQR of minimum model size over 200 replications for Experiment 1.c. 56

3.4 The median and IQR of minimum model size over 200 replications for Experiment 1.d. 57

3.5 FDR control performances over 200 replications with $p = 1000$. 58

3.6 FDR control performances over 200 replications with $p = 2000$. 59

4.1 Top 10 selected variables by DR-WD_1-SIS. 84

4.2 Top 10 selected variables by DR-WD_2-SIS. 85

4.3 Selected variables at quantile level $\alpha = 0.25$. 87

A.1 Results of bias adjustification term 162

A.2 Variable names of U.S. 2020 economic data 163

Chapter 1

Introduction

1.1 Challenges in High-Dimensional Data Analysis

Recently, there is a rapid increase of the demand to high-dimensional data analysis in various domains, from genomics to finance. However, analyzing high-dimensional datasets are usually challenging.

On one hand, high-dimensional data often suffer from redundancy, multicollinearity, and spurious dependence, which can obscure model specification and reduce model interpretability (Fan and Li, 2006). These problems also result in overfitting in statistical models, which could reduce the effectiveness of classical statistical methods (Hastie, 2009) in practice. Feature screening plays a vital role in resolving the challenges in analyzing high-dimensional data. Feature screening usually ranks features based on marginal statistical criteria, such as correlation coefficients. By selecting the most "informative" variables, feature screening could improve model interpretability, and enhances generalization performance (Fan and Lv, 2008; Guyon and Elisseeff, 2003).

On the other hand, due to the curse of dimensionality (Bellman, 1966), it could possibly lead to severe computational bottlenecks and storage demands (Donoho et al., 2000). Traditional computing methods struggle to handle high-dimensional data analysis efficiently since they are built on classical computer with classical bits. In contrast, utilizing quantum computer becomes an attractive alternative. Quantum computation, constructed on quantum bits (qubits), provides tremendous speedup for plenty of computational tasks (Montanaro and Pallister, 2016). Shor's factorization algorithm (Shor, 1994) and Grover's search algorithm (Grover, 1996) reveal that quantum computing outperforms classical methods in optimization and search problems.

1.2 Literature Review for Feature Screening

Datasets with high dimensional features characterize many contemporary research areas. When the features contain redundant or noisy information, estimating their functional relationship with the response may become quite challenging (Fan et al., 2009; Hall and Miller, 2009; Lv and Liu, 2014). To address these challenges arising from high dimensionality, Fan and Lv, 2008 proposed the sure independence screening (SIS) method, which aims to remove redundant features by ranking their marginal Pearson correlations. Motivated by the success of SIS, the idea of feature screening has been extended to analyze various high-dimensional datasets (Fan and Lv, 2010; J. Liu et al., 2015).

The idea of feature screening is popular for the following two reasons.

Sure screening. The selected subset of features contains all the active ones with probability approaching one.

Computationally efficient. Selection of features is processed by ranking the marginal "significance" of the features. The computational complexity is linearly proportional to the dimension.

However, most existing feature screening methods rely on model assumptions such as linearity or other specific parametric forms. In high-dimensional regimes, it is challenging to correctly specify a model before discarding a large number of redundant features. Consequently, it is crucial to develop model-free feature screening methods that can be implemented without pre-specifying a model. The model-free property is essential because it ensures the effectiveness of the screening method even when the underlying model is misspecified, which is crucial for obtaining a reliable and parsimonious set of active feature for downstream analyses. Over the past decade, the development of model-free feature screening methods has become a hot topic in statistics (Zhu et al., 2011; R. Li et al., 2012; Mai and Zou, 2013; Y. Zhou and Zhu, 2018; W. Liu et al., 2022). Additionally, model-free feature screening techniques have been applied to discriminant analysis (Cui et al., 2015), censored data analysis (T. Zhou and Zhu, 2017), survival data analysis (Lin et al., 2018), multi-class classification (Ni and Fang, 2016), among many other applications.

To apply feature screening methods, one need to choose a threshold parameter to separate active from inactive features. Under some pre-specified model assumption, this threshold can be determined using cross-validation or information-criterion approaches. However, in a model-free context, such approaches are not directly applicable because goodness-of-fit is not well-defined. In practice, one may choose a conservative threshold to ensure that all active features are likely to be included, although this can admit many inactive features and inflate the false discovery rate (FDR). Consequently, balancing the sure screening property with FDR control is inherently challenging in model-free feature screening. Unfortunately, only limited works tried to integrate FDR control with feature screening methods (W. Liu et al., 2022; Guo et al., 2023; Tong et al., 2023).

1.2.1 Distribution-Robust feature screening

Another crucial property of feature screening is *data adaptivity*, which implies that the performance of the feature screening methodology should not rely on strong assumptions, such as independence or sub-Gaussianity. These assumptions are often unreasonable due to the presence of spurious correlations and heterogeneity in high-dimensional datasets (Fan and Zhou, 2016; Fan et al., 2018). If not properly addressed, these challenges can lead to misleading conclusions. Unfortunately, only a limited number of studies have focused on these complex issues, see McKennan and Nicolae, 2019; Wang et al., 2012. Achieving data adaptivity for feature screening is particularly difficult, and only a few works have tackled this problem. For example, Xie et al., 2020 proposed a category-adaptive screening method for high-dimensional heterogeneous data to identify category-specific important covariates, while He et al., 2013 introduced a quantile-adaptive screening method for such data. In this context, we refer to feature screening methods that are insensitive to the underlying data distributions as *Distribution-Robust feature screening*. Such methods would be better equipped to handle high-dimensional heterogeneous data.

1.2.2 Feature screening with quantile regression

Quantiles provide informative snapshots for summarizing a probability distribution. In contrast to moments, which capture global characteristics and can be heavily influenced by extreme values, quantiles focus on localized distributional properties and are less sensitive to small perturbations in the data. Building on this concept, quantile regressions have become fundamental tools in statistics and data science (Stigler, 1984). Rather than focusing solely on the mean, quantile regressions examine various quantiles of the conditional distribution of a response variable, allowing researchers to explore a specific percentile of interest or to investigate how predictors influence the response across lower, median, and upper quantiles. This flexibility is particularly valuable in applications where extreme values or specific regions of the distribution are of primary concern, such as assessing tail risk in finance (Linton and Xiao, 2017; Nguyen et al., 2020), studying treatment effects in epidemiology (Wei et al., 2019; Powell, 2020), and detecting anomalies in sensor networks (Xu et al., 2019, Z. Li and Van Leeuwen, 2023).

The applications of quantile regressions continue to expand in the era of big data, offering a nuanced understanding of how features influence various segments of the outcome distribution (Yu et al., 2003; Koenker, 2017). Despite the importance of model-free feature screening for quantile regressions, this area remains under-explored. A quantile-adaptive nonlinear feature screening method was proposed for high-dimensional heterogeneous data, which employs splines to model marginal effects at a specific quantile of interest (He et al., 2013). More recently, a sure independence screening procedure based on quantile correlation was proposed, which is robust against outliers and can capture nonlinear relationships between the response variable and features (Ma and Zhang, 2016).

1.3 Literature Review for Quantum Sampling

Quantum Walks (QWs) are quantum counterparts to classical random walks, offering a rich set of concepts in this area. Like classical random walks, QWs can be defined as an evolution process on a graph. In classical random walks, the walker occupies definite locations (states), and the transition process stochastically depends on a probability distribution. In contrast, in QWs, the walker exists in a superposition of states, and the evolution is determined by unitary operators, which are deterministic. For a thorough introduction to QWs, we refer to works such as Aharonov et al., 2001; Kempe, 2003; S. Venegas-Andraca, 2022, among others. Limit theorems for QWs (Attal et al., 2015; Chisaki et al., 2009; Konno, 2005) have also inspired the development of quantum sampling methods.

A key concept in QWs is the QW on the Line, which corresponds to the evolution process over the set of all integers $\mathbb{Z}$, as defined in Nayak and Vishwanath, 2000; S. E. Venegas-Andraca, 2012. In Machida, 2013, the author proved limit theorems for the 2-state QW on the line, under various initial states, showing its ability to sample from a variety of target distributions, including the Semicircle distribution, Uniform distribution, Truncated Gaussian distribution, and Arcsine distribution. Despite its strong theoretical properties, the 2-state QW on the line, at any given time t, generates samples from a discrete distribution that does not match the target distributions in the limiting case, which limits its finite-sample performance in QW-based sampling methods.

1.4 Organization of this dissertation

The structure of this proposal is as follows. In this chapter, we provide the background on feature screening and Quantum Walks, along with a review of existing research in the literature. In Chapter 2, we introduce a Distribution-Robust feature screening method designed for high-dimensional heterogeneous data. This method is based on a novel dependence measure derived from Wasserstein distance. In Chapter 3, we present a model-free feature screening approach and an associated FDR control method for high-dimensional quantile regressions. Chapter 4 focuses on analyzing U.S. 2020 economic data using the proposed feature screening and downstream methods, exploring strategies for maintaining economic stability during major crises such as COVID-19. Finally, in Chapter 5, we propose a novel quantum sampling algorithm that combines the principles of 2-state Quantum Walk on the line with Kernel smoothing techniques.

1.5 Notation

Let $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{C}$ denote the set of integers, natural numbers, real numbers, and complex numbers respectively. $\Re(z)$, and $\Im(z)$ denote the real part, and imaginary part of $z \in \mathbb{C}$ respectively. Let $|S|$ denotes the cardinality of a set S. The superscript $^\intercal$ denotes the transpose of a matrix or a vector. Given a vector $\boldsymbol{x} = (x_1, \ldots, x_d)^\intercal \in \mathbb{R}^d$, we write the vector ℓ_q-norm as $\|\boldsymbol{x}\|_q = \left(\sum_{j=1}^d |x_j|^q\right)^{1/q}$ for $1 \leqslant q < \infty$ and the vector ℓ_∞-norm as $\|\boldsymbol{x}\|_\infty = \max_{j\in\{1,\cdots,d\}} |x_j|$. Given two vectors $|x\rangle_c$, $|y\rangle_c$ in a Hilbert space $\mathcal{H}_c$, and $|z\rangle_p$ in a Hilbert space $\mathcal{H}_p$, we write the inner product between $|x\rangle_c$ and $|y\rangle_c$ as $\langle x|y\rangle_c$, outer product between $|x\rangle_c$ and $|y\rangle_c$ as $|x\rangle_c\langle y|$, tensor product between $|x\rangle_c$ and $|z\rangle_p$ as $|x\rangle_c \otimes |z\rangle_p = |x\rangle_c |z\rangle_p$. Given a matrix $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$, if $d_1 = d_2 = d$, $\mathrm{tr}(\mathbf{A})$ denote the trace of $\mathbf{A}$. For $a \in \mathbb{R}$, let $\lfloor a \rfloor = \max\{z \in \mathbb{Z}, z \leqslant a\}$ and $\lceil a \rceil = \min\{z \in \mathbb{Z}, z \geqslant a\}$. Function $f(a) = (a)_+$ is defined as $f(a) = a$ if $a \geqslant 0$, and $f(a) = 0$ if $a < 0$. $\mathbb{1}_{\{.\}}$ denotes indicator function, i.e. for event $\mathcal{B}$, $\mathbb{1}_\mathcal{B} = 1$ if $\mathcal{B}$ holds, and $\mathbb{1}_\mathcal{B} = 0$ if $\mathcal{B}$ does not hold. For a set $D \subset \mathbb{R}$, and $r > 0$, $C^r(D)$ denotes the set of functions that are differentiable up to order r on the set D; $C^\infty(D)$ denotes the set of functions that are infinitely differentiable on the set D. For a set $D \subset \mathbb{R}$, D^+ denotes the positive subset of D. For event $\mathcal{B}$, $\mathcal{P}(\mathcal{B})$ denotes the probability that event $\mathcal{B}$ happens. For random variable X, $\mathbb{E}(X)$ denotes the expectation of X. $\xrightarrow{d}$, and $\xrightarrow{p}$ denote convergence in distribution, and convergence in probability respectively. For random variables X_n, and sequence a_n, $X_n = \mathcal{O}_P(a_n)$ as $n \to \infty$ means that $\forall \epsilon > 0$, there exists a finite $M > 0$, s.t. $\mathcal{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \epsilon$, for large enough n; $X_n = o_p(a_n)$ as $n \to \infty$ means that

$\forall \epsilon > 0, \lim_{n\to\infty} \mathcal{P}\left(\left|\frac{X_n}{a_n}\right| \geqslant \epsilon\right) = 0$; $X_n = \mathcal{O}_{a.s.}(a_n)$ as $n \to \infty$ means that there exists a finite $M > 0$, s.t. $\left|\frac{X_n}{a_n}\right| \leqslant M$ holds almost surely for large enough n; $X_n = o_{a.s.}(a_n)$ as $n \to \infty$ means that $X_n/a_n \to 0$ almost surely as $n \to \infty$. For sequence a_n, and b_n, $b_n = \mathcal{O}(a_n)$ as $n \to \infty$ means that there exists a finite $M > 0$, s.t. $\left|\frac{b_n}{a_n}\right| \leqslant M$ for large enough n; $b_n = o(a_n)$ as $n \to \infty$ means that $b_n/a_n \to 0$ as $n \to \infty$.

Chapter 2

Distribution-Robust High-dimensional Feature Screening

In this chapter, we propose a feature screening method by a novel dependence measure constructed on Wasserstein distance, which is fit to high-dimensional heterogeneous data. In Section 2.1, we introduce the dependence measure constructed on Wasserstein distance, and reviewed its limitations in previous works. In Section 2.2, we propose a Wasserstein dependence measure via Gaussianization and establish its non-asymptotic properties, which is not requiring restrictive conditions on the data. In Section 2.3, we formalize the feature screening methodology and show that our proposed method enjoys sure screening property and even stronger rank consistency property with strong enough signal strength and proper selected screening threshold,. We name this method as *Distribution-Robust* r*-Wasserstein Dependence Sure Independence Screening (DR-*WD_r*-SIS)*. In Section 2.4, we present the results by simulation studies.

2.1 Wasserstein Dependence

2.1.1 Dependence measure with Wasserstein distance

Wasserstein distances are metrics on spaces of probability measures that possess finite moments of a certain order. They quantify the distance between two distributions by determining the minimal cost required to transport probability mass by transforming one distribution into the other. The formal definition of the Wasserstein distance is provided below.

Definition 2.1.1. *Let $\mathcal{P}\left(\mathbb{R}^d\right)$ denote the set of Borel probability measures on $\mathbb{R}^d$, and let $\mathcal{P}_r\left(\mathbb{R}^d\right)$ be the subset comprising measures with finite moments of order $r \in [1, \infty)$. For measures $\mu, \rho \in \mathcal{P}\left(\mathbb{R}^d\right)$, let $\Gamma(\mu, \rho)$ denote the set of probability measures γ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ρ; that is,*

$$\gamma\left(B \times \mathbb{R}^d\right) = \mu(B) \quad \textit{and} \quad \gamma\left(\mathbb{R}^d \times B\right) = \rho(B)$$

for all Borel sets $B \subseteq \mathbb{R}^d$. The r-Wasserstein distance between measures μ and ρ in $\mathcal{P}_r\left(\mathbb{R}^d\right)$ is defined by

$$\mathcal{W}_r(\mu, \rho) \doteq \left(\inf_{\gamma \in \Gamma(\mu, \rho)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^r \, \mathrm{d}\gamma(x, y)\right)^{1/r},$$

where $\|\cdot\|$ denotes some norm in $\mathbb{R}^d$. For simplicity, we considered $\|\cdot\|$ as ℓ_q norm, $\|\cdot\|_q$, for $q \geqslant 1$, and $r = 1, 2$ in this Chapter.

By Definition 2.1.1, it is obvious that $\mathcal{W}_r(\mu, \rho) = 0$ if and only if $\mu = \rho$, which motivates studies on constructing dependence measures between variables using Wasserstein distances. Consider two continuous random variables X and Y with joint distribution $f_{X,Y}(x, y)$. The *Wasserstein dependence of order*

r, is defined as the r-Wasserstein distance between the joint distribution $f_{X,Y}(x,y)$ and the product of their marginal distributions $f_X(x)\, f_Y(y)$. Mathematically, Wasserstein dependence of order r is defined as

$$\mathcal{W}_r\left(f_{X,Y},\ f_X \otimes f_Y\right),$$

where we use $f_{X,Y}$, f_X, and f_Y to represent their corresponding measures, and $f_X \otimes f_Y$ represents the measure of (X, Y) when they are independent with each other. This dependence measure captures the degree to which (X, Y) deviates from independence.

2.1.2 Limitation of Wasserstein Dependence

There has been a few works studying constructing dependence measure by Wasserstein distance. However, Hallin et al., 2021; Ozair et al., 2019 lack the theoretical results of proposed methods. De Keyser and Gijbels, 2025; Mordant and Segers, 2022 impose strong assumptions based on Multivariate Gaussian distributions or Gaussian Copula, which limits their general applicability. These methods often normalize the proposed Wasserstein dependence statistic by the supreme of the Wasserstein dependence over all possible measures to slightly mitigate the impact of heterogeneous data. However, constructing Wasserstein dependence which is not sensitive to the conditions of marginal distributions is still unsolved.

Despite these drawbacks on previous works, estimation of Wasserstein dependence, or empirical Wasserstein distance is usually needed in statistical applications, since one typically does not have direct access to the distributions of interest $f_{X,Y}(x,y)$, and its margins $f_X(x)$, and $f_Y(y)$. Instead, statisticians usually only have access to the sample or equivalently, to their empirical measures, $\widehat{f}_{X,Y}(x,y)$, $\widehat{f}_X(x)$, and $\widehat{f}_Y(y)$. The convergence properties, in Wasserstein distance, of these empirical measures has been studied extensively; In Rippl et al., 2016, limiting distributions of empirical Wasserstein distance is proved under Gaussian assumptions. Del Barrio and Loubes, 2019; Del Barrio et al., 2024; Fournier and Guillin, 2015; Lei, 2020; Weed and Bach, 2019 exhaustively studied both asymptotic and non-asymptotic properties of empirical Wasserstein distance under more general settings. Nevertheless, these works are always based on strong marginal conditions such as existence of moments to a specific order. These assumptions can be restrictive in high-dimensional settings, especially for capturing complex dependencies; requiring all features to satisfy such conditions often degrades performance in real-world applications.

In addition, construct the estimation of Wasserstein dependence usually involves the estimation of $f_{X,Y}(x, y)$, and its margins $f_X(x)$, and $f_Y(y)$ simultaneously using the same sample. As proposed in Nies et al., 2022, there are mainly two estimators. The first methodology is to apply sample splitting scheme. Formally, suppose we have a sample of size $2n$, $\{(x_i, y_i)\}_{i=1}^{2n}$, then the first half $\{(x_i, y_i)\}_{i=1}^{n}$ is used to estimate the joint measure $f_{X,Y}$, and the second half $\{(x_i, y_i)\}_{i=n+1}^{2n}$ is used to estimate the margins f_X, and f_Y. However, this method has flaws in sample efficiency. On the other hand, a permutation estimator can be an alternative by pretending the permuted sample $\{\left(x_i, y_{\sigma(i)}\right)\}$ is from the independent version of the measure, $f_X \otimes f_Y$, and apply it for estimation, where σ is a permutation over the samples. However, this method performs poor in practice, and it will be shown in Section 2.4.

2.2 Wasserstein Dependence via Guassianization

To overcome the disadvantages discussed in Section 2.1, we propose to construct Wasserstein Dependence via Guassianization. Mathematically, let $\{(x_i, y_i)\}_{i=1}^{n}$ be a random sample observed from (X, Y). The marginal CDFs of X, and Y can be estimated by empirical CDFs,

$$\widehat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i \leqslant x\}} \quad \text{and} \quad \widehat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i \leqslant y\}}.$$

For simplicity, we ignore the case when ties appear in the samples. Then, the Guassianized random sample

$$(\widehat{s}_i, \widehat{t}_i) \doteq \left(\Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_X(x_i)\right), \Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_Y(y_i)\right)\right), \tag{2.2.1}$$

is the counterpart of its genuine Gaussianized random sample

$$(s_i, t_i) \doteq \left(\Phi^{-1}\left(F_X(x_i)\right), \Phi^{-1}\left(F_Y(y_i)\right)\right), \tag{2.2.2}$$

for $i = 1, \cdots, n$. Consequently, we construct a sample version dependence measure

$$\widehat{I}_{\mathcal{W}}\left(X, Y; r\right) \doteq \mathcal{W}_r\left(\widehat{f}_{\widehat{S},\widehat{T}}, \ \phi \otimes \phi\right), \tag{2.2.3}$$

where $\widehat{f}_{\widehat{S},\widehat{T}}$ represents the empirical measure on Gaussianized random sample (2.2.1), $\phi \otimes \phi$ represents the measure of independent bivariate Gaussian distribution. In this definition, the second measure is replaced with $\phi \otimes \phi$ since the margins of Gaussianized sample are mimicking the standard Gaussian distribution. Clearly, (2.2.3) is the empirical version of

$$I_{\mathcal{W}}\left(X, Y; r\right) \doteq \mathcal{W}_r\left(f_{S,T}, \ \phi \otimes \phi\right), \tag{2.2.4}$$

where $f_{S,T}$ represents the joint measure of $(S,T) \doteq \left(\Phi^{-1}\left(F_X(X)\right), \Phi^{-1}\left(F_Y(Y)\right)\right)$. Correspondingly, we name this dependence measure as *Distribution-Robust r-Wasserstein Dependence (DR-WD$_r$)*. Then we study the property of the sample version estimator (2.2.3) in the following theorem.

Theorem 2.2.1. *For Distribution-Robust r-Wasserstein Dependence, and its sample version defined in* (2.2.3), *and* (2.2.4),

1. *when $r = 1$, there exists some positive constants M, C, n_0, and ε_0, which are not related to margins of (X, Y) or n, s.t. when $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{\log n}{\sqrt{n}}, \varepsilon_0)$,*

$$\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}\left(X,Y;1\right) - I_{\mathcal{W}}\left(X,Y;1\right)\right| \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right).$$

   More specifically, for some $K > 0$, and $\kappa \in [0, \frac{1}{2})$, there exists some positive constants n_0, and C, which are not related to margins of (X, Y) or n, s.t. when $n \geqslant n_0$,

$$\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}\left(X,Y;1\right) - I_{\mathcal{W}}\left(X,Y;1\right)\right| \geqslant Kn^{-\kappa}\right) \leqslant C\exp\left(-\frac{Cn^{1-2\kappa}}{\log^2 n}\right).$$

2. *when $r = 2$, there exists some positive constants M, C, n_0, and ε_0, which are not related to margins of (X, Y) or n, and $\forall \beta \in (0, 1)$, s.t. when $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{\log^{\frac{3}{4}} n}{n^{\frac{1}{4}}}, \varepsilon_0)$,*

$$\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}\left(X,Y;2\right) - I_{\mathcal{W}}\left(X,Y;2\right)\right| \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^4}{\log^2 n}\right) + C\exp\left(-C\left(n\varepsilon^2\right)^{\beta}\right).$$

More specifically, for some $K > 0$, *and* $\kappa \in [0, \frac{1}{4})$, *there exists some positive constants* n_0, *and* C, *which are not related to margins of* (X, Y) *or* n, *and* $\forall \beta \in (0, 1)$, *s.t. when* $n \geqslant n_0$,

$$
\begin{aligned}
&\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}(X, Y; 2) - I_{\mathcal{W}}(X, Y; 2)\right| \geqslant Kn^{-\kappa}\right) \\
\leqslant &C \exp\left(-\frac{Cn^{1-4\kappa}}{\log^2 n}\right) + C \exp\left(-Cn^{\beta(1-2\kappa)}\right).
\end{aligned}
$$

Theorem 2.2.1 discussed the non-asymptotic property of DR-WD$_r$, where the exponential convergence rate is fast and guarantees superior theoretical performance. In addition, the results are not established on restricted moment conditions on the margins or even stronger distributional assumptions, which shows robustness to the underlying distribution of (X, Y), and suggests DR-WD$_r$ is naturally fit to highly heterogeneous data.

2.3 Distribution Robust Feature Screening

2.3.1 Screening Methodology

The advantages of DR-WD$_r$ motivates us to construct a model-free feature screening method which is fit to heterogeneous data. Consider a high-dimensional data set, where Y is the response, and $X_1, \cdots, X_p$ are predictors. Feature screening is to identify the true active set for modeling Y with $\mathbf{X} \doteq (X_1, \cdots, X_p)$, which is defined as

$$
\mathcal{A} \doteq \left\{j \in \{1, \cdots, p\} : F_{Y|\mathbf{X}}(y|\mathbf{X}) \text{ functionally depends on } X_j \text{ for some } y\right\}.
$$

The natural idea is to reserve the features which have large DR-WD$_r$ with Y. Specifically, we estimate $\mathcal{A}$ via

$$\mathcal{A}^*(t) \doteq \{j \in \{1, \cdots, p\} : I_{\mathcal{W}}(X_j, Y; r) \geqslant t\},$$

where $t > 0$ is a screening threshold.

Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ be an i.i.d. random sample drawn from $(\mathbf{X}, Y)$, where $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top$. We estimate $I_{\mathcal{W}}(X_j, Y; r)$ by $\widehat{I}_{\mathcal{W}}(X_j, Y; r)$ as defined in (2.2.3). Then the active set is estimated by

$$\widehat{\mathcal{A}}(t_n) = \left\{j \in \{1, \cdots, p\} : \widehat{I}_{\mathcal{W}}(X_j, Y; r) \geqslant t_n\right\}, \tag{2.3.1}$$

with screening threshold, t_n. We name this screening approach as *Distribution-Robust r-Wasserstein Dependence Sure Independence Screening (DR-WD$_r$-SIS)*.

2.3.2 Theoretical properties

DR-WD$_r$-SIS achieves both the sure screening property and a stronger rank consistency property under some relatively mild conditions, and proper selected threshold t_n. These conditions are introduced below.

Condition 2.1. *Let $c_1 > 0$, $c_2 > 0$, $\kappa_1 \in [0, \frac{1}{2})$, and $\kappa_2 \in [0, \frac{1}{2r})$ be some constants.*

(a) $\min_{j \in \mathcal{A}} \{[I_{\mathcal{W}}(X_j, Y; r)]^r\} \geqslant c_1 n^{-\kappa_1}$.

(b) $\min_{j \in \mathcal{A}} \{I_{\mathcal{W}}(X_j, Y; r)\} - \max_{j \in \mathcal{A}^c} \{I_{\mathcal{W}}(X_j, Y; r)\} \geqslant c_2 n^{-\kappa_2}$.

Condition 2.1(a) is a minimum signal strength condition which indeed requires the DR-WD$_r$ between active features and the response variable to be uniformly bounded below. This condition allows the signal strength to converge to zero when n diverge, and it also requires the decaying speed of the signal strength is not too quick. Secondly, Condition 2.1(b), states the requirement on the difference between the signal strengths of active and inactive features. Because DR-WD$_r$ is non-negative, Condition 2.1(b) is generally stronger than Condition 2.1(a). In general, these minimum signal strength conditions allowing active features to be distinguished from inactive ones, and Condition 2.1 is not too restrictive, as it permits the minimum signal strength to diminish toward zero as n increases.

Theorem 2.3.1 (Sure screening property)**.** *Suppose Condition 2.1(a) holds, we take threshold value* $t_n^r = c_t n^{-\tau}$*, with* $\forall c_t \in (0, c_1)$*, and* $\tau \geqslant \kappa_1$*, then there exists some positive constants* n_1 *and* C_1*, which are not related to margins of* $(\mathbf{X}, Y)$ *or n, s.t. when* $n \geqslant n_1$*,*

1. *with* $r = 1$*,*

$$\mathcal{P}\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}(t_n)\right) \geqslant 1 - \mathcal{O}\left(\mathcal{S}_{\mathcal{A}} \exp\left(-\frac{C_1 n^{1-2\kappa_1}}{\log^2 n}\right)\right),$$

   where $\mathcal{S}_{\mathcal{A}}$ *is the cardinality of the active set* $\mathcal{A}$*.*

2. *with* $r = 2$*,* $\forall \beta \in (0, 1)$*,*

$$\mathcal{P}\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}(t_n)\right) \geqslant 1 - \mathcal{O}\left(\mathcal{S}_{\mathcal{A}}\left[\exp\left(-\frac{C_1 n^{1-2\kappa_1}}{\log^2 n}\right) + \exp\left(-C_1 n^{\beta(1-\kappa_1)}\right)\right]\right),$$

   where $\mathcal{S}_{\mathcal{A}}$ *is the cardinality of the active set* $\mathcal{A}$*.*

Theorem 2.3.2 (Rank consistency property)**.** *Suppose Condition 2.1(b) holds. There exists some positive constants* n_2 *and* C_2*, which are not related to margins of* $(\mathbf{X}, Y)$ *or n, s.t. when* $n \geqslant n_2$*,*

1. *with* $r = 1$,

$$\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 1\right)\right\} - \max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 1\right)\right\} > 0\right)$$
$$\geqslant 1 - \mathcal{O}\left(p\exp\left(-\frac{C_2 n^{1-2\kappa_2}}{\log^2 n}\right)\right).$$

If $\log p = o\left(\frac{n^{1-2\kappa_2}}{\log^2 n}\right)$, *we also have*

$$\liminf_{n\to\infty}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 1\right)\right\} - \max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 1\right)\right\}\right) > 0, \text{ almost surely.}$$

2. *with* $r = 2, \forall \beta \in (0, 1)$,

$$\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 2\right)\right\} - \max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 2\right)\right\} > 0\right)$$
$$\geqslant 1 - \mathcal{O}\left(p\left[\exp\left(-\frac{C_2 n^{1-4\kappa_2}}{\log^2 n}\right) + \exp\left(-C_2 n^{\beta(1-2\kappa_2)}\right)\right]\right).$$

If $\log p = o\left(\frac{n^{1-4\kappa_2}}{\log^2 n}\right)$, *and* $\log p = o\left(n^{\beta(1-2\kappa_2)}\right)$, *we also have*

$$\liminf_{n\to\infty}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 2\right)\right\} - \max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j, Y; 2\right)\right\}\right) > 0, \text{ almost surely.}$$

Condition 2.1(a) guarantees the sure screening property in Theorem 2.3.1, ensuring that all active features are included in the selected subset of variables when an appropriate threshold is applied to DR-WD$_r$-SIS. Under the stronger Condition 2.1(b), the rank consistency property in Theorem 2.3.2 maintains the stability of feature rankings based on their Distribution-Robust Wasserstein Dependences, effectively separating active features from inactive ones. Collectively, these theorems validate that ranking variables by their marginal Distribution-Robust Wasserstein Dependences with the response and selecting the highest-ranked features serves as a reliable method for feature screening.

2.4 Simulation Study

Table 2.1: The quantiles of minimum model size for Experiment 1 over 200 replications.

	Experiment 1.a					Experiment 1.b				
Quantiles	**5%**	**25%**	**50%**	**75%**	**95%**	**5%**	**25%**	**50%**	**75%**	**95%**
					$p = 500$					
SIS	4.00	5.00	8.00	22.25	121.25	4.00	7.00	12.00	34.25	148.10
DC-SIS	4.00	5.00	10.00	34.00	199.65	4.95	7.00	16.50	47.75	207.50
bcDC-SIS	4.00	5.75	11.00	39.00	218.10	5.00	8.00	18.00	50.25	265.10
BCor-SIS	5.00	17.00	51.00	166.25	315.55	8.00	35.75	106.50	209.00	362.20
MDC-SIS	4.00	5.00	9.00	26.50	159.25	5.00	7.00	18.00	43.25	205.80
WD_1-SIS	112.00	213.00	296.50	383.25	473.00	273.70	321.25	376.50	436.00	483.05
WD_2-SIS	87.85	203.50	270.50	390.00	462.00	278.85	328.50	378.50	431.25	486.05
DR-WD_1-SIS	4.00	6.00	14.50	67.25	292.25	5.00	12.75	49.00	141.25	340.50
DR-WD_2-SIS	4.00	6.00	25.00	93.25	331.00	5.00	15.75	76.00	179.75	307.05
					$p = 2000$					
SIS	4.00	5.00	6.00	9.00	35.10	4.00	5.00	6.00	10.00	33.15
DC-SIS	4.00	5.00	6.00	12.00	66.00	5.00	5.00	7.00	11.00	51.50
bcDC-SIS	4.00	5.00	6.00	12.00	66.00	5.00	5.00	7.00	11.00	69.10
BCor-SIS	4.00	6.75	23.00	104.25	762.45	5.00	13.00	48.00	187.25	853.05
MDC-SIS	4.00	5.00	6.00	9.00	52.05	5.00	5.00	7.00	12.00	65.20
WD_1-SIS	110.95	357.50	725.50	1204.25	1849.40	1035.95	1184.50	1337.00	1582.50	1901.85
WD_2-SIS	79.50	228.50	623.00	1111.50	1811.80	913.70	1255.75	1383.50	1568.75	1847.05
DR-WD_1-SIS	4.00	5.00	6.5	17.00	217.55	5.00	6.00	9.00	25.00	334.70
DR-WD_2-SIS	4.00	5.00	6.0	22.00	309.30	5.00	6.00	9.50	27.00	372.05

In this section, three simulated experiments were utilized to compare our proposal (DR-WD_r-SIS with $r = 1, 2$) with SIS in Fan and Lv, 2008, DC-SIS in R. Li et al., 2012, bcDC-SIS in Székely and Rizzo, 2014, MDC-SIS in Shao and Zhang, 2014, BCor-SIS in Pan et al., 2019. We also present the screening performance using Wasserstein dependence without Gaussianization and name it as WD_r-SIS with $r =$

$1, 2$. In this method, we implement permutation estimator as mentioned in Section 2.1. In addition, the calculation of (2.2.3) involves semi-discete Wasserstein distance, for which multiple methods were proposed (Dieci and Omarov, 2024; Hartmann and Schuhmacher, 2020) to approximate the semi-discrete Wasserstein distance. In this section, we consider constructing 2-dimensional low-discrepancy sequence (Sobol', 1967) of size n on $[0, 1]^2$, and then apply Probit transformations to estimate the measure $\phi \otimes \phi$. For each experiment, we have two settings $n = 100, p = 500$, and $n = 200, p = 2000$. We repeat the experiments for 200 replicates. Within each replication, we rank the features descendingly by the above nine screening metrics and record the minimum model size that contains all active variables. The screening performance is measured by quantiles (at levels $5\%, 25\%, 50\%, 75\%, 95\%$) of the minimum model size ($\mathcal{M}_{min}$) over 200 replications. Throughout this subsection, we denote $\Sigma = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 0.5^{|i-j|}$. We consider the following three regression models.

1. $Y = 5X_1 + 3X_{12} + 4X_{26} + 6X_{39} + \sqrt{20}\varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, 1)$. This model is an additive linear model, which has 4 active variables.

2. $Y = 5X_1 + 2\sin\left(\frac{\pi}{2}X_2\right) + 2|X_3| + 2\exp(5X_4) + \varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, 1)$. This model is an additive nonlinear model, which has 4 active variables.

3. $Y = 4X_1 + 3\log\left(\left|\frac{X_2}{1-X_1}\right|\right)\sin(2\pi|X_3|) + 4|X_3|\exp(5X_4 + 5X_5) + \varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, 1)$. This model has more complex nonlinear structure, which has 5 active variables.

In each experiment, ε is independent of the covariate vector $\mathbf{X} = (X_1, \cdots, X_p)^\intercal$, and we consider two subexperiments, which come from different data generation processes, and stand for different data structure.

a. $\mathbf{X} \sim MVN(\mathbf{0}_p, \Sigma)$, which represents homogeneous data.

Table 2.2: The quantiles of minimum model size for Experiment 2 over 200 replications.

	Experiment 2.a					**Experiment 2.b**				
Quantiles	**5%**	**25%**	**50%**	**75%**	**95%**	**5%**	**25%**	**50%**	**75%**	**95%**
					$p = 500$					
SIS	50.55	194.75	309.50	403.75	479.10	62.70	207.75	325.00	431.25	480.00
DC-SIS	50.95	167.50	294.50	391.25	464.20	47.00	117.25	248.50	380.50	478.10
bcDC-SIS	40.65	160.75	271.50	372.50	475.10	33.90	140.50	266.00	379.00	482.05
BCor-SIS	4.00	4.00	6.00	10.00	57.10	4.00	5.00	9.00	25.00	83.10
MDC-SIS	52.70	183.50	327.00	413.50	490.05	33.95	147.25	272.00	390.50	480.00
WD_1-SIS	48.75	126.00	234.00	355.50	463.15	276.00	332.75	394.00	445.50	490.10
WD_2-SIS	151.90	326.75	407.00	469.25	494.00	232.75	355.75	424.00	463.75	493.05
DR-WD_1-SIS	4.00	4.00	5.00	5.00	11.05	4.00	4.00	5.00	6.00	10.10
DR-WD_2-SIS	4.00	4.00	5.00	5.00	16.10	4.00	4.00	5.00	5.00	9.00
					$p = 2000$					
SIS	192.75	771.50	1217.00	1647.75	1936.00	194.55	712.25	1233.00	1653.75	1957.35
DC-SIS	161.00	657.00	1023.00	1544.50	1876.15	92.75	515.00	910.00	1420.25	1868.20
bcDC-SIS	105.40	463.00	880.00	1465.00	1887.60	93.50	423.00	858.50	1444.25	1852.60
BCor-SIS	4.00	4.00	5.00	5.00	7.00	4.00	5.00	6.00	7.00	29.15
MDC-SIS	167.75	707.75	1093.50	1588.00	1958.05	89.75	547.25	1097.50	1545.00	1939.05
WD_1-SIS	27.75	169.50	421.00	834.25	1484.25	1065.90	1246.50	1434.50	1718.25	1913.05
WD_2-SIS	823.85	1439.00	1747.50	1881.00	1970.10	1061.55	1401.50	1718.00	1868.00	1971.00
DR-WD_1-SIS	4.00	4.00	4.00	5.00	5.00	4.00	4.75	5.00	6.00	6.00
DR-WD_2-SIS	4.00	4.00	4.00	5.00	5.00	4.00	4.00	4.00	5.00	6.00

b. $X_i^* \sim F_{i \bmod 4}$ independently, for $i = 1, \cdots, p$, where (1) $F_0(x)$ is the standard normal distribution; (2) $F_1(x)$ is the Student's t distribution with degrees of freedom 5; (3) $F_2(x)$ is Pareto distribution with shape parameter $k = 3$, and scale parameter $\alpha = 1$; (4) $F_3(x)$ is Weibull distribution with shape parameter $k = 1.5$, and scale parameter $\lambda = 1$. Then with $\mathbf{X}^* = \left(X_1^*, \cdots, X_p^*\right)^\intercal$, we have $\mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{X}^*$, which represents heterogeneous data.

The results of the above experiments are given in Tables 2.1 – 2.3. Firstly, across the 3 experiments, it is obvious that without Gaussianization, feature screening with Wasserstein Dependence measure (WD_r-SIS)

Table 2.3: The quantiles of minimum model size for Experiment 3 over 200 replications.

	Experiment 3.a					Experiment 3.b				
Quantiles	**5%**	**25%**	**50%**	**75%**	**95%**	**5%**	**25%**	**50%**	**75%**	**95%**
					$p = 500$					
SIS	63.60	214.75	322.00	413.50	484.15	81.00	255.25	344.50	442.00	490.05
DC-SIS	71.85	208.50	330.50	419.50	482.10	68.85	228.25	323.00	427.75	485.05
bcDC-SIS	68.40	222.75	311.00	413.75	485.00	59.75	177.00	310.50	421.00	485.00
BCor-SIS	5.00	8.00	17.00	77.75	251.05	7.00	25.00	73.50	172.00	309.25
MDC-SIS	85.55	218.75	335.00	424.25	486.05	70.45	195.25	330.50	430.75	483.10
WD_1-SIS	141.85	305.25	396.50	463.00	491.05	242.00	373.25	436.00	473.00	494.05
WD_2-SIS	280.65	371.00	430.00	476.25	496.00	264.00	383.25	435.00	470.00	494.05
DR-WD_1-SIS	5.00	5.00	8.50	24.25	155.20	6.00	7.00	10.00	28.50	134.20
DR-WD_2-SIS	5.00	6.00	9.00	34.50	154.70	6.00	6.00	8.00	22.25	157.15
					$p = 2000$					
SIS	265.85	842.75	1330.00	1668.25	1897.25	358.70	1020.25	1556.00	1821.25	1961.15
DC-SIS	248.65	772.75	1216.50	1626.50	1918.00	286.00	829.75	1373.50	1682.50	1945.05
bcDC-SIS	149.25	729.25	1230.50	1616.75	1955.05	220.70	779.50	1252.00	1633.25	1915.50
BCor-SIS	5.00	6.00	10.0	39.00	206.65	7.95	20.75	65.00	177.25	471.00
MDC-SIS	190.70	858.50	1318.50	1658.75	1967.20	278.75	835.50	1294.00	1676.50	1917.35
WD_1-SIS	563.20	1274.50	1667.50	1843.00	1972.10	1072.85	1496.00	1752.50	1881.75	1979.10
WD_2-SIS	1145.35	1529.25	1704.00	1882.25	1972.10	1061.80	1505.25	1716.00	1887.50	1980.05
DR-WD_1-SIS	5.00	5.00	6.00	7.00	13.05	6.00	6.00	6.00	8.00	27.20
DR-WD_2-SIS	5.00	5.00	6.00	6.00	16.10	6.00	6.00	6.00	6.00	8.05

needs hundreds of variables to identify the active variables, which means it could hardly find the active variables regardless of the true model. More specifically, Experiment 1 involves a linear model, where in ideal cases (quantile levels 5% and 25%), most methods (except for WD_r-SIS) could successfully select the 4 active variables with a relatively small amount of features. However, in extreme cases (quantile levels 95%), SIS outperforms all other methods. In Experiments 2 and 3, two nonlinear models are implemented, one is additive and the other one is even more complex. In these cases, only BCor-SIS and DR-WD_r-SIS could succeed in finding the active variables. Comparing subexperiments a & b in nonlinear cases, BCor-SIS is not as robust as our proposal, DR-WD$_r$-SIS. DR-WD$_r$-SIS's performance is not changing much between homogeneous data and heterogeneous data. However, sensitivity to heterogeneous data is actually happening to all of the methods except for DR-WD$_r$-SIS across different models.

2.5 Proofs for Chapter 2

Proposition 2.5.1. *McDiarmid, 1989. Let $(z_1, \cdots, z_n)$ be independent random variables. Suppose a measurable function g satisfies $|g(\mathbf{x}) - g(\tilde{\mathbf{x}})| \leqslant c_i$, where $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^n$ only differ on the ith coordinate. Then for any $\varepsilon > 0$, we have*

$$\mathcal{P}\left(|g(z_1, \cdots, z_n) - \mathbb{E}g(z_1, \cdots, z_n)| \geqslant \varepsilon\right) \leqslant 2\exp\left(-2\frac{\varepsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

Proposition 2.5.2. *Lemmas 12.1 and 12.3 from Abramovich et al., 2006. For $z \geqslant 1$ and $0 < \eta \leqslant 0.01$, we have*

$$\frac{\phi(z)}{2z} \leqslant 1 - \Phi(z) \leqslant \frac{\phi(z)}{z},$$
$$\Phi^{-1}(1-\eta) \leqslant \sqrt{2\log\eta^{-1}}.$$

Proposition 2.5.3. *Proposition 3 from Mai et al., 2023. For any random sample ξ_i, $i = 1, \cdots, n$, and its empirical CDF $\widehat{F}_\xi(\cdot)$, there exists some positive constants C, and n_0, which are not related to ξ or n, s.t. when $n \geqslant n_0$, $\forall t \in \mathbb{R}$,*

$$\Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_\xi(t)\right) \leqslant C\sqrt{\log n}.$$

Lemma 2.5.1. *Corrected Lemma 5 from Mai et al., 2023. For i.i.d. standard Gaussian sample* $z_i \overset{i.i.d.}{\sim} N(0,1)$*, and* $z_i^* \doteq z_i \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}} + \operatorname{sign}(z_i)\sqrt{2\log n}\,\mathbb{1}_{\left\{|z_i| > \sqrt{2\log n}\right\}}$, $i = 1, \cdots, n$*, there exists some positive constants* M*,* C*,* n_0*, and* ε_0*, which are not related to* n*, s.t. when* $n \geqslant n_0$*,* $\forall \varepsilon \in (M\frac{1}{n}, \varepsilon_0)$*,*

$$\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n} |z_i^* - z_i| \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log n}\right).$$

Proof. Let $g(\mathbf{x}) \doteq \frac{1}{n}\sum_{i=1}^{n} |x_i| \mathbb{1}_{\left\{|x_i| \leqslant \sqrt{2\log n}\right\}}$, where $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. Consider $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^n$, which only differ on the ith coordinate. W.L.O.G., we may assume $\mathbf{x} = (x_1, \cdots, x_n)$, and $\tilde{\mathbf{x}} = (\tilde{x}_1, \cdots, x_n)$. Then we have

$$|g(\mathbf{x}) - g(\tilde{\mathbf{x}})| = \left|\frac{1}{n}|x_1| \mathbb{1}_{\left\{|x_1| \leqslant \sqrt{2\log n}\right\}} - \frac{1}{n}|\tilde{x}_1| \mathbb{1}_{\left\{|\tilde{x}_1| \leqslant \sqrt{2\log n}\right\}}\right| \leqslant C\frac{\sqrt{\log n}}{n},$$

for some positive constant C, which is not related to n. Since $z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, applying Proposition 2.5.1, we have

$$\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} |z_i| \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}} - \mathbb{E}\left[|z_i| \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right]\right| \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log n}\right), \qquad (2.5.1)$$

for any $\varepsilon > 0$ and some positive constant C, which is not related to n.

We also observe the truth that

$$\mathbb{E}\left[|z_i| \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right] = \int_{-\sqrt{2\log n}}^{\sqrt{2\log n}} |x| \phi(x) dx = \sqrt{\frac{2}{\pi}} - \sqrt{\frac{2}{\pi}}\frac{1}{n}, \qquad (2.5.2)$$

when $n \geqslant 3$. Then, following (2.5.1) and (2.5.2), and assuming that $\varepsilon > \frac{2}{n}\sqrt{\frac{2}{\pi}}$, we have

$$
\begin{aligned}
&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}-\sqrt{\frac{2}{\pi}}\right|\geqslant\varepsilon\right)\\
=&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}-\mathbb{E}\left[|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}\right]-\sqrt{\frac{2}{\pi}}\frac{1}{n}\right|\geqslant\varepsilon\right)\\
\leqslant&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}-\mathbb{E}\left[|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}\right]\right|+\sqrt{\frac{2}{\pi}}\frac{1}{n}\geqslant\varepsilon\right)\\
\leqslant&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}-\mathbb{E}\left[|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}\right]\right|\geqslant\frac{\varepsilon}{2}\right)\\
\leqslant&C\exp\left(-\frac{Cn\varepsilon^2}{\log n}\right).
\end{aligned}
\tag{2.5.3}
$$

Finally, we have

$$
\begin{aligned}
\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}|z_i^*-z_i|\geqslant\varepsilon\right)=&\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(|z_i|-\sqrt{2\log n}\right)\mathbb{1}_{\left\{|z_i|>\sqrt{2\log n}\right\}}\geqslant\varepsilon\right)\\
\leqslant&\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|>\sqrt{2\log n}\right\}}\geqslant\varepsilon\right)\\
=&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|-\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}\right|\geqslant\varepsilon\right)\\
\leqslant&\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|-\sqrt{\frac{2}{\pi}}\right|\geqslant\frac{\varepsilon}{2}\right) \qquad (2.5.4)\\
&+\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|z_i|\,\mathbb{1}_{\left\{|z_i|\leqslant\sqrt{2\log n}\right\}}-\sqrt{\frac{2}{\pi}}\right|\geqslant\frac{\varepsilon}{2}\right), \qquad (2.5.5)\\
\leqslant&C\exp\left(-\frac{Cn\varepsilon^2}{\log n}\right),
\end{aligned}
$$

where (2.5.4) follows Lemma 2 from Mai et al., 2023 with $\varepsilon \in (0, \varepsilon_0)$ for some positive constant ε_0 not related to n, and (2.5.5) follows from (2.5.3). □

Lemma 2.5.2. *(Modified Lemma 12 from Mai et al., 2023) With the same condition in Lemma 2.5.1, there exists some positive constants M, C, n_0, and ε_0, which are all not related to n, s.t. when $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{\sqrt{\log n}}{n}, \varepsilon_0)$,*

$$\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(z_i^* - z_i\right)^2 \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right).$$

Proof. The proof is similar to Lemma 2.5.1. Let $g\left(\mathbf{x}\right) \doteq \frac{1}{n}\sum_{i=1}^{n} x_i^2 \mathbb{1}_{\left\{|x_i| \leqslant \sqrt{2\log n}\right\}}$, $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. Consider $\mathbf{x}, \tilde{\mathbf{x}} \in \mathbb{R}^n$, which only differ on the ith coordinate. W.L.O.G., we may assume $\mathbf{x} = (x_1, \cdots, x_n)$, and $\tilde{\mathbf{x}} = (\tilde{x}_1, \cdots, x_n)$. Then we have

$$\left|g\left(\mathbf{x}\right) - g\left(\tilde{\mathbf{x}}\right)\right| = \left|\frac{1}{n}x_1^2\mathbb{1}_{\left\{|x_1| \leqslant \sqrt{2\log n}\right\}} - \frac{1}{n}\tilde{x}_1^2\mathbb{1}_{\left\{|\tilde{x}_1| \leqslant \sqrt{2\log n}\right\}}\right| \leqslant C\frac{\log n}{n},$$

for some positive constant C, which is not related to n. Since $z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, applying Proposition 2.5.1, we have

$$\mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2\mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}} - \mathbb{E}\left[z_i^2\mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right]\right| \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right), \tag{2.5.6}$$

for some positive constant C, which is not related to n.

We have the similar result as (2.5.1),

$$\mathbb{E}\left[z_i^2\mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right] = \int_{-\sqrt{2\log n}}^{\sqrt{2\log n}} x^2\phi(x)dx = 1 - \frac{2}{\sqrt{\pi}}\frac{\sqrt{\log n}}{n} - 2\Phi\left(-\sqrt{2\log n}\right). \tag{2.5.7}$$

When $n \geqslant 100$, by Proposition 2.5.2, we have

$$
\begin{aligned}
\Phi\left(\sqrt{2\log n}\right) \geqslant & 1-\frac{1}{n}, \\
\Phi\left(-\sqrt{2\log n}\right) = 1-\Phi\left(\sqrt{2\log n}\right) \leqslant & \frac{1}{n}, \\
1 \geqslant \mathbb{E}\left[z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right] \geqslant & 1-\frac{2}{\sqrt{\pi}}\frac{\sqrt{\log n}}{n}-\frac{2}{n} \geqslant 1-\left(\frac{2}{\sqrt{\pi}}+2\right)\frac{\sqrt{\log n}}{n}.
\end{aligned}
$$

Then, following that (2.5.6), and (2.5.7), and assuming that $M \geqslant 2\left(\frac{2}{\sqrt{\pi}}+2\right)$, and $\varepsilon > M\frac{\sqrt{\log n}}{n}$, we have

$$
\begin{aligned}
& \mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}-1\right| \geqslant \varepsilon\right) \\
\leqslant & \mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}-\mathbb{E}\left[z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right]\right| \geqslant \frac{\varepsilon}{2}\right) \\
\leqslant & C\exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right).
\end{aligned}
\tag{2.5.8}
$$

Finally, we have

$$
\begin{aligned}
\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(z_i^*-z_i\right)^2 \geqslant \varepsilon\right) = & \mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}\left(|z_i|-\sqrt{2\log n}\right)^2 \mathbb{1}_{\left\{|z_i| > \sqrt{2\log n}\right\}} \geqslant \varepsilon\right) \\
\leqslant & \mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n} z_i^2 \mathbb{1}_{\left\{|z_i| > \sqrt{2\log n}\right\}} \geqslant \varepsilon\right) \\
= & \mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2-\frac{1}{n}\sum_{i=1}^{n} z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}\right| \geqslant \varepsilon\right) \\
\leqslant & \mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2-1\right| \geqslant \frac{\varepsilon}{2}\right) \qquad (2.5.9) \\
& + \mathcal{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} z_i^2 \mathbb{1}_{\left\{|z_i| \leqslant \sqrt{2\log n}\right\}}-1\right| \geqslant \frac{\varepsilon}{2}\right), \qquad (2.5.10)
\end{aligned}
$$

$$\leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right),$$

where (2.5.9) follows from Lemma 2 from Mai et al., 2023 with $\varepsilon \in (0, \varepsilon_0)$ for some positive constant ε_0 not related to n, and (2.5.10) follows from (2.5.8). □

Theorem 2.5.4. *(Modified Theorem 1 from Mai et al., 2023) For any continuous random sample* $\xi_i \overset{i.i.d.}{\sim} F_\xi(\cdot)$, $i = 1, \cdots, n$, *and its empirical CDF* $\widehat{F}_\xi(\cdot)$, *consider* $z_i \doteq \Phi^{-1}(F_\xi(\xi_i))$, $\widehat{z}_i \doteq \Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_\xi(\xi_i)\right)$, $i = 1, \cdots, n$. *There exists some positive constants* M, C, n_0, *and* ε_0, *which are not related to* ξ *or* n, *s.t. when* $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{\log n}{\sqrt{n}}, \varepsilon_0)$,

$$\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}|\widehat{z}_i - z_i| \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log n}\right).$$

Proof. Theorem 2.5.4 follows directly from Lemma 8 in Mai et al., 2023 and Lemma 2.5.1, noting that $z_i \overset{\text{i.i.d.}}{\sim} N(0,1), i = 1, \ldots, n$. □

Theorem 2.5.5. *(Modified Lemma 14 from Mai et al., 2023) With the same conditions in Theorem 2.5.4, there exists some positive constants* M, C, n_0, *and* ε_0, *which are not related to* ξ *or* n, *s.t. when* $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{(\log n)^{\frac{3}{2}}}{\sqrt{n}}, \varepsilon_0)$,

$$\mathcal{P}\left(\frac{1}{n}\sum_{i=1}^{n}(\widehat{z}_i - z_i)^2 \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right).$$

Proof. Firstly, it is easy to observe that $z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$. With the same definition of z_i^*, $i = 1, \cdots, n$, in Lemma 2.5.1, when $n \geqslant n_0$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{z}_i - z_i\right)^2 &\leqslant \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{z}_i - z_i\right|\left(\left|\widehat{z}_i - z_i^*\right| + \left|z_i^* - z_i\right|\right) \\
&\leqslant C\frac{\sqrt{\log n}}{n}\sum_{i=1}^{n}\left|\widehat{z}_i - z_i\right| + \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{z}_i - z_i\right|\left|z_i^* - z_i\right| \qquad (2.5.11) \\
&\leqslant C\frac{\sqrt{\log n}}{n}\sum_{i=1}^{n}\left|\widehat{z}_i - z_i\right| + \frac{1}{2n}\sum_{i=1}^{n}\left(\widehat{z}_i - z_i\right)^2 + \frac{1}{2n}\sum_{i=1}^{n}\left(z_i^* - z_i\right)^2,
\end{aligned}$$

for some positive constants C, and n_0, which are not related to ξ or n. Here (2.5.11) follows from Proposition 2.5.3. Hence, we have

$$\frac{1}{2n}\sum_{i=1}^{n}\left(\widehat{z}_i - z_i\right)^2 \leqslant C\frac{\sqrt{\log n}}{n}\sum_{i=1}^{n}\left|\widehat{z}_i - z_i\right| + \frac{1}{n}\sum_{i=1}^{n}\left(z_i^* - z_i\right)^2,$$

and the rest proof follows directly from Lemma 2.5.2 and Theorem 2.5.4. □

Proof of Theorem 2.2.1. Let $\widehat{f}_{S,T}$ be the empirical measure on the genuine Gaussianized sample (2.2.2). Then for $r \geqslant 1$, by triangular inequality and property of distance, it is easy to have

$$\begin{aligned}
&\left|\widehat{I}_{\mathcal{W}}\left(X, Y; r\right) - I_{\mathcal{W}}\left(X, Y; r\right)\right| \\
=&\left|\mathcal{W}_r(\widehat{f}_{\widehat{S},\widehat{T}}, \phi\otimes\phi) - \mathcal{W}_r(f_{S,T}, \phi\otimes\phi)\right| \\
\leqslant&\mathcal{W}_r(\widehat{f}_{\widehat{S},\widehat{T}}, f_{S,T}) \\
\leqslant&\mathcal{W}_r(\widehat{f}_{\widehat{S},\widehat{T}}, \widehat{f}_{S,T}) + \mathcal{W}_r(\widehat{f}_{S,T}, f_{S,T}), \qquad (2.5.12)
\end{aligned}$$

Then we consider the two terms separately.

1. For $\mathcal{W}_r(\widehat{f}_{\widehat{S},\widehat{T}}, \widehat{f}_{S,T})$, with $\mathbf{\Pi}$ as the set of all permutations on $\{1, \cdots, n\}$, we have

$$
\begin{aligned}
\mathcal{W}_r^r(\widehat{f}_{\widehat{S},\widehat{T}}, \widehat{f}_{S,T}) = \inf_{\boldsymbol{\pi}\in\mathbf{\Pi}} &\left\{ \frac{1}{n}\sum_{i=1}^{n} \left\| \begin{pmatrix} s_i \\ t_i \end{pmatrix} - \begin{pmatrix} \widehat{s}_{\boldsymbol{\pi}(i)} \\ \widehat{t}_{\boldsymbol{\pi}(i)} \end{pmatrix} \right\|_q^r \right\} \\
\leqslant & \frac{1}{n}\sum_{i=1}^{n} \left\| \begin{pmatrix} s_i \\ t_i \end{pmatrix} - \begin{pmatrix} \widehat{s}_i \\ \widehat{t}_i \end{pmatrix} \right\|_q^r \\
\leqslant & \frac{1}{n}\sum_{i=1}^{n} \left[|s_i - \widehat{s}_i| + \left|t_i - \widehat{t}_i\right| \right]^r \qquad (2.5.13) \\
\leqslant & C_r \left[\frac{1}{n}\sum_{i=1}^{n} |s_i - \widehat{s}_i|^r + \frac{1}{n}\sum_{i=1}^{n} \left|t_i - \widehat{t}_i\right|^r \right], \qquad (2.5.14)
\end{aligned}
$$

where (2.5.13) follows from $\|x\|_q \leqslant \|x\|_1$, for $q \geqslant 1$ and $x \in \mathbb{R}^d$, and (2.5.14) follows from $|a+b|^r \leqslant C_r (|a|^r + |b|^r)$ for $C_r = 2^{(r-1)_+}$ and $r \geqslant 0$.

With $r = 1$, by Theorem 2.5.4, there exists some positive constants M, C, n_0, and ε_0, which are not related to margins of (X, Y) or n, s.t. when $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{\log n}{\sqrt{n}}, \varepsilon_0)$,

$$
\mathcal{P}\left(\mathcal{W}_1(\widehat{f}_{\widehat{S},\widehat{T}}, \widehat{f}_{S,T}) \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^2}{\log n}\right). \qquad (2.5.15)
$$

With $r = 2$, by Theorem 2.5.5, there exists some positive constants M, C, n_0, and ε_0, which are not related to margins of (X, Y) or n, s.t. when $n \geqslant n_0$, $\forall \varepsilon \in (M\frac{(\log n)^{\frac{3}{4}}}{n^{\frac{1}{4}}}, \varepsilon_0)$,

$$
\mathcal{P}\left(\mathcal{W}_2(\widehat{f}_{\widehat{S},\widehat{T}}, \widehat{f}_{S,T}) \geqslant \varepsilon\right) \leqslant C\exp\left(-\frac{Cn\varepsilon^4}{\log^2 n}\right). \qquad (2.5.16)
$$

2. For $\mathcal{W}_r(\widehat{f}_{S,T}, f_{S,T})$, we mainly refer to the proof of Theorem 2 in Fournier and Guillin, 2015. Since margins of (S, T) are both standard Gaussian, it is easy to find the conditions in Fournier and Guillin, 2015 could always be satisfied.

   With $r = 1$, Condition (1) in Theorem 2 in Fournier and Guillin, 2015 is satisfied with $\alpha = 2$ and any $\gamma < \frac{1}{8}$. Correspondingly, there exist some positive constant C, which is not related to margins of (X, Y) or n, s.t. when $n \geqslant 1, \forall \varepsilon \in (0, 1)$,

$$\mathcal{P}\left(\mathcal{W}_1(\widehat{f}_{S,T}, f_{S,T}) \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log^2\left(2+\frac{1}{\varepsilon}\right)}\right). \tag{2.5.17}$$

   With $r = 2$, Condition (2) in Theorem 2 in Fournier and Guillin, 2015 is satisfied with $\forall \alpha \in (0, 2)$ and any $\gamma > 0$. Correspondingly, $\forall \alpha \in (0, 2)$, and $\forall \delta \in (0, \alpha)$, there exist some positive constant C, which is not related to margins of (X, Y) or n, s.t. when $n \geqslant 1, \forall \varepsilon \in (0, 1)$,

$$\mathcal{P}\left(\mathcal{W}_2(\widehat{f}_{S,T}, f_{S,T}) \geqslant \varepsilon\right) \leqslant C \exp\left(-Cn\varepsilon^4\right) + C \exp\left(-C\left(n\varepsilon^2\right)^{\frac{\alpha-\delta}{2}}\right). \tag{2.5.18}$$

Finally, when $r = 1$, and $\varepsilon > M\frac{\log n}{\sqrt{n}}$, it is easy to find $\log^2\left(2+\frac{1}{\varepsilon}\right) \leqslant C\log^2 n$ for some positive constant C. Hence combining (2.5.15), (2.5.17), and (2.5.12), it induces

$$\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}(X, Y; 1) - I_{\mathcal{W}}(X, Y; 1)\right| \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^2}{\log^2 n}\right).$$

When $r = 2$, combining (2.5.16), (2.5.18), and (2.5.12), it induces $\forall \beta = \frac{\alpha-\delta}{2} \in (0, 1)$,

$$\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}(X, Y; 2) - I_{\mathcal{W}}(X, Y; 2)\right| \geqslant \varepsilon\right) \leqslant C \exp\left(-\frac{Cn\varepsilon^4}{\log^2 n}\right) + C \exp\left(-C\left(n\varepsilon^2\right)^{\beta}\right).$$

The second statement follows directly from the first statement. □

Proof of Theorem 2.3.1. Under Condition 2.1(a), we can observe that

$$t_n \leqslant c_t^{\frac{1}{r}} n^{-\frac{\kappa_1}{r}} < c_1^{\frac{1}{r}} n^{-\frac{\kappa_1}{r}} \leqslant \min_{j\in\mathcal{A}} \{I_{\mathcal{W}}(X_j, Y; r)\}.$$

Then we have

$$\begin{aligned} \mathcal{P}\left(\mathcal{A} \nsubseteq \widehat{\mathcal{A}}(t_n)\right) =& \mathcal{P}\left(\bigcup_{j\in\mathcal{A}} \left\{ j \notin \widehat{\mathcal{A}}(t_n)\right\}\right) \\ \leqslant& \sum_{j\in\mathcal{A}} \mathcal{P}\left(\widehat{I}_{\mathcal{W}}(X_j, Y; r) \leqslant t_n\right) \\ \leqslant& \sum_{j\in\mathcal{A}} \mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}(X_j, Y; r) - I_{\mathcal{W}}(X_j, Y; r)\right| \geqslant I_{\mathcal{W}}(X_j, Y; r) - t_n\right) \\ \leqslant& \sum_{j\in\mathcal{A}} \mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}(X_j, Y; r) - I_{\mathcal{W}}(X_j, Y; r)\right| \geqslant (c_1^{\frac{1}{r}} - c_t^{\frac{1}{r}}) n^{-\frac{\kappa_1}{r}}\right). \end{aligned}$$

Then there exist some positive constants C_1, and n_1, which are not related to margins of $(\mathbf{X}, Y)$ or n. Let $n \geqslant n_1$, by Theorem 2.2.1, and consider the complementary set,

1. when $r = 1$, it follows that

$$\mathcal{P}\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}(t_n)\right) \geqslant 1 - C_1 \mathcal{S}_{\mathcal{A}} \exp\left(-\frac{C_1 n^{1-2\kappa_1}}{\log^2 n}\right).$$

2. when $r = 2, \forall \beta \in (0, 1)$, it follows that

$$\mathcal{P}\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}(t_n)\right) \geqslant 1 - \mathcal{S}_{\mathcal{A}} \left[C_1 \exp\left(-\frac{C_1 n^{1-2\kappa_1}}{\log^2 n}\right) + C_1 \exp\left(-C_1 n^{\beta(1-\kappa_1)}\right)\right].$$

□

Proof of Theorem 2.3.2. Under Condition 2.1(b), we have

$$
\begin{aligned}
&\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}\leqslant 0\right)\\
\leqslant&\mathcal{P}\Bigg(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}\\
\leqslant&\min_{j\in\mathcal{A}}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-c_2n^{-\kappa_2}\Bigg)\\
=&\mathcal{P}\Bigg(\left[\min_{j\in\mathcal{A}}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}\right]\\
&+\left[\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\}\right]\geqslant c_2n^{-\kappa_2}\Bigg).
\end{aligned}
$$

Let $j_1\doteq\operatorname*{argmin}_{j\in\mathcal{A}}\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)$ and $j_2\doteq\operatorname*{argmax}_{j\in\mathcal{A}^c}\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)$, then we could have

$$
\begin{aligned}
I_{\mathcal{W}}\left(X_{j_1},Y;r\right)&\geqslant\min_{j\in\mathcal{A}}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\},\\
I_{\mathcal{W}}\left(X_{j_2},Y;r\right)&\leqslant\max_{j\in\mathcal{A}^c}\left\{I_{\mathcal{W}}\left(X_j,Y;r\right)\right\}.
\end{aligned}
$$

It induces

$$
\begin{aligned}
&\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)\right\}\leqslant 0\right)\\
\leqslant&\mathcal{P}\left(\left[I_{\mathcal{W}}\left(X_{j_1},Y;r\right)-\widehat{I}_{\mathcal{W}}\left(X_{j_1},Y;r\right)\right]+\left[\widehat{I}_{\mathcal{W}}\left(X_{j_2},Y;r\right)-I_{\mathcal{W}}\left(X_{j_2},Y;r\right)\right]\geqslant c_2n^{-\kappa_2}\right)\\
\leqslant&\mathcal{P}\left(\left\{\left|\widehat{I}_{\mathcal{W}}\left(X_{j_1},Y;r\right)-I_{\mathcal{W}}\left(X_{j_1},Y;r\right)\right|\geqslant\frac{1}{2}c_2n^{-\kappa_2}\right\}\right)\\
&+\mathcal{P}\left(\left\{\left|\widehat{I}_{\mathcal{W}}\left(X_{j_2},Y;r\right)-I_{\mathcal{W}}\left(X_{j_2},Y;r\right)\right|\geqslant\frac{1}{2}c_2n^{-\kappa_2}\right\}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant 2\mathcal{P}\left(\max_{j\in\{1,\cdots,p\}}\left\{\left|\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)-I_{\mathcal{W}}\left(X_j,Y;r\right)\right|\right\}\geqslant\frac{1}{2}c_2n^{-\kappa_2}\right)\\
&=2\mathcal{P}\left(\bigcup_{j=1}^{p}\left\{\left|\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)-I_{\mathcal{W}}\left(X_j,Y;r\right)\right|\geqslant\frac{1}{2}c_2n^{-\kappa_2}\right\}\right)\\
&\leqslant 2\sum_{j=1}^{p}\mathcal{P}\left(\left|\widehat{I}_{\mathcal{W}}\left(X_j,Y;r\right)-I_{\mathcal{W}}\left(X_j,Y;r\right)\right|\geqslant\frac{1}{2}c_2n^{-\kappa_2}\right).
\end{aligned}
$$

Then by Theorem 2.2.1, there exists some positive constants n_2 and C_2, which are not related to margins of $(\mathbf{X}, Y)$ or n, s.t. when $n \geqslant n_2$,

1. with $r = 1$,

$$
\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;1\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;1\right)\right\}>0\right)\geqslant 1-2C_2p\exp\left(-\frac{C_2n^{1-2\kappa_2}}{\log^2 n}\right),
$$

2. with $r = 2, \forall\beta \in (0, 1)$,

$$
\begin{aligned}
&\mathcal{P}\left(\min_{j\in\mathcal{A}}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;2\right)\right\}-\max_{j\in\mathcal{A}^c}\left\{\widehat{I}_{\mathcal{W}}\left(X_j,Y;2\right)\right\}>0\right)\\
\geqslant&1-2C_2p\left[\exp\left(-\frac{C_2n^{1-4\kappa_2}}{\log^2 n}\right)+\exp\left(-C_2n^{\beta(1-2\kappa_2)}\right)\right],
\end{aligned}
$$

The rest of the proof follows directly from the proof of Theorem 3 in W. Liu et al., 2022. □

Chapter 3

Model-free Feature Screening and False Discovery Control for High-dimensional Quantile Regressions

In this chapter, we establish a novel model-free feature screening and corresponding FDR control method for high-dimensional quantile regressions. In Section 3.1, we introduce a Copula Quantile Dependence as a novel measure of quantile dependence. In Sections 3.2 – 3.4, we discuss the method to estimate Copula Quantile Dependence through the estimation of Copula. We also prove the asymptotic properties of the estimations. In Section 3.5, we propose a model-free feature screening procedure for quantile regression and establish its sure screening and rank consistency properties under mild conditions. In Section 3.6, we develop a data-driven threshold selection method to control the false discovery rate for the proposed screening procedure. In Section 3.7, we use simulation experiments to assess the empirical performance of the proposed methods.

3.1 Quantile Dependence and Conditional Copula Density

Let X and Y be two continuous random variables with marginal cumulative distribution functions (CDFs): $U \doteq F_X(X)$ and $V \doteq F_Y(Y)$, respectively. By the probability integral transform, U and V are uniformly distributed over $[0, 1]$. According to Sklar's theorem (Sklar, 1959), the joint probability distribution of X and Y can be represented by

$$F_{X,Y}(x, y) = \mathcal{P}(X \leqslant x, Y \leqslant y) = \mathcal{P}(U \leqslant F_X(x), V \leqslant F_Y(y)) \doteq \mathcal{C}\left(u, v\right),$$

where $u \doteq F_X(x)$, $v \doteq F_Y(y)$, and $\mathcal{C}(u, v) \doteq \mathcal{P}(U \leqslant u, V \leqslant v)$ is the copula function of X and Y.

Suppose that the joint density of X and Y exists and that the copula function $\mathcal{C}(\cdot, \cdot)$ is differentiable. A classical result in probability theory suggests that the conditional probability of V given $U = u$ can be calculated through the first-order derivative of the copula function with respect to u. To be specific, we have

$$\mathcal{P}(V \leqslant v \mid U = u) = \frac{\partial}{\partial u}\mathcal{C}(u, v) \doteq \mathcal{C}_{2|1}(v \mid u).$$

In the rest of the paper, we call the derivative $\mathcal{C}_{2|1}(v \mid u)$ the conditional copula function of V given U. We can define $\mathcal{C}_{1|2}(u \mid v)$ in a similar fashion. Denote $F_Y^{-1}(\cdot)$ as the (generalized) inverse function of $F_Y(\cdot)$ and $C_{2|1}^{-1}(\cdot \mid u)$ as the (generalized) partial inverse function of $\mathcal{C}_{2|1}(\cdot \mid u) : v \mapsto \mathcal{C}_{2|1}\left(v \mid u\right)$. The conditional α-th quantile of Y given $X = x$ with $\alpha \in (0, 1)$, i.e. $Q_\alpha(Y|X = x)$, can be expressed by

$$Q_\alpha(Y|X = x) = F_Y^{-1}\left(\mathcal{C}_{2|1}^{-1}\left(\alpha \mid F_X(x)\right)\right). \tag{3.1.1}$$

We say Y is quantile independent of X at the quantile level α if and only if the conditional quantile $Q_\alpha(Y|X = x)$ equals the marginal α-th quantile of Y (i.e. $F_Y^{-1}(\alpha)$) almost surely with respect to the measure of X. This is equivalent to the condition that $\mathcal{C}_{2|1}\left(\alpha \mid u\right) \equiv \alpha$ for almost every $u \in [0, 1]$ under some mild conditions to be specified below. This observation motivates us to measure the quantile dependence between X and Y through an integrated (weighted) square deviation.

Definition 3.1.1 (Copula Quantile Dependence)**.** *For continuous random variables X and Y, we define their quantile dependence at the quantile level $\alpha \in (0, 1)$ as*

$$D_\alpha(X, Y; \omega) = \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right]^2 \omega\left(u\right) du, \tag{3.1.2}$$

where $\omega(\cdot)$ is an almost everywhere positive integrable weight function.

Next, we introduce two conditions and a lemma to study the properties of the quantile dependence defined above.

Condition 3.1. *The joint density of X and Y exists and the inverse of $F_Y(\cdot)$ exists on the domain of Y.*

Condition 3.2. *The copula function $\mathcal{C}(\cdot, \cdot)$ is differentiable. The partial inverse of $\mathcal{C}_{2|1}(\cdot \mid u)$ exists for almost every $u \in (0, 1)$.*

Lemma 3.1.1. *Suppose Conditions 3.1 and 3.2 hold. The following properties hold $\forall\alpha \in (0, 1)$.*

(a) $D_\alpha(X, Y; \omega) \geqslant 0$*.*

(b) $D_\alpha(X, Y; \omega) = 0$ *if and only if* Y *is quantile independent of* X *at the quantile level* α*.*

(c) $\mathcal{C}_{2|1}(\alpha \mid u) \equiv \alpha$ *for almost every* $u \in [0, 1]$ *if and only if* Y *is quantile independent of* X *at the quantile level* α*.*

Lemma 3.1.1 introduces the nice statistical properties of Copula Quantile Dependence as a measure of the association between X and Y at a given quantile level α. This measure is always nonnegative. Moreover, if X and Y are independent at quantile level α, then the Copula Quantile Dependence is exactly zero. Hence, the magnitude of this metric naturally serves as a statistic to evaluate the quantile dependence. In addition, Condition 3.1 and 3.2 are mild, since they hold for most commonly used distributions and parametric copulas.

3.2 Estimation of Copula Density

Accurate copula density estimation is essential for constructing the Copula Quantile Dependence, as defined in (3.1.2). Traditional parametric approaches (e.g., Gaussian or Clayton copulas) tend to be rigid and susceptible to misspecification. In contrast, non-parametric methods offer flexibility but often struggle with boundary bias and inconsistencies when densities are unbounded (Chen and Huang, 2007). To overcome these limitations, we propose estimating the copula density using a Probit transformation. This transformation idea has been explored by existing literature (Segers, 2012; Geenens, 2014; Geenens et al., 2017).

Let $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and density of the standard Gaussian distribution. We apply the Probit transformation to U and V by defining

$$S = \Phi^{-1}(U) \quad \text{and} \quad T = \Phi^{-1}(V).$$

Because copulas are invariant under monotonic transformations, the joint CDF of (S, T) is

$$F_{S,T}(s,t) = \mathcal{P}(S \leqslant s, T \leqslant t) = \mathcal{C}(\Phi(s), \Phi(t)), \quad \forall (s,t) \in \mathbb{R}^2.$$

Differentiating $F_{S,T}(s,t)$ with respect to s and t yields the joint density of S and T, i.e.

$$f_{S,T}(s,t) = \frac{\partial^2}{\partial u \partial v}\mathcal{C}(\Phi(s), \Phi(t))\phi(s)\phi(t).$$

Therefore, the copula density satisfies

$$\mathcal{C}_{1,2}(u,v) \doteq \frac{\partial^2}{\partial u \partial v}\mathcal{C}(u,v) = \frac{f_{S,T}\big(\Phi^{-1}(u), \Phi^{-1}(v)\big)}{\phi\big(\Phi^{-1}(u)\big)\,\phi\big(\Phi^{-1}(v)\big)}. \tag{3.2.1}$$

Equation (3.2.1) offers a practical way to estimate the copula density. Let $\{(x_i, y_i)\}_{i=1}^n$ be a sample drawn from (X, Y). We estimate the marginal CDFs of X and Y by

$$\widehat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i \leqslant x\}} \quad \text{and} \quad \widehat{F}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{y_i \leqslant y\}}.$$

We then apply the Probit transformation to obtain the sample

$$(\widehat{s}_i, \widehat{t}_i) \doteq \left(\Phi^{-1}\left(\widehat{u}_i\right), \Phi^{-1}\left(\widehat{v}_i\right)\right) = \left(\Phi^{-1}\left(\tfrac{n}{n+1}\widehat{F}_X(x_i)\right), \Phi^{-1}\left(\tfrac{n}{n+1}\widehat{F}_Y(y_i)\right)\right). \tag{3.2.2}$$

Using this transformed sample, we estimate $f_{S,T}(s,t)$ with a kernel density estimator:

$$\widehat{f}_{S,T}(s,t) = \frac{1}{n|\mathbf{H}|}\sum_{i=1}^{n}\mathbf{K}\left(\mathbf{H}^{-1/2}\begin{pmatrix} s-\widehat{s}_i \\ t-\widehat{t}_i \end{pmatrix}\right),$$

where $\mathbf{K}(\cdot,\cdot)$ is a kernel function, and $\mathbf{H}$ is a bandwidth matrix.

Choosing $\mathbf{K}(s,t) = \phi(s)\,\phi(t)$ as a product Gaussian kernel and $\mathbf{H} = h^2\mathbf{I}$ for some bandwidth $h > 0$, we substitute $\widehat{f}_{S,T}(\cdot,\cdot)$ into (3.2.1) to estimate the copula density by

$$\widehat{\mathcal{C}}_{1,2}(u,v) = \frac{1}{nh^2\phi\left(s\right)\phi\left(t\right)}\sum_{i=1}^{n}\phi\left(\frac{s-\widehat{s}_i}{h}\right)\phi\left(\frac{t-\widehat{t}_i}{h}\right), \tag{3.2.3}$$

where $s = \Phi^{-1}(u)$ and $t = \Phi^{-1}(v)$. The asymptotic normality of $\widehat{\mathcal{C}}_{1,2}(u,v)$, along with other properties, was studied in Geenens et al., 2017.

The proposed transformation from (X,Y) to (S,T) can be viewed through the lens of optimal transport. Consider the p-Wasserstein distance

$$\mathcal{W}_p(\mu,\rho) = \inf_{\gamma\in\Gamma(\mu,\rho)}\left(\int_{\mathbb{R}^2\times\mathbb{R}^2}\|\mathbf{x}-\mathbf{y}\|_p^p\,d\gamma(\mathbf{x},\mathbf{y})\right)^{\frac{1}{p}},$$

where μ is the joint distribution of (X,Y), ρ is the joint distribution of (S,T), and $\Gamma(\mu,\rho)$ is the collection of all couplings of μ and ρ. Among all transformations from μ to ρ that preserve the underlying copula, the marginal transformation

$$(S,T) = \left(\Phi^{-1}(F_X(X)), \Phi^{-1}(F_Y(Y))\right)$$

minimizes the Wasserstein distance. For further details, see Alfonsi and Jourdain, 2014.

3.3 Estimation of Conditional Copula

Using the copula density estimator introduced in (3.2.3), we propose the following estimator for the conditional copula function:

$$\begin{aligned}
\widehat{\mathcal{C}}_{2|1}(v \mid u) \doteq & \int_0^v \widehat{\mathcal{C}}_{1,2}(u,v')dv' \\
= & \frac{1}{nh^2\phi\left[\Phi^{-1}(u)\right]} \sum_{i=1}^{n} \phi\left(\frac{\Phi^{-1}(u) - \widehat{s}_i}{h}\right) \int_0^v \frac{\phi\left(\frac{\Phi^{-1}(v')-\widehat{t}_i}{h}\right)}{\phi\left[\Phi^{-1}(v')\right]} dv' \\
= & \frac{1}{nh^2\phi\left[\Phi^{-1}(u)\right]} \sum_{i=1}^{n} \phi\left(\frac{\Phi^{-1}(u) - \widehat{s}_i}{h}\right) \int_{-\infty}^{\Phi^{-1}(v)} \phi\left(\frac{v'' - \widehat{t}_i}{h}\right) dv'' \\
= & \frac{1}{nh\phi\left(\Phi^{-1}(u)\right)} \sum_{i=1}^{n} \phi\left(\frac{\Phi^{-1}(u) - \widehat{s}_i}{h}\right) \Phi\left(\frac{\Phi^{-1}(v) - \widehat{t}_i}{h}\right). \qquad (3.3.1)
\end{aligned}$$

In Theorem 3.3.1, we establish the asymptotic normality of this conditional copula estimator under mild conditions from Geenens et al., 2017 and Segers, 2012. These conditions ensure both the existence and uniqueness of the copula, and they hold for many commonly used copula families. Further details can be found in Segers, 2012.

Condition 3.3. *Suppose* $\{(x_i, y_i)\}_{i=1}^n$ *is an i.i.d. sample from* (X,Y) *with joint CDF* $F_{X,Y}$*. We require the following conditions to hold.*

(a) *The marginal distributions* F_X *and* F_Y *of* $F_{X,Y}$ *are continuous.*

(b) *The copula* $\mathcal{C}(\cdot,\cdot)$ *associated with* $F_{X,Y}$ *has continuous first and second partial derivatives in the interior of the unit square. Specifically,* $\frac{\partial}{\partial u}\mathcal{C}(u,v)$*, and* $\frac{\partial^2}{\partial v^2}$ *are continuous on* $(0,1)\times[0,1]$*; and* $\frac{\partial}{\partial v}\mathcal{C}(u,v)$ *and* $\frac{\partial^2}{\partial v^2}\mathcal{C}(u,v)$ *are continuous on* $[0,1]\times(0,1)$*. Moreover, there exists positive constants* K_1 *and* K_2 *such that*

$$\begin{cases} \left|\frac{\partial^2}{\partial u^2}\mathcal{C}(u,v)\right| \leqslant \frac{K_1}{u(1-u)}, & \text{for } (u,v)\in(0,1)\times[0,1]; \\ \left|\frac{\partial^2}{\partial v^2}\mathcal{C}(u,v)\right| \leqslant \frac{K_2}{v(1-v)}, & \text{for } (u,v)\in[0,1]\times(0,1). \end{cases}$$

(c) *The copula density* $\mathcal{C}_{1,2}(\cdot,\cdot)$ *exists, is strictly positive, and has continuous second partial derivatives on* $(0,1)^2$*. Furthermore, there is a positive constant* K_{00} *such that*

$$\mathcal{C}_{1,2}(u,v) \leqslant K_{00}\min\left\{\frac{1}{u(1-u)},\frac{1}{v(1-v)}\right\}.$$

Theorem 3.3.1. *Suppose Condition 3.3 holds. Let* $\mathbf{K}(s,t)=\phi(s)\phi(t)$*,* $\mathbf{H}=h^2\mathbf{I}$*, and* $h=\mathcal{O}(n^{-\beta})$ *for some* $\beta\in(0,\frac{1}{2})$*. When* $n\to\infty$*, the following result holds* $\forall(u,v)\in(0,1)^2$*:*

$$\sqrt{nh}\left[\widehat{\mathcal{C}}_{2|1}\left(v\mid u\right)-\mathcal{C}_{2|1}\left(v\mid u\right)-h^2B(u,v)-o(h^2)\right]\xrightarrow{d}N\left(0,\sigma^2(u,v)\right).$$

The expressions of $B(\cdot,\cdot)$ *and* $\sigma^2(\cdot,\cdot)$ *are specified in Section 3.8.*

The bandwidth parameter h is crucial in estimating the conditional copula function and understanding its asymptotic behavior. Next, we investigate the optimal choice of h that minimizes the asymptotic estimation error.

Definition 3.3.1. *Let $\widehat{\mathcal{C}}_{2|1}(v \mid u)$ be the conditional copula estimator in* (3.3.1) *for $v \in (0,1)$. We define the mean integrated (weighted) square error (MISE) and the asymptotic mean integrated (weighted) square error (AMISE). as follows.*

$$\text{MISE}(v, h, \omega) \doteq \mathbb{E}\left\{ \int_0^1 \left[\widehat{\mathcal{C}}_{2|1}(v \mid u) - \mathcal{C}_{2|1}(v \mid u)\right]^2 \omega(u)\, du \right\},$$

$$\text{AMISE}(v, h, \omega) \doteq \int_0^1 \left[h^4\, B^2(u, v) + \frac{1}{n\,h}\, \sigma^2(u, v)\right] \omega(u)\, du,$$

where $B^2(u, v)$ and $\sigma^2(u, v)$ represent the squared asymptotic bias and asymptotic variance of the estimator, respectively.

The AMISE in Definition 3.3.1 is derived from the asymptotic normality of the estimator in Theorem 3.3.1, where $h^4 B^2(u, v)$ captures the squared bias and $(nh)^{-1}\sigma^2(u, v)$ captures the variance. Condition 3.4 guarantees that $\omega(\cdot)$ is integrable in the MISE and AMISE definitions, and Corollary 3.3.1 studies the optimal bandwidth that minimizes AMISE.

Condition 3.4. *The weight function $\omega(u)$ is differentiable at any $u \in (0,1)$, and $\frac{\omega(u)}{\phi^2(\Phi^{-1}(u))}$ is integrable over $u \in (0,1)$.*

Corollary 3.3.1. *Suppose Conditions 3.3 and 3.4 hold. The bandwidth for estimator* (3.3.1) *that minimizes AMISE satisfies*

$$h^*(v,\omega) = \underset{h>0}{\operatorname{argmin}}\, AMISE\,(v,h,\omega) = \left[\frac{\int_0^1 \sigma^2(u,v)\omega(u)du}{4\int_0^1 B^2(u,v)\omega(u)du}\right]^{\frac{1}{5}} \cdot n^{-\frac{1}{5}}, \quad \forall v \in (0,1).$$

3.4 Estimation of Copula Quantile Dependence

Substituting the conditional copula estimator in (3.3.1) into (3.1.2) yields the following estimator for the Copula Quantile Dependence:

$$\widehat{D}_\alpha(X,Y;\omega,h) \doteq \int_0^1 \left[\widehat{\mathcal{C}}_{2|1}\left(\alpha \mid u\right) - \alpha\right]^2 \omega(u)du. \tag{3.4.1}$$

In practice, the integral in (3.4.1) can be approximated by a finite-sample weighted average over the transformed observations $\{\widehat{u}_i\}_{i=1}^n$, giving

$$\widehat{D}^*_\alpha(X,Y;\omega,h) \doteq \frac{1}{n}\sum_{i=1}^n \left[\widehat{\mathcal{C}}_{2|1}\left(\alpha \mid \widehat{u}_i\right) - \alpha\right]^2 \omega(\widehat{u}_i). \tag{3.4.2}$$

Theorem 3.4.1. *Suppose that Conditions 3.1–3.4 hold. Let* $\mathbf{K}(s,t) = \phi(s)\phi(t)$, $\mathbf{H} = h^2\mathbf{I}$*, and choose* $h = \mathcal{O}(n^{-\beta})$ *for some* $\beta \in (0, \frac{1}{4})$*. Let* Z_1*,* Z_2*, and* Z_3 *be i.i.d. standard normal random variables. For any quantile level* $\alpha \in (0,1)$*, the following results hold as* $n \to \infty$*:*

(a) If Y *is* ***quantile independent of*** X *at level* α*, then*

$$\widehat{D}_\alpha(X,Y;\omega,h) \xrightarrow{d} n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4\, M_{\perp_\alpha,4}(\omega)$$

$$+\ 2\,n^{-\frac{1}{2}}h^2\,\sigma_{\perp_\alpha,1}(\omega)\,Z_1\ +\ \sqrt{2}\,n^{-1}h^{-\frac{1}{2}}\,\sigma_{\perp_\alpha,3}(\omega)\,Z_2.$$

*(b) If Y is **quantile dependent on** X at level α, then*

$$\widehat{D}_\alpha(X,Y;\omega,h) \xrightarrow{d} D_\alpha(X,Y;\omega) + h^2\,M_{\not\perp_\alpha,4}(\omega) + 2\,n^{-\frac{1}{2}}\,\sigma_{\not\perp_\alpha,1}(\omega)\,Z_3.$$

Explicit expressions for $M_{\perp_\alpha,2}(\omega)$, $M_{\perp_\alpha,4}(\omega)$, $\sigma_{\perp_\alpha,1}(\omega)$, $\sigma_{\perp_\alpha,3}(\omega)$, $M_{\not\perp_\alpha,4}(\omega)$, and $\sigma_{\not\perp_\alpha,1}(\omega)$ are given in Section 3.8.

The above theorem presents the asymptotic properties of the sample Copula Quantile Dependence statistic, highlighting how it behaves under various conditions. When Y is quantile independent of X at quantile level α, the asymptotic distribution of $\widehat{D}_\alpha(X,Y;\omega,h)$ is driven by two independent Gaussian components. Depending on the choice of β, one of these components may dominate. At $\beta=\frac{1}{5}$, both components converge at the same rate, and the asymptotic distribution is the sum of two independent Gaussian random variables.

3.5 Model-free Feature Screening for Quantile Regressions

3.5.1 Screening Procedure

Let $Y\in\mathbb{R}$ be a response variable and $\mathbf{X}=(X_1,\ldots,X_p)^\top\in\mathbb{R}^p$ be a covariate vector of p features. For a quantile level $\alpha\in(0,1)$, suppose there is a sparse set of active features

$$\mathcal{M}_\alpha \doteq \{j\in\{1,\cdots,p\} : Q_\alpha(Y|\mathbf{X}) \text{ functionally depends on } X_j\},$$

where $Q_\alpha(Y|\mathbf{X})$ is the α-th conditional quantile of Y given $\mathbf{X}$. Please note that the active set $\mathcal{M}_\alpha$ may vary across different quantile levels. We also define $\mathcal{M}_\alpha^c$, the complement set of $\mathcal{M}_\alpha$, as the inactive set.

Motivated by the desirable properties of Copula Quantile Dependence (see Lemma 3.1.1), we propose a feature screening approach that retains features exhibiting high Copula Quantile Dependence with Y. Specifically, we estimate $\mathcal{M}_\alpha$ via

$$\mathcal{M}_\alpha^*(t_\alpha) \doteq \{j \in \{1, \cdots, p\} : D_\alpha(X_j, Y; \omega) \geqslant t_\alpha\},$$

where ω is a weight function satisfying Condition 3.4, and $t_\alpha > 0$ is a screening threshold.

Let $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ be an i.i.d. random sample drawn from $(\mathbf{X}, Y)$, where $\boldsymbol{x}_i = (x_{i,1}, \ldots, x_{i,p})^\top$. With a suitably chosen bandwidth h, we estimate $D_\alpha(X_j, Y; \omega)$ by $\widehat{D}_\alpha(X_j, Y; \omega, h)$ as defined in (3.4.1). We then estimate the active set at the quantile level α by

$$\widehat{\mathcal{M}}_\alpha(t_{n,\alpha}) = \left\{j \in \{1, \cdots, p\} : \widehat{D}_\alpha(X_j, Y; \omega, h) \geqslant t_{n,\alpha}\right\}, \tag{3.5.1}$$

where $t_{n,\alpha}$ is a screening threshold. We refer to this screening approach as *Quantile Copula–based Screening (QC-Screening)*, and summarize it in Algorithm 3.1.

The computational complexity of Algorithm 3.1 is of order $\mathcal{O}(n^2 p)$ since either (3.4.1) or (3.4.2) incurs a computational cost of order $\mathcal{O}(n^2)$. For (3.4.1), note that $\widehat{\mathcal{C}}_{2|1}(\cdot \mid \cdot)$ introduces an extra summation over n. To analyze the computational cost of (3.4.2), we can use the definition of $\widehat{\mathcal{C}}_{2|1}(\cdot \mid \cdot)$ in (3.3.1), expand the quadratic terms, and swap the integral with the summation to derive an analytic expression for $\widehat{D}_\alpha(X, Y; \omega, h)$. The cost is dominated by the $\mathcal{O}(n^2)$ summation over interaction terms. Specifically,

Algorithm 3.1 Quantile Copula-based Screening (QC-Screen)

Input: An observed sample $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, a quantile level $\alpha \in (0,1)$, a bandwidth parameter h, a weight function $\omega(\cdot)$, and a selection threshold $t > 0$.
Calculate Empirical CDF of Y by $\widehat{F}_Y(\cdot) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{y_i \leqslant \cdot\}}$.
Calculate pseudo sample by $\left\{\widehat{t}_i\right\}_{i=1}^n = \left\{\Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_Y(y_i)\right)\right\}_{i=1}^n$.
for $j \in \{1, \cdots, p\}$ **do**
    Calculate Empirical CDF of X_j by $\widehat{F}_{X_j}(\cdot) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{x_{i,j} \leqslant \cdot\}}$.
    Calculate pseudo sample by $\{\widehat{s}_{i,j}\}_{i=1}^n = \left\{\Phi^{-1}\left(\frac{n}{n+1}\widehat{F}_{X_j}(x_{i,j})\right)\right\}_{i=1}^n$.
    Calculate conditional copula function estimator between X_j and Y by (3.3.1).
    Calculate Copula Quantile Dependence estimator, $\widehat{D}_\alpha(X_j, Y; \omega, h)$ by (3.4.1).
Find the selected set $\widehat{\mathcal{M}}_\alpha(t)$ by (3.5.1).
Output: Selected set $\widehat{\mathcal{M}}_\alpha(t)$.

with $\omega(u) = \phi^2\left(\Phi^{-1}(u)\right)$, the analytic form $\widehat{D}_\alpha(X, Y; \omega, h)$ admits

$$\begin{aligned}\widehat{D}_\alpha(X, Y; \omega, h) = {} & \frac{2}{n^2 h\sqrt{2+h^2}} \sum_{1 \le k < l \le n} \phi\Big(\tfrac{\widehat{s}_k - \widehat{s}_l}{h\sqrt{2+h^2}}\Big)\, \phi\Big(\tfrac{\sqrt{\widehat{s}_k^2 + \widehat{s}_l^2}}{\sqrt{2+h^2}}\Big)\, \Phi\Big(\tfrac{\Phi^{-1}(\alpha) - \widehat{t}_k}{h}\Big)\, \Phi\Big(\tfrac{\Phi^{-1}(\alpha) - \widehat{t}_l}{h}\Big) \\ & + \frac{1}{\sqrt{2\pi}\, n^2\, h\, \sqrt{2+h^2}} \sum_{i=1}^n \phi\Big(\tfrac{\widehat{s}_i}{\sqrt{1+\frac{h^2}{2}}}\Big)\, \Phi^2\Big(\tfrac{\Phi^{-1}(\alpha) - \widehat{t}_i}{h}\Big) \\ & - \frac{\alpha}{\sqrt{\pi}\, n \sqrt{\frac{1}{2} + h^2}} \sum_{i=1}^n \phi\Big(\tfrac{\widehat{s}_i}{\sqrt{\frac{1}{2}+h^2}}\Big)\, \Phi\Big(\tfrac{\Phi^{-1}(\alpha) - \widehat{t}_i}{h}\Big) \; + \; \frac{\alpha^2}{2\sqrt{3}\,\pi}. \end{aligned} \tag{3.5.2}$$

3.5.2 Theoretical Properties

Under some mild conditions and with an appropriately chosen threshold $t_{n,\alpha}$, QC-Screen achieves both the sure screening property and a stronger rank consistency property. We first state these conditions.

Condition 3.5. *Let* $c_{\alpha 1} > 0$, $c_{\alpha 2} > 0$, $\kappa_{\alpha 1} \in [0, 2\beta)$, *and* $\kappa_{\alpha 2} \in [0, 2\beta)$ *be some constants.*

(a) $\min_{j \in \mathcal{M}_\alpha} \{D_\alpha(X_j, Y; \omega)\} \geqslant c_{\alpha 1} n^{-\kappa_{\alpha 1}}$.

(b) $\min_{j\in\mathcal{M}_\alpha}\{D_\alpha(X_j,Y;\omega)\} - \max_{j\in\mathcal{M}_\alpha^c}\{D_\alpha(X_j,Y;\omega)\} \geqslant c_{\alpha 2}n^{-\kappa_{\alpha 2}}$.

Condition 3.5(a) is a minimum signal strength condition that requires the Copula Quantile Dependence between active features and the response variable to be uniformly bounded below and does not decay to zero too quickly as the sample size n increases. Condition 3.5(b), on the other hand, imposes an assumption on the gap between the signal strengths of active and inactive features. Since the Copula Quantile Dependence is always non-negative, Condition 3.5(a) is weaker than Condition 3.5(b). These minimum signal strength conditions can be viewed as sparsity assumptions, allowing active features to be distinguished from inactive ones. In general, Condition 3.5 is very mild, as it allows the minimum signal strength to approach zero as the sample size grows.

Theorem 3.5.1 (Sure screening property)**.** *Suppose Conditions 3.1 – 3.4, and 3.5(a) hold. Choose* $h = \mathcal{O}(n^{-\beta})$ *for some* $\beta \in (0, \frac{1}{4})$*. For any quantile level* $\alpha \in (0,1)$*, let* $t_{n,\alpha} = c_{t,\alpha}n^{-\tau_\alpha}$ *for some* $c_{t,\alpha} \in (0, c_{\alpha 1})$ *and* $\tau_\alpha \geqslant \kappa_{\alpha 1}$*. Then, there exists positive constants* $n_{\alpha 1}$ *and* $C_{\alpha 1}$*, such that*

$$\mathcal{P}\left(\mathcal{M}_\alpha \subseteq \widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right) \geqslant 1 - \mathcal{O}\left(\mathcal{S}_\alpha \exp\left(-C_{\alpha 1}n^{1-2\kappa_{\alpha 1}}\right)\right), \quad \text{when } n \geqslant n_{\alpha 1},$$

where $\mathcal{S}_\alpha$ *is the cardinality of* $\mathcal{M}_\alpha$*. If* $\mathcal{S}_\alpha = \mathcal{O}(n^{\nu_\alpha})$ *for a positive constant* ν_α*, we also have*

$$\lim_{n\to\infty} \mathcal{P}\left(\mathcal{M}_\alpha \subseteq \widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right) = 1.$$

Theorem 3.5.2 (Rank consistency property)**.** *Suppose Conditions 3.1 – 3.4, and 3.5(b) hold. Choose* $h = \mathcal{O}(n^{-\beta})$ *for some* $\beta \in (0, \frac{1}{4})$*. Then, for any quantile level* $\alpha \in (0,1)$*, there exists positive constants* $n_{\alpha 2}$

and $C_{\alpha 2}$, such that

$$\mathcal{P}\left(\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}>0\right)$$
$$>1-\mathcal{O}\left(p\exp\left(-C_{\alpha 2}n^{1-2\kappa_{\alpha 2}}\right)\right),\quad \textit{when } n\geqslant n_{\alpha 2}.$$

If $\log p = o(n^{1-2\kappa_{\alpha 2}})$, *we also have*

$$\liminf_{n\to\infty}\left(\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}\right)>0, \textit{ almost surely.}$$

Based on Condition 3.5(a), the sure screening property in Theorem 3.5.1 ensures that all active features are reserved in the selected subset of features with a well selected threshold for the Copula Quantile Dependence. Based on the stronger Condition 3.5(b), the rank consistency property in Theorem 3.5.2 ensures that the ranking of features by their Copula Quantile Dependence remains stable, which indicates the separability among active features and inactive features. Both theorems guarantee that ranking variables based on their marginal Copula Quantile Dependence with the response and selecting the top-ranked features would be a feasible solution for feature screening.

3.6 False Discovery Control

In the QC-Screen procedure, selecting an appropriate screening threshold parameter $t_{n,\alpha}$ is crucial for distinguishing between active and inactive features. Theorem 3.5.1 suggests us to choice $t_{n,\alpha} = c_{t,\alpha}n^{-\tau_\alpha}$ for some unknown constants $c_{t,\alpha}$ and τ_α. When certain model assumptions hold, $c_{t,\alpha}$ and τ_α (and thus $t_{n,\alpha}$) can be determined through cross-validation or information criterion methods. However, in a model-free context, these approaches are not directly applicable because loss functions that assess goodness-of-fit are not well-defined. In practice, one may opt for an arbitrary but conservative screening threshold to ensure that all active features are included with high probability. For instance, following Fan and Lv, 2008 and He et al., 2013, one can rank the dependence measures and retain the top $\lfloor n/\log n \rfloor$ features, where $\lfloor \cdot \rfloor$ denotes the floor function. However, such a rule-of-thumb screening threshold is not data-adaptive and may include too many inactive features, thereby inflating the false discovery rate (FDR).

In this section, we introduce a data-driven method for selecting the screening threshold in QC-Screen, designed to adaptively balance the inclusion of active features with FDR control. We rigorously demonstrate that the threshold selected by this method ensures asymptotic control of the FDR, making it a practical and theoretically sound approach for model-free feature screening.

Our proposed method applies similar symmetric scheme with Barber and Candès, 2015; Guo et al., 2023; W. Liu et al., 2022; Tong et al., 2023. Via Theorem 3.4.1, we notice our proposed Copula Quantile Dependence is asymptotic symmetric around its mean values. To increase the FDR performance with increasing symmetry, we propose the following Corollary 3.6.1.

Corollary 3.6.1. *Under the same conditions in Theorem 3.4.1, for any quantile level* $\alpha \in (0,1)$*, we define*

$$\begin{aligned} &\widehat{D}^{\dagger}_{\alpha}(X,Y;\omega,h) \\ &\doteq \frac{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4\, M_{\perp_\alpha,4}(\omega)}{\sqrt{4n^{-1}h^4\sigma^2_{\perp_\alpha,1}(\omega) + 2n^{-2}h^{-1}\sigma^2_{\perp_\alpha,3}(\omega)}} \log\left[\frac{\widehat{D}_{\alpha}(X,Y;\omega,h)}{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4\, M_{\perp_\alpha,4}(\omega)}\right] + \delta_n(\alpha), \end{aligned} \tag{3.6.1}$$

where $\delta_n(\alpha)$ is a sequence converging to 0 as $n \to \infty$. If Y is ***quantile independent of*** *X at level α, the following result holds as $n \to \infty$:*

$$\widehat{D}_\alpha^\dagger(X, Y; \omega, h) \xrightarrow{d} Z.$$

Since the mapping from $\widehat{D}_\alpha(X, Y; \omega, h)$ to $\widehat{D}_\alpha^\dagger(X, Y; \omega, h)$ is one to one monotonic, we may focus on an equivalent problem setup introduced in Section 3.5. For a given screening threshold $t > 0$, the active set $\mathcal{M}_\alpha(t)$ is estimated by QC-Screen as

$$\widehat{\mathcal{M}}_\alpha(t) = \left\{ j \in \{1, \cdots, p\} : \widehat{D}_\alpha^\dagger(X_j, Y; \omega, h) \geqslant t \right\}. \tag{3.6.2}$$

Then, the false discovery proportion (FDP) of the screening procedure is defined by

$$\mathrm{FDP}(t; \alpha) \doteq \frac{\#\{j : j \in \widehat{\mathcal{M}}_\alpha(t) \cap \mathcal{M}_\alpha^c\}}{\#\{j : j \in \widehat{\mathcal{M}}_\alpha(t)\}} = \frac{\#\{j \in \mathcal{M}_\alpha^c : \widehat{D}_\alpha^\dagger(X_j, Y; \omega, h) \geqslant t\}}{\#\{j : \widehat{D}_\alpha^\dagger(X_j, Y; \omega, h) \geqslant t\}}, \tag{3.6.3}$$

where $\#\{\cdot\}$ is the cardinality of a set, and we follow the convention that $0/0 = 0$. Consequently, we define FDR as the expectation of FDP, i.e. $\mathrm{FDR}(t; \alpha) \doteq \mathbb{E}[\mathrm{FDP}(t; \alpha)]$.

Corollary 3.6.1 has shown that $\widehat{D}_\alpha^\dagger(X_j, Y; \omega, h)$ converges in distribution to a standard Gaussian distribution when Y is quantile independent with X_j at the quantile level α. Therefore, $\widehat{D}_\alpha^\dagger(X_j, Y; \omega, h)$ is nearly symmetric towards 0. Additionally, it is common to make a sparsity assumption that the number of active features $\mathcal{S}_\alpha$ is much smaller than the dimensionality p for screening problems. These two results together motivate us us to approximate the unknown numerator in (3.6.3) as follows

$$\#\{j \in \mathcal{M}_\alpha^c : \widehat{D}_\alpha^\dagger(X_j, Y; \omega, h) \geqslant t\} \approx \#\{j \in \mathcal{M}_\alpha^c : \widehat{D}_\alpha^\dagger(X_j, Y; \omega, h) \leqslant -t\}$$

$$\leqslant \#\{j : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \leqslant -t\},$$

which yields a slightly conservative approximation of $\mathrm{FDP}(t;\alpha)$, i.e.

$$\widehat{\mathrm{FDP}}(t;\alpha) = \frac{\#\{j : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \leqslant -t\}}{\#\{j : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \geqslant t\}}. \tag{3.6.4}$$

To control FDR at a pre-specified level $\gamma \in (0,1)$, we choose the threshold $T_{\gamma}(\alpha)$ by

$$T_{\gamma}(\alpha) = \inf\left\{t > 0 : \frac{1 + \#\{j : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \leqslant -t\}}{\#\{j : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \geqslant t\}} \leqslant \gamma\right\}, \tag{3.6.5}$$

where the extra term 1 in the numerator makes the choice of $T_{\gamma}(\alpha)$ slightly more conservative. Then, the selected set is given by

$$\widehat{\mathcal{M}}_{\alpha}\left(T_{\gamma}(\alpha)\right) = \left\{j \in \{1, \cdots, p\} : \widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) \geqslant T_{\gamma}(\alpha)\right\}. \tag{3.6.6}$$

When the solution of (3.6.5) does not exist, we set $\widehat{\mathcal{M}}_{\alpha}\left(T_{\gamma}(\alpha)\right)$ to be an empty set. We summarize this FDR control approach in Algorithm 3.2.

Next, we introduce a condition for the dependence among features and a theorem to demonstrate that the QC-Screen procedure with the screening threshold selected by (3.6.5) can asymptotically control FDR at the level γ.

Condition 3.6. *Denote* $p_{0,\alpha} = |\mathcal{M}_{\alpha}^{c}|$ *the cardinality of the inactive set,* $B_j(\alpha) \doteq \mathbb{1}_{\left\{\widehat{D}_{\alpha}^{\dagger}(X_j, Y; \omega, h) < 0\right\}}$ *and* $S_{\alpha}^{c} = \sum_{j \in \mathcal{M}_{\alpha}^{c}} B_j(\alpha)$*. We assume* $p_{0,\alpha} \to \infty$ *as* $n \to \infty$ *and* $Var\left(S_{\alpha}^{c}\right) = o(p_{0,\alpha}^{2})$*.*

Algorithm 3.2 FDR control for QC-Screen

Input: An observed sample $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, a quantile level $\alpha \in (0,1)$, a bandwidth parameter h, a weight function $\omega(\cdot)$, a FDR level γ.
for $j \in \{1, \cdots, p\}$ **do**
    Calculate Copula Quantile Dependence estimator, $\widehat{D}_\alpha(X_j, Y; \omega, h)$ through Algorithm 3.1.
    Calculate $\widehat{D}^{\dagger}_\alpha(X_j, Y; \omega, h)$ by (3.6.1), and our suggested $\delta_n(\alpha)$ in Appendices A.3.
Sort $\left|\widehat{D}^{\dagger}_\alpha(X_j, Y; \omega, h)\right|$'s from the smallest to the largest as $t_1, \cdots, t_p$, and let $t_0 = 0$ and $t_{p+1} = \infty$.
Set $m = 0$.
while $\frac{1+\#\{j:\widehat{D}^{\dagger}_\alpha(X_j,Y;\omega,h)\leqslant -t_m\}}{\#\{j:\widehat{D}^{\dagger}_\alpha(X_j,Y;\omega,h)\geqslant t_m\}} > \gamma$ **and** $m \leqslant p$ **do**
    $m = m + 1$.
Find the selected set $\widehat{\mathcal{M}}_\alpha(t_m)$ by (3.6.2).
Output: Threshold t_m, and selected set $\widehat{\mathcal{M}}_\alpha(t_m)$.

Remark 3.6.1. *The above condition is inherited from Dai et al., 2023, which restricts the dependency among inactive features. This condition would fail to hold only under some extreme cases, e.g. the Copula Quantile Dependences corresponding to null features have constant pairwise correlations, or can be clustered into a fixed number of groups so that their within-group correlation is a constant. For further details, see Sections 2 and 3 in Dai et al., 2023.*

Theorem 3.6.1 (Asymptotic FDR control). *Suppose Conditions 3.1 – 3.6 hold. We choose $h = \mathcal{O}\left(n^{-\beta}\right)$ for some $\beta \in (0, \frac{1}{4})$. For any quantile level $\alpha \in (0,1)$, and $\forall\gamma \in [0,1]$, let $T_\gamma(\alpha)$ be a screening threshold selected by* (3.6.5). *Then we have*

$$\limsup_{n\to\infty} \mathrm{FDR}\left(T_\gamma(\alpha); \alpha\right) \leqslant \gamma.$$

3.7 Simulation Study

In this section, we illustrate the methodoloies we proposed in by several experiments. The choice of the weight function is actually flexible. For simplicity of calculation and to guarantee integrability in $\widehat{D}_\alpha(X, Y; \omega, h)$, we would select $\omega(u) = \phi^2(\Phi^{-1}(u))$ in this section.

3.7.1 Feature Screening Performance

Table 3.1: The median and IQR of minimum model size over 200 replications for Experiment 1.a.

	$p = 1000$		$p = 5000$	
	median	**IQR**	**median**	**IQR**
		$\alpha = 0.5$		
QC-Screen	9.00	4.00	9.00	5.00
SIS	6.00	4.00	7.00	4.00
DC-SIS	7.00	4.00	7.00	3.00
Qa-SIS	8.00	4.00	8.00	4.00
QC-SIS	8.00	4.00	8.00	3.00
		$\alpha = 0.75$		
QC-Screen	10.00	5.00	11.00	5.25
SIS	6.00	4.00	7.00	4.00
DC-SIS	7.00	4.00	7.00	3.00
Qa-SIS	9.00	4.25	8.00	5.00
QC-SIS	8.00	4.00	9.00	5.00

In this section, we use simulated examples to evaluate the finite-sample performance of the proposed Quantile Copula-based screening procedure (QC-Screen). We compare its performance against several existing methods, including sure independence screening (SIS; Fan and Lv, 2008), distance correlation-based screening (DC-SIS; R. Li et al., 2012), quantile-adaptive variable screening (Qa-SIS; He et al., 2013), quantile correlation-based screening (QC-SIS; G. Li et al., 2015, Ma and Zhang, 2016).

We consider the following four regression models.

Model 1.a: $Y = 5X_1 + 3X_4 + 4X_{10} + 6X_{15} + \sqrt{50}\varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0,1)$. The number of active features is 4 at all quantile levels.

Model 1.b: $Y = 5\log\left(15\left|\frac{X_1}{1-X_1}\right|\right) + 3(2X_2-1)^2 + 4\frac{\sin(2\pi X_3)}{2-\sin(2\pi X_3)} + 6\exp(5X_4) + \sqrt{1.74}\varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0,1)$. The number of active features is 4 at all quantile levels.

Model 1.c: $Y = 4(X_1)^2 + 7\sin(2\pi X_2)\cdot|X_3| + 2\exp(5X_4)\cdot\log\left(7\left|\frac{X_5}{1-X_5}\right|\right) + \exp(0.2X_6 + 0.3X_7)\cdot\varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0,1)$. The number of active features is 5 at the quantile level $\alpha = 0.5$ and 7 at other quantile levels.

Model 1.d: $Y = 2X_1^2 + 2X_2^2 + \exp(5X_3 + 5X_4 + 5X_5 + 5X_6 + 5X_7)\cdot\varepsilon$, where $\varepsilon \overset{\text{i.i.d.}}{\sim} N(-q_\alpha, 1)$, and q_α is the α quantile of standard normal distribution. The number of active variables is 2 across different α values.

Table 3.2: The median and IQR of minimum model size over 200 replications for Experiment 1.b.

	$p = 1000$		$p = 5000$	
	median	**IQR**	**median**	**IQR**
	$\alpha = 0.5$			
QC-Screen	6.00	1.00	6.00	2.00
SIS	28.00	72.75	114.50	517.75
DC-SIS	8.00	22.00	13.00	170.00
Qa-SIS	7.00	1.00	7.00	1.00
QC-SIS	40.00	107.75	237.00	816.25
	$\alpha = 0.75$			
QC-Screen	6.00	1.00	7.00	2.00
SIS	28.00	72.75	114.50	517.75
DC-SIS	8.00	22.00	13.00	170.00
Qa-SIS	7.00	1.00	7.00	2.00
QC-SIS	6.00	1.00	7.00	1.00

In all models, the feature vector is $\mathbf{X} = (X_1, \ldots, X_p)^\intercal \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ij} = 0.8^{|i-j|}$. The error term ε is independently sampled from $N(0,1)$ and is independent of $\mathbf{X}$. We set the sample size $n = 400$, the number of features $p = 1000, 5000$, and generate 200 replicates for each model.

Table 3.3: The median and IQR of minimum model size over 200 replications for Experiment 1.c.

	$p = 1000$		$p = 5000$	
	median	**IQR**	**median**	**IQR**
$\alpha = 0.5$				
QC-Screen	6.00	1.00	6.00	1.25
SIS	31.00	80.50	108.50	465.25
DC-SIS	8.00	21.25	11.00	125.00
Qa-SIS	7.00	2.00	7.00	2.00
QC-SIS	11.00	12.75	20.00	81.25
$\alpha = 0.75$				
QC-Screen	7.00	1.00	8.00	3.00
SIS	70.50	169.25	249.50	827.75
DC-SIS	13.50	49.00	42.50	234.25
Qa-SIS	7.00	1.00	7.00	1.00
QC-SIS	7.00	1.00	7.00	1.00

For each model, the quantile levels $\alpha = 0.5$, and $\alpha = 0.75$ are analyzed. In each simulation, features are ranked in descending order according to the screening criteria of the respective methods. The minimum model size that includes all active features is recorded for each replication. Screening performance is assessed using the median and inter quartile range (IQR) of the minimum model size over 200 replications.

The results of the experiments are summarized in Tables 3.1 – 3.4. For Model 1.a, all five methods perform comparably, successfully identifying the four active variables within about 10 features. Model 1.b, an additive nonlinear model, highlights the limitations of SIS in handling nonlinearity, resulting in significantly worse performance compared to other methods. Among quantile-based methods, QC-SIS shows inconsistent results across quantile levels, notably failing to identify active features efficiently at the 0.5 quantile level. Model 1.c examines a complex nonlinear model with varying active variables across quantile levels. Here, SIS and DC-SIS perform poorly due to their inability to address quantile-specific screening, with performance declining further in this challenging setting. Model 1.d, a highly complex nonlinear model with only two active variables, shows QC-Screen as the only effective method. SIS and DC-SIS are severely impacted by nonlinearity and quantile-specific features. Qa-SIS overestimates model size and struggles at extreme quantile levels, while QC-SIS remains better overall but has limitations near the 0.5 quantile level.

Table 3.4: The median and IQR of minimum model size over 200 replications for Experiment 1.d.

	$p = 1000$		$p = 5000$	
	median	**IQR**	**median**	**IQR**
$\alpha = 0.5$				
QC-Screen	2.00	0.00	2.00	0.00
SIS	176.00	421.75	872.00	1846.00
DC-SIS	138.50	281.00	694.00	1734.25
Qa-SIS	6.00	2.00	7.00	2.00
QC-SIS	3.00	8.00	4.00	23.25
$\alpha = 0.75$				
QC-Screen	2.00	0.00	2.00	2.00
SIS	157.50	341.75	821.50	1829.00
DC-SIS	137.50	275.25	705.00	1701.75
Qa-SIS	12.00	13.00	35.50	51.00
QC-SIS	2.00	1.00	2.00	2.00

In summary, across all models, the proposed QC-Screen method demonstrates robust feature screening performance, consistently outperforming other methods. These results emphasize the strengths of QC-Screen in handling complex, nonlinear, and quantile-dependent relationships.

3.7.2 False Discovery Control Performance

In this subsection, we evaluate the empirical performance of the proposed QC-FDR procedure through simulated experiments. We conduct simulated experiments to assess its ability to control FDR and to verify its effectiveness in achieving the sure screening property. The implementation details are outlined in Algorithm 3.2.

Table 3.5: FDR control performances over 200 replications with $p = 1000$.

Model	α	γ	$\lvert\widehat{\mathcal{S}}\rvert$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_9$	$\mathcal{P}_{10}$	$\mathcal{P}_a$	$\widehat{\text{FDR}}$
2.a	0.5	0.10	12.00	0.965	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.965	0.1284
		0.15	12.00	0.975	0.985	0.990	0.990	0.990	0.990	0.990	0.990	0.985	0.970	0.965	0.1575
		0.20	12.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.1934
		0.25	13.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.2198
	0.75	0.10	11.00	0.885	0.905	0.905	0.905	0.905	0.905	0.905	0.905	0.905	0.900	0.880	0.1188
		0.15	11.00	0.910	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.990	0.950	0.880	0.1470
		0.20	12.00	0.985	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.985	0.2076
		0.25	13.00	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.2448
2.b	0.5	0.10	11.00	0.935	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.935	0.1268
		0.15	12.00	0.955	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.955	0.935	0.1500
		0.20	12.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.1911
		0.25	13.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.2152
	0.75	0.10	11.00	0.820	0.835	0.835	0.835	0.835	0.835	0.835	0.835	0.835	0.835	0.820	0.0994
		0.15	11.00	0.895	0.980	0.980	0.980	0.980	0.980	0.980	0.980	0.960	0.900	0.845	0.1448
		0.20	12.00	0.980	0.995	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.985	0.975	0.2068
		0.25	13.00	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.2367
2.c	0.5	0.10	11.00	0.955	0.955	0.955	0.955	0.955	0.955	0.955	0.955	0.955	0.955	0.955	0.1286
		0.15	12.00	0.985	0.990	0.990	0.990	0.990	0.990	0.990	0.990	0.980	0.955	0.955	0.1509
		0.20	12.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.995	0.1906
		0.25	13.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.2155
	0.75	0.10	11.00	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.0913
		0.15	11.00	0.990	0.995	0.995	0.995	0.995	0.995	0.995	0.975	0.935	0.850	0.840	0.1300
		0.20	12.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.985	0.970	0.970	0.1971
		0.25	13.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.970	0.970	0.2280

We consider three models, each has 10 active features, defined as follows:

Model 2.a: $Y = 5 \sum_{j=1}^{10} X_j + \varepsilon.$

Model 2.b: $Y = 5\sum_{j=1}^{10} \exp(\frac{1}{2}X_j) + \varepsilon.$

Model 2.c: $Y = 5\sum_{j=1}^{2} \log\left(\frac{|X_j|}{2|1-X_j|}\right) + 5\sum_{j=3}^{5} \sin\left(\frac{\pi}{2}X_j\right) + 5\sum_{j=6}^{8} \exp(\frac{1}{2}X_j) + 5\sum_{j=9}^{10} X_j + \varepsilon.$

In the above models, the feature vector $\mathbf{X} = (X_1, \ldots, X_p)^\top$ and the error term ε are generated as described in Section 3.7.1. The sample size is set to $n = 400$, the number of features is $p = 1000, 2000$, and each model is simulated for 200 replications.

Table 3.6: FDR control performances over 200 replications with $p = 2000$.

Model	α	γ	$\lvert\widehat{\mathcal{S}}\rvert$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_9$	$\mathcal{P}_{10}$	$\mathcal{P}_a$	$\widehat{\text{FDR}}$
2.a	0.5	0.10	11.00	0.895	0.905	0.905	0.905	0.905	0.905	0.905	0.905	0.905	0.905	0.895	0.0981
		0.15	11.00	0.940	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.935	0.900	0.1116
		0.20	12.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.1704
		0.25	12.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.2026
	0.75	0.10	10.00	0.780	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.810	0.805	0.775	0.0811
		0.15	11.00	0.880	0.985	0.990	0.990	0.990	0.990	0.990	0.990	0.970	0.890	0.815	0.1100
		0.20	11.00	0.980	0.995	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.995	0.980	0.1713
		0.25	13.00	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.2370
2.b	0.5	0.10	11.00	0.920	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.925	0.920	0.0946
		0.15	11.00	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.945	0.920	0.1146
		0.20	12.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.1652
		0.25	13.00	0.995	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.2099
	0.75	0.10	10.00	0.675	0.715	0.715	0.715	0.715	0.715	0.715	0.715	0.715	0.715	0.675	0.0661
		0.15	10.00	0.810	0.965	0.980	0.980	0.980	0.980	0.980	0.980	0.965	0.800	0.695	0.0957
		0.20	11.00	0.950	0.990	0.995	1.000	1.000	1.000	1.000	1.000	0.980	0.930	0.910	0.1629
		0.25	12.00	0.970	0.990	0.995	1.000	1.000	1.000	1.000	1.000	0.985	0.950	0.935	0.2219
2.c	0.5	0.10	11.00	0.915	0.915	0.915	0.915	0.915	0.915	0.915	0.915	0.915	0.915	0.915	0.0901
		0.15	11.00	0.980	0.995	0.995	0.995	0.995	0.995	0.990	0.985	0.970	0.930	0.915	0.1088
		0.20	12.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.990	0.990	0.1586
		0.25	12.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.990	0.990	0.1993
	0.75	0.10	10.00	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.660	0.660	0.0692
		0.15	10.00	0.995	0.995	0.995	0.995	0.995	0.995	0.990	0.965	0.920	0.695	0.680	0.0942
		0.20	11.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.965	0.885	0.885	0.1647
		0.25	13.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.925	0.925	0.2322

We evaluate the proposed method under four FDR levels: $\gamma = 0.1, 0.15, 0.2, 0.25$. For each replication, we compute the false discovery proportion (FDP), defined as the ratio of false discoveries to total discoveries. The FDR is estimated as the average FDP over 200 replications ($\widehat{\text{FDR}}$). The performance of feature screening is assessed using the median of the selected model sizes ($|\widehat{\mathcal{S}}|$), as well as the proportion of replications in which each active feature is selected ($\mathcal{P}_i, i = 1, \cdots, 10$) and all active features are correctly selected simultaneously ($\mathcal{P}_a$). The QC-FDR method is implemented at two quantile levels: $\alpha = 0.5$ and $\alpha = 0.75$.

The results, summarized in Tables 3.5 and 3.6, show that the proposed QC-FDR method effectively controls the empirical FDR at or below the specified FDR levels in almost all scenarios, across linear and nonlinear models. Additionally, the QC-FDR method achieves the sure screening property with high probability. For simpler settings with $p = 1000$, the probability of selecting all active features exceeds 80%. For the higher dimension with $p = 2000$, this probability exceeds 65%. These results hold for both linear models, and even more complex nonlinear models. These findings confirm that the QC-FDR procedure is robust and effective for controlling FDR across a wide range of settings. The method balances false discovery control with sure screening property, making it a practical and powerful tool for high-dimensional quantile regression analysis.

3.8 Important Parameters

We have the following parameters used in the Theorem 3.3.1.

$$\begin{aligned} B(u,v) =& \frac{1}{2}\Bigg[\left(\left[\Phi^{-1}(u)\right]^2 - 1\right)\mathcal{C}_{2|1}(v \mid u) + 3\phi'\left(\Phi^{-1}(u)\right)\frac{\partial}{\partial u}\mathcal{C}_{2|1}(v \mid u) \\ &+ \phi^2\left(\Phi^{-1}(u)\right)\frac{\partial^2}{\partial u^2}\mathcal{C}_{2|1}(v \mid u) + \phi'\left(\Phi^{-1}(v)\right)\frac{\partial}{\partial v}\mathcal{C}_{2|1}(v \mid u) \\ &+ \phi^2\left(\Phi^{-1}(v)\right)\frac{\partial^2}{\partial v^2}\mathcal{C}_{2|1}(v \mid u)\Bigg], \end{aligned} \tag{3.8.1}$$

$$\sigma^2(u,v) = \frac{1}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}(v \mid u) - \mathcal{C}_{2|1}^2(v \mid u)}{\phi\left(\Phi^{-1}(u)\right)}. \tag{3.8.2}$$

We have the following parameters used in the Theorem 3.4.1.

$$\sigma^2_{\perp\alpha,1}(\omega) = \left(\alpha - \alpha^2\right)\Bigg\{\int_0^1 \left[\frac{1}{2}\left(\left[\Phi^{-1}(u)\right]^2 - 1\right)\alpha + \frac{1}{2}\phi'\left(\Phi^{-1}(\alpha)\right)\right]^2 \omega^2(u)\,du$$

$$- \left(\int_0^1 \left[\frac{1}{2}\left(\left[\Phi^{-1}(u)\right]^2 - 1\right)\alpha + \frac{1}{2}\phi'\left(\Phi^{-1}(\alpha)\right)\right]\omega(u)\,du\right)^2\Bigg\}, \tag{3.8.3}$$

$$\sigma^2_{\not\perp\alpha,1}(\omega) = \int_0^1 \left[\mathcal{C}_{2|1}(\alpha \mid u) - \mathcal{C}^2_{2|1}(\alpha \mid u)\right]\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]^2 \omega^2(u)\,du$$

$$+ \left(\alpha - \alpha^2\right)\left\{\int_0^1 \frac{\partial^2 \mathcal{C}}{\partial u \partial v}(u,\alpha)\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du\right\}^2$$

$$- \left\{\int_0^1 u\frac{\partial^2 \mathcal{C}}{\partial u^2}(u,\alpha)\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du\right\}^2$$

$$- 2\int_0^1 \frac{\partial^2 \mathcal{C}}{\partial u \partial v}(u,\alpha)\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du$$

$$\int_0^1 \left\{\mathcal{C}_{2|1}(\alpha \mid u) - \mathcal{C}^2_{2|1}(\alpha \mid u) - \left[\mathcal{C}(u,\alpha) - \alpha u\right]\frac{\partial^2 \mathcal{C}}{\partial u^2}(u,\alpha)\right\}$$

$$\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du, \tag{3.8.4}$$

$$M_{\perp\alpha,2}(\omega) = \frac{\alpha - \alpha^2}{2\sqrt{\pi}}\int_0^1 \frac{\omega(u)}{\phi\left(\Phi^{-1}(u)\right)}du, \tag{3.8.5}$$

$$\sigma^2_{\perp\alpha,3}(\omega) = \frac{\left(\alpha - \alpha^2\right)^2}{2\sqrt{2\pi}}\int_0^1 \frac{\omega^2(u)}{\phi\left(\Phi^{-1}(u)\right)}du, \tag{3.8.6}$$

$$M_{\perp\alpha,4}(\omega) = \int_0^1 \left[\frac{1}{2}\left(\left[\Phi^{-1}(u)\right]^2 - 1\right)\alpha + \frac{1}{2}\phi'\left(\Phi^{-1}(\alpha)\right)\right]^2 \omega(u)\,du, \tag{3.8.7}$$

$$M_{\not\perp\alpha,4}(\omega) = 2\int_0^1 \left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]B(u,\alpha)\omega(u)du. \tag{3.8.8}$$

When we select $\omega(u) = \phi^2(\Phi^{-1}(u))$, we have

$$\sigma^2_{\perp\alpha,1}(\omega) = \left(\alpha - \alpha^2\right)\left(\frac{18\alpha^2 - 40\alpha\phi'(\Phi^{-1}(\alpha)) + 25\left[\phi'(\Phi^{-1}(\alpha))\right]^2}{400\sqrt{5}\pi^2} - \frac{\left[3\phi'(\Phi^{-1}(\alpha)) - 2\alpha\right]^2}{432\pi^2}\right),$$

$$M_{\perp\alpha,2}(\omega) = \frac{\alpha - \alpha^2}{4\pi},$$

$$\sigma^2_{\perp_\alpha,3}(\omega) = \frac{[\alpha - \alpha^2]^2}{16\pi^2},$$
$$M_{\perp_\alpha,4}(\omega) = \frac{2\alpha^2 - 4\alpha\phi'(\Phi^{-1}(\alpha)) + 3\left[\phi'(\Phi^{-1}(\alpha))\right]^2}{24\sqrt{3}\pi}.$$

3.9 Proof of Key Theorems in Chapter 3

3.9.1 Proof for Section 3.3

For the following proofs, we define the genuine sample as $\{(s_i, t_i)\}_{i=1}^n \doteq \{(\Phi^{-1}(u_i), \Phi^{-1}(v_i))\}_{i=1}^n$, where $\{(u_i, v_i)\}_{i=1}^n \doteq \{(F_X(x_i), F_Y(y_i))\}_{i=1}^n$ is the genuine copula sample (or true copula sample) observed from the distribution of (U, V), i.e. $\mathcal{C}(u, v)$, pretending the marginal distributions are available. Then, we define the genuine sample based empirical copula as

$$\mathcal{C}_n(u, v) \doteq \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{u_i \leqslant u, v_i \leqslant v\}}, \tag{3.9.1}$$

and the corresponding genuine sample based empirical copula process as

$$\mathbb{B}_n(u, v) \doteq \sqrt{n}\left(\mathcal{C}_n(u, v) - \mathcal{C}(u, v)\right). \tag{3.9.2}$$

Besides, we also define the pseudo sample based empirical copula as

$$\widehat{\mathcal{C}}_n(u, v) \doteq \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\widehat{u}_i \leqslant u, \widehat{v}_i \leqslant v\}}, \tag{3.9.3}$$

and the corresponding pseudo sample based empirical copula process as

$$\mathbb{C}_n(u,v) \doteq \sqrt{n}\left(\widehat{\mathcal{C}}_n(u,v) - \mathcal{C}(u,v)\right). \tag{3.9.4}$$

We also define the following counterpart of $\widehat{\mathcal{C}}_{2|1}(v \mid u)$, which is the kernel density estimator using genuine sample Mathematically, we consider

$$\widehat{\mathcal{C}}^*_{2|1}(v \mid u) \doteq \frac{1}{nh\phi\left(\Phi^{-1}(u)\right)} \sum_{i=1}^{n} \phi\left(\frac{\Phi^{-1}(u) - s_i}{h}\right) \Phi\left(\frac{\Phi^{-1}(v) - t_i}{h}\right). \tag{3.9.5}$$

Proof of Theorem 3.3.1. By the definition of pseudo sample based empirical copula in (3.9.3), and considering substitution $(s,t) = \left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$, we have

$$\widehat{\mathcal{C}}_{2|1}\left(\Phi(t) \mid \Phi(s)\right) = \frac{1}{h\phi(s)} \iint_{\mathcal{I}^2} \phi\left(\frac{s - \Phi^{-1}(u^*)}{h}\right) \Phi\left(\frac{t - \Phi^{-1}(v^*)}{h}\right) d\widehat{\mathcal{C}}_n(u^*, v^*).$$

Then by the definition of pseudo sample based empirical copula process in (3.9.4), we have

$$\begin{aligned} &\sqrt{nh}\left\{\widehat{\mathcal{C}}_{2|1}\left(\Phi(t) \mid \Phi(s)\right) - \mathbb{E}\left[\widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi(s)\right)\right]\right\} \\ =&\frac{1}{\sqrt{h}\phi(s)} \iint_{\mathcal{I}^2} \phi\left(\frac{s - \Phi^{-1}(u^*)}{h}\right) \Phi\left(\frac{t - \Phi^{-1}(v^*)}{h}\right) d\mathbb{C}_n(u^*, v^*). \end{aligned} \tag{3.9.6}$$

With genuine sample based empirical copula process defined in (3.9.2), we define the random process $\{\mathbb{G}_n(u,v) : (u,v) \in \mathcal{I}^2\}$ as

$$\mathbb{G}_n(u,v) \doteq \mathbb{B}_n(u,v) - \frac{\partial \mathcal{C}}{\partial u}(u,v)\mathbb{B}_n(u,1) - \frac{\partial \mathcal{C}}{\partial v}(u,v)\mathbb{B}_n(1,v). \tag{3.9.7}$$

In Segers, 2012, it is shown that when Condition 3.3 holds, $\mathbb{G}_n(u,v)$ and $\mathbb{C}_n(u,v)$ are such that

$$\sup_{(u,v)\in\mathcal{I}^2} |\mathbb{C}_n(u,v) - \mathbb{G}_n(u,v)| = \mathcal{O}_{\text{a.s.}}\left(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\right), \tag{3.9.8}$$

as $n \to \infty$. Mimicking the integration by parts in proof of Theorem 6 in Fermanian et al., 2004 for (3.9.6), we have

$$\begin{aligned}
&\sqrt{nh}\left\{\widehat{\mathcal{C}}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \mathbb{E}\left[\hat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)\right]\right\} \\
=&\frac{1}{\sqrt{h}\phi\left(s\right)}\iint_{\mathcal{I}^2} \mathbb{C}_n(u^*,v^*)\phi'\left(\frac{s-\Phi^{-1}(u^*)}{h}\right)\phi\left(\frac{t-\Phi^{-1}(v^*)}{h}\right)\frac{du^*}{h\phi(\Phi^{-1}(u^*))}\frac{dv^*}{h\phi(\Phi^{-1}(v^*))} \\
=&\frac{1}{\sqrt{h}\phi\left(s\right)}\iint_{\mathcal{I}^2} \mathbb{G}_n(u^*,v^*)\phi'\left(\frac{s-\Phi^{-1}(u^*)}{h}\right)\phi\left(\frac{t-\Phi^{-1}(v^*)}{h}\right)\frac{du^*}{h\phi(\Phi^{-1}(u^*))}\frac{dv^*}{h\phi(\Phi^{-1}(v^*))} \\
&+ R_n(s,t).
\end{aligned} \tag{3.9.9}$$

Then with the result in (3.9.8), we have

$$\begin{aligned}
|R_n\left(s,t\right)| \leqslant&\frac{1}{\sqrt{h}\phi\left(s\right)}\sup_{(u^*,v^*)\in\mathcal{I}^2}|\mathbb{C}_n(u^*,v^*) - \mathbb{G}_n(u^*,v^*)| \\
&\left\{\int_0^1 \left|\phi'\left(\frac{s-\Phi^{-1}(u^*)}{h}\right)\right|\frac{du^*}{h\phi(\Phi^{-1}(u^*))}\right\}\left\{\int_0^1 \phi\left(\frac{t-\Phi^{-1}(v^*)}{h}\right)\frac{dv^*}{h\phi\left(\Phi^{-1}(v^*)\right)}\right\} \\
=&\frac{1}{\sqrt{h}\phi\left(s\right)}\sup_{(u^*,v^*)\in\mathcal{I}^2}|\mathbb{C}_n(u^*,v^*) - \mathbb{G}_n(u^*,v^*)|\int_{\mathbb{R}}|\phi'(z)|\,dz\int_{\mathbb{R}}\phi(z)dz \\
=&\mathcal{O}_{\text{a.s.}}\left(n^{-1/4}h^{-1/2}(\log n)^{1/2}(\log\log n)^{1/4}\right) \\
=&o_{\text{a.s.}}(1),
\end{aligned} \tag{3.9.10}$$

when $h = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{2})$. Then for short, let

$$\begin{aligned}
J_{st,h}\left(u^*, v^*\right) =& \phi\left(\frac{s - \Phi^{-1}(u^*)}{h}\right) \Phi\left(\frac{t - \Phi^{-1}(v^*)}{h}\right), \\
dJ_{st,h}\left(u^*, v^*\right) =& \phi'\left(\frac{s - \Phi^{-1}(u^*)}{h}\right) \phi\left(\frac{t - \Phi^{-1}(v^*)}{h}\right) \frac{du^*}{h\phi(\Phi^{-1}(u^*))} \frac{dv^*}{h\phi(\Phi^{-1}(v^*))},
\end{aligned}$$

and consider the definition of $\mathbb{G}_n(u, v)$ in (3.9.7), we have

$$\begin{aligned}
&\sqrt{nh}\left\{\widehat{\mathcal{C}}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \mathbb{E}\left[\widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)\right]\right\} \\
=& \frac{1}{\sqrt{h}\phi\left(s\right)} \iint_{\mathcal{I}^2} \mathbb{B}_n(u^*, v^*) dJ_{st,h}\left(u^*, v^*\right) \\
&- \frac{1}{\sqrt{h}\phi\left(s\right)} \iint_{\mathcal{I}^2} \frac{\partial \mathcal{C}}{\partial u}(u^*, v^*) \mathbb{B}_n(u^*, 1) dJ_{st,h}\left(u^*, v^*\right) \\
&- \frac{1}{\sqrt{h}\phi\left(s\right)} \iint_{\mathcal{I}^2} \frac{\partial \mathcal{C}}{\partial v}(u^*, v^*) \mathbb{B}_n(1, v^*) dJ_{st,h}\left(u^*, v^*\right) \\
&+ R_n\left(s, t\right) \\
\doteq& K^*_n\left(s, t\right) + B_{n,1}\left(s, t\right) + B_{n,2}\left(s, t\right) + R_n\left(s, t\right). \qquad (3.9.11)
\end{aligned}$$

By the definition of $\mathbb{B}_n(u, v)$ in (3.9.2), we may notice the truth that

$$K^*_n\left(s, t\right) = \sqrt{nh}\left\{\widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \mathbb{E}\left[\widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)\right]\right\}.$$

Hence we have

$$\widehat{\mathcal{C}}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) = \widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) + \frac{B_{n,1}\left(s, t\right)}{\sqrt{nh}} + \frac{B_{n,2}\left(s, t\right)}{\sqrt{nh}} + \frac{R_n\left(s, t\right)}{\sqrt{nh}}$$

$$=A_n(s,t)+\frac{R_n\left(s,t\right)}{\sqrt{nh}}, \tag{3.9.12}$$

where by (3.9.5) and Lemmas A.1.1 and A.1.2, we have

$$\begin{aligned}
&A_n(s,t)\\
=&\frac{1}{nh}\sum_{i=1}^{n}\left\{\frac{\phi\left(\frac{s-s_i}{h}\right)}{\phi\left(s\right)}\Phi\left(\frac{t-t_i}{h}\right)+B_{n,1,i}\left(s,t\right)+B_{n,2,i}\left(s,t\right)-\mathbb{E}\left[B_{n,1,i}\left(s,t\right)\right]-\mathbb{E}\left[B_{n,2,i}\left(s,t\right)\right]\right\}\\
\doteq&\frac{1}{nh}\sum_{i=1}^{n}\mathcal{B}_n\left[s,t;\left(s_i,t_i\right)\right].
\end{aligned} \tag{3.9.13}$$

Then, by Lemma A.1.5, and (3.9.10), when $h=\mathcal{O}(n^{-\beta})$, $\beta\in(0,\frac{1}{2})$, for any $(s,t)\in\mathbb{R}^2$, we have

$$\begin{aligned}
&\sqrt{nh}\left[\widehat{\mathcal{C}}_{2|1}\left(\Phi(t)\mid\Phi\left(s\right)\right)-\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s))-h^2B(\Phi(s),\Phi(t))-o(h^2)\right]\\
&\xrightarrow{d}N\left(0,\sigma^2\left(\Phi(s),\Phi(t)\right)\right),
\end{aligned}$$

as $n\to\infty$. The rest of the proof follows directly from applying back substitution $(u,v)=(\Phi(s),\Phi(t))$.

□

3.9.2 Proof for Section 3.4

Proof of Theorem 3.4.1. By (3.9.12), it is obvious that $A_n\left(\Phi^{-1}(u),\Phi^{-1}(\alpha)\right)$ is part of $\widehat{\mathcal{C}}_{2|1}\left(\alpha\mid u\right)$. By definition in (3.4.1), we have

$$\begin{aligned}
&\widehat{D}_\alpha(X,Y;\omega,h)\\
=&\int_0^1\left[A_n(\Phi^{-1}(u),\Phi^{-1}(\alpha))+\frac{R_n\left(\Phi^{-1}(u),\Phi^{-1}(\alpha)\right)}{\sqrt{nh}}-\alpha\right]^2\omega\left(u\right)du
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 \left[A_n(\Phi^{-1}(u), \Phi^{-1}(\alpha)) - \alpha\right]^2 \omega\left(u\right) du \\
&\quad + 2\int_0^1 \left[A_n(\Phi^{-1}(u), \Phi^{-1}(\alpha)) - \alpha\right] \frac{R_n\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)}{\sqrt{nh}} \omega\left(u\right) du \\
&\quad + \int_0^1 \left[\frac{R_n\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)}{\sqrt{nh}}\right]^2 \omega\left(u\right) du \\
&\doteq \widehat{D}_{\alpha 1} + \widehat{D}_{\alpha 2} + \widehat{D}_{\alpha 3}.
\end{aligned}
$$

We consider $\widehat{D}_{\alpha 1}$ first. Mimicking the proof in Sections 3 & 4 in Hall, 1984, with substitution $(s,t) = (\Phi^{-1}(u), \Phi^{-1}(\alpha))$, and by (3.9.13), we have

$$
\begin{aligned}
&\int_0^1 \left[A_n\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right) - \alpha\right]^2 \omega(u) du \\
=& \int_{\mathbb{R}} \left[A_n\left(s,t\right) - \alpha\right]^2 \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
=& 2\int_{\mathbb{R}} \left\{A_n\left(s,t\right) - \mathbb{E}\left[A_n\left(s,t\right)\right]\right\} \left\{\mathbb{E}\left[A_n\left(s,t\right)\right] - \alpha\right\} \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
&+ \int_{\mathbb{R}} \left\{A_n\left(s,t\right) - \mathbb{E}\left[A_n\left(s,t\right)\right]\right\}^2 \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
&+ \int_{\mathbb{R}} \left\{\mathbb{E}\left[A_n\left(s,t\right)\right] - \alpha\right\}^2 \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
=& 2\int_{\mathbb{R}} \left\{\frac{1}{nh}\sum_{i=1}^n \left(\mathcal{B}_n\left[s,t;(s_i,t_i)\right] - \mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right)\right\} \\
&\left\{\mathbb{E}\left[A_n\left(s,t\right)\right] - \alpha\right\} \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
&+ \int_{\mathbb{R}} \left\{\frac{1}{nh}\sum_{i=1}^n \left(\mathcal{B}_n\left[s,t;(s_i,t_i)\right] - \mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right)\right\}^2 \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
&+ \int_{\mathbb{R}} \left\{\mathbb{E}\left[A_n\left(s,t\right)\right] - \alpha\right\}^2 \phi\left(s\right) \omega\left(\Phi(s)\right) ds \\
=& \frac{2}{nh}\sum_{i=1}^n \int_{\mathbb{R}} \left\{\mathcal{B}_n\left[s,t;(s_i,t_i)\right] - \mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right\} \\
&\left\{\mathbb{E}\left[A_n\left(s,t\right)\right] - \alpha\right\} \phi\left(s\right) \omega\left(\Phi(s)\right) ds
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{n^2h^2}\sum_{i=1}^{n}\int_{\mathbb{R}}\left\{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right\}^2\phi\left(s\right)\omega\left(\Phi(s)\right)ds\\
&+\frac{2}{n^2h^2}\sum_{1\leqslant k<l\leqslant n}\int_{\mathbb{R}}\left\{\mathcal{B}_n\left[s,t;(s_k,t_k)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_k,t_k)\right]\right]\right\}\\
&\left\{\mathcal{B}_n\left[s,t;(s_l,t_l)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_l,t_l)\right]\right]\right\}\phi\left(s\right)\omega\left(\Phi(s)\right)ds\\
&+\int_{\mathbb{R}}\left\{\mathbb{E}\left[A_n\left(s,t\right)\right]-\alpha\right\}^2\phi\left(s\right)\omega\left(\Phi(s)\right)ds\\
\doteq&2I_{n1}+I_{n2}+2I_{n3}+I_{n4}.
\end{aligned}
\tag{3.9.14}
$$

By Lemmas A.1.6 – A.1.9, and noticing that: 1) when $nh\to\infty$ as $n\to\infty$, the variance of I_{n2} is neglectable comparing to the variance of I_{n3}, i.e. $\mathcal{O}_P(n^{-\frac{3}{2}}h^{-1})=o_P(n^{-1}h^{-\frac{1}{2}})$; 2) when $nh\to\infty$ as $n\to\infty$, and $Y\not\perp_\alpha X$, the variance of I_{n3} is neglectable comparing to the variance of I_{n1}, i.e. $\mathcal{O}_P(n^{-1}h^{-\frac{1}{2}})=o_P(n^{-\frac{1}{2}})$; 3) when $h=\mathcal{O}(n^{-\beta})$, $\beta\in(0,\frac{1}{3})$, and $Y\not\perp_\alpha X$, the mean of I_{n2} is neglectable comparing to the second term in mean of I_{n4}, i.e. $\mathcal{O}(n^{-1}h^{-1})=o(h^2)$, we have

$$
\begin{aligned}
\widehat{D}_{\alpha 1}=&\mathcal{O}_P(n^{-\frac{1}{2}}h^2)\\
&+n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega)+o(n^{-1}h^{-1})+\mathcal{O}_P(n^{-\frac{3}{2}}h^{-1})\\
&+\mathcal{O}_P(n^{-1}h^{-\frac{1}{2}})\\
&+h^4M_{\perp_\alpha,4}(\omega)+o(h^4)\\
=&\mathcal{O}(n^{-1}h^{-1})+\mathcal{O}(h_n^4)+\mathcal{O}_P(n^{-1}h_n^{-\frac{1}{2}})+\mathcal{O}_P(n^{-\frac{1}{2}}h_n^2),\qquad \text{if } Y\perp_\alpha X
\end{aligned}
\tag{3.9.15}
$$

$$
\begin{aligned}
\widehat{D}_{\alpha 1}=&\mathcal{O}_P(n^{-\frac{1}{2}})\\
&+n^{-1}h^{-1}M_2(\omega)+o(n^{-1}h^{-1})+\mathcal{O}_P(n^{-\frac{3}{2}}h^{-1})\\
&+\mathcal{O}_P(n^{-1}h^{-\frac{1}{2}})
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]^2 \omega\left(u\right) du + h^2 M_{\not\perp_\alpha,4}(\omega) + o(h_n^2) \\
=&\int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]^2 \omega(u) du + \mathcal{O}(h^2) + \mathcal{O}_P(n^{-\frac{1}{2}}), \qquad \text{if } Y \not\perp_\alpha X \qquad (3.9.16)
\end{aligned}
$$

as $n \to \infty$, where $Y \perp_\alpha X$ ($Y \not\perp_\alpha X$) represents Y is quantile independent (dependent) with X at quantile level α.

Besides, by Lemma A.1.10, and considering "convergence almost surely" implies "convergence in probability", we may have

$$
\begin{aligned}
\widehat{D}_{\alpha 3} =&\mathcal{O}_{\text{a.s.}}\left(n^{-3/2}h^{-2}\log n(\log\log n)^{1/2}\right) \\
=&\mathcal{O}_P\left(n^{-3/2}h^{-2}\log n(\log\log n)^{1/2}\right) \\
=&o_P\left(n^{-1}\right), \qquad (3.9.17)
\end{aligned}
$$

as $n \to \infty$, when $h = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{4})$. Obviously, $\widehat{D}_{\alpha 3}$ is neglectable comparing to $\mathcal{O}_P$ terms of $\widehat{D}_{\alpha 1}$ in (3.9.15) and (3.9.16), for both cases $Y \perp_\alpha X$ and $Y \not\perp_\alpha X$, when $h = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{4})$.

Finally, for $\widehat{D}_{\alpha 2}$, by applying Cauchy-Schwarz inequality, we have

$$
\left|\widehat{D}_{\alpha 2}\right| \leqslant 2\left(\widehat{D}_{\alpha 1} \times \widehat{D}_{\alpha 3}\right)^{\frac{1}{2}}.
$$

Correspondingly, we have two cases.

1. If $Y \perp_\alpha X$, by results in (3.9.15) and (3.9.17), we have

$$
\left|\widehat{D}_{\alpha 2}\right| \leqslant \left\{o_P(n^{-2}h^{-1}) + o_P(n^{-1}h^4) + o_P(n^{-2}h^{-\frac{1}{2}}) + o_P(n^{-\frac{3}{2}}h^2)\right\}^{\frac{1}{2}}
$$

$$= \left\{ o_P(n^{-2}h^{-1}) + o_P(n^{-1}h^4) \right\}^{\frac{1}{2}}, \tag{3.9.18}$$

as $n \to \infty$, when $h = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{4})$. In (3.9.18), by applying continuous mapping theorem, we have

$$\begin{aligned} \left| \widehat{D}_{\alpha 2} \right| \leqslant & o_P \left(\left[\max\{n^{-2}h^{-1}, n^{-1}h^4\} \right]^{\frac{1}{2}} \right) \\ = & o_P \left(\max\{n^{-1}h^{-\frac{1}{2}}, n^{-\frac{1}{2}}h^2\} \right), \end{aligned}$$

as $n \to \infty$. This means $\left| \widehat{D}_{\alpha 2} \right|$ is always bounded in order by the $\mathcal{O}_P$ terms in (3.9.15). Hence $\widehat{D}_{\alpha 2}$ is neglectable comparing to $\widehat{D}_{\alpha 1}$.

2. If $Y \not\angle_\alpha X$, by results in (3.9.16) and (3.9.17), we have

$$\begin{aligned} \left| \widehat{D}_{\alpha 2} \right| & \leqslant \left\{ o_P(n^{-1}) + o_P(n^{-1}h^2) + o_P(n^{-\frac{3}{2}}) \right\}^{\frac{1}{2}} \\ & = \left\{ o_P(n^{-1}) \right\}^{\frac{1}{2}}, \end{aligned} \tag{3.9.19}$$

as $n \to \infty$, when $h_n = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{4})$. In (3.9.19), by applying continuous mapping theorem, we have

$$\left| \widehat{D}_{\alpha 2} \right| \leqslant o_P(n^{-\frac{1}{2}}),$$

as $n \to \infty$. This means $\left| \widehat{D}_{\alpha 2} \right|$ is always bounded in order by the $\mathcal{O}_P$ term in (3.9.16). Hence $\widehat{D}_{\alpha 2}$ is neglectable comparing to $\widehat{D}_{\alpha 1}$.

Combining all above analyses, we conclude $\widehat{D}_\alpha(X, Y; \omega, h)$ and $\widehat{D}_{\alpha 1}$ have the same asymptotic distribution.

Therefore, on one hand, when $Y \perp_\alpha X$, as in the analysis above (3.9.15), and using the results in Lemmas A.1.6 – A.1.9, we could have

$$\begin{aligned}\widehat{D}_\alpha(X, Y; \omega, h) \xrightarrow{d} & n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4 M_{\perp_\alpha,4}(\omega) \\ & + 2n^{-\frac{1}{2}}h^2\sigma_{\perp_\alpha,1}(\omega)Z_1 + \sqrt{2}n^{-1}h^{-\frac{1}{2}}\sigma_{\perp_\alpha,3}(\omega)Z_2,\end{aligned}$$

as $n \to \infty$, where Z_1 and Z_2 as standard normal distributions are naturally independent by the way how I_{n1} and I_{n3} are defined in (3.9.14). On the other hand, when $Y \not\perp_\alpha X$, as in the analysis above (3.9.16), and using the results in Lemmas A.1.6 – A.1.9, we could have

$$\begin{aligned}\widehat{D}_\alpha(X, Y; \omega, h) \xrightarrow{d} & \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right]^2 \omega(u)du + h^2 M_{\not\perp_\alpha,4}(\omega) \\ & + 2n^{-\frac{1}{2}}\sigma_{\not\perp_\alpha,1}(\omega)Z_3,\end{aligned}$$

as $n \to \infty$, where Z_3 is a standard normal distribution. □

3.9.3 Proof for Section 3.5

Proof of Theorem 3.5.1. $\forall j \in \mathcal{M}_\alpha$, noticing that $h^2 M_{\not\perp_\alpha,4}(\omega) + o(h^2) = o(n^{-\kappa_{\alpha 1}})$, by the selection of $t_{n,\alpha}$, we have $0 < t_{n,\alpha} \leqslant \frac{c_{t,\alpha}}{c_{\alpha 1}} D_\alpha(X_j, Y; \omega)$, then we have

$$\left[D_\alpha(X_j, Y; \omega) + h^2 M_{\not\perp_\alpha,4}(\omega) + o(h^2)\right] - t_{n,\alpha}$$

$$
\begin{aligned}
&\geqslant \left(1-\frac{c_{t,\alpha}}{c_{\alpha 1}}\right) D_\alpha(X_j,Y;\omega)+h^2 M_{\not\perp_\alpha,4}(\omega)+o(h^2)\\
&\geqslant \frac{1}{2}\left(1-\frac{c_{t,\alpha}}{c_{\alpha 1}}\right) D_\alpha(X_j,Y;\omega)>0,
\end{aligned}
$$

with large enough n. When $j \in \mathcal{M}_\alpha$, we have $Y \not\perp_\alpha X_j$, then by results in Theorem 3.4.1 and normal tail bound, we have

$$
\begin{aligned}
&\mathcal{P}\left(j \in \widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right)\\
=&\mathcal{P}\left(\widehat{D}_\alpha(X_j,Y;\omega,h) \geqslant t_{n,\alpha}\right)\\
\geqslant& 1-C_{*1}\exp\left[-\frac{1}{2}\left(\frac{t_{n,\alpha}-[D_\alpha(X_j,Y;\omega)+h^2M_{\not\perp_\alpha,4}(\omega)+o(h^2)]}{2n^{-\frac{1}{2}}\sigma_{\not\perp_\alpha,1}(\omega)}\right)^2\right]\\
\geqslant& 1-C_{*1}\exp\left[-\frac{1}{32}\left(1-\frac{c_{t,\alpha}}{c_{\alpha 1}}\right)^2 n\left(\frac{D_\alpha(X_j,Y;\omega)}{\sigma_{\not\perp_\alpha,1}(\omega)}\right)^2\right],
\end{aligned}
$$

for some positive constant C_{*1}. Then, from result in (3.8.4), we have

$$
\begin{aligned}
\sigma^2_{\not\perp_\alpha,1}(\omega) \leqslant& \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\mathcal{C}^2_{2|1}\left(\alpha \mid u\right)\right]\left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]^2\omega^2\left(u\right)du\\
&+\left\{\int_0^1 \frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(u,\alpha\right)\left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]\omega\left(u\right)du\right\}^2\\
&+2\int_0^1 \frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(u,\alpha\right)\left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]\omega\left(u\right)du\\
&\int_0^1\left\{\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\mathcal{C}^2_{2|1}\left(\alpha \mid u\right)-\left[\mathcal{C}\left(u,\alpha\right)-\alpha u\right]\frac{\partial^2\mathcal{C}}{\partial u^2}\left(u,\alpha\right)\right\}\\
&\left[\mathcal{C}_{2|1}\left(\alpha \mid u\right)-\alpha\right]\omega\left(u\right)du\\
\leqslant& \int_0^1 \omega^2\left(u\right)du+\left[K_{00}\sup_{u\in(0,1)}\frac{\phi^2\left(\Phi^{-1}(u)\right)}{u(1-u)}\int_0^1\frac{\omega\left(u\right)}{\phi^2\left(\Phi^{-1}(u)\right)}du\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&+2\left[K_{00}\sup_{u\in(0,1)}\frac{\phi^2\left(\Phi^{-1}(u)\right)}{u(1-u)}\int_0^1\frac{\omega\left(u\right)}{\phi^2\left(\Phi^{-1}(u)\right)}du\right]\\
&\left[\sup_{u\in(0,1)}\left\{\phi^2\left(\Phi^{-1}(u)\right)\left[1+\frac{K_1}{u(1-u)}\right]\right\}\int_0^1\frac{\omega\left(u\right)}{\phi^2\left(\Phi^{-1}(u)\right)}du\right] \qquad (3.9.20)\\
\leqslant&\int_0^1\omega^2\left(u\right)du+C_{*2}\left[\int_0^1\frac{\omega\left(u\right)}{\phi^2\left(\Phi^{-1}(u)\right)}du\right]^2 \qquad (3.9.21)\\
\leqslant&C_{*3}, \qquad (3.9.22)
\end{aligned}
$$

where C_{*2} and C_{*3} are some positive constants, (3.9.20) follows from $\mathcal{C}_{2|1}\left(\alpha \mid u\right) \in [0,1]$ and Condition 3.3; (3.9.21) follows from $\frac{\phi^2\left(\Phi^{-1}(u)\right)}{u(1-u)}$ and $\phi^2\left(\Phi^{-1}(u)\right)$ are bounded over $(0,1)$; (3.9.22) follows from Condition 3.4 and $\omega\left(u\right)$ is bounded over $(0,1)$. With some positive constants C_{*4} and C_{*5}, under Condition 3.5(a), this induces

$$
\begin{aligned}
\mathcal{P}\left(j\in\widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right)\geqslant&1-C_{*1}\exp\left[-C_{*4}nD_\alpha^2(X_j,Y;\omega)\right]\\
\geqslant&1-C_{*1}\exp\left(-C_{*5}n^{1-2\kappa_{\alpha1}}\right).
\end{aligned}
$$

Finally, by Fréchet inequality, when n is large enough, we have

$$
\begin{aligned}
\mathcal{P}\left(\mathcal{M}_\alpha\subseteq\widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right)=&\mathcal{P}\left(\bigcap_{j\in\mathcal{M}_\alpha}\left\{j\in\widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right\}\right)\\
\geqslant&\sum_{j\in\mathcal{M}_\alpha}\mathcal{P}\left(j\in\widehat{\mathcal{M}}_\alpha\left(t_{n,\alpha}\right)\right)-\left(\mathcal{S}_\alpha-1\right)\\
\geqslant&1-\mathcal{O}\left(\mathcal{S}_\alpha\exp\left(-C_{\alpha1}n^{1-2\kappa_{\alpha1}}\right)\right),
\end{aligned}
$$

for some positive constant $C_{\alpha1}$. The rest of the proof follows directly from the above results. □

Proof of Theorem 3.5.2. Under Condition 3.5(b), we have

$$
\begin{aligned}
&\mathcal{P}\left(\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}\leqslant 0\right)\\
\leqslant&\mathcal{P}\Bigg(\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}\\
\leqslant&\min_{j\in\mathcal{M}_\alpha}\left\{D_\alpha(X_j,Y;\omega)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{D_\alpha(X_j,Y;\omega)\right\}-c_{\alpha 2}n^{-\kappa_{\alpha 2}}\Bigg)\\
=&\mathcal{P}\Bigg(\left[\min_{j\in\mathcal{M}_\alpha}\left\{D_\alpha(X_j,Y;\omega)\right\}-\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}\right]\\
&+\left[\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{D_\alpha(X_j,Y;\omega)\right\}\right]\geqslant c_{\alpha 2}n^{-\kappa_{\alpha 2}}\Bigg).
\end{aligned}
$$

Let $j_1\doteq\operatorname{argmin}_{j\in\mathcal{M}_\alpha}\widehat{D}_\alpha(X_j,Y;\omega,h)$ and $j_2\doteq\operatorname{argmax}_{j\in\mathcal{M}_\alpha^c}\widehat{D}_\alpha(X_j,Y;\omega,h)$, then we could have

$$
\begin{aligned}
D_\alpha(X_{j_1},Y;\omega)&\geqslant\min_{j\in\mathcal{M}_\alpha}\left\{D_\alpha(X_j,Y;\omega)\right\},\\
D_\alpha(X_{j_2},Y;\omega)&\leqslant\max_{j\in\mathcal{M}_\alpha^c}\left\{D_\alpha(X_j,Y;\omega)\right\}.
\end{aligned}
$$

It induces

$$
\begin{aligned}
&\mathcal{P}\left(\min_{j\in\mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}-\max_{j\in\mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j,Y;\omega,h)\right\}\leqslant 0\right)\\
\leqslant&\mathcal{P}\Bigg(\left[D_\alpha(X_{j_1},Y;\omega)-\widehat{D}_\alpha(X_{j_1},Y;\omega,h)\right]\\
&+\left[\widehat{D}_\alpha(X_{j_2},Y;\omega,h)-D_\alpha(X_{j_2},Y;\omega)\right]\geqslant c_{\alpha 2}n^{-\kappa_{\alpha 2}}\Bigg)\\
\leqslant&\mathcal{P}\Bigg(\left\{\left|\widehat{D}_\alpha(X_{j_1},Y;\omega,h)-D_\alpha(X_{j_1},Y;\omega)\right|\geqslant\frac{1}{2}c_{\alpha 2}n^{-\kappa_{\alpha 2}}\right\}\\
&\bigcup\left\{\left|\widehat{D}_\alpha(X_{j_2},Y;\omega,h)-D_\alpha(X_{j_2},Y;\omega)\right|\geqslant\frac{1}{2}c_{\alpha 2}n^{-\kappa_{\alpha 2}}\right\}\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
\leqslant & \mathcal{P}\left(\left\{\left|\widehat{D}_\alpha(X_{j_1}, Y; \omega, h) - D_\alpha(X_{j_1}, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\}\right) \\
& + \mathcal{P}\left(\left\{\left|\widehat{D}_\alpha(X_{j_2}, Y; \omega, h) - D_\alpha(X_{j_2}, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\}\right) \\
\leqslant & 2\mathcal{P}\left(\max_{j \in [p]} \left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right|\right\} \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right) \\
= & 2\mathcal{P}\left(\bigcup_{j=1}^{p} \left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\}\right) \\
\leqslant & 2\sum_{j=1}^{p} \mathcal{P}\left(\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right). \qquad (3.9.23)
\end{aligned}
$$

Then we consider two different cases

1. When $Y \not\perp_\alpha X_j$, noticing that $h^2 M_{\not\perp_\alpha,4}(\omega) + o(h^2) = o(n^{-\kappa_{\alpha 2}})$, we could have

$$
\begin{aligned}
& \left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\} \\
\subseteq & \left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - \left[D_\alpha(X_j, Y; \omega) + h^2 M_{\not\perp_\alpha,4}(\omega) + o(h^2)\right]\right| \geqslant \frac{1}{4} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\},
\end{aligned}
$$

with large enough n. Correspondingly, by results in Theorem 3.4.1 and normal tail bound, we have

$$
\begin{aligned}
& \mathcal{P}\left(\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right) \\
\leqslant & C_{*6} \exp\left[-\frac{1}{2}\left(\frac{1}{8} c_{\alpha 2} \frac{n^{\frac{1}{2} - \kappa_{\alpha 2}}}{\sigma_{\not\perp_\alpha,1}(\omega)}\right)^2\right],
\end{aligned}
$$

where C_{*6} is some positive constant. Besides, we have $\sigma^2_{\not\perp_\alpha,1}(\omega)$ is bounded by (3.9.22). This induces

$$
\mathcal{P}\left(\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right) \leqslant C_{*6} \exp\left(-C_{*7} n^{1 - 2\kappa_{\alpha 2}}\right), \qquad (3.9.24)
$$

where C_{*7} is some positive constant.

2. When $Y \perp_\alpha X_j$, noticing that $n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4 M_{\perp_\alpha,4}(\omega) + o(n^{-1}h^{-1}) + o(h^4) = o(n^{-\kappa_{\alpha 2}})$, we could have

$$\begin{aligned}
&\left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\} \\
\subseteq &\left\{\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - \left[D_\alpha(X_j, Y; \omega) + n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4 M_{\perp_\alpha,4}(\omega)\right.\right.\right. \\
&\left.\left.\left. + o(n^{-1}h^{-1}) + o(h^4)\right]\right| \geqslant \frac{1}{4} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right\},
\end{aligned}$$

with large enough n. Correspondingly, define

$$\sigma_{n,\perp_\alpha} = \begin{cases} 2n^{-\frac{1}{2}} h^2 \sigma_{\perp_\alpha,1}(\omega) & \text{if } \beta \in (0, \frac{1}{5}), \\ \sqrt{2} n^{-1} h^{-\frac{1}{2}} \sigma_{\perp_\alpha,3}(\omega) & \text{if } \beta \in (\frac{1}{5}, \frac{1}{4}), \\ n^{-\frac{9}{10}} \sqrt{4c_h^4 \sigma^2_{\perp_\alpha,1}(\omega) + 2c_h^{-1} \sigma^2_{\perp_\alpha,3}(\omega)} & \text{if } \beta = \frac{1}{5}, h \sim c_h n^{-\frac{1}{5}}, \end{cases} \tag{3.9.25}$$

where c_h is some positive constant. Correspondingly, by results in Theorem 3.4.1 and normal tail bound, we have

$$\mathcal{P}\left(\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right) \leqslant C_{*8} \exp\left[-\frac{1}{32}\left(c_{\alpha 2}\right)^2 \frac{n^{-2\kappa_{\alpha 2}}}{\sigma^2_{n,\perp_\alpha}}\right],$$

where C_{*8} is some positive constant. Besides, we could have $\sigma_{n,\perp_\alpha} = o(n^{-\frac{1}{2}})$, which induces

$$\mathcal{P}\left(\left|\widehat{D}_\alpha(X_j, Y; \omega, h) - D_\alpha(X_j, Y; \omega)\right| \geqslant \frac{1}{2} c_{\alpha 2} n^{-\kappa_{\alpha 2}}\right) \leqslant C_{*8} \exp\left(-C_{*9} n^{1-2\kappa_{\alpha 2}}\right), \tag{3.9.26}$$

where C_{*9} is some positive constant.

Combining the results in (3.9.24) and (3.9.26) back into (3.9.23), we have

$$\mathcal{P}\left(\min_{j \in \mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j, Y; \omega, h)\right\} - \max_{j \in \mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j, Y; \omega, h)\right\} \leqslant 0\right) \leqslant \mathcal{O}\left(p \exp\left(-C_{\alpha 2} n^{1-2\kappa_{\alpha 2}}\right)\right),$$

for some positive constant $C_{\alpha 2}$, or equivalently,

$$\begin{aligned}&\mathcal{P}\left(\min_{j \in \mathcal{M}_\alpha}\left\{\widehat{D}_\alpha(X_j, Y; \omega, h)\right\} - \max_{j \in \mathcal{M}_\alpha^c}\left\{\widehat{D}_\alpha(X_j, Y; \omega, h)\right\} > 0\right)\\ >&1 - \mathcal{O}\left(p \exp\left(-C_{\alpha 2} n^{1-2\kappa_{\alpha 2}}\right)\right).\end{aligned}$$

The rest of the proof follows directly from the proof of Theorem 3 in W. Liu et al., 2022. □

3.9.4 Proofs for Section 3.6

Proof of Corollary 3.6.1. By Theorem 3.4.1, when $Y \perp_\alpha X$, we have

$$\frac{\widehat{D}_\alpha(X, Y; \omega, h)}{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4 M_{\perp_\alpha,4}(\omega)} \xrightarrow{d} \frac{\sqrt{4n^{-1}h^4\sigma^2_{\perp_\alpha,1}(\omega) + 2n^{-2}h^{-1}\sigma^2_{\perp_\alpha,3}(\omega)}}{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega) + h^4 M_{\perp_\alpha,4}(\omega)} Z + 1,$$

where Z represents some standard Gaussian random variable. Then with Delta method, it is direct to have

$$\log\left[\frac{\widehat{D}_\alpha(X,Y;\omega,h)}{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega)+h^4M_{\perp_\alpha,4}(\omega)}\right]\xrightarrow{d}\frac{\sqrt{4n^{-1}h^4\sigma^2_{\perp_\alpha,1}(\omega)+2n^{-2}h^{-1}\sigma^2_{\perp_\alpha,3}(\omega)}}{n^{-1}h^{-1}M_{\perp_\alpha,2}(\omega)+h^4\,M_{\perp_\alpha,4}(\omega)}Z,$$

which completes the proof. □

Proof of Theorem 3.6.1. For FDR control, we mainly consider $\widetilde{\mathcal{M}}^c_\alpha = \{j\in[p]:Y\perp_\alpha X_j\}$, $\tilde{p}_{0,\alpha} = \left|\widetilde{\mathcal{M}}^c_\alpha\right|$, the cardinality of $\widetilde{\mathcal{M}}^c_\alpha$, and $\tilde{p}_{0,\alpha}/p_{0,\alpha}\to 1$ instead. By definition, we have

$$\begin{aligned}
\mathrm{FDR}\left(T_\gamma(\alpha);\alpha\right) =&\mathbb{E}\left[\mathrm{FDP}\left(T_\gamma(\alpha);\alpha\right)\right]\\
=&\mathbb{E}\left[\frac{\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}{\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}\right]\\
=&\mathbb{E}\left[\frac{\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}{1+\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}\right.\\
&\left.\cdot\frac{1+\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}{\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}\right]\\
\leqslant&\mathbb{E}\left[\frac{\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}{1+\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}\right.\\
&\left.\cdot\frac{1+\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}{\#\{j:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}\right]\\
\leqslant&\gamma\mathbb{E}\left[\frac{\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}{1+\#\{j\in\widetilde{\mathcal{M}}^c_\alpha:\widehat{D}^\dagger_\alpha(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}\right],
\end{aligned}\tag{3.9.27}$$

where the last inequality follows from how $T_\gamma(\alpha)$ is selected.

In the rest of the proof, we find the upper bound of expectation term in (3.9.27) by super-martingale theories.

W.L.O.G., we assume

$$\left|\widehat{D}^{\dagger}_{\alpha}(X_1, Y; \omega, h)\right| \geqslant \left|\widehat{D}^{\dagger}_{\alpha}(X_2, Y; \omega, h)\right| \geqslant \cdots \geqslant \left|\widehat{D}^{\dagger}_{\alpha}(X_p, Y; \omega, h)\right| > 0,$$

and we further define $\left|\widehat{D}^{\dagger}_{\alpha}(X_{p+1}, Y; \omega, h)\right| = 0$. To determine threshold $T_\gamma(\alpha)$, different values of t starting from the smallest, $\left|\widehat{D}^{\dagger}_{\alpha}(X_{p+1}, Y; \omega, h)\right|$, to the largest, $\left|\widehat{D}^{\dagger}_{\alpha}(X_1, Y; \omega, h)\right|$, are tried in (3.6.5). In this process, $T_\gamma(\alpha)$ is actually the stopping time. Moreover, since value of $\frac{\#\{j \in \widetilde{\mathcal{M}}^c_\alpha : \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) \geqslant t\}}{1 + \#\{j \in \widetilde{\mathcal{M}}^c_\alpha : \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) \leqslant -t\}}$ would not change for $t = \left|\widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h)\right|$, $j \notin \widetilde{\mathcal{M}}^c_\alpha$, it would not matter if we only consider $t = \left|\widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h)\right|$, $j \in \widetilde{\mathcal{M}}^c_\alpha$. Correspondingly, W.L.O.G., we denote the index set of $\widetilde{\mathcal{M}}^c_\alpha$ as $\{1, \ldots, \tilde{p}_{0,\alpha}\}$, and assume

$$\left|\widehat{D}^{\dagger}_{\alpha}(X_1, Y; \omega, h)\right| \geqslant \left|\widehat{D}^{\dagger}_{\alpha}(X_2, Y; \omega, h)\right| \geqslant \cdots \geqslant \left|\widehat{D}^{\dagger}_{\alpha}(X_{\tilde{p}_{0,\alpha}}, Y; \omega, h)\right| > 0,$$

and further define $\left|\widehat{D}^{\dagger}_{\alpha}(X_{\tilde{p}_{0,\alpha}+1}, Y; \omega, h)\right| = 0$, and consider $T_\gamma(\alpha)$ as the stopping time. In other words, for $k = \tilde{p}_{0,\alpha} + 1, \tilde{p}_{0,\alpha}, \cdots, 1$, we define

$$\begin{aligned} L_k(\alpha) &= \frac{\#\left\{j \in \widetilde{\mathcal{M}}^c_\alpha : \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) \geqslant \left|\widehat{D}^{\dagger}_{\alpha}(X_k, Y; \omega, h)\right|\right\}}{1 + \#\left\{j \in \widetilde{\mathcal{M}}^c_\alpha : \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) \leqslant -\left|\widehat{D}^{\dagger}_{\alpha}(X_k, Y; \omega, h)\right|\right\}} \\ &= \frac{\#\left\{j \in \widetilde{\mathcal{M}}^c_\alpha : j \leqslant k \text{ and } \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) \geqslant 0\right\}}{1 + \#\left\{j \in \widetilde{\mathcal{M}}^c_\alpha : j \leqslant k \text{ and } \widehat{D}^{\dagger}_{\alpha}(X_j, Y; \omega, h) < 0\right\}}. \end{aligned}$$

$\forall j \in \widetilde{\mathcal{M}}^c_\alpha$, by definition of $B_j(\alpha)$, and $\widetilde{S}^c_k(\alpha) \doteq \sum_{j=1}^k B_j(\alpha)$, we could have

$$L_k(\alpha) = \frac{\sum_{j=1}^k [1 - B_j(\alpha)]}{1 + \sum_{j=1}^k B_j(\alpha)}$$

$$
\begin{aligned}
=&\frac{k-\widetilde{S}_k^c(\alpha)}{1+\widetilde{S}_k^c(\alpha)}\\
=&\frac{k+1}{1+\widetilde{S}_k^c(\alpha)}-1.
\end{aligned}
$$

Denote the σ-field generated by $\left\{\sum_{j=1}^{k} B_j(\alpha), B_{k+1}(\alpha), \cdots, B_{\tilde{p}_{0,\alpha}+1}(\alpha)\right\}$ as $\mathcal{F}_{k,\alpha}$. Correspondingly, $T_\gamma(\alpha)$ would be a stopping time in reverse time (from $\tilde{p}_{0,\alpha}+1$ to 1) with respect to the random process $\{L_k(\alpha)\}_{k=\tilde{p}_{0,\alpha}+1}^{1}$ and its backward filtration $\mathcal{F}_{\tilde{p}_{0,\alpha}+1,\alpha} \subset \cdots \subset \mathcal{F}_{1,\alpha}$.

By Corollary 3.6.1, when $j \in \widetilde{\mathcal{M}}_\alpha^c$, it is easy to notice $\widehat{D}_\alpha^\dagger(X_j, Y; \omega, h)$'s are identically distributed, hence $\{B_1(\alpha), \cdots, B_k(\alpha)\}$ are identically distributed with respect to $\mathcal{F}_{k,\alpha}$. So, via Lemma 8 in Tong et al., 2023, we could have $\mathcal{P}\left(B_k(\alpha)=1 \mid \mathcal{F}_{k,\alpha}\right)=\frac{\widetilde{S}_k^c(\alpha)}{k}$. Consequently, on one hand, if $\widetilde{S}_k^c(\alpha)=0$, it means $\widetilde{S}_{k-1}^c(\alpha)=0$, hence $L_{k-1}(\alpha)=k-1<k=L_k(\alpha)$. On the other hand, if $\widetilde{S}_k^c(\alpha)>0$, then we have

$$
\begin{aligned}
&\mathbb{E}\left[L_{k-1}(\alpha) \mid \mathcal{F}_{k,\alpha}\right]\\
=&\left[\frac{k}{1+\widetilde{S}_k^c(\alpha)}-1\right]\mathcal{P}\left(B_k(\alpha)=0 \mid \mathcal{F}_{k,\alpha}\right)+\left[\frac{k}{1+\widetilde{S}_k^c(\alpha)-1}-1\right]\mathcal{P}\left(B_k(\alpha)=1 \mid \mathcal{F}_{k,\alpha}\right)\\
=&\left[\frac{k}{1+\widetilde{S}_k^c(\alpha)}-1\right]\frac{k-\widetilde{S}_k^c(\alpha)}{k}+\left[\frac{k}{\widetilde{S}_k^c(\alpha)}-1\right]\frac{\widetilde{S}_k^c(\alpha)}{k}\\
=&\frac{k+1}{1+\widetilde{S}_k^c(\alpha)}-1\\
=&L_k(\alpha).
\end{aligned}
$$

Therefore, $\{L_k(\alpha)\}_{k=\tilde{p}_{0,\alpha}+1}^{1}$ is a super-martingale with respect to $\{\mathcal{F}_{k,\alpha}\}$. Then, by Optional Stopping Time theorem, we could have

$$\begin{aligned}\mathbb{E}\left[\frac{\#\{j\in\widetilde{\mathcal{M}}_\alpha^c:\widehat{D}_\alpha^\dagger(X_j,Y;\omega,h)\geqslant T_\gamma(\alpha)\}}{1+\#\{j\in\widetilde{\mathcal{M}}_\alpha^c:\widehat{D}_\alpha^\dagger(X_j,Y;\omega,h)\leqslant -T_\gamma(\alpha)\}}\right]\leqslant&\mathbb{E}\left[L_{\tilde{p}_{0,\alpha}}\right]\\ =&\mathbb{E}\left[\frac{\tilde{p}_{0,\alpha}-\widetilde{S}_\alpha^c}{1+\widetilde{S}_\alpha^c}\right]\\ =&\mathbb{E}\left[\frac{1-\frac{\widetilde{S}_\alpha^c}{\tilde{p}_{0,\alpha}}}{\frac{1}{\tilde{p}_{0,\alpha}}+\frac{\widetilde{S}_\alpha^c}{\tilde{p}_{0,\alpha}}}\right],\end{aligned}\tag{3.9.28}$$

where $\widetilde{S}_\alpha^c \doteq \sum_{j=1}^{\tilde{p}_{0,\alpha}} B_j(\alpha)$. Then, by letting $\Delta S_\alpha^c = S_\alpha^c - \widetilde{S}_\alpha^c$, we have

$$\begin{aligned}\text{Var}\left[\frac{\widetilde{S}_\alpha^c}{p_{0,\alpha}}\right]=&\text{Var}\left[\frac{S_\alpha^c-\Delta S_\alpha^c}{p_{0,\alpha}}\right]\\ =&\text{Var}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]+\text{Var}\left[\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]-2\mathbb{E}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]+2\mathbb{E}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]\mathbb{E}\left[\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]\\ \leqslant&\text{Var}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]+\text{Var}\left[\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]+2\mathbb{E}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]+2\mathbb{E}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]\mathbb{E}\left[\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]\\ \leqslant&\text{Var}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]+\text{Var}\left[\frac{\Delta S_\alpha^c}{p_{0,\alpha}}\right]+4\frac{p_{0,\alpha}-\tilde{p}_{0,\alpha}}{p_{0,\alpha}} \qquad (3.9.29)\\ \leqslant&\text{Var}\left[\frac{S_\alpha^c}{p_{0,\alpha}}\right]+\frac{1}{4}\left[\frac{p_{0,\alpha}-\tilde{p}_{0,\alpha}}{p_{0,\alpha}}\right]^2+4\frac{p_{0,\alpha}-\tilde{p}_{0,\alpha}}{p_{0,\alpha}}\\ =&o(1)\end{aligned}$$

where (3.9.29) follows from the truth $S_\alpha^c \leqslant p_{0,\alpha}$ and $\Delta S_\alpha^c \leqslant p_{0,\alpha} - \tilde{p}_{0,\alpha}$, and the last equality follows from Condition 3.6. Hence, $\text{Var}\left[\frac{\widetilde{S}_\alpha^c}{\tilde{p}_{0,\alpha}}\right] = \left(\frac{p_{0,\alpha}}{\tilde{p}_{0,\alpha}}\right)^2 \text{Var}\left[\frac{\widetilde{S}_\alpha^c}{p_{0,\alpha}}\right] = o(1)$. Besides, by Corollary 3.6.1, it is obvious that $\mathbb{E}\left[\frac{\widetilde{S}_\alpha^c}{\tilde{p}_{0,\alpha}}\right] \to \frac{1}{2}$, and by Markov's inequality, we have $\frac{\widetilde{S}_\alpha^c}{\tilde{p}_{0,\alpha}} \xrightarrow{p} \frac{1}{2}$, as $n \to \infty$. Then since

$f(x) = \frac{1-x}{\frac{1}{\tilde{p}_{0,\alpha}}+x}$ is bounded and continuous, we have

$$\mathbb{E}\left[\frac{1-\frac{\widetilde{S}_{\alpha}^{c}}{\tilde{p}_{0,\alpha}}}{\frac{1}{\tilde{p}_{0,\alpha}}+\frac{\widetilde{S}_{\alpha}^{c}}{\tilde{p}_{0,\alpha}}}\right] \to \mathbb{E}\left[\frac{1-\frac{1}{2}}{0+\frac{1}{2}}\right] = 1. \tag{3.9.30}$$

The rest of the proof follows directly from (3.9.27), (3.9.28), and (3.9.30),

$$\limsup_{n\to\infty} \mathrm{FDR}\left(T_{\gamma}(\alpha);\alpha\right) \leqslant \gamma \limsup_{n\to\infty} \mathbb{E}\left[\frac{1-\frac{\widetilde{S}_{\alpha}^{c}}{\tilde{p}_{0,\alpha}}}{\frac{1}{\tilde{p}_{0,\alpha}}+\frac{\widetilde{S}_{\alpha}^{c}}{\tilde{p}_{0,\alpha}}}\right] = \gamma.$$

□

Chapter 4

Applications to the U.S. 2020 Economic Data

The economic landscape of the United States in 2020 underwent significant shifts due to the unprecedented impact of the COVID-19 pandemic (Polyakova et al., 2020). This period was marked by substantial economic disruptions, with varying degrees of resilience and recovery across states (Altig et al., 2020; Thakur et al., 2020). Understanding the underlying factors driving these disparities is essential for formulating effective economic policies and strategies (Abedi et al., 2021). Moreover, the inherent disparities among states, such as differences in population density, urbanization, health infrastructure, and economic resources, provide critical context for analyzing the structure of the U.S. economy during this tumultuous period (Carethers, 2021; Paul et al., 2021).

To address these issues, we analyze a dataset containing macroeconomic variables for all 50 states and the District of Columbia ($n = 51$). This dataset comprises 163 variables ($p = 163$) categorized into six domains: (1) housing information, (2) employment status, (3) income and benefits, (4) health insurance

Table 4.1: Top 10 selected variables by DR-WD_1-SIS.

Variable	Name
X_{17}	Proportion of population 16 years and over in unemployed civilian labor force
X_{36}	Proportion of workers 16 years and over working from home
X_{42}	Proportion of civilian employed population 16 years and over in natural resources, construction, and maintenance occupations
X_{45}	Proportion of civilian employed population 16 years and over in construction industries
X_{48}	Proportion of civilian employed population 16 years and over in retail trade industries
X_{57}	Proportion of civilian employed population 16 years and over private wage and salary workers
X_{79}	Proportion of civilian noninstitutionalized population with no health insurance coverage
X_{85}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with health insurance coverage
X_{88}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with no health insurance coverage
X_{93}	Proportion of civilian noninstitutionalized population 19 to 64 years in unemployed labor force with no health insurance coverage

coverage, (5) poverty levels, and (6) demographic characteristics, which is available at the official website of U.S. Census Bureau, https://data.census.gov. Details of the variables are provided in Table A.2 in the appendices. The primary objective of this study is to identify the most influential factors associated with the growth rate of Gross Domestic Product (GDP) from 2019 to 2020, which is available at the official website of Bureau of Economic Analysis, U.S. Department of Commerce, https://www.bea.gov.

4.1 Feature Screening with DR-WD_r-SIS

As discussed in Chapter 2, DR-WD_r-SIS is a general feature screening method fit to high-dimensional heterogeneous data. We apply DR-WD_r-SIS to GDP growth rate and each variable, ranking their marginal

Table 4.2: Top 10 selected variables by DR-WD$_2$-SIS.

Variable	Name
X_{17}	Proportion of population 16 years and over in unemployed civilian labor force
X_{32}	Proportion of workers 16 years and over commuting to work with car, truck, or van (carpooled)
X_{36}	Proportion of workers 16 years and over working from home
X_{42}	Proportion of civilian employed population 16 years and over in natural resources, construction, and maintenance occupations
X_{45}	Proportion of civilian employed population 16 years and over in construction industries
X_{46}	Proportion of civilian employed population 16 years and over in manufacturing industries
X_{48}	Proportion of civilian employed population 16 years and over in retail trade industries
X_{79}	Proportion of civilian noninstitutionalized population with no health insurance coverage
X_{85}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with health insurance coverage
X_{88}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with no health insurance coverage

significance and select top 10 variables which is highly related to GDP growth. We select variables with both DR-WD$_1$-SIS and DR-WD$_2$-SIS, and the results are summarized in Table 4.1 and Table 4.2. There are 8 variables which are selected by both methods, which fall into three categories: (1) employment status ($X_{17}, X_{42}, X_{45}, X_{48}$), (2) commuting to work (X_{36}), and (3) health insurance coverage (X_{79}, X_{85}, X_{88}). As a general screening method, DR-WD$_r$-SIS could screen out redundant features and reduce the dimension effectively. However, with data resources from 51 states, it would be meaningful if we construct comparison between the states with high GDP growth rate and low GDP growth rate during the pandemic. This helps understand the vital factors stablize economy during tumultuous period. Hence we consider feature screening with different quantile levels of GDP growth rate in Section 4.2.

Low GDP growth rate state	High GDP growth rate state
Hawaii	Idaho
Louisiana	Arizona
Wyoming	Utah
North Dakota	Maine
Nevada	Arkansas
Illinois	South Dakota
Connecticut	Washington
Pennsylvania	Montana
Oklahoma	North Carolina
Alaska	Tennessee
Michigan	Nebraska
New Jersey	Mississippi

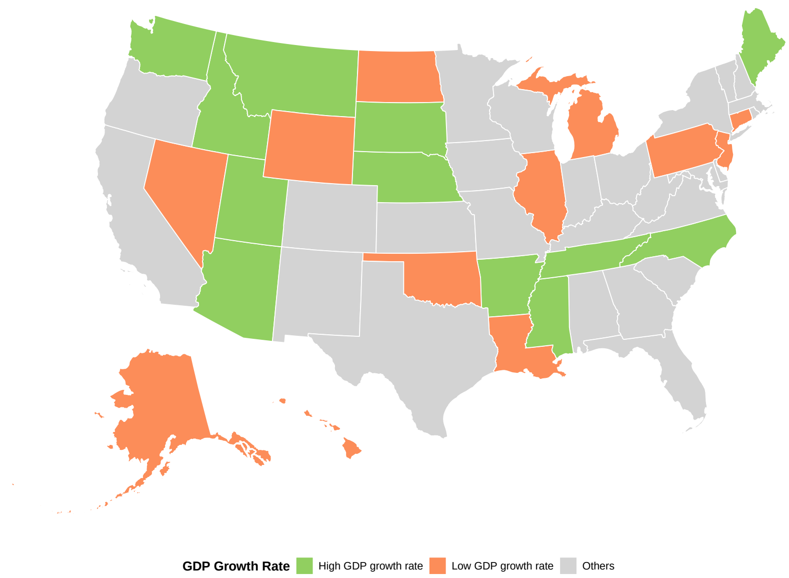


Figure 4.1: Top 12 and bottom 12 states in GDP growth rate from 2019 to 2020.

4.2 Feature Screening with QC-FDR

We focus on the $\alpha = 0.25, 0.75$ quantile levels among 51 samples, which characterizes top 12 states and the bottom 12 states in terms of GDP growth rate as shown in Figure 4.1. States labeled in green (orange) are the highest (lowest) 12 among 51 samples in GDP growth rate. Specifically, we employ the proposed QC-FDR method to screen important macroeconomic variables at quantile levels $\alpha = 0.25$ and $\alpha = 0.75$. The weight function is set as $\omega(u) = \phi^2(\Phi^{-1}(u))$, and the target FDR level is chosen as $\gamma = 0.15$. The bandwidth parameter is determined using the optimal bandwidth selection procedure detailed in Corollary3.3.1.

Using the QC-FDR at the $\alpha = 0.25$ quantile level, we identified seven key macroeconomic variables for the 12 states with the lowest GDP growth rate, as summarized in Table 4.3. These variables fall into

Table 4.3: Selected variables at quantile level $\alpha = 0.25$.

Variable	Name
X_{17}	Proportion of population 16 years and over in unemployed civilian labor force
X_{36}	Proportion of workers 16 years and over working from home
X_{42}	Proportion of civilian employed population 16 years and over in natural resources, construction, and maintenance occupations
X_{49}	Proportion of civilian employed population 16 years and over in transportation and warehousing, and utilities industries
X_{57}	Proportion of civilian employed population 16 years and over private wage and salary workers
X_{58}	Proportion of civilian employed population 16 years and over government workers
X_{107}	Percentage of families and people whose income in the past 12 months is below the poverty level among families with female householder, no spouse present with related children of the householder under 5 years only

three main categories: (1) employment status (X_{17}, X_{42}, X_{49}, X_{57}, X_{58}), (2) commuting to work (X_{36}), and (3) poverty level (X_{107}). However, when implementing QC-FDR at the $\alpha = 0.75$ quantile level, none variable is selected. To investigate the relationship between GDP growth rate and these variables, we employed local linear quantile regressions (Yu and Jones, 1998) at the $\alpha = 0.25, 0.75$ quantile level. The results are illustrated in Figure 4.2. Seven patterns are plot for seven selected variables. Lines in red represent the fitted local linear quantile regression at quantile level $\alpha = 0.25$; lines in green represent the fitted local linear quantile regression at quantile level $\alpha = 0.75$.

Based on the results in Figure 4.2, we can conclude that during the COVID-19 pandemic from 2019 to 2020, the GDP growth rate in the United States exhibited notable relationships with key labor force characteristics, reflecting the economic disruptions and structural shifts triggered by the crisis.

First, for X_{17}, there is a negative relationship between GDP growth and the proportion of the population in the unemployed civilian labor force. The result highlights the widespread unemployment during the pandemic, especially in industries which are tremendously impacted by lockdown and declining demand. High unemployment led to low household income and consumption, further exacerbating the economy.

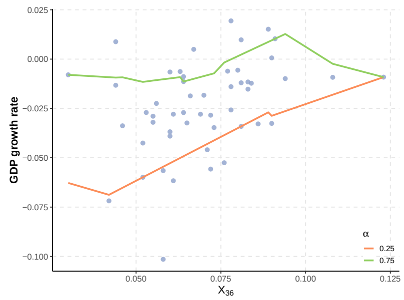
Second, for X_{36}, there is a positive relationship between GDP growth and the proportion of workers working from home. This underscores the elasticity of the industries which are capable of remote work. With remote working, businesses would be able to sustain productivity and making revenue, which contributes to the economic stability.

Thirdly, for X_{49}, it shows a negative relationship between GDP growth and the proportion of the employed population in transportation, warehousing, and utilities. This reflects severe disruptions in supply chains. Low mobility, decreased trade activity, and reduced industrial demand for utilities contributed to economic deceleration.

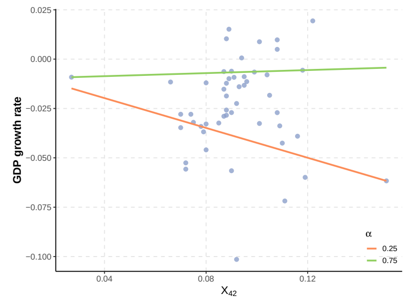
One of the most interesting variables is X_{42}. The relationship between GDP growth and the proportion of the workforce in natural resources, construction, and maintenance occupations varied across states with high and low GDP growth rates. This divergence probably reflects differences in industry resilience and economic structures during the pandemic. For states with high GDP growth rates, the positive correlation suggests that these industries played a stabilizing or even growth-promoting role. Conversely, in states with low GDP growth rates, the negative correlation suggests that these occupations suffered greater disruptions. This contrast underscores how the same industry can have different impacts depending on a state's underlying economic conditions and industrial composition.

Finally, the relationships between GDP growth from 2019 to 2020 and the remaining selected variables (X_{57}, X_{58}, X_{107}) appear more complex and potentially nonlinear, indicating the need for further research and in-depth economic analysis.

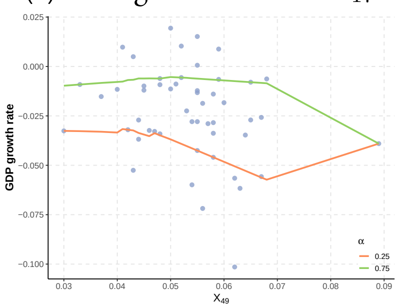
(a) GPD growth rate v.s. X_{17}.



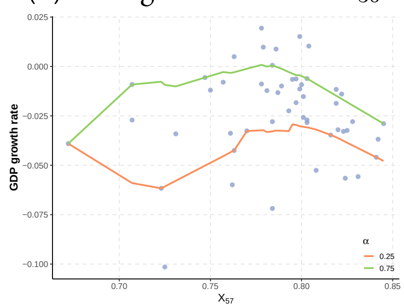
(b) GPD growth rate v.s. X_{36}.



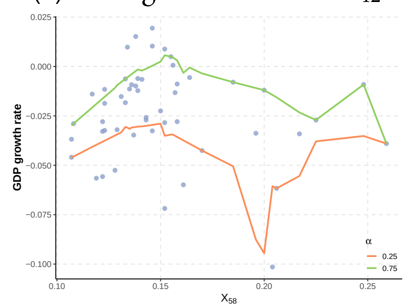
(c) GPD growth rate v.s. X_{42}.



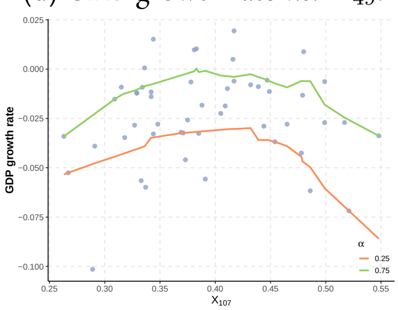
(d) GPD growth rate v.s. X_{49}.



(e) GPD growth rate v.s. X_{57}.



(f) GPD growth rate v.s. X_{58}.



(g) GPD growth rate v.s. X_{107}.

Figure 4.2: Scatter plots and fitted local linear quantile regressions between GDP growth rate and selected variables at quantile level $\alpha = 0.25, 0.75$.

Chapter 5

Consistent Sampling With Smoothed Quantum Walk

In this chapter, we introduce a novel sampling technique grounded in the dynamics of a 2-state Quantum Walk (QW) in a one-dimensional space. The content of this chapter is primarily derived from our publication (Zhang and Ke, 2024). In Section 5.1, we revisit the 2-state QW on the line and pinpoint its limitations in sampling. In Section 5.2, we review the kernel smoothing method and elucidate the reasons why it serves as a complementary approach to the 2-state QW on the line sampling method. These reasons include the presence of discontinuities in sampling distributions and potential inaccuracies in limiting distributions, which are issues that the kernel smoothing method can address. In Section 5.3, by applying the Epanechnikov kernel and transformation method, we develop innovative smoothed quantum sampling methods. These methods effectively mitigate the limitations associated with sampling using the 2-state QW, thereby enhancing the overall sampling quality. In Section 5.4, we demonstrate the superior empirical properties of these smoothed quantum sampling methods through extensive experiments. The results clearly show significant improvements in density estimation and sampling efficacy when compared to traditional Quantum Walk distributions and sampling techniques, highlighting the practical advantages of our proposed methods.

5.1 Sampling with 2-state Quantum Walk on the Line

Based on the state - space postulate presented in Machida, 2013, the discrete-time 2-state QW on the line is defined within the tensor space $\mathcal{H}_p \otimes \mathcal{H}_c$. Here, $\mathcal{H}_p$ represents the position Hilbert space, while $\mathcal{H}_c$ denotes the coin Hilbert space. Specifically, the position space can be expressed as the span of its basis states, $\mathcal{H}_p = \text{Span}\left(\{|x\rangle_p : x \in \mathbb{Z}\}\right)$. And the coin space can be written as $\mathcal{H}_c = \text{Span}\left(\{|0\rangle_c, |1\rangle_c\}\right)$ where $\langle 0|_c = [1, 0]$ and $\langle 1|_c = [0, 1]$. Let $|\Psi_t\rangle$ be the superposition of 2-state QWs on the line at time $t \in \{0, 1, 2, \ldots\}$. $|\Psi_t\rangle$ can be decomposed as

$$|\Psi_t\rangle = \sum_{x \in \mathbb{Z}} |x\rangle_p \otimes |\psi_t(x)\rangle_c ,$$

where $|\psi_t(x)\rangle_c \in \mathcal{H}_c$. Upon the measurement of $|\Psi_t\rangle$, the squared modulus of $|\psi_t(x)\rangle_c$ quantifies the probability mass function associated with the observation of the quantum walker at position x at time t.

The evolution of QWs can be regarded as a stochastic process (S. E. Venegas-Andraca, 2012) that is contingent upon two quantum operators. First, a Hadamard operator denoted as H_c, is applied to the coin state. It is defined as

$$H_c \doteq \cos\theta|0\rangle_c\langle 0| + \sin\theta|0\rangle_c\langle 1| + \sin\theta|1\rangle_c\langle 0| - \cos\theta|1\rangle_c\langle 1|$$

$$= \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \tag{5.1.1}$$

where $\theta \in [0, 2\pi)$. It is straightforward to verify that H_c is a unitary matrix. Subsequently, a conditional shift operator acts on the position state. If the coin state is $|1\rangle_c$, this operator moves the quantum walker one step forward; if the coin state is $|0\rangle_c$, it moves the walker one step backward. Specifically, the conditional shift operator can be defined as

$$S_p \doteq \sum_{i\in\mathbb{Z}} |i+1\rangle_p\langle i| \otimes |1\rangle_c\langle 1| + \sum_{i\in\mathbb{Z}} |i-1\rangle_p\langle i| \otimes |0\rangle_c\langle 0|. \tag{5.1.2}$$

By the two quantum operators, QWs update the current superposition $|\Psi_t\rangle$ to a new superposition $|\Psi_{t+1}\rangle$ at time t through

$$|\Psi_{t+1}\rangle = S_p \left(I_p \otimes H_c\right) |\Psi_t\rangle\,,$$

where I_p is the identity operator on the position state. Then, the probability of finding the quantum walker X_t in position x at time t is calculated by

$$\mathcal{P}\left(X_t = x\right) = \langle \psi_t(x) | \psi_t(x) \rangle_c\,. \tag{5.1.3}$$

By employing the discrete-time Fourier transformation, we establish the following definition:

$$|\hat{\Psi}_t(k)\rangle_c \doteq \sum_{x\in\mathbb{Z}} e^{-ikx}\, |\psi_t(x)\rangle_c\,, \quad \text{for } k \in [-\pi, \pi).$$

Subsequently, the inverse transformation can be defined as

$$|\psi_t(x)\rangle_c \doteq \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{\Psi}_t(k)\rangle_c \, e^{ikx} dk.$$

As elaborated in Machida, 2013, the mathematical expressions presented above lay the groundwork for a quantum sampling scheme to generate a random sample of the target distribution by matching the moments in the quantum sampling process.

Theorem 5.1.1 (Machida, 2013). *Let $F : \mathbb{R} \mapsto \mathbb{R}$ be a real function satisfying*

1. $F(k + 2\pi) = F(k)$,
2. $\int_{-\pi}^{\pi} F(k)^2 dk = 2\pi$,
3. $F(k) \in C^\infty([-\pi, \pi])$ *almost everywhere,*
4. $|F(k - \pi)| = |F(-k)| = |F(k)|$.

Construct the following non-localized initial state

$$|\hat{\Psi}_0(k)\rangle_c = F(k)(\alpha\,|0\rangle_c + \beta\,|1\rangle_c),$$

or equivalently

$$|\psi_0(x)\rangle_c = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} F(k) e^{ikx} dk\right] (\alpha\,|0\rangle_c + \beta\,|1\rangle_c)$$

with $\alpha, \beta \in \mathbb{C}$ *and* $|\alpha|^2 + |\beta|^2 = 1$.

Let r be a non-negative integer, $c = \cos\theta$, and $s = \sin\theta$. The limiting moments of X_t/t satisfy

$$\lim_{t\to\infty} \mathbb{E}\left[\left(\frac{X_t}{t}\right)^r\right] = \int_{\mathbb{R}} x^r f(x;\alpha,\beta) F(\kappa(x))^2 \mathbb{1}_{\{x\in(-|c|,|c|)\}} dx,$$

where

$$\begin{aligned} f(x;\alpha,\beta) =& \frac{|s|}{\pi\left(1-x^2\right)\sqrt{c^2-x^2}}\left[1-\left\{|\alpha|^2-|\beta|^2+\frac{2s\Re(\alpha\bar{\beta})}{c}\right\}x\right], \\ \kappa(x) =& \arccos\left(\frac{|s|x}{c\sqrt{1-x^2}}\right). \end{aligned}$$

Furthermore, when α, β are selected in a way that $|\alpha|^2 - |\beta|^2 + \frac{2s\Re(\alpha\bar{\beta})}{c} = 0$ (e.g. $\alpha = \frac{\sqrt{2}}{2}, \beta = \frac{\sqrt{2}}{2}i$), it becomes possible to select the form of F to regulate the (scaled) limiting moments of X_t. Provided that an appropriately chosen initial state $|\psi_0(x)\rangle_c$, we are able to generate $\frac{X_{t,i}}{t}$, $i = 1, \cdots, N$, as a random sample drawn from a target distribution. This process allows us to obtain a set of samples that follow the desired statistical characteristics. Here we list a few examples.

1. **Wigner semicircle law**:

   Let $F(k) = \frac{\sqrt{2|s|^3}\sin k}{1-c^2\sin^2 k}$. Then,

$$\lim_{t\to\infty} \mathbb{E}\left[\left(\frac{X_t}{t}\right)^r\right] = \int_{(-|c|,|c|)} x^r \frac{2\sqrt{c^2-x^2}}{\pi c^2} dx;$$

2. **Uniform distribution**:

Let $F(k)=\sqrt{\frac{\pi s^2|\sin k|}{2(1-c^2\sin^2 k)^{\frac{3}{2}}}}$. Then,

$$\lim_{t\to\infty}\mathbb{E}\left[\left(\frac{X_t}{t}\right)^r\right]=\int_{(-|c|,|c|)}x^r\frac{1}{2|c|}dx;$$

3. **Truncated Gaussian distribution**:

Let $F(k)=\sqrt{\frac{\sqrt{2\pi}|c|s^2|\sin k|}{2\sigma\operatorname{erf}\left(\frac{|c|}{\sqrt{2}\sigma}\right)\left(1-c^2\sin^2 k\right)^{\frac{3}{2}}}}\exp\left\{-\frac{c^2\cos^2 k}{4\sigma^2\left(1-c^2\sin^2 k\right)}\right\}$. Then,

$$\lim_{t\to\infty}\mathbb{E}\left[\left(\frac{X_t}{t}\right)^r\right]=\int_{(-|c|,|c|)}x^r\frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma\operatorname{erf}\left(\frac{|c|}{\sqrt{2}\sigma}\right)}dx,$$

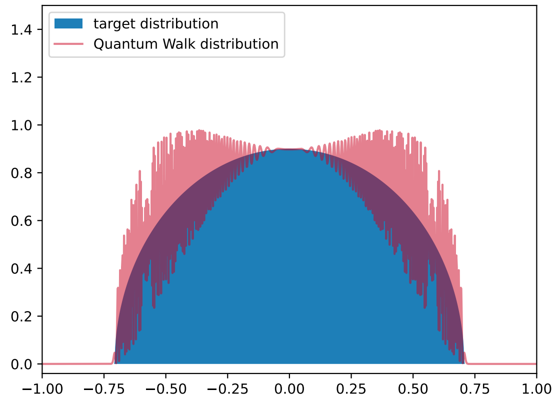
where $\operatorname{erf}(\cdot)$ is the Gaussian error function, and $\sigma>0$ stands for the standard deviation;

4. **Arcsine law**:

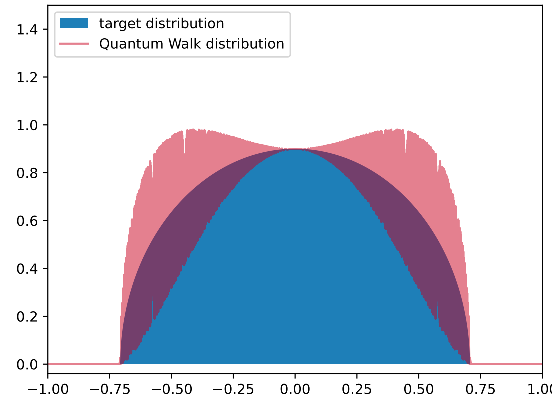
Let $F(k)=\sqrt{\frac{|s|}{1-c^2\sin^2 k}}$. Then,

$$\lim_{t\to\infty}\mathbb{E}\left[\left(\frac{X_t}{t}\right)^r\right]=\int_{(-|c|,|c|)}x^r\frac{1}{\pi\sqrt{c^2-x^2}}dx.$$

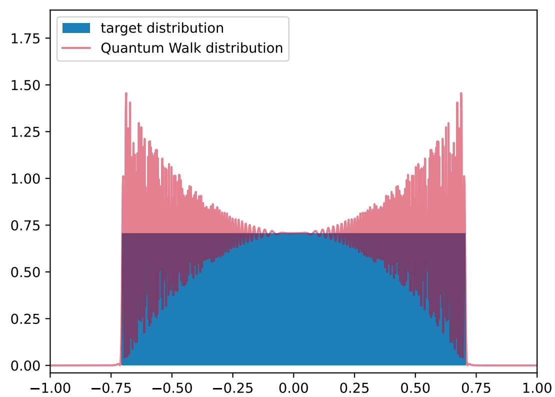
In most scenarios, the preparation of the initial state $|\psi_0(x)\rangle_c$ is not overly complex. For example, when dealing with the Wigner semicircle law, Uniform distribution, or Truncated Gaussian distribution, even if the non-localized initial state cannot be generated accurately, preparing the initial state $|\psi_0(x)\rangle_c$ solely for those x around 0 would be enough. This approach can approximate the true initial state on $\mathbb{Z}$ with arbitrarily tiny difference. We plot the Quantum Walk distributions with target distributions in Figure 5.1. In this figure, Four patterns are illustrated for different approximation laws and t values. The



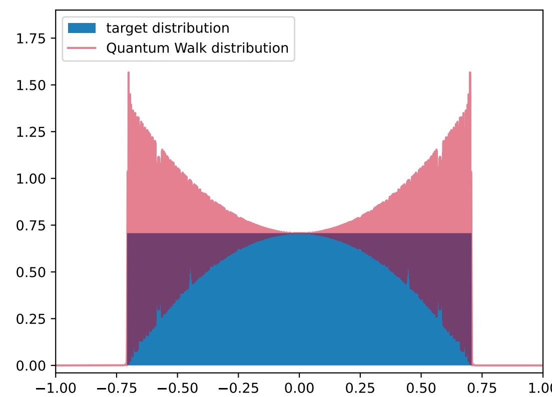
(a) Semicircle law, $t = 1000$.



(b) Semicircle law, $t = 5000$.



(c) Uniform distribution, $t = 1000$.



(d) Uniform distribution, $t = 5000$.

Figure 5.1: Quantum Walk distributions and target distributions.

red lines are the true probability distributions of the Quantum Walk sample $\frac{X_t}{t}$, while the blue areas denote the probability density functions (PDFs) of the target distributions. Nevertheless, this quantum sampling scheme has two notable limitations. Firstly, at a finite time t, $\frac{X_t}{t}$ actually approximates a continuous target distribution on the set $(-|c|, |c|)$ by an empirical discrete distribution on the set $\left\{\frac{i}{t} : i \in \mathbb{Z} \cap (-|c| \cdot t, |c| \cdot t)\right\}$. This approximation error may introduce bias, consequently resulting in sub-optimal finite-sample performance. Secondly, the previous analysis merely offers the asymptotic behavior of the (scaled) moments. A more desirable property, the limiting distribution of X_t (or of $\frac{X_t}{t}$) is still lacking. Denote $f_{\text{target}}(\cdot)$ as the PDF of the target distribution. It is not guaranteed that $\forall x \in (-|c|, |c|)$, $\lim_{t\to\infty} \mathcal{P}\left(\frac{X_t}{t} = x\right) = f_{\text{target}}(x)$. In fact, $\lim_{t\to\infty} \mathcal{P}\left(\frac{X_t}{t} = x\right)$ does not even exist in general. To overcome these two limitations, we propose innovating the quantum sampling method by drawing on the insights from statistical kernel smoothing techniques.

5.2 Kernel Smoothing

In non-parametric statistics, kernel smoothing stands as a prevalent technique for deriving estimates through a weighted average of a "localized" neighborhood in the random sample (Fan, 1996; Nadaraya, 1964; Wand and Jones, 1994; Watson, 1964). A Kernel function $K(\cdot)$ is commonly employed to assign weights. The extent of the local neighborhood is regulated by a bandwidth (or smoothing) parameter h that can converge to 0 as the sample size n diverges. For example, in non-parametric regression, the Epanechnikov kernel is a highly favored choice. It is renowned for attaining a high minimax efficiency (Fan, 1992). To be specific, we define Epanechnikov kernel as

$$K(u) = \frac{3}{4}\left(1 - u^2\right)_+ . \tag{5.2.1}$$

Suppose $X_1 \ \ldots, \ X_n$ form an independent and identically distributed (i.i.d.) sample obtained from a probability density function $f(\cdot)$. The kernel smoothing density estimator of $f(\cdot)$ then defined in the following way.

$$\hat{f}(x) \doteq \frac{1}{n}\sum_{i=1}^{n} K_h\left(x, X_i\right), \tag{5.2.2}$$

where $K_h(x, y) = h^{-1}K\{(x-y)/h\}$.

Under mild conditions, the kernel density estimator possesses the following two desirable statistical properties.

1. $\lim_{n\to\infty} \mathbb{E}\left[\hat{f}(x)\right] = f(x)$,

2. $\lim_{n\to\infty} \text{Var}\left[\hat{f}(x)\right] = 0$.

Furthermore, it can be proved that kernel density estimator converges in probability to underlying true density function, i.e. $\forall x, \hat{f}(x) \xrightarrow{p} f(x)$. Therefore, we intend to leverage the insights of kernel smoothing to address the limitations of the quantum sampling method, as discussed in Section 5.1.

5.3 Kernel Smoothed Quantum Sampling

In this section, we propose a novel quantum sampling method that combines quantum walk with kernel smoothing. We define $f^*(\cdot)$ as the density function of the target distribution, $X_{t,i}$ as the location of the ith quantum walker at time t, and $Y_{t,i} \doteq \frac{X_{t,i}}{t}$ for $i = 1, \ldots, N$ and $t \in \mathbb{Z}^+$.

We then define a kernel smoothing density estimator of the target distribution $f^*(\cdot)$ using Epanechnikov kernel. Mathematically,

$$\hat{f}_0(x) \doteq \frac{1}{Nh}\frac{3}{4}\sum_{i=1}^{N}\left(1-\left(\frac{x-Y_{t,i}}{h}\right)^2\right)_+ .$$

As illustrated in Section 5.1, the target distribution for quantum sampling usually has a bounded domain, $(-|c|, |c|)$. However, the support of $\hat{f}_0(\cdot)$ is the whole real line. To bridge the definitions, $\hat{f}_0(\cdot)$ can be rescaled by

$$\hat{f}_1(x) \doteq \frac{\hat{f}_0(x)}{\int_{(-|c|,|c|)} \hat{f}_0(z)dz}. \tag{5.3.1}$$

We refer to $\hat{f}_1(x)$ as *Smoothed Quantum Sampling (SQS)*. As demonstrated by the empirical experiments in Section 5.4, SQS exhibits biases near the boundary regions.

To reduce the boundary bias problem, we propose a *Transformed Smoothed Quantum Sampling (TSQS)*. Let $T_c : (-|c|, |c|) \mapsto \mathbb{R}$ be a monotonically increasing function that is three times continuously differentiable. Such a $T_c(\cdot)$ function is straightforward to construct. For example, one can select $T_c(\cdot)$ as the inverse of a Gaussian distribution function

$$T_{c,1}(x) = \Phi^{-1}\left(\frac{x + |c|}{2|c|}\right), \tag{5.3.2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard Gaussian random variable. Alternatively, $T_c(\cdot)$ can be chosen as logit function

$$T_{c,2}(x) = \text{logit}\left(\frac{x + |c|}{2|c|}\right) = \ln \frac{x + |c|}{|c| - x}.$$

Using an appropriately selected transformation function $T_c(\cdot)$, we define the TSQS estimator by

$$\hat{f}_2(x) \doteq T_c'(x) \cdot \frac{1}{Nh}\frac{3}{4}\sum_{i=1}^{N}\left(1 - \left(\frac{T_c(x) - T_{t,i}}{h}\right)^2\right)_+, \tag{5.3.3}$$

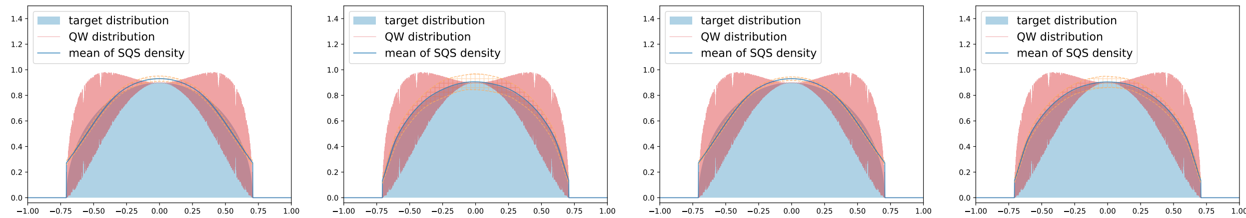
where $T_{t,i} = T_c(Y_{t,i})$, and $T_c'(\cdot)$ represents the first order derivative of $T_c(\cdot)$.

5.4 Experiments

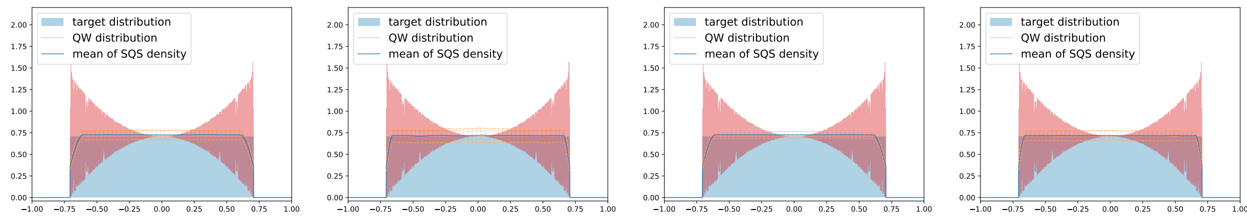
In this section, we conduct a series of numerical experiments to validate the concepts presented in this paper. Additionally, we evaluate and compare the empirical performance of Quantum Walk (QW), Smoothed Quantum Sampling (SQS), and Transformed Smoothed Quantum Sampling (TSQS) across different scenarios. Throughout this section, we set $\alpha = \frac{\sqrt{2}}{2}$, $\beta = \frac{\sqrt{2}}{2}i$, and $\theta = \frac{\pi}{4}$.

5.4.1 Empirical analysis for Quantum Sampling Performance

In the first experiment, we evaluate the empirical performance of SQS and TSQS across different settings. Additionally, we compare SQS and TSQS with QW and the target distribution. As target distributions, we consider the Wigner semicircle law and the Uniform distribution, as introduced in Section 5.1. The Quantum Walk (QW) runs for $t = 5000$ steps, and the sample size is set to be $N = 5000$ and 10000. For SQS and TSQS, we use the Epanechnikov kernel and analyze their performance under both large and small smoothing parameters. Specifically, we choose $h = 0.4$ and 0.1 for the Wigner semicircle law and choose $h = 0.1$ and 0.05 for the Uniform distribution. For TSQS, we apply the inverse Gaussian transformation function $T_{c,1}(\cdot)$ as defined in (5.3.2). For each scenario, we repeat 500 replications. The experiment results for SQS and TSQS are presented in Figures 5.2 and 5.3. In each figure, eight patterns are plotted for different setting of approximation law, sample size N, and bandwidth h. The light blue lines are the mean functions of Epanechnikov kernel estimators with 500 replications. The orange cross-shaded areas represent the area given by $\text{mean} \pm 2 \times \text{standard error}$. The red lines are the true probability distributions of the Quantum Walk sample $Y_{t,i}$. The blue areas are PDFs of the target distributions.



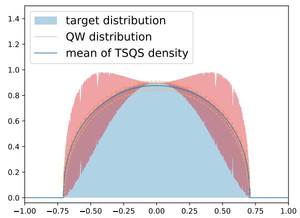
(a) Semicircle law, $t = 5000$, $N = 5000$, $h = 0.4$.
(b) Semicircle law, $t = 5000$, $N = 5000$, $h = 0.1$.
(c) Semicircle law, $t = 5000$, $N = 10000$, $h = 0.4$.
(d) Semicircle law, $t = 5000$, $N = 10000$, $h = 0.1$.



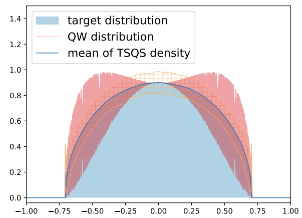
(e) Uniform distribution, $t = 5000$, $N = 5000$, $h = 0.1$.
(f) Uniform distribution, $t = 5000$, $N = 5000$, $h = 0.05$.
(g) Uniform distribution, $t = 5000$, $N = 10000$, $h = 0.1$.
(h) Uniform distribution, $t = 5000$, $N = 10000$, $h = 0.05$.

Figure 5.2: SQS densities.

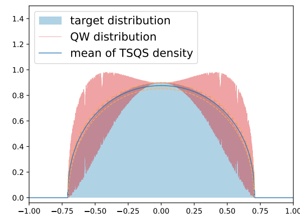
The experimental results clearly demonstrate that, for a given time t, the samples generated by both SQS and TSQS, with properly selected N and smoothing parameter h, are much closer to the target density than those generated by QW. This improvement is due to the bias correction introduced by kernel smoothing. Additionally, as t and N increase, the samples from SQS and TSQS gradually converge to the target density. However, the choice of h plays a crucial role in balancing the bias-variance trade-off. A large h results in a smoother estimate with lower variance but higher bias, whereas a smaller h produces a more “localized" estimate with lower bias but higher variance. In practice, we can choose h through a multi-fold cross-validate approach.



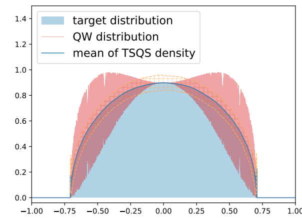
(a) Semicircle law, $t = 5000$, $N = 5000$, $h = 0.4$.



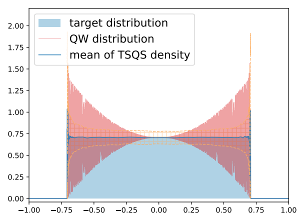
(b) Semicircle law, $t = 5000$, $N = 5000$, $h = 0.1$.



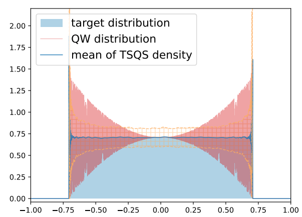
(c) Semicircle law, $t = 5000$, $N = 10000$, $h = 0.4$.



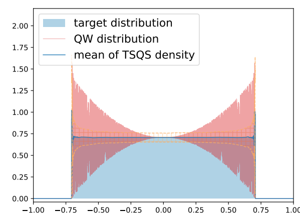
(d) Semicircle law, $t = 5000$, $N = 10000$, $h = 0.1$.



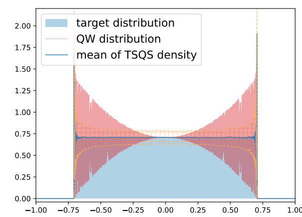
(e) Uniform distribution, $t = 5000$, $N = 5000$, $h = 0.1$.



(f) Uniform distribution, $t = 5000$, $N = 5000$, $h = 0.05$.



(g) Uniform distribution, $t = 5000$, $N = 10000$, $h = 0.1$.



(h) Uniform distribution, $t = 5000$, $N = 10000$, $h = 0.05$.

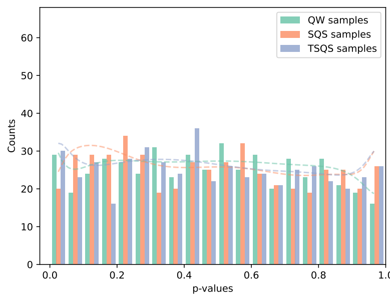
Figure 5.3: TSQS densities.

Furthermore, as illustrated in Figure 5.2, both QW and SQS exhibit boundary bias. In the Wigner semicircle law experiment, $\hat{f}_1(\cdot)$ shows densities outside $(-|c|, |c|)$, indicating that $\hat{f}_1(x) > 0$ when $x = -|c|$ or $x = |c|$, by the continuity of $\hat{f}_1(x)$. This results in a positive bias at the boundary for the samples generated by SQS. For the Uniform distribution case, when x is close to the boundaries, $\hat{f}_1(x)$ only uses interior observations, leading to an underestimation of the target density and hence a negative bias. Although this issue diminishes as $N \to \infty$ and $h \to 0$, it remains significant in practice with a finite sample. Encouragingly, as shown in Figure 5.3, this boundary bias issue is mitigated by TSQS as we expected.

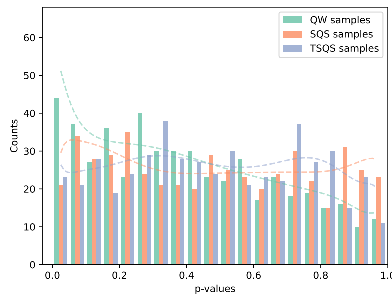
5.4.2 Statistical analysis for Quantum Sampling Performance

Next, we conduct a statistical inference analysis for quantum sampling methods. Specifically, we apply the Kolmogorov–Smirnov (KS) test to assess whether a sample of size m generated by a quantum sampling method originates from the target distribution. The KS test is a non-parametric method for comparing two continuous one-dimensional probability distributions. In our context, it measures the goodness-of-fit between the sampled data and the target distribution by analyzing their cumulative distribution functions. The resulting p-values from the KS test indicate the likelihood that the observed differences between the sample and target distributions arise by chance. A high p-value provides little evidence against the null hypothesis, suggesting that the sample distribution aligns well with the target distribution.

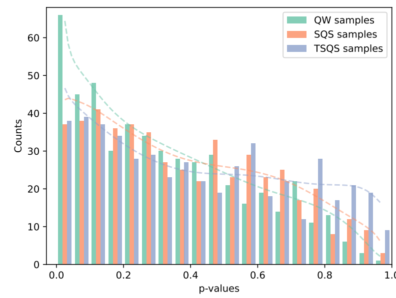
We consider a similar experiment setting as Section 5.4.1. For QW, we set $t = 5000$ and $N = 200000$. We also set the sample size of the obsevations in KS tests to be $m = 50000$, 200000 and 500000. For the experiment of Wigner semicircle law, we choose $h = 0.02$ for SQS and $h = 0.05$ for TSQS. For the experiment of Uniform distribution, we choose $h = 0.006$ for SQS and $h = 0.08$ for TSQS. The histograms of p-values over 500 replications are reported in Figure 5.4. In this figure, six patterns are plotted for different approximation laws and different sample sizes m of observations in KS tests. Green bars and lines are from QW samples. Orange bars and lines are from SQS samples. Blue bars and lines are from TSQS samples. When the null hypothesis of the KS test holds true, the p-value of the test statistic should follow a uniform distribution between 0 and 1. In Figure 5.4, the QW samples and SQS samples exhibit comparable performance when N is small, with all methods displaying evenly distributed p-values across $[0, 1]$. However, as N increases, all methods tend to perform poorly, with their p-values concentrate increasingly towards 0. Notably, TSQS samples outperform the others, as their p-values are less skewed



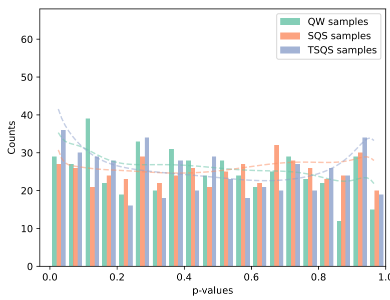
(a) Semicircle law, $m = 50000$.



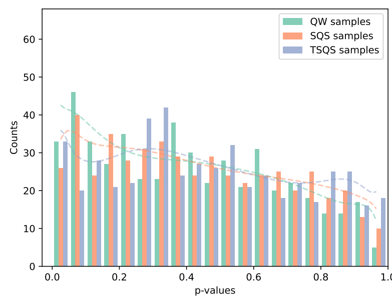
(b) Semicircle law, $m = 200000$.



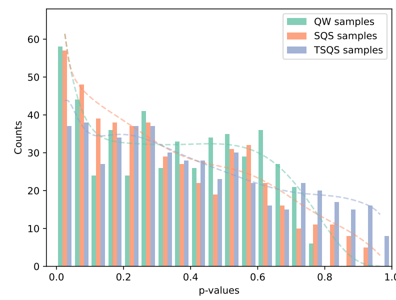
(c) Semicircle law, $m = 500000$.



(d) Uniform distribution, $m = 50000$.



(e) Uniform distribution, $m = 200000$.



(f) Uniform distribution, $m = 500000$.

Figure 5.4: Histograms and distributions of p-values of KS tests on QW samples and SQS samples.

towards 0 compared to those of QW and SQS. This suggests that TSQS samples better approximate the target distributions as the sample size grows.

Appendices

A.1 Additional Lemmas for Chapter 3

Lemma A.1.1. *For $B_{n,1}(s,t)$ defined in* (3.9.11)*, we have*

$$B_{n,1}(s,t) = n^{-\frac{1}{2}} h^{-\frac{1}{2}} \sum_{i=1}^{n} \left(B_{n,1,i}(s,t) - \mathbb{E}\left[B_{n,1,i}(s,t) \right] \right),$$

where

$$B_{n,1,i}(s,t) \doteq \frac{-1}{\phi(s)} \iint_{\mathbb{R}^2} \mathbb{1}_{\{s_i \leqslant s - hw_u\}} \phi'(w_u) \frac{\partial \mathcal{C}}{\partial u} \left(\Phi(s - hw_u), \Phi(t - hw_v) \right) \phi(w_v) \, dw_v dw_u.$$

Let $h \to 0$ as $n \to \infty$. When $n \to \infty$, we have the following results hold for any $(s,t) \in \mathbb{R}^2$,

$$\begin{aligned}
\mathbb{E}\left[B_{n,1,i}(s,t) \right] =& - h \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s) \right) + \Phi(s) \frac{\partial^2 \mathcal{C}}{\partial u^2} \left(\Phi(s), \Phi(t) \right) \right] + o(h), \\
\mathbb{E}\left[B_{n,1,i}^2(s,t) \right] =& h \frac{1}{2\sqrt{\pi}} \frac{\mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi(s) \right)}{\phi(s)} + o(h), \\
\mathbb{E}\left[B_{n,1}(s,t) \right] =& 0, \\
\mathit{Var}[B_{n,1}(s,t)] =& \frac{1}{2\sqrt{\pi}} \frac{\mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi(s) \right)}{\phi(s)} + o(1).
\end{aligned}$$

Proof. By definition and change of variables, $w_u = \frac{s-\Phi^{-1}(u^*)}{h}$ and $w_v = \frac{t-\Phi^{-1}(v^*)}{h}$, we have

$$\begin{aligned} B_{n,1}(s,t) = &- \frac{1}{\sqrt{h}\phi(s)} \iint_{\mathbb{R}^2} \mathbb{B}_n(\Phi(s-hw_u),1)\phi'(w_u) \\ &\frac{\partial \mathcal{C}}{\partial u}(\Phi(s-hw_u), \Phi(t-hw_v))\,\phi(w_v)\,dw_v dw_u, \end{aligned}$$

By definition of $\mathbb{B}_n(\cdot,\cdot)$, we have

$$\mathbb{B}_n(\Phi(s-hw_u),1) = \sqrt{n}\left(\mathcal{C}_n(\Phi(s-hw_u),1) - \mathcal{C}(\Phi(s-hw_u),1)\right),$$

and

$$\begin{aligned} \mathcal{C}_n(\Phi(s-hw_u),1) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{u_i \leqslant \Phi(s-hw_u), v_i \leqslant 1\}} \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{u_i \leqslant \Phi(s-hw_u)\}} \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{s_i \leqslant s-hw_u\}} \\ &= F_{n,S}(s-hw_u), \\ \mathcal{C}(\Phi(s-hw_u),1) &= F_{U,V}(\Phi(s-hw_u),1) \\ &= F_U(\Phi(s-hw_u)) \\ &= F_S(s-hw_u), \end{aligned}$$

where $F_{n,S}(\cdot)$ is the empirical CDF on sample $\{s_i\}_{i=1}^{n}$. Hence we have

$$\begin{aligned}\mathbb{B}_n(\Phi(s-hw_u),1)=&\sqrt{n}\left(F_{n,S}\left(s-hw_u\right)-F_S\left(s-hw_u\right)\right)\\=&\sqrt{n}\left(F_{n,S}\left(s-hw_u\right)-\Phi\left(s-hw_u\right)\right)\\=&\mathbb{P}_n\left(s-hw_u\right),\end{aligned}$$

which is the standard Gaussian empirical process. Then we have

$$\begin{aligned}&B_{n,1}\left(s,t\right)\\=&-\frac{1}{\phi\left(s\right)}n^{-\frac{1}{2}}h^{-\frac{1}{2}}\sum_{i=1}^{n}\\&\left[\iint_{\mathbb{R}^2}\mathbb{1}_{\{s_i\leqslant s-hw_u\}}\phi'\left(w_u\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_vdw_u\right.\\&\left.-\iint_{\mathbb{R}^2}\Phi\left(s-hw_u\right)\phi'\left(w_u\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_vdw_u\right]\\=&n^{-\frac{1}{2}}h^{-\frac{1}{2}}\sum_{i=1}^{n}\left(B_{n,1,i}\left(s,t\right)-\mathbb{E}\left[B_{n,1,i}\left(s,t\right)\right]\right)\end{aligned}$$

For $\mathbb{E}\left[B_{n,1,i}\left(s,t\right)\right]$, by Taylor expansion at (s,t), we have

$$\begin{aligned}&\mathbb{E}\left[B_{n,1,i}\left(s,t\right)\right]\\=&\frac{-1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\left\{\Phi\left(s\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s\right),\Phi\left(t\right)\right)-\frac{\partial\left[\Phi\left(s\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s\right),\Phi\left(t\right)\right)\right]}{\partial(s,t)^{\intercal}}(hw_u,hw_v)^{\intercal}\right\}\\&\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u+o(h)\\=&h\left[\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s\right),\Phi\left(t\right)\right)+\Phi\left(s\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s\right),\Phi\left(t\right)\right)\right]\\&\iint_{\mathbb{R}^2}w_u\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u+o(h)\end{aligned}$$

$$= -h\left[\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right) + \Phi\left(s\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right] + o(h).$$

For $\mathbb{E}\left[B_{n,1,i}^2\left(s,t\right)\right]$, by Taylor expansions at (s,t), we have

$$\begin{aligned}
&\mathbb{E}\left[B_{n,1,i}^2\left(s,t\right)\right] \\
=&\mathbb{E}\Bigg[\frac{1}{\phi^2\left(s\right)}\iiiint_{\mathbb{R}^4}\mathbb{1}_{\{s_i\leqslant s-hw_{u1}\}}\mathbb{1}_{\{s_i\leqslant s-hw_{u2}\}} \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}\Bigg] \\
=&\frac{1}{\phi^2\left(s\right)}\iiiint_{\mathbb{R}^4}\mathbb{E}\left[\mathbb{1}_{\{s_i\leqslant s-hw_{u1}\}}\mathbb{1}_{\{s_i\leqslant s-hw_{u2}\}}\right] \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2} \\
=&\frac{1}{\phi^2\left(s\right)}\iiiint_{\mathbb{R}^4}\Phi\left(s-h\max\{w_{u1},w_{u2}\}\right) \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2} \\
=&\frac{2}{\phi^2\left(s\right)}\iiint_{\mathbb{R}^3}\int_{w_{u1}}^{\infty}\Phi\left(s-hw_{u2}\right) \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u2}dw_{u1}dw_{v1}dw_{v2} \\
=&\frac{2}{\phi^2\left(s\right)}\iiint_{\mathbb{R}^3}\int_{w_{u1}}^{\infty}\Bigg\{\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\Phi\left(s\right) \\
&-\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\Phi\left(s\right)\frac{\partial\left[\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]}{\partial(s,t)^{\intercal}}(hw_{u1},hw_{v1})^{\intercal}
\end{aligned}$$

$$
- \frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right) \frac{\partial\left[\Phi\left(s\right) \frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]}{\partial(s,t)^{\intercal}}(hw_{u2}, hw_{v2})^{\intercal}\Bigg\}
$$

$$
\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right) dw_{u2}dw_{u1}dw_{v1}dw_{v2} + o(h)
$$

$$
\begin{aligned}
= & -h\frac{2}{\phi^2\left(s\right)}\iiint_{\mathbb{R}^3}\int_{w_{u1}}^{\infty}\Bigg\{\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\Phi\left(s\right)\phi\left(s\right)w_{u1} \\
& + \left[\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]^2\phi\left(s\right)w_{u2} \\
& + \frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\Phi\left(s\right)\phi\left(s\right)w_{u2}\Bigg\} \\
& \phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right) dw_{u2}dw_{u1}dw_{v1}dw_{v2} + o(h) \\
= & h\frac{1}{2\sqrt{\pi}}\frac{1}{\phi\left(s\right)}\left[\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]^2 + o(h).
\end{aligned}
$$

Then, the rest of the proof follows directly from the above results, and the definition of $\mathcal{C}_{2|1}(\cdot \mid \cdot)$. □

Lemma A.1.2. *For $B_{n,2}\left(s,t\right)$ defined in (3.9.11), we have*

$$
B_{n,2}\left(s,t\right) = n^{-\frac{1}{2}}h^{-\frac{1}{2}}\sum_{i=1}^{n}\left(B_{n,2,i}\left(s,t\right) - \mathbb{E}\left[B_{n,2,i}\left(s,t\right)\right]\right),
$$

where

$$
B_{n,2,i}\left(s,t\right) \doteq \frac{-1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\mathbb{1}_{\{t_i\leqslant t-hw_v\}}\phi'\left(w_u\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_v dw_u.
$$

Let $h \to 0$ as $n \to \infty$. When $n \to \infty$, we have the following results hold for any $(s,t) \in \mathbb{R}^2$,

$$
\begin{aligned}
\mathbb{E}\left[B_{n,2,i}\left(s,t\right)\right] = & -h\Phi\left(t\right)\frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s\right), \Phi\left(t\right)\right) + o(h), \\
\mathbb{E}\left[B_{n,2,i}^2\left(s,t\right)\right] = & o(h),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[B_{n,2}\left(s,t\right)\right] =& 0,\\
\mathit{Var}[B_{n,2}\left(s,t\right)] =& o(1).
\end{aligned}
$$

Proof. By definition and change of variables, $w_u = \frac{s-\Phi^{-1}(u^*)}{h}$ and $w_v = \frac{t-\Phi^{-1}(v^*)}{h}$, we have

$$
\begin{aligned}
B_{n,2}\left(s,t\right) = & -\frac{1}{\sqrt{h}\phi\left(s\right)}\iint_{\mathbb{R}^2}\mathbb{B}_n(1,\Phi\left(t-hw_v\right))\phi'\left(w_u\right)\\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_v dw_u,
\end{aligned}
$$

Similar with the analysis in Lemma A.1.1, we have

$$
\begin{aligned}
\mathbb{B}_n(1,\Phi\left(t-hw_v\right)) =& \sqrt{n}\left(F_{n,T}\left(t-hw_v\right)-\Phi\left(t-hw_v\right)\right)\\
=& \mathbb{P}_n\left(t-hw_v\right),
\end{aligned}
$$

which is the standard Gaussian empirical process, where $F_{n,T}(\cdot)$ is the empirical CDF on sample $\{t_i\}_{i=1}^n$. Then we have

$$
\begin{aligned}
& B_{n,2}\left(s,t\right)\\
=& -\frac{1}{\phi\left(s\right)}n^{-\frac{1}{2}}h^{-\frac{1}{2}}\sum_{i=1}^n\\
& \left[\iint_{\mathbb{R}^2}\mathbb{1}_{\{t_i\leqslant t-hw_v\}}\phi'\left(w_u\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_v dw_u\right.\\
& \left.-\iint_{\mathbb{R}^2}\Phi\left(t-hw_v\right)\phi'\left(w_u\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi\left(w_v\right)dw_v dw_u\right]\\
=& n^{-\frac{1}{2}}h^{-\frac{1}{2}}\sum_{i=1}^n\left(B_{n,2,i}\left(s,t\right)-\mathbb{E}\left[B_{n,2,i}\left(s,t\right)\right]\right)
\end{aligned}
$$

For $\mathbb{E}\left[B_{n,2,i}(s,t)\right]$, by Taylor expansion at (s,t), we have

$$
\begin{aligned}
&\mathbb{E}\left[B_{n,2,i}(s,t)\right]\\
=&\frac{-1}{\phi(s)}\iint_{\mathbb{R}^2}\left\{\Phi(t)\frac{\partial\mathcal{C}}{\partial v}(\Phi(s),\Phi(t))-\frac{\partial\left[\Phi(t)\frac{\partial\mathcal{C}}{\partial v}(\Phi(s),\Phi(t))\right]}{\partial(s,t)^{\intercal}}(hw_u,hw_v)^{\intercal}\right\}\\
&\phi'(w_u)\phi(w_v)\,dw_v dw_u+o(h)\\
=&h\Phi(t)\frac{\partial^2\mathcal{C}}{\partial u\partial v}(\Phi(s),\Phi(t))\iint_{\mathbb{R}^2}w_u\phi'(w_u)\phi(w_v)\,dw_v dw_u+o(h)\\
=&-h\Phi(t)\frac{\partial^2\mathcal{C}}{\partial u\partial v}(\Phi(s),\Phi(t))+o(h).
\end{aligned}
$$

For $\mathbb{E}\left[B^2_{n,2,i}(s,t)\right]$, by Taylor expansions at (s,t), we have

$$
\begin{aligned}
&\mathbb{E}\left[B^2_{n,2,i}(s,t)\right]\\
=&\mathbb{E}\left[\frac{1}{\phi^2(s)}\iiiint_{\mathbb{R}^4}\mathbb{1}_{\{t_i\leqslant t-hw_{v1}\}}\mathbb{1}_{\{t_i\leqslant t-hw_{v2}\}}\right.\\
&\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u1}),\Phi(t-hw_{v1}))\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u2}),\Phi(t-hw_{v2}))\\
&\left.\phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2}\right]\\
=&\frac{1}{\phi^2(s)}\iiiint_{\mathbb{R}^4}\mathbb{E}\left[\mathbb{1}_{\{t_i\leqslant t-hw_{v1}\}}\mathbb{1}_{\{t_i\leqslant t-hw_{v2}\}}\right]\\
&\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u1}),\Phi(t-hw_{v1}))\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u2}),\Phi(t-hw_{v2}))\\
&\phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2}\\
=&\frac{1}{\phi^2(s)}\iiiint_{\mathbb{R}^4}\Phi(t-h\max\{w_{v1},w_{v2}\})\\
&\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u1}),\Phi(t-hw_{v1}))\frac{\partial\mathcal{C}}{\partial v}(\Phi(s-hw_{u2}),\Phi(t-hw_{v2}))\\
&\phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2}
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{2}{\phi^2(s)}\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}\Phi\left(t-hw_{v2}\right)\\
&\frac{\partial\mathcal{C}}{\partial v}\left(\Phi\left(s-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial v}\left(\Phi\left(s-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)\\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}\\
=&\frac{2}{\phi^2(s)}\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}\left\{\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\Phi(t)\right.\\
&-\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\Phi(t)\frac{\partial\left[\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\right]}{\partial(s,t)^\intercal}(hw_{u1},hw_{v1})^\intercal\\
&\left.-\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\frac{\partial\left[\Phi(t)\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\right]}{\partial(s,t)^\intercal}(hw_{u2},hw_{v2})^\intercal\right\}\\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}+o(h)\\
=&-h\frac{2}{\phi(s)}\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s),\Phi(t)\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi(s),\Phi(t)\right)\Phi(t)\iiint_{\mathbb{R}^3}\\
&\int_{w_{v1}}^{\infty}\left(w_{u1}+w_{u2}\right)\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}+o(h)\\
=&o(h).
\end{aligned}
$$

Then, the rest of the proof follows directly from the above results. □

Lemma A.1.3. *Let $h\to 0$ as $n\to\infty$. When $n\to\infty$, we have the following result holds for $A_n(s,t)$ defined in* (3.9.13)*, for any $(s,t)\in\mathbb{R}^2$,*

$$
\mathbb{E}\left[A_n(s,t)\right]=\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s))+h^2B(\Phi(s),\Phi(t))+o(h^2),
$$

where $B(\cdot,\cdot)$ is given in Section 3.8.

Proof. By definition of $A_n(\cdot,\cdot)$, and $(s_i,t_i) \overset{\text{i.i.d.}}{\sim} F_{S,T}$, $i=1,\cdots,n$, we have

$$\begin{aligned}\mathbb{E}\left[A_n(s,t)\right] =&\mathbb{E}\left[\widehat{\mathcal{C}}^*_{2|1}\left(\Phi(t) \mid \Phi(s)\right)\right]\\ =&\frac{1}{h\phi\left(s\right)}\mathbb{E}\left[\phi\left(\frac{s-s_1}{h}\right)\Phi\left(\frac{t-t_1}{h}\right)\right]\\ =&\frac{1}{h\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(\frac{s-s^*}{h}\right)\Phi\left(\frac{t-t^*}{h}\right)f_{S,T}\left(s^*,t^*\right)dt^*ds^*\\ =&\frac{1}{h\phi\left(s\right)}\int_{\mathbb{R}}\phi\left(\frac{s-s^*}{h}\right)\Bigg\{\left[\Phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx\right]\Bigg|_{t^*=-\infty}^{\infty}\\ &+\int_{\mathbb{R}}\frac{1}{h}\phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dxdt^*\Bigg\}ds^*,\end{aligned} \tag{A.1.1}$$

where (A.1.1) follows from integration by parts. In (A.1.1), for any $(s,t)\in\mathbb{R}^2$, $s^*\in\mathbb{R}$, and $h>0$, we have

$$\begin{aligned}\lim_{t^*\to+\infty}\Phi\left(\frac{t-t^*}{h}\right)&=0,\\ \lim_{t^*\to-\infty}\Phi\left(\frac{t-t^*}{h}\right)&=1,\\ \lim_{t^*\to-\infty}\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx&=0,\end{aligned}$$

and

$$g\left(s^*,t^*\right)\doteq\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx=\frac{\partial F_{S,T}}{\partial s}\left(s^*,t^*\right)=\phi\left(s^*\right)\mathcal{C}_{2|1}\left(\Phi(t^*)\mid\Phi(s^*)\right), \tag{A.1.2}$$

which is bounded for any $t^* \in \mathbb{R}$, hence $\left[\Phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*} f_{S,T}\left(s^*, x\right)dx\right]\Big|_{t^*=-\infty}^{\infty} = 0$. Then by change of variables $u_s = \frac{s-s*}{h}$, and $u_t = \frac{t-t*}{h}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[A_n(s,t)\right] \\
=&\frac{1}{h^2\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(\frac{s-s^*}{h}\right)\phi\left(\frac{t-t^*}{h}\right)\left\{\int_{-\infty}^{t^*} f_{S,T}\left(s^*,x\right)dx\right\}dt^*ds^* \\
=&\frac{1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)\left\{\int_{-\infty}^{t-hu_t} f_{S,T}\left(s-hu_s,x\right)dx\right\}du_t du_s \qquad \text{(A.1.3)} \\
=&\frac{1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)g\left(s,t\right)du_t du_s \\
&-\frac{1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)\left[\frac{\partial g\left(s,t\right)}{\partial(s,t)^\intercal}(hu_s,hu_t)^\intercal\right]du_t du_s \\
&+\frac{1}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)\left[\frac{1}{2}(hu_s,hu_t)\frac{\partial^2 g\left(s,t\right)}{\partial(s,t)^\intercal\partial(s,t)}(hu_s,hu_t)^\intercal\right]du_t du_s \\
&+o\left(h^2\right) \qquad \text{(A.1.4)} \\
=&\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)du_t du_s \\
&+\frac{h^2}{2\phi\left(s\right)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)\operatorname{tr}\left[\frac{\partial^2 g\left(s,t\right)}{\partial(s,t)^\intercal\partial(s,t)}\mathbf{u}\mathbf{u}^\intercal\right]du_t du_s \\
&+o\left(h^2\right) \qquad \text{(A.1.5)} \\
=&\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)+\frac{h^2}{2\phi\left(s\right)}\operatorname{tr}\left[\frac{\partial^2 g\left(s,t\right)}{\partial(s,t)^\intercal\partial(s,t)}\iint_{\mathbb{R}^2}\phi\left(u_s\right)\phi\left(u_t\right)\mathbf{u}\mathbf{u}^\intercal du_t du_s\right] \\
&+o\left(h^2\right) \\
=&\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)+\frac{h^2}{2\phi\left(s\right)}\operatorname{tr}\left[\frac{\partial^2 g\left(s,t\right)}{\partial(s,t)^\intercal\partial(s,t)}\mathbf{I}_2\right]+o\left(h^2\right) \qquad \text{(A.1.6)} \\
=&\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)+\frac{h^2}{2\phi\left(s\right)}\left[\frac{\partial^2 g\left(s,t\right)}{\partial s^2}+\frac{\partial^2 g\left(s,t\right)}{\partial t^2}\right]+o\left(h^2\right), \qquad \text{(A.1.7)}
\end{aligned}
$$

where (A.1.4) follows from Taylor expansion of $g\left(s-hu_s, t-hu_t\right)$ at (s,t); (A.1.5) follows from the fact in (A.1.2), the second term in (A.1.4) integrates to 0, and $\mathbf{u}=(u_s,u_t)^\intercal$; (A.1.6) follows from

$$\iint_{\mathbb{R}^2} \phi\left(u_s\right)\phi\left(u_t\right)\mathbf{u}\mathbf{u}^\intercal du_s du_t = \mathbf{I}_2,$$

which is 2 by 2 identity matrix.

The rest of the proof follows directly from substituting the second order partial derivatives of $g\left(s,t\right) = \phi\left(s\right)\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s)\right)$ into (A.1.7). □

Lemma A.1.4. *Let $h\rightarrow 0$ as $n\rightarrow\infty$. When $n\rightarrow\infty$, we have the following result holds for $A_n(s,t)$ defined in* (3.9.13)*, for any $(s,t)\in\mathbb{R}^2$,*

$$\mathit{Var}\left[A_n(s,t)\right] = \frac{1}{nh}\sigma^2\left(\Phi(s),\Phi(t)\right) + o\left(\frac{1}{nh}\right),$$

where $\sigma^2(\cdot,\cdot)$ is given in Section 3.8.

Proof. By Lemma A.1.3, when $n\rightarrow\infty$, we have

$$\mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right] = h\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s)) + h^3 B(\Phi(s),\Phi(t)) + o(h^3). \tag{A.1.8}$$

Besides, by definition of $A_n(\cdot,\cdot)$, and Lemmas A.1.1 and A.1.2, we have

$$\begin{aligned}&\mathbb{E}\left[\mathcal{B}_n^2\left[s,t;(s_i,t_i)\right]\right]\\ =&\mathbb{E}\left[\frac{\phi^2\left(\frac{s-s_i}{h}\right)\Phi^2\left(\frac{t-t_i}{h}\right)}{\phi^2\left(s\right)}\right] + \mathbb{E}\left[B_{n,1,i}^2\left(s,t\right)\right] + 2\mathbb{E}\left[B_{n,1,i}\left(s,t\right)B_{n,2,i}\left(s,t\right)\right]\end{aligned}$$

$$+ 2\mathbb{E}\left[\frac{\phi\left(\frac{s-s_i}{h}\right)\Phi\left(\frac{t-t_i}{h}\right)}{\phi\left(s\right)}B_{n,1,i}\left(s,t\right)\right] + 2\mathbb{E}\left[\frac{\phi\left(\frac{s-s_i}{h}\right)\Phi\left(\frac{t-t_i}{h}\right)}{\phi\left(s\right)}B_{n,2,i}\left(s,t\right)\right] + o(h)$$

$$\doteq \sum_{k=1}^{5} I_{n2k}(s,t) + o(h). \tag{A.1.9}$$

Then we consider the five terms separately.

1. For $I_{n21}(s,t)$, similar with the proof of Lemma A.1.3, when $h \to 0$ as $n \to \infty$, we have

$$\begin{aligned}
& I_{n21}(s,t) \\
=& \frac{1}{\phi^2\left(s\right)} \iint_{\mathbb{R}^2} \phi^2\left(\frac{s-s^*}{h}\right)\Phi^2\left(\frac{t-t^*}{h}\right) f_{S,T}\left(s^*,t^*\right) dt^* ds^* \\
=& \frac{1}{\phi^2\left(s\right)} \int_{\mathbb{R}} \phi^2\left(\frac{s-s^*}{h}\right)\Bigg\{\left[\Phi^2\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*} f_{S,T}\left(s^*,x\right)dx\right]\Bigg|_{t^*=-\infty}^{\infty} \\
& + \int_{\mathbb{R}} \frac{2}{h}\Phi\left(\frac{t-t^*}{h}\right)\phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*} f_{S,T}\left(s^*,x\right)dx dt^*\Bigg\} ds^* \qquad \text{(A.1.10)} \\
=& \frac{2}{h\phi^2\left(s\right)} \iint_{\mathbb{R}^2} \phi^2\left(\frac{s-s^*}{h}\right)\Phi\left(\frac{t-t^*}{h}\right)\phi\left(\frac{t-t^*}{h}\right) \\
& \int_{-\infty}^{t^*} f_{S,T}\left(s^*,x\right)dx dt^* ds^* \qquad \text{(A.1.11)} \\
=& \frac{2h}{\phi^2\left(s\right)} \iint_{\mathbb{R}^2} \phi^2\left(u_s\right)\phi\left(u_t\right)\Phi\left(u_t\right) g\left(s-hu_s, t-hu_t\right) du_t du_s \qquad \text{(A.1.12)} \\
=& \frac{2h}{\phi\left(s\right)}\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right) \iint_{\mathbb{R}^2} \phi^2\left(u_s\right)\phi\left(u_t\right)\Phi\left(u_t\right) du_t du_s + o(h) \qquad \text{(A.1.13)} \\
=& \frac{h}{2\sqrt{\pi}} \frac{\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)}{\phi\left(s\right)} + o(h), \qquad \text{(A.1.14)}
\end{aligned}$$

where (A.1.10) follows from similar integration by parts with (A.1.1); (A.1.11) follows from

$$\left[\Phi^2\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*} f_{S,T}\left(s^*,x\right)dx\right]\Bigg|_{t^*=-\infty}^{\infty} = 0,$$

which is induced by similar analysis in Lemma A.1.3; (A.1.12) follows from the same change of variables in (A.1.3); (A.1.13) follows from Taylor expansion at (s,t), and the truth in (A.1.2); (A.1.14) follows from

$$\iint_{\mathbb{R}^2} \phi^2(u_s) \cdot \phi(u_t)\,\Phi(u_t)\,du_t du_s = \frac{1}{4\sqrt{\pi}}.$$

2. When $h \to 0$ as $n \to \infty$, the result of $I_{n22}(s,t)$ is given in Lemma A.1.1.

3. For $I_{n23}(s,t)$, similar with the proof of Lemmas A.1.1 and A.1.2, when $h \to 0$ as $n \to \infty$, by definition and Taylor expansions at (s,t), we have

$$
\begin{aligned}
& I_{n23}(s,t) \\
=& 2\mathbb{E}\Bigg[\frac{1}{\phi^2(s)} \iiiint_{\mathbb{R}^4} \mathbb{1}_{\{s_i \leqslant s - hw_{u1}\}} \mathbb{1}_{\{t_i \leqslant t - hw_{v2}\}} \\
& \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s-hw_{u1}), \Phi(t-hw_{v1})\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi(s-hw_{u2}), \Phi(t-hw_{v2})\right) \\
& \phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2}\Bigg] \\
=& \frac{2}{\phi^2(s)} \iiiint_{\mathbb{R}^4} \mathbb{E}\left[\mathbb{1}_{\{s_i \leqslant s - hw_{u1}\}} \mathbb{1}_{\{t_i \leqslant t - hw_{v2}\}}\right] \\
& \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s-hw_{u1}), \Phi(t-hw_{v1})\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi(s-hw_{u2}), \Phi(t-hw_{v2})\right) \\
& \phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2} \\
=& \frac{2}{\phi^2(s)} \iiiint_{\mathbb{R}^4} F_{S,T}(s-hw_{u1}, t-hw_{v2}) \\
& \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s-hw_{u1}), \Phi(t-hw_{v1})\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi(s-hw_{u2}), \Phi(t-hw_{v2})\right) \\
& \phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2} \\
=& \frac{2}{\phi^2(s)} \iiiint_{\mathbb{R}^4} \Bigg\{ F_{S,T}(s,t) \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s), \Phi(t)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi(s), \Phi(t)\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s\right), \Phi\left(t\right)\right) \frac{\partial F_{S,T}\left(s,t\right)}{\partial (s,t)^\intercal}(hw_{u1}, hw_{v2})^\intercal \\
&- F_{S,T}\left(s,t\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s\right), \Phi\left(t\right)\right) \frac{\partial\left[\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]}{\partial (s,t)^\intercal}(hw_{u1}, hw_{v1})^\intercal \\
&- F_{S,T}\left(s,t\right) \frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s\right), \Phi\left(t\right)\right) \frac{\partial\left[\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right]}{\partial (s,t)^\intercal}(hw_{u2}, hw_{v2})^\intercal \Bigg\} \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right) dw_{v1}dw_{u1}dw_{v2}dw_{u2} + o(h) \\
=& o(h).
\end{aligned}
$$

4. For $I_{n24}(s,t)$, similar with the proof of Lemma A.1.3, when $h \to 0$ as $n \to \infty$, we have

$$
\begin{aligned}
&I_{n24}(s,t) \\
=&-\frac{2}{\phi^2\left(s\right)}\mathbb{E}\Bigg[\iint_{\mathbb{R}^2} \phi\left(\frac{s-s_i}{h}\right)\Phi\left(\frac{t-t_i}{h}\right)\mathbb{1}_{\{s_i \leqslant s-hw_u\}} \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u\Bigg] \\
=&-\frac{2}{\phi^2\left(s\right)}\iint_{\mathbb{R}^2}\mathbb{E}\left[\phi\left(\frac{s-s_i}{h}\right)\Phi\left(\frac{t-t_i}{h}\right)\mathbb{1}_{\{s_i \leqslant s-hw_u\}}\right] \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \\
=&-\frac{2}{\phi^2\left(s\right)}\iint_{\mathbb{R}^2}\int_{-\infty}^{s-hw_u}\phi\left(\frac{s-s^*}{h}\right)\int_{\mathbb{R}}\Phi\left(\frac{t-t^*}{h}\right)f_{S,T}\left(s^*,t^*\right)dt^*ds^* \\
&\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \\
=&-\frac{2}{\phi^2\left(s\right)}\iint_{\mathbb{R}^2}\int_{-\infty}^{s-hw_u}\phi\left(\frac{s-s^*}{h}\right) \\
&\left\{\left[\Phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx\right]\Bigg|_{t^*=-\infty}^{\infty}\right. \\
&\left.+\frac{1}{h}\int_{\mathbb{R}}\phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dxdt^*\right\}ds^*
\end{aligned}
$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \tag{A.1.15}$$

$$=-\frac{1}{h}\frac{2}{\phi^2\left(s\right)}\iint_{\mathbb{R}^2}\int_{-\infty}^{s-hw_u}\phi\left(\frac{s-s^*}{h}\right)\int_{\mathbb{R}}\phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)$$

$$dxdt^*ds^*\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \tag{A.1.16}$$

$$=-h\frac{2}{\phi^2\left(s\right)}\iint_{\mathbb{R}^2}\int_{w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s-hz_s,t-hz_t\right)dz_t dz_s$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \tag{A.1.17}$$

$$=-h\frac{2\mathcal{C}_{2|1}^2\left(\Phi\left(t\right)\mid\Phi\left(s\right)\right)}{\phi\left(s\right)}\iint_{\mathbb{R}^2}\int_{w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)dz_t dz_s$$

$$\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u+o(h) \tag{A.1.18}$$

$$=-h\frac{\mathcal{C}_{2|1}^2\left(\Phi\left(t\right)\mid\Phi\left(s\right)\right)}{\sqrt{\pi}\phi\left(s\right)}+o(h), \tag{A.1.19}$$

where (A.1.15) follows from similar integration by parts with (A.1.1); (A.1.16) follows from

$$\left[\Phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx\right]\Big|_{t^*=-\infty}^{\infty}=0,$$

which is analyzed in Lemma A.1.3; (A.1.17) follows from change of variables, $z_t=\frac{t-t^*}{h}$, and $z_s=\frac{s-s^*}{h}$; (A.1.18) follows from Taylor expansions at (s,t), the truth in (A.1.2), and the definition of $\mathcal{C}_{2|1}(\cdot\mid\cdot)$; (A.1.19) follows from

$$\iint_{\mathbb{R}^2}\int_{w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)dz_t dz_s\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u=\frac{1}{2\sqrt{\pi}}.$$

5. For $I_{n25}(s,t)$, similar with the proof of Lemma A.1.3, when $h \to 0$ as $n \to \infty$, we have

$$
\begin{aligned}
& I_{n25}(s,t) \\
= & -\frac{2}{\phi^2(s)} \mathbb{E}\Bigg[\iint_{\mathbb{R}^2} \phi\left(\frac{s-s_i}{h}\right) \Phi\left(\frac{t-t_i}{h}\right) \mathbb{1}_{\{t_i \leqslant t-hw_v\}} \\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \Bigg] \\
= & -\frac{2}{\phi^2(s)} \iint_{\mathbb{R}^2} \mathbb{E}\left[\phi\left(\frac{s-s_i}{h}\right) \Phi\left(\frac{t-t_i}{h}\right) \mathbb{1}_{\{t_i \leqslant t-hw_v\}} \right] \\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \\
= & -\frac{2}{\phi^2(s)} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(\frac{s-s^*}{h}\right) \int_{-\infty}^{t-hw_v} \Phi\left(\frac{t-t^*}{h}\right) f_{S,T}\left(s^*, t^*\right) dt^* ds^* \\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \\
= & -\frac{2}{\phi^2(s)} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(\frac{s-s^*}{h}\right) \\
& \Bigg\{ \left[\Phi\left(\frac{t-t^*}{h}\right) \int_{-\infty}^{t^*} f_{S,T}\left(s^*, x\right) dx \right] \Bigg|_{t^*=-\infty}^{t-hw_v} \\
& + \frac{1}{h} \int_{-\infty}^{t-hw_v} \phi\left(\frac{t-t^*}{h}\right) \int_{-\infty}^{t^*} f_{S,T}\left(s^*, x\right) dx dt^* \Bigg\} ds^* \\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \qquad \text{(A.1.20)} \\
= & -\frac{2}{\phi^2(s)} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(\frac{s-s^*}{h}\right) \\
& \left\{ \Phi\left(w_v\right) g\left(s^*, t-hw_v\right) + \frac{1}{h} \int_{-\infty}^{t-hw_v} \phi\left(\frac{t-t^*}{h}\right) g(s^*, t^*) dt^* \right\} ds^* \\
& \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \\
= & -h \frac{2}{\phi^2(s)} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(z_s\right) \\
& \left\{ \Phi\left(w_v\right) g\left(s-hz_s, t-hw_v\right) + \int_{w_v}^{\infty} \phi\left(z_t\right) g(s-hz_s, t-hz_t) dz_t \right\} dz_s
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s-hw_u\right), \Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u \qquad \text{(A.1.21)}\\
=&-h\frac{2\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s\right)\right)\mathcal{C}_{1|2}\left(\Phi\left(s\right) \mid \Phi\left(t\right)\right)}{\phi\left(s\right)}\\
&\iiint_{\mathbb{R}^3}\phi\left(z_s\right)\Phi\left(w_v\right)dz_s\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u\\
&-h\frac{2\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s\right)\right)\mathcal{C}_{1|2}\left(\Phi\left(s\right) \mid \Phi\left(t\right)\right)}{\phi\left(s\right)}\\
&\iiint_{\mathbb{R}^3}\phi\left(z_s\right)\int_{w_v}^{\infty}\phi\left(z_t\right)dz_t dz_s\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u + o(h) \qquad \text{(A.1.22)}\\
=&o(h),
\end{aligned}
$$

where (A.1.20) follows from similar integration by parts with (A.1.1); (A.1.21) follows from change of variables, $z_t = \frac{t-t^*}{h}$, and $z_s = \frac{s-s^*}{h}$; (A.1.22) follows from Taylor expansions at (s,t), the truth in (A.1.2), and the definition of $\mathcal{C}_{1|2}(\cdot \mid \cdot)$.

Combining all above results and by (A.1.9), we have

$$
\mathbb{E}\left[\mathcal{B}_n^2\left[s,t;\left(s_i,t_i\right)\right]\right] = \frac{h}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)} - \frac{h}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)} + o(h), \qquad \text{(A.1.23)}
$$

as $n \to \infty$. Results in (A.1.8) and (A.1.23) induce

$$
\begin{aligned}
\operatorname{Var}\left[\mathcal{B}_n\left[s,t;\left(s_i,t_i\right)\right]\right] =& \mathbb{E}\left[\mathcal{B}_n^2\left[s,t;\left(s_i,t_i\right)\right]\right] - \left\{\mathbb{E}\left[\mathcal{B}_n\left[s,t;\left(s_i,t_i\right)\right]\right]\right\}^2\\
=& h\left[\frac{1}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)} - \frac{1}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)}\right]\\
&+ o(h), \qquad \text{(A.1.24)}
\end{aligned}
$$

as $n \to \infty$. Consequently, by definition of $A_n(\cdot, \cdot)$, and $(s_i, t_i) \overset{\text{i.i.d.}}{\sim} F_{S,T}$, $i = 1, \cdots, n$, we have

$$\begin{aligned}
\text{Var}\left[A_n(s,t)\right] =& n^{-1}h^{-2}\text{Var}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right] \\
=& \frac{1}{nh}\left[\frac{1}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)} - \frac{1}{2\sqrt{\pi}}\frac{\mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi\left(s\right)\right)}{\phi\left(s\right)}\right] \\
& + o\left(\frac{1}{nh}\right). \qquad \text{(A.1.25)}
\end{aligned}$$

□

Lemma A.1.5. *Let $h \to 0$, and $nh \to \infty$ as $n \to \infty$. When $n \to \infty$, we have the following result holds for $A_n(s,t)$ defined in* (3.9.13)*, for any $(s,t) \in \mathbb{R}^2$,*

$$\sqrt{nh}\left[A_n(s,t) - \mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s)) - h^2 B(\Phi(s), \Phi(t)) - o(h^2)\right] \xrightarrow{d} N\left(0, \sigma^2\left(\Phi(s), \Phi(t)\right)\right),$$

where $B(\cdot, \cdot)$ and $\sigma^2(\cdot, \cdot)$ are given in Section 3.8.

Proof. By similar with the proof in Lemmas A.1.3 and A.1.4, it is easy to find $\mathbb{E}\left[\left|\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right|^3\right] = \mathcal{O}(h)$. Hence, together with the definition of $A_n(\cdot, \cdot)$, and (A.1.23), we have

$$\begin{aligned}
\sigma_{in}^2 \doteq& \text{Var}\left[\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h}\right] = \mathcal{O}\left(\frac{1}{h}\right), \\
\rho_{in} \doteq& \mathbb{E}\left[\left|\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h} - \mathbb{E}\left[\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h}\right]\right|^3\right] \\
\leqslant& \mathbb{E}\left[\left(\left|\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h}\right| + \left|\mathbb{E}\left[\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h}\right]\right|\right)^3\right] \\
\leqslant& 8\mathbb{E}\left[\left|\frac{\mathcal{B}_n\left[s,t;(s_i,t_i)\right]}{h}\right|^3\right]
\end{aligned}$$

$$=\mathcal{O}\left(\frac{1}{h^2}\right),$$

as $n \to \infty$. Then, when $nh \to \infty$ as $n \to \infty$, we have

$$\begin{aligned}\frac{\left(\sum_{i=1}^{n}\rho_{in}\right)^{\frac{1}{3}}}{\left(\sum_{i=1}^{n}\sigma_{in}^2\right)^{\frac{1}{2}}} \leqslant & \mathcal{O}\left(n^{\frac{1}{3}}h^{-\frac{2}{3}}\right)\cdot\mathcal{O}\left(n^{-\frac{1}{2}}h^{\frac{1}{2}}\right)\\ =&\mathcal{O}\left((nh)^{-\frac{1}{6}}\right)\to 0,\end{aligned}$$

as $n \to \infty$. Then by Liapunov's CLT for Triangular Arrays, we have asymptotic normality of $A_n(s,t) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathcal{B}_n[s,t;(s_i,t_i)]}{h}$. The rest of the proof follows directly from Lemmas A.1.3 and A.1.4. □

Lemma A.1.6. *Suppose Condition 3.4 holds. Let $h \to 0$ as $n \to \infty$. For I_{n1} defined in (3.9.14), consider $I_{n1} \doteq \frac{1}{nh}\sum_{i=1}^{n} Z_{n1i}$. For any quantile level $\alpha \in (0,1)$, we have the following results hold as $n \to \infty$.*

*(a) When Y is **quantile independent** with X at the quantile level α, we have*

$$\begin{aligned}\mathbb{E}\left[Z_{n1i}\right] =&0,\\ \mathbb{E}\left[Z_{n1i}^2\right] = \mathit{Var}\left[Z_{n1i}\right] =&h^6\sigma^2_{\perp_\alpha,1}(\omega)+o(h^6),\end{aligned}$$

where $\sigma^2_{\perp_\alpha,1}(\omega)$ is defined in Section 3.8. Correspondingly, we have $I_{n1} \xrightarrow{d} N\left(0, n^{-1}h^4\sigma^2_{\perp_\alpha,1}(\omega)\right)$.

*(b) When Y is **quantile dependent** with X at the quantile level α, we have*

$$\begin{aligned}\mathbb{E}\left[Z_{n1i}\right] =&0,\\ \mathbb{E}\left[Z_{n1i}^2\right] = \mathit{Var}\left[Z_{n1i}\right] =&h^2\sigma^2_{\not\perp_\alpha,1}(\omega)+o(h^2),\end{aligned}$$

where $\sigma^2_{\not\perp_\alpha,1}(\omega)$ *is defined in Section 3.8. Correspondingly, we have* $I_{n1} \xrightarrow{d} N\left(0, n^{-1}\sigma^2_{\not\perp_\alpha,1}(\omega)\right)$.

Proof. With $t = \Phi^{-1}(\alpha)$, we define $Z_{n1i} \doteq Y_{n1i} - \mathbb{E}\left[Y_{n1i}\right]$, where

$$Y_{n1i} = \int_{\mathbb{R}} \mathcal{B}_n\left[s, t; (s_i, t_i)\right]\left\{\mathbb{E}\left[A_n\left(s, t\right)\right] - \alpha\right\}\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds.$$

Define $t_n^{(k)} = \mathbb{E}\left[Y_{n1i}^k\right]$ for $k \in \mathbb{Z}^+$, then we have

$$t_n^{(1)} = \int_{\mathbb{R}} \mathbb{E}\left[\mathcal{B}_n\left[s, t; (s_i, t_i)\right]\right]\left\{\mathbb{E}\left[A_n\left(s, t\right)\right] - \alpha\right\}\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds.$$

By Lemma A.1.3, and (A.1.8), we have

$$t_n^{(1)} = h^3\alpha\int_{\mathbb{R}} B(\Phi\left(s\right), \Phi(t))\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds + o(h^3), \quad \text{if } Y \perp_\alpha X, \tag{A.1.26}$$

$$t_n^{(1)} = h\int_{\mathbb{R}} \mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right)\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \alpha\right]\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds + \mathcal{O}(h^3), \quad \text{if } Y \not\perp_\alpha X, \tag{A.1.27}$$

as $n \to \infty$. Then, we consider

$$\begin{aligned}
&t_n^{(2)} \\
=&\mathbb{E}\left[Y_{n1i}^2\right] \\
=&\mathbb{E}\Bigg[\iint_{\mathbb{R}^2} \mathcal{B}_n\left[s_{*1}, t; (s_i, t_i)\right]\left\{\mathbb{E}\left[A_n\left(s_{*1}, t\right)\right] - \alpha\right\}\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right) \\
&\mathcal{B}_n\left[s_{*2}, t; (s_i, t_i)\right]\left\{\mathbb{E}\left[A_n\left(s_{*2}, t\right)\right] - \alpha\right\}\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*2}ds_{*1}\Bigg]
\end{aligned}$$

$$= \iint_{\mathbb{R}^2} \mathbb{E}\left(\mathcal{B}_n\left[s_{*1}, t; (s_i, t_i)\right] \mathcal{B}_n\left[s_{*2}, t; (s_i, t_i)\right]\right)$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*1}\right)\right) - \alpha\right] + h^2 B(\Phi\left(s_{*1}\right), \Phi(t)) + o(h^2)\right\} \phi\left(s_{*1}\right) \omega\left(\Phi\left(s_{*1}\right)\right)$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*2}\right)\right) - \alpha\right] + h^2 B(\Phi\left(s_{*2}\right), \Phi(t)) + o(h^2)\right\} \phi\left(s_{*2}\right) \omega\left(\Phi\left(s_{*2}\right)\right)$$

$$ds_{*2} ds_{*1}, \tag{A.1.28}$$

when $h \to 0$ as $n \to \infty$. Then, by definition of $A_n(\cdot, \cdot)$, we have

$$\begin{aligned}
&\mathbb{E}\left(\mathcal{B}_n\left[s_{*1}, t; (s_i, t_i)\right] \mathcal{B}_n\left[s_{*2}, t; (s_i, t_i)\right]\right) \\
=&\mathbb{E}\Bigg\{\Bigg[\frac{\phi\left(\frac{s_{*1}-s_i}{h}\right)}{\phi\left(s_{*1}\right)} \Phi\left(\frac{t-t_i}{h}\right) + B_{n,1,i}\left(s_{*1}, t\right) + B_{n,2,i}\left(s_{*1}, t\right) \\
&- \mathbb{E}\left[B_{n,1,i}\left(s_{*1}, t\right)\right] - \mathbb{E}\left[B_{n,2,i}\left(s_{*1}, t\right)\right]\Bigg] \\
&\Bigg[\frac{\phi\left(\frac{s_{*2}-s_i}{h}\right)}{\phi\left(s_{*2}\right)} \Phi\left(\frac{t-t_i}{h}\right) + B_{n,1,i}\left(s_{*2}, t\right) + B_{n,2,i}\left(s_{*2}, t\right) \\
&- \mathbb{E}\left[B_{n,1,i}\left(s_{*2}, t\right)\right] - \mathbb{E}\left[B_{n,2,i}\left(s_{*2}, t\right)\right]\Bigg]\Bigg\} \\
=&\mathbb{E}\left[\frac{\phi\left(\frac{s_{*1}-s_i}{h}\right)}{\phi\left(s_{*1}\right)} \Phi\left(\frac{t-t_i}{h}\right) \frac{\phi\left(\frac{s_{*2}-s_i}{h}\right)}{\phi\left(s_{*2}\right)} \Phi\left(\frac{t-t_i}{h}\right)\right] \\
&+ \mathbb{E}\left[\frac{\phi\left(\frac{s_{*1}-s_i}{h}\right)}{\phi\left(s_{*1}\right)} \Phi\left(\frac{t-t_i}{h}\right) B_{n,1,i}\left(s_{*2}, t\right)\right] \\
&+ \mathbb{E}\left[\frac{\phi\left(\frac{s_{*1}-s_i}{h}\right)}{\phi\left(s_{*1}\right)} \Phi\left(\frac{t-t_i}{h}\right) B_{n,2,i}\left(s_{*2}, t\right)\right] \\
&- \mathbb{E}\left[\frac{\phi\left(\frac{s_{*1}-s_i}{h}\right)}{\phi\left(s_{*1}\right)} \Phi\left(\frac{t-t_i}{h}\right)\right] \left(\mathbb{E}\left[B_{n,1,i}\left(s_{*2}, t\right)\right] + \mathbb{E}\left[B_{n,2,i}\left(s_{*2}, t\right)\right]\right) \\
&+ \mathbb{E}\left[\frac{\phi\left(\frac{s_{*2}-s_i}{h}\right)}{\phi\left(s_{*2}\right)} \Phi\left(\frac{t-t_i}{h}\right) B_{n,1,i}\left(s_{*1}, t\right)\right] \\
&+ \mathbb{E}\left[B_{n,1,i}\left(s_{*1}, t\right) B_{n,1,i}\left(s_{*2}, t\right)\right]
\end{aligned}$$

$$+\,\mathbb{E}\left[B_{n,1,i}\left(s_{*1},t\right)B_{n,2,i}\left(s_{*2},t\right)\right]$$

$$+\,\mathbb{E}\left[\frac{\phi\left(\frac{s_{*2}-s_i}{h}\right)}{\phi\left(s_{*2}\right)}\Phi\left(\frac{t-t_i}{h}\right)B_{n,2,i}\left(s_{*1},t\right)\right]$$

$$+\,\mathbb{E}\left[B_{n,1,i}\left(s_{*2},t\right)B_{n,2,i}\left(s_{*1},t\right)\right]$$

$$+\,\mathbb{E}\left[B_{n,2,i}\left(s_{*1},t\right)B_{n,2,i}\left(s_{*2},t\right)\right]$$

$$-\,\mathbb{E}\left[\frac{\phi\left(\frac{s_{*2}-s_i}{h}\right)}{\phi\left(s_{*2}\right)}\Phi\left(\frac{t-t_i}{h}\right)\right]\left(\mathbb{E}\left[B_{n,1,i}\left(s_{*1},t\right)\right]+\mathbb{E}\left[B_{n,2,i}\left(s_{*1},t\right)\right]\right)$$

$$-\left(\mathbb{E}\left[B_{n,1,i}\left(s_{*1},t\right)\right]+\mathbb{E}\left[B_{n,2,i}\left(s_{*1},t\right)\right]\right)\left(\mathbb{E}\left[B_{n,1,i}\left(s_{*2},t\right)\right]+\mathbb{E}\left[B_{n,2,i}\left(s_{*2},t\right)\right]\right)$$

$$\doteq\sum_{k=1}^{12}I_{n1k}(s_{*1},s_{*2},t), \tag{A.1.29}$$

where we consider the twelve terms separately.

1. For $I_{n11}(s_{*1},s_{*2},t)$, by similar integration by parts in (A.1.1), we have

$$\begin{aligned}
&I_{n11}(s_{*1},s_{*2},t)\\
=&\iint_{\mathbb{R}^2}\frac{\phi\left(\frac{s_{*1}-s^*}{h}\right)}{\phi\left(s_{*1}\right)}\frac{\phi\left(\frac{s_{*2}-s^*}{h}\right)}{\phi\left(s_{*2}\right)}\Phi^2\left(\frac{t-t^*}{h}\right)f_{S,T}\left(s^*,t^*\right)dt^*ds^*\\
=&\frac{2}{h}\frac{1}{\phi\left(s_{*1}\right)}\frac{1}{\phi\left(s_{*2}\right)}\int_{\mathbb{R}}\phi\left(\frac{s_{*1}-s^*}{h}\right)\phi\left(\frac{s_{*2}-s^*}{h}\right)\\
&\int_{\mathbb{R}}\phi\left(\frac{t-t^*}{h}\right)\Phi\left(\frac{t-t^*}{h}\right)\left\{\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dx\right\}dt^*ds^*.
\end{aligned}$$

Then, with change of variables $z_s=\frac{s_{*1}-s^*}{h}$, and $z_t=\frac{t-t^*}{h}$, and the fact in (A.1.2), we have

$$\begin{aligned}
I_{n11}(s_{*1},s_{*2},t)=&2h\frac{1}{\phi\left(s_{*1}\right)}\frac{1}{\phi\left(s_{*2}\right)}\int_{\mathbb{R}}\phi\left(z_s\right)\phi\left(z_s+\frac{s_{*2}-s_{*1}}{h}\right)\\
&\int_{\mathbb{R}}\phi\left(z_t\right)\Phi\left(z_t\right)g(s_{*1}-hz_s,t-hz_t)dz_tdz_s.
\end{aligned}$$

2. For $I_{n12}(s_{*1}, s_{*2}, t)$, by similar integration by parts in (A.1.1), we have

$$
\begin{aligned}
& I_{n12}(s_{*1}, s_{*2}, t) \\
=& \frac{-1}{\phi(s_{*1})\phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{-\infty}^{s_{*2}-hw_u} \phi\left(\frac{s_{*1}-s^*}{h}\right) \int_{\mathbb{R}} \Phi\left(\frac{t-t^*}{h}\right) f_{S,T}(s^*, t^*) \\
& dt^* ds^* \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s_{*2}-hw_u), \Phi(t-hw_v)\right) \phi'(w_u)\phi(w_v)\, dw_v dw_u \\
=& \frac{-1}{h\phi(s_{*1})\phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{-\infty}^{s_{*2}-hw_u} \phi\left(\frac{s_{*1}-s^*}{h}\right) \int_{\mathbb{R}} \phi\left(\frac{t-t^*}{h}\right) \\
& \int_{-\infty}^{t^*} f_{S,T}(s^*, x)\, dx dt^* ds^* \\
& \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s_{*2}-hw_u), \Phi(t-hw_v)\right) \phi'(w_u)\phi(w_v)\, dw_v dw_u.
\end{aligned}
$$

Then, with change of variables $z_t = \frac{t-t^*}{h}$ and $z_s = \frac{s_{*1}-s^*}{h}$, and the fact in (A.1.2), we have

$$
\begin{aligned}
& I_{n12}(s_{*1}, s_{*2}, t) \\
=& -\frac{h}{\phi(s_{*1})\phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{\frac{s_{*1}-s_{*2}}{h}+w_u}^{\infty} \phi(z_s) \int_{\mathbb{R}} \phi(z_t)\, g(s_{*1}-hz_s, t-hz_t) \\
& dz_t dz_s \frac{\partial \mathcal{C}}{\partial u}\left(\Phi(s_{*2}-hw_u), \Phi(t-hw_v)\right) \phi'(w_u)\phi(w_v)\, dw_v dw_u.
\end{aligned}
$$

3. For $I_{n13}(s_{*1}, s_{*2}, t)$, by similar integration by parts in (A.1.1), we have

$$
\begin{aligned}
& I_{n13}(s_{*1}, s_{*2}, t) \\
=& \frac{-1}{\phi(s_{*1})\phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(\frac{s_{*1}-s^*}{h}\right) \int_{-\infty}^{t-hw_v} \Phi\left(\frac{t-t^*}{h}\right) f_{S,T}(s^*, t^*) \\
& dt^* ds^* \frac{\partial \mathcal{C}}{\partial v}\left(\Phi(s_{*2}-hw_u), \Phi(t-hw_v)\right) \phi'(w_u)\phi(w_v)\, dw_v dw_u
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{-1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi\left(\frac{s_{*1}-s^*}{h}\right)\Bigg\{\Phi\left(w_v\right)\int_{-\infty}^{t-hw_v}f_{S,T}\left(s^*,x\right)dx\\
&+\frac{1}{h}\int_{-\infty}^{t-hw_v}\phi\left(\frac{t-t^*}{h}\right)\int_{-\infty}^{t^*}f_{S,T}\left(s^*,x\right)dxdt^*\Bigg\}ds^*\\
&\frac{\partial\mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u.
\end{aligned}
$$

Then, with change of variables $z_t=\frac{t-t^*}{h}$ and $z_s=\frac{s_{*1}-s^*}{h}$, and the fact in (A.1.2), we have

$$
\begin{aligned}
&I_{n13}(s_{*1},s_{*2},t)\\
=&\frac{-h}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi\left(z_s\right)\Bigg[\Phi\left(w_v\right)g\left(s_{*1}-hz_s,t-hw_v\right)\\
&+\int_{w_v}^{\infty}\phi\left(z_t\right)g\left(s_{*1}-hz_s,t-hz_t\right)dz_t\Bigg]dz_s\\
&\frac{\partial\mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u.
\end{aligned}
$$

4. For $I_{n14}(s_{*1},s_{*2},t)$, by Lemmas A.1.1, A.1.2 and A.1.3, we have

$$
\begin{aligned}
&I_{n14}(s_{*1},s_{*2},t)\\
=&h^2\left[\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*1}))\right]\\
&\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*2}\right)\right)+\Phi\left(s_{*2}\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\\
&+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\Bigg]+o(h^2).
\end{aligned}
$$

5. For $I_{n15}(s_{*1},s_{*2},t)$, it is easy to find that $I_{n15}(s_{*1},s_{*2},t)=I_{n12}(s_{*2},s_{*1},t)$.

6. For $I_{n16}(s_{*1}, s_{*2}, t)$, by normality of s_i, we have

$$
\begin{aligned}
&I_{n16}(s_{*1}, s_{*2}, t) \\
=&\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiiint_{\mathbb{R}^4}\Phi\left(\min\{s_{*1}-hw_{u1}, s_{*2}-hw_{u2}\}\right) \\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2} \\
=&\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}^{\infty}\Phi\left(s_{*1}-hw_{u1}\right) \\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2} \\
&+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{-\infty}^{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}\Phi\left(s_{*2}-hw_{u2}\right) \\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}.
\end{aligned}
$$

7. For $I_{n17}(s_{*1}, s_{*2}, t)$, by definition, we have

$$
\begin{aligned}
&I_{n17}(s_{*1}, s_{*2}, t) \\
=&\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiiint_{\mathbb{R}^4}F_{S,T}\left(s_{*1}-hw_{u1}, t-hw_{v2}\right) \\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right), \Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right), \Phi\left(t-hw_{v2}\right)\right) \\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}.
\end{aligned}
$$

8. For $I_{n18}(s_{*1}, s_{*2}, t)$, it is easy to find that $I_{n18}(s_{*1}, s_{*2}, t) = I_{n13}(s_{*2}, s_{*1}, t)$.

9. For $I_{n19}(s_{*1}, s_{*2}, t)$, it is easy to find that $I_{n19}(s_{*1}, s_{*2}, t) = I_{n17}(s_{*2}, s_{*1}, t)$.

10. For $I_{n110}(s_{*1}, s_{*2}, t)$, by normality of t_i, we have

$$
\begin{aligned}
& I_{n110}(s_{*1}, s_{*2}, t) \\
=& \frac{1}{\phi(s_{*1})\phi(s_{*2})} \iiiint_{\mathbb{R}^4} \Phi(t - h\max\{w_{v1}, w_{v2}\}) \\
& \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*1} - hw_{u1}), \Phi(t - hw_{v1})) \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*2} - hw_{u2}), \Phi(t - hw_{v2})) \\
& \phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\, dw_{v1}dw_{u1}dw_{v2}dw_{u2} \\
=& \frac{1}{\phi(s_{*1})\phi(s_{*2})} \iiint_{\mathbb{R}^3} \int_{w_{v1}}^{\infty} \Phi(t - hw_{v2}) \\
& \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*1} - hw_{u1}), \Phi(t - hw_{v1})) \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*2} - hw_{u2}), \Phi(t - hw_{v2})) \\
& \phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\, dw_{v2}dw_{v1}dw_{u1}dw_{u2} \\
& + \frac{1}{\phi(s_{*1})\phi(s_{*2})} \iiint_{\mathbb{R}^3} \int_{-\infty}^{w_{v1}} \Phi(t - hw_{v1}) \\
& \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*1} - hw_{u1}), \Phi(t - hw_{v1})) \frac{\partial \mathcal{C}}{\partial v}(\Phi(s_{*2} - hw_{u2}), \Phi(t - hw_{v2})) \\
& \phi'(w_{u1})\phi(w_{v1})\phi'(w_{u2})\phi(w_{v2})\, dw_{v2}dw_{v1}dw_{u1}dw_{u2}.
\end{aligned}
$$

11. For $I_{n111}(s_{*1}, s_{*2}, t)$, it is easy to find that $I_{n111}(s_{*1}, s_{*2}, t) = I_{n14}(s_{*2}, s_{*1}, t)$.

12. For $I_{n112}(s_{*1}, s_{*2}, t)$, by Lemmas A.1.1 and A.1.2, we have

$$
\begin{aligned}
& I_{n112}(s_{*1}, s_{*2}, t) \\
=& -h^2\Big[\mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*1})) + \Phi(s_{*1}) \frac{\partial^2 \mathcal{C}}{\partial u^2}(\Phi(s_{*1}), \Phi(t))
\end{aligned}
$$

$$+ \Phi(t) \frac{\partial^2 \mathcal{C}}{\partial u \partial v} (\Phi(s_{*1}), \Phi(t)) \Big]$$

$$\Big[\mathcal{C}_{2|1} (\Phi(t) \mid \Phi(s_{*2})) + \Phi(s_{*2}) \frac{\partial^2 \mathcal{C}}{\partial u^2} (\Phi(s_{*2}), \Phi(t))$$

$$+ \Phi(t) \frac{\partial^2 \mathcal{C}}{\partial u \partial v} (\Phi(s_{*2}), \Phi(t)) \Big]$$

$$+ o(h^2).$$

Substituting the above results and (A.1.29) back into (A.1.28), we could have

$$\begin{aligned}
& t_n^{(2)} \\
= & \iint_{\mathbb{R}^2} \sum_{k=1}^{12} I_{n1k}(s_{*1}, s_{*2}, t) \\
& \left\{ \left[\mathcal{C}_{2|1} (\Phi(t) \mid \Phi(s_{*1})) - \alpha \right] + h^2 B(\Phi(s_{*1}), \Phi(t)) + o(h^2) \right\} \phi(s_{*1}) \omega(\Phi(s_{*1})) \\
& \left\{ \left[\mathcal{C}_{2|1} (\Phi(t) \mid \Phi(s_{*2})) - \alpha \right] + h^2 B(\Phi(s_{*2}), \Phi(t)) + o(h^2) \right\} \phi(s_{*2}) \omega(\Phi(s_{*2})) \\
& ds_{*2} ds_{*1} \\
= & \iint_{\mathbb{R}^2} \Bigg[2h \int_{\mathbb{R}} \phi(z_s) \phi\left(z_s + \frac{s_{*2} - s_{*1}}{h} \right) \\
& \int_{\mathbb{R}} \phi(z_t) \Phi(z_t) g(s_{*1} - hz_s, t - hz_t) dz_t dz_s \\
& - h \iint_{\mathbb{R}^2} \int_{\frac{s_{*1} - s_{*2}}{h} + w_u}^{\infty} \phi(z_s) \int_{\mathbb{R}} \phi(z_t) g(s_{*1} - hz_s, t - hz_t) \\
& dz_t dz_s \frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*2} - hw_u), \Phi(t - hw_v)) \phi'(w_u) \phi(w_v) dw_v dw_u \\
& - h \iint_{\mathbb{R}^2} \int_{\frac{s_{*2} - s_{*1}}{h} + w_u}^{\infty} \phi(z_s) \int_{\mathbb{R}} \phi(z_t) g(s_{*2} - hz_s, t - hz_t) \\
& dz_t dz_s \frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1} - hw_u), \Phi(t - hw_v)) \phi'(w_u) \phi(w_v) dw_v dw_u \\
& + \iiint_{\mathbb{R}^3} \int_{\frac{s_{*1} - s_{*2}}{h} + w_{u2}}^{\infty} \Phi(s_{*1} - hw_{u1})
\end{aligned}$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}$$

$$+\iiint_{\mathbb{R}^3}\int_{-\infty}^{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}\Phi\left(s_{*2}-hw_{u2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\left.\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}\right]$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*1}\right)\right)-\alpha\right]+h^2B(\Phi\left(s_{*1}\right),\Phi(t))+o(h^2)\right\}\omega\left(\Phi\left(s_{*1}\right)\right)$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*2}\right)\right)-\alpha\right]+h^2B(\Phi\left(s_{*2}\right),\Phi(t))+o(h^2)\right\}\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*2}ds_{*1} \tag{A.1.30}$$

$$+\iint_{\mathbb{R}^2}\left\{-\frac{h}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi\left(z_s\right)\left[\Phi\left(w_v\right)g\left(s_{*1}-hz_s,t-hw_v\right)\right.\right.$$

$$\left.+\int_{w_v}^{\infty}\phi\left(z_t\right)g\left(s_{*1}-hz_s,t-hz_t\right)dz_t\right]dz_s$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u$$

$$-\frac{h}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi\left(z_s\right)\left[\Phi\left(w_v\right)g\left(s_{*2}-hz_s,t-hw_v\right)\right.$$

$$\left.+\int_{w_v}^{\infty}\phi\left(z_t\right)g\left(s_{*2}-hz_s,t-hz_t\right)dz_t\right]dz_s$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u$$

$$+h^2\left[\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*1}))\right]\left[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*2}\right)\right)+\Phi\left(s_{*2}\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\right.$$

$$\left.+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\right]+o(h^2)$$

$$+h^2\left[\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*2}))\right]\left[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)+\Phi\left(s_{*1}\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\right.$$

$$\left.+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\right]+o(h^2)$$

$$+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiiint_{\mathbb{R}^4}F_{S,T}\left(s_{*1}-hw_{u1},t-hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiiint_{\mathbb{R}^4}F_{S,T}\left(s_{*2}-hw_{u1},t-hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}\Phi\left(t-hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}$$

$$+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{v1}}\Phi\left(t-hw_{v1}\right)$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}$$

$$-h^2\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)+\Phi\left(s_{*1}\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)$$

$$+\Phi\left(t\right)\frac{\partial^2 \mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\Bigg]$$

$$\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*2}\right)\right)+\Phi\left(s_{*2}\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)$$

$$+\Phi\left(t\right)\frac{\partial^2 \mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\Bigg]+o(h^2)\Bigg\}$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*1}\right)\right)-\alpha\right]+h^2B(\Phi\left(s_{*1}\right),\Phi(t))+o(h^2)\right\}\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)$$

$$\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*2}\right)\right)-\alpha\right]+h^2B(\Phi\left(s_{*2}\right),\Phi(t))+o(h^2)\right\}\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)$$

$$ds_{*2}ds_{*1}$$

$$= \iint_{\mathbb{R}^2} \left[2h^2 \int_{\mathbb{R}} \phi(z_s) \phi(z_s+u) \int_{\mathbb{R}} \phi(z_t) \Phi(z_t) g(s_{*1}-hz_s, t-hz_t) dz_t dz_s \right.$$

$$- h^2 \iint_{\mathbb{R}^2} \int_{w_u-u}^{\infty} \phi(z_s) \int_{\mathbb{R}} \phi(z_t) g(s_{*1}-hz_s, t-hz_t) dz_t dz_s$$

$$\frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}+hu-hw_u), \Phi(t-hw_v)) \phi'(w_u) \phi(w_v) dw_v dw_u$$

$$- h^2 \iint_{\mathbb{R}^2} \int_{w_u+u}^{\infty} \phi(z_s) \int_{\mathbb{R}} \phi(z_t) g(s_{*1}+hu-hz_s, t-hz_t) dz_t dz_s$$

$$\frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}-hw_u), \Phi(t-hw_v)) \phi'(w_u) \phi(w_v) dw_v dw_u$$

$$+ h \iiint_{\mathbb{R}^3} \int_{w_{u2}-u}^{\infty} \Phi(s_{*1}-hw_{u1})$$

$$\frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}-hw_{u1}), \Phi(t-hw_{v1})) \frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}+hu-hw_{u2}), \Phi(t-hw_{v2}))$$

$$\phi'(w_{u1}) \phi(w_{v1}) \phi'(w_{u2}) \phi(w_{v2}) dw_{u1} dw_{u2} dw_{v1} dw_{v2}$$

$$+ h \iiint_{\mathbb{R}^3} \int_{-\infty}^{w_{u2}-u} \Phi(s_{*1}+hu-hw_{u2})$$

$$\frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}-hw_{u1}), \Phi(t-hw_{v1})) \frac{\partial \mathcal{C}}{\partial u} (\Phi(s_{*1}+hu-hw_{u2}), \Phi(t-hw_{v2}))$$

$$\left. \phi'(w_{u1}) \phi(w_{v1}) \phi'(w_{u2}) \phi(w_{v2}) dw_{u1} dw_{u2} dw_{v1} dw_{v2} \right]$$

$$\left\{ \left[\mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*1})) - \alpha \right] + h^2 B(\Phi(s_{*1}), \Phi(t)) + o(h^2) \right\} \omega(\Phi(s_{*1}))$$

$$\left\{ \left[\mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*1}+hu)) - \alpha \right] + h^2 B(\Phi(s_{*1}+hu), \Phi(t)) + o(h^2) \right\}$$

$$\omega(\Phi(s_{*1}+hu)) du ds_{*1}$$

$$+ \iint_{\mathbb{R}^2} \left\{ - \frac{h}{\phi(s_{*1}) \phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi(z_s) \left[\Phi(w_v) g(s_{*1}-hz_s, t-hw_v) \right. \right.$$

$$\left. + \int_{w_v}^{\infty} \phi(z_t) g(s_{*1}-hz_s, t-hz_t) dz_t \right] dz_s$$

$$\frac{\partial \mathcal{C}}{\partial v} (\Phi(s_{*2}-hw_u), \Phi(t-hw_v)) \phi'(w_u) \phi(w_v) dw_v dw_u$$

$$- \frac{h}{\phi(s_{*1}) \phi(s_{*2})} \iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi(z_s) \left[\Phi(w_v) g(s_{*2}-hz_s, t-hw_v) \right.$$

$$+ \int_{w_v}^{\infty} \phi\left(z_t\right) g\left(s_{*2} - hz_s, t - hz_t\right) dz_t \Bigg] dz_s$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1} - hw_u\right), \Phi\left(t - hw_v\right)\right) \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u$$

$$+ h^2 \left[\mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*1}))\right] \Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*2}\right)\right) + \Phi\left(s_{*2}\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right)$$

$$+ \Phi\left(t\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \Bigg] + o(h^2)$$

$$+ h^2 \left[\mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*2}))\right] \Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) + \Phi\left(s_{*1}\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right)$$

$$+ \Phi\left(t\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \Bigg] + o(h^2)$$

$$+ \frac{1}{\phi\left(s_{*1}\right) \phi\left(s_{*2}\right)} \iiiint_{\mathbb{R}^4} F_{S,T}\left(s_{*1} - hw_{u1}, t - hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1} - hw_{u1}\right), \Phi\left(t - hw_{v1}\right)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2} - hw_{u2}\right), \Phi\left(t - hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{v1} dw_{u1} dw_{v2} dw_{u2}$$

$$+ \frac{1}{\phi\left(s_{*1}\right) \phi\left(s_{*2}\right)} \iiiint_{\mathbb{R}^4} F_{S,T}\left(s_{*2} - hw_{u1}, t - hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*2} - hw_{u1}\right), \Phi\left(t - hw_{v1}\right)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1} - hw_{u2}\right), \Phi\left(t - hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{v1} dw_{u1} dw_{v2} dw_{u2}$$

$$+ \frac{1}{\phi\left(s_{*1}\right) \phi\left(s_{*2}\right)} \iiint_{\mathbb{R}^3} \int_{w_{v1}}^{\infty} \Phi\left(t - hw_{v2}\right)$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1} - hw_{u1}\right), \Phi\left(t - hw_{v1}\right)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2} - hw_{u2}\right), \Phi\left(t - hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{v2} dw_{v1} dw_{u1} dw_{u2}$$

$$+ \frac{1}{\phi\left(s_{*1}\right) \phi\left(s_{*2}\right)} \iiint_{\mathbb{R}^3} \int_{-\infty}^{w_{v1}} \Phi\left(t - hw_{v1}\right)$$

$$\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1} - hw_{u1}\right), \Phi\left(t - hw_{v1}\right)\right) \frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2} - hw_{u2}\right), \Phi\left(t - hw_{v2}\right)\right)$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{v2} dw_{v1} dw_{u1} dw_{u2}$$

$$
\begin{aligned}
&- h^2\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) + \Phi\left(s_{*1}\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \\
&+ \Phi\left(t\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right)\Bigg] \\
&\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*2}\right)\right) + \Phi\left(s_{*2}\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \\
&+ \Phi\left(t\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right)\Bigg] + o(h^2)\Bigg\} \\
&\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*1}\right)\right) - \alpha\right] + h^2 B(\Phi\left(s_{*1}\right), \Phi(t)) + o(h^2)\right\} \phi\left(s_{*1}\right) \omega\left(\Phi\left(s_{*1}\right)\right) \\
&\left\{\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*2}\right)\right) - \alpha\right] + h^2 B(\Phi\left(s_{*2}\right), \Phi(t)) + o(h^2)\right\} \phi\left(s_{*2}\right) \omega\left(\Phi\left(s_{*2}\right)\right) \\
&ds_{*2} ds_{*1},
\end{aligned}
$$

where the last equality follows from change of variable $u = \frac{s_{*2} - s_{*1}}{h}$ to (A.1.30). Then we consider two different cases.

1. If $Y \perp_{\alpha} X$, by the fact that $\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) = \mathcal{C}_{2|1}\left(\alpha \mid \Phi\left(s\right)\right) \equiv \alpha$, applying Taylor expansions at (s_{*1}, t), and the fact in (A.1.2), we have

$$
\begin{aligned}
&t_n^{(2)} \\
=&h^6 \iint_{\mathbb{R}^2} \Bigg\{2\alpha\phi\left(s_{*1}\right) \int_{\mathbb{R}} \phi\left(z_s\right) \phi\left(z_s + u\right) \int_{\mathbb{R}} \phi\left(z_t\right) \Phi\left(z_t\right) dz_t dz_s \\
&- \alpha^2 \phi\left(s_{*1}\right) \iint_{\mathbb{R}^2} \int_{w_u - u}^{\infty} \phi\left(z_s\right) \int_{\mathbb{R}} \phi\left(z_t\right) dz_t dz_s \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \\
&- \alpha^2 \phi\left(s_{*1}\right) \iint_{\mathbb{R}^2} \int_{w_u + u}^{\infty} \phi\left(z_s\right) \int_{\mathbb{R}} \phi\left(z_t\right) dz_t dz_s \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u \\
&- \alpha^2 \phi\left(s_{*1}\right) \iiint_{\mathbb{R}^3} \int_{w_{u2} - u}^{\infty} w_{u1} \phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) \\
&dw_{u1} dw_{u2} dw_{v1} dw_{v2}
\end{aligned}
$$

$$
\begin{aligned}
&+\alpha^2\phi\left(s_{*1}\right)\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{u2}-u}(u-w_{u2})\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)\\
&dw_{u1}dw_{u2}dw_{v1}dw_{v2}\Bigg\}B^2(\Phi\left(s_{*1}\right),\Phi(t))\omega^2\left(\Phi\left(s_{*1}\right)\right)duds_{*1}\\
&+h^6\iint_{\mathbb{R}^2}\Bigg\{\alpha\iiint_{\mathbb{R}^3}\phi\left(z_s\right)dz_s w_u\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u\\
&+\alpha\iiint_{\mathbb{R}^3}\phi\left(z_s\right)dz_s w_u\phi'\left(w_u\right)\phi\left(w_v\right)dw_v dw_u\\
&+2\alpha^2\iiiint_{\mathbb{R}^4}w_{u1}w_{u2}\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}\\
&+\alpha\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}w_{u1}w_{u2}\\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}\\
&+\alpha\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{v1}}w_{u1}w_{u2}\\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}\Bigg\}\\
&B(\Phi\left(s_{*1}\right),\Phi(t))\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\\
&B(\Phi\left(s_{*2}\right),\Phi(t))\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*2}ds_{*1}+o(h^6)\\
=&h^6\left(\alpha-\alpha^2\right)\int_{\mathbb{R}}B^2(\Phi\left(s\right),\Phi(t))\phi\left(s\right)\omega^2\left(\Phi\left(s\right)\right)ds\\
&+h^6\left(2\alpha^2-\alpha\right)\left[\int_{\mathbb{R}}B(\Phi\left(s\right),\Phi(t))\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds\right]^2+o(h^6),
\end{aligned}
\tag{A.1.31}
$$

as $n\to\infty$.

2. If $Y\not\angle_\alpha X$, applying Taylor expansions at (s_{*1},t), and the fact in (A.1.2), we have

$$
t_n^{(2)}
$$

$$=h^2 \iint_{\mathbb{R}^2} \Bigg\{$$

$$2\phi\left(s_{*1}\right) \mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s_{*1})\right) \int_{\mathbb{R}} \phi\left(z_s\right) \phi\left(z_s+u\right) \int_{\mathbb{R}} \phi\left(z_t\right) \Phi\left(z_t\right) dz_t dz_s$$

$$-\phi\left(s_{*1}\right) \mathcal{C}^2_{2|1}\left(\Phi(t) \mid \Phi(s_{*1})\right)$$

$$\iint_{\mathbb{R}^2} \int_{w_u-u}^{\infty} \phi\left(z_s\right) \int_{\mathbb{R}} \phi\left(z_t\right) dz_t dz_s \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u$$

$$-\phi\left(s_{*1}\right) \mathcal{C}^2_{2|1}\left(\Phi(t) \mid \Phi(s_{*1})\right)$$

$$\iint_{\mathbb{R}^2} \int_{w_u+u}^{\infty} \phi\left(z_s\right) \int_{\mathbb{R}} \phi\left(z_t\right) dz_t dz_s \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u$$

$$+\iiint_{\mathbb{R}^3} \int_{w_{u2}-u}^{\infty} \Bigg[-w_{u1}\phi\left(s_{*1}\right) \mathcal{C}^2_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right)$$

$$+(u-w_{u1}-w_{u2})\phi\left(s_{*1}\right) \Phi\left(s_{*1}\right) \mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right)\Bigg]$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{u1} dw_{u2} dw_{v1} dw_{v2}$$

$$+\iiint_{\mathbb{R}^3} \int_{-\infty}^{w_{u2}-u} \Bigg[(u-w_{u2})\phi\left(s_{*1}\right) \mathcal{C}^2_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right)$$

$$+(u-w_{u1}-w_{u2})\phi\left(s_{*1}\right) \Phi\left(s_{*1}\right) \mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right)\Bigg]$$

$$\phi'\left(w_{u1}\right) \phi\left(w_{v1}\right) \phi'\left(w_{u2}\right) \phi\left(w_{v2}\right) dw_{u1} dw_{u2} dw_{v1} dw_{v2} \Bigg\}$$

$$\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s_{*1})\right) - \alpha\right]^2 \omega^2\left(\Phi\left(s_{*1}\right)\right) du ds_{*1}$$

$$+h^2 \iint_{\mathbb{R}^2} \Bigg\{\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi(s_{*2}), \Phi(t)\right)$$

$$\iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(z_s\right) dz_s w_u \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u$$

$$+\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*2}\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi(s_{*1}), \Phi(t)\right)$$

$$\iint_{\mathbb{R}^2} \int_{\mathbb{R}} \phi\left(z_s\right) dz_s w_u \phi'\left(w_u\right) \phi\left(w_v\right) dw_v dw_u$$

$$+\mathcal{C}^2_{2|1}\left(\Phi(t) \mid \Phi(s_{*1})\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \iiiint_{\mathbb{R}^4}$$

$$w_{u1}w_{u2}\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\mathcal{C}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\iiiint_{\mathbb{R}^4}$$

$$w_{u1}w_{u2}\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\mathcal{C}^2_{2|1}\left(\Phi(t)\mid\Phi(s_{*2})\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\iiiint_{\mathbb{R}^4}$$

$$w_{u1}w_{u2}\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\mathcal{C}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\iiiint_{\mathbb{R}^4}$$

$$w_{u1}w_{u2}\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}w_{u1}w_{u2}$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}$$

$$+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{v1}}w_{u1}w_{u2}$$

$$\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}$$

$$+\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*2}))\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)$$

$$-\left[\Phi\left(s_{*1}\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\right]$$

$$\left.\left[\Phi\left(s_{*2}\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)+\Phi\left(t\right)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\right]\right\}$$

$$\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*1}\right)\right)-\alpha\right]\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)$$

$$\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi\left(s_{*2}\right)\right)-\alpha\right]\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*2}ds_{*1}+o(h^2)$$

$$=h^2\int_{\mathbb{R}}\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)-\mathcal{C}^2_{2|1}\left(\Phi(t)\mid\Phi(s)\right)\right]$$

$$\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)-\alpha\right]^2\phi\left(s\right)\omega^2\left(\Phi\left(s\right)\right)ds$$

$$
\begin{aligned}
&- h^2 \iint_{\mathbb{R}^2} \Bigg\{ \\
&\left[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) - \mathcal{C}_{2|1}^2\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right)\right] \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi(s_{*2}), \Phi(t)\right) \\
&+ \left[\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*2}\right)\right) - \mathcal{C}_{2|1}^2\left(\Phi\left(t\right) \mid \Phi\left(s_{*2}\right)\right)\right] \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi(s_{*1}), \Phi(t)\right) \\
&- \left[\mathcal{C}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) - \alpha \Phi\left(s_{*1}\right)\right] \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \\
&- \left[\mathcal{C}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) - \alpha \Phi\left(s_{*2}\right)\right] \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \\
&- \left(\alpha - \alpha^2\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \\
&- \mathcal{C}_{2|1}(\Phi(t) \mid \Phi(s_{*2}))\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s_{*1}\right)\right) \\
&+ \Phi\left(s_{*1}\right) \Phi\left(s_{*2}\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right), \Phi\left(t\right)\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right), \Phi\left(t\right)\right) \Bigg\} \\
&\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*1}\right)\right) - \alpha\right] \phi\left(s_{*1}\right) \omega\left(\Phi\left(s_{*1}\right)\right) \\
&\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s_{*2}\right)\right) - \alpha\right] \phi\left(s_{*2}\right) \omega\left(\Phi\left(s_{*2}\right)\right) ds_{*2} ds_{*1} + o(h^2), \qquad \text{(A.1.32)}
\end{aligned}
$$

as $n \to \infty$.

Actually, by similar proofs with those of $t_n^{(1)}$ and $t_n^{(2)}$ in (A.1.26), (A.1.27), (A.1.31), and (A.1.32), we could have, $\forall k \geqslant 1$,

$$
t_n^{(k)} = \mathcal{C}_{\perp_\alpha, k} h^{3k} + o(h^{3k}), \qquad \text{if } Y \perp_\alpha X, \qquad \text{(A.1.33)}
$$

$$
t_n^{(k)} = \mathcal{C}_{\not\perp_\alpha, k} h^{k} + o(h^{k}), \qquad \text{if } Y \not\perp_\alpha X, \qquad \text{(A.1.34)}
$$

as $n \to \infty$, for some constants $\mathcal{C}_{\perp_\alpha,k}$, and $\mathcal{C}_{\not\perp_\alpha,k}$.

Correspondingly, with all above results, we have the following two conditions.

1. If $Y \perp_\alpha X$, by definition of Z_{n1i}, and results in (A.1.26), (A.1.31),and (A.1.33), change of variable $u = \Phi(s)$, and back substitution $\alpha = \Phi(t)$, we could have

$$\begin{aligned}
\mathbb{E}\left[Z_{n1i}\right] =& 0, \\
\mathbb{E}\left[Z_{n1i}^2\right] = \operatorname{Var}\left[Z_{n1i}\right] =& t_n^{(2)} - \left(t_n^{(1)}\right)^2 \\
=& h^6\left(\alpha - \alpha^2\right)\left\{\int_{\mathbb{R}} B^2(\Phi\left(s\right), \Phi(t))\phi\left(s\right)\omega^2\left(\Phi\left(s\right)\right)ds \right.\\
& \left. - \left[\int_{\mathbb{R}} B(\Phi\left(s\right), \Phi(t))\phi\left(s\right)\omega\left(\Phi\left(s\right)\right)ds\right]^2\right\} + o(h^6) \\
=& h^6\left(\alpha - \alpha^2\right)\left\{\int_0^1 B^2(u, \alpha)\omega^2\left(u\right)du \right.\\
& \left. - \left[\int_0^1 B(u, \alpha)\omega\left(u\right)du\right]^2\right\} + o(h^6), \\
\mathbb{E}\left[Z_{n1i}^4\right] =& t_n^{(4)} - 4t_n^{(3)}t_n^{(1)} + 6t_n^{(2)}\left(t_n^{(1)}\right)^2 - 3\left(t_n^{(1)}\right)^4 \\
=& \mathcal{O}(h^{12}),
\end{aligned}$$

as $n \to \infty$. Then with the definition of $B(\cdot, \cdot)$, and noticing that, when $Y \perp_\alpha X$,

$$B(u, \alpha) = \frac{1}{2}\left(\left[\Phi^{-1}(u)\right]^2 - 1\right)\alpha + \frac{1}{2}\phi'\left(\Phi^{-1}(\alpha)\right), \tag{A.1.35}$$

we could have the result of $\mathbb{E}\left[Z_{n1i}^2\right]$ and $\operatorname{Var}\left[Z_{n1i}\right]$.

2. If $Y \not\perp_\alpha X$, by definition of Z_{n1i}, and results in (A.1.27), (A.1.32), and (A.1.34), change of variable $u = \Phi(s)$, and back substitution $\alpha = \Phi(t)$, we could have

$$\mathbb{E}\left[Z_{n1i}\right] = 0,$$

$$
\begin{aligned}
\mathbb{E}\left[Z_{n1i}^2\right] = \mathrm{Var}\left[Z_{n1i}\right] =& t_n^{(2)} - \left(t_n^{(1)}\right)^2 \\
=& h^2 \int_{\mathbb{R}} \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s)\right) - \mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi(s)\right)\right] \\
& \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi(s)\right) - \alpha\right]^2 \phi\left(s\right) \omega^2\left(\Phi\left(s\right)\right) ds \\
& + h^2\left(\alpha - \alpha^2\right)\left\{\int_{\mathbb{R}} \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right. \\
& \left.\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \alpha\right] \phi\left(s\right) \omega\left(\Phi\left(s\right)\right) ds\right\}^2 \\
& - h^2\left\{\int_{\mathbb{R}} \Phi\left(s\right) \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right. \\
& \left.\left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \alpha\right] \phi\left(s\right) \omega\left(\Phi\left(s\right)\right) ds\right\}^2 \\
& - 2h^2 \int_{\mathbb{R}} \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi(s), \Phi(t)\right) \\
& \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \alpha\right] \phi\left(s\right) \omega\left(\Phi\left(s\right)\right) ds \\
& \int_{\mathbb{R}} \left\{\mathcal{C}_{2|1}\left(\Phi\left(t\right) \mid \Phi\left(s\right)\right) - \mathcal{C}_{2|1}^2\left(\Phi\left(t\right) \mid \Phi\left(s\right)\right)\right. \\
& \left. - \left[\mathcal{C}\left(\Phi\left(s\right), \Phi\left(t\right)\right) - \alpha \Phi\left(s\right)\right] \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s\right), \Phi\left(t\right)\right)\right\} \\
& \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \alpha\right] \phi\left(s\right) \omega\left(\Phi\left(s\right)\right) ds + o(h^2) \\
=& h^2 \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \mathcal{C}_{2|1}^2\left(\alpha \mid u\right)\right] \\
& \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right]^2 \omega^2\left(u\right) du \\
& + h^2\left(\alpha - \alpha^2\right)\left\{\int_0^1 \frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(u, \alpha\right)\right. \\
& \left.\left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right] \omega\left(u\right) du\right\}^2 \\
& - h^2\left\{\int_0^1 u \frac{\partial^2 \mathcal{C}}{\partial u^2}\left(u, \alpha\right) \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right] \omega\left(u\right) du\right\}^2
\end{aligned}
$$

$$
\begin{aligned}
&- 2h^2 \int_0^1 \frac{\partial^2 \mathcal{C}}{\partial u \partial v}(u,\alpha)\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du \\
&\int_0^1 \Big\{ \mathcal{C}_{2|1}(\alpha \mid u) - \mathcal{C}_{2|1}^2(\alpha \mid u) \\
&- \left[\mathcal{C}(u,\alpha) - \alpha u\right] \frac{\partial^2 \mathcal{C}}{\partial u^2}(u,\alpha) \Big\} \\
&\left[\mathcal{C}_{2|1}(\alpha \mid u) - \alpha\right]\omega(u)\,du + o(h^2)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[Z_{n1i}^4\right] =& t_n^{(4)} - 4t_n^{(3)}t_n^{(1)} + 6t_n^{(2)}\left(t_n^{(1)}\right)^2 - 3\left(t_n^{(1)}\right)^4 \\
=& \mathcal{O}(h^4),
\end{aligned}
$$

as $n \to \infty$.

Then, the rest proof of asymptotic property of I_{n1} in both conditions would be similar with the proof of Lemma 1 in Hall, 1984 by using Lindeberg's condition and noticing

$$
s_n^{-2} \sum_{i=1}^{n} \mathbb{E}\left[Z_{n1i}^2 \mathbb{1}_{\{|Z_{n1i}| > \varepsilon s_n\}}\right] \leqslant \varepsilon^{-2} s_n^{-4} \sum_{i=1}^{n} \mathbb{E}\left[Z_{n1i}^4\right] \to 0,
$$

as $n \to \infty$, where $s_n^2 = \sum_{i=1}^n \mathbb{E}\left[Z_{n1i}^2\right]$. □

Lemma A.1.7. *Suppose Condition 3.4 holds. Let $h \to 0$ as $n \to \infty$. For I_{n2} defined in (3.9.14), consider $I_{n2} \doteq \frac{1}{n^2 h^2} \sum_{i=1}^n Z_{n2i}$. For any quantile level $\alpha \in (0,1)$, we have the following results hold as $n \to \infty$,*

$$
\begin{aligned}
\mathbb{E}\left[Z_{n2i}\right] =& hM_2(\omega) + o(h), \\
\mathbb{E}\left[Z_{n2i}^2\right] =& \mathcal{O}(h^2),
\end{aligned}
$$

where

$$M_2(\omega) = \frac{1}{2\sqrt{\pi}} \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \mathcal{C}_{2|1}^2\left(\alpha \mid u\right)\right] \frac{\omega\left(u\right)}{\phi\left(\Phi^{-1}(u)\right)} du.$$

Correspondingly, we have $I_{n2} = n^{-1}h^{-1}M_2(\omega) + o(n^{-1}h^{-1}) + \mathcal{O}_P(n^{-\frac{3}{2}}h^{-1})$*.*

Further when Y *is* ***quantile independent*** *with* X *at the quantile level* α*, we have* $M_{\perp_\alpha,2}(\omega) \doteq M_2(\omega)$ *as defined in Section 3.8.*

Proof. With $t = \Phi^{-1}(\alpha)$, and by definition, we have

$$\mathbb{E}\left[Z_{n2i}\right] = \int_{\mathbb{R}} \mathbb{E}\left\{\mathcal{B}_n\left[s,t;(s_i,t_i)\right] - \mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right\}^2 \phi\left(s\right)\omega\left(\Phi(s)\right) ds,$$

where the result of $\mathbb{E}\left\{\mathcal{B}_n\left[s,t;(s_i,t_i)\right] - \mathbb{E}\left[\mathcal{B}_n\left[s,t;(s_i,t_i)\right]\right]\right\}^2$ is given in (A.1.24). Hence, we have,

$$\mathbb{E}\left[Z_{n2i}\right] = h\frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} \left[\mathcal{C}_{2|1}\left(\Phi(t) \mid \Phi\left(s\right)\right) - \mathcal{C}_{2|1}^2\left(\Phi(t) \mid \Phi\left(s\right)\right)\right] \omega\left(\Phi(s)\right) ds + o(h),$$

as $n \to \infty$. Then, by change of variable $u = \Phi(s)$, and back substitution $\alpha = \Phi(t)$, we have

$$\mathbb{E}\left[Z_{n2i}\right] = h\frac{1}{2\sqrt{\pi}} \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \mathcal{C}_{2|1}^2\left(\alpha \mid u\right)\right] \frac{\omega\left(u\right)}{\phi\left(\Phi^{-1}(u)\right)} du + o(h),$$

as $n \to \infty$. Further if $Y \perp_\alpha X$, by the truth $\mathcal{C}_{2|1}\left(\alpha \mid u\right) \equiv \alpha$, we have the corresponding result. Then, using similar method in the proof of Lemma A.1.6, we could also find that $\mathbb{E}\left[Z_{n2i}^2\right] = \mathcal{O}(h^2)$, as $n \to \infty$.

The rest of the proof follows directly from the above results. □

Lemma A.1.8. *Suppose Condition 3.4 holds. Let $h \to 0$, and $nh \to \infty$ as $n \to \infty$. For I_{n3} defined in* (3.9.14)*, consider $I_{n3} \doteq \frac{1}{n^2h^2} U_n$ with*

$$
\begin{aligned}
U_n &= \sum_{1 \leqslant k < l \leqslant n} H_n \left[(s_k, t_k), (s_l, t_l)\right], \\
H_n \left[(s_k, t_k), (s_l, t_l)\right] &= \int_{\mathbb{R}} V_n(s_k, t_k; s, t) V_n(s_l, t_l; s, t) \phi(s) \omega(\Phi(s))\, ds, \\
V_n(s^*, t^*; s, t) &= \mathcal{B}_n \left[s, t; (s^*, t^*)\right] - \mathbb{E}\left[\mathcal{B}_n \left[s, t; (s_1, t_1)\right]\right],
\end{aligned}
\tag{A.1.36}
$$

where $\mathcal{B}_n[\cdot, \cdot; (\cdot, \cdot)]$ is defined in (3.9.13)*. For any quantile level $\alpha \in (0,1)$, we have the following result holds as $n \to \infty$,*

$$
\mathbb{E}\left\{H_n^2 \left[(s_1, t_1), (s_2, t_2)\right]\right\} = h^3 \sigma_3^2(\omega) + o(h^3),
$$

where

$$
\sigma_3^2(\omega) = \frac{1}{2\sqrt{2\pi}} \int_0^1 \left[\mathcal{C}_{2|1}(\alpha \mid u) - \mathcal{C}_{2|1}^2(\alpha \mid u)\right]^2 \frac{\omega^2(u)}{\phi(\Phi^{-1}(u))} du.
$$

Correspondingly, we have $I_{n3} \xrightarrow{d} N\left(0, \frac{1}{2} n^{-2} h^{-1} \sigma_3^2(\omega)\right)$.

Further when Y is ***quantile independent*** *with X at the quantile level α, we have $\sigma_{\perp_\alpha,3}^2(\omega) \doteq \sigma_3^2(\omega)$ as defined in Section 3.8.*

Proof. With $t = \Phi^{-1}(\alpha)$, and by definition, we have

$$
\begin{aligned}
& H_n^2 \left[(s_1, t_1), (s_2, t_2)\right] \\
= & \iint_{\mathbb{R}^2} \left\{\mathcal{B}_n \left[s_{*1}, t; (s_1, t_1)\right] - \mathbb{E}\left[\mathcal{B}_n \left[s_{*1}, t; (s_1, t_1)\right]\right]\right\}
\end{aligned}
$$

$$\{\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right]\}$$

$$\{\mathcal{B}_n\left[s_{*1},t;(s_2,t_2)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\right]\}$$

$$\{\mathcal{B}_n\left[s_{*2},t;(s_2,t_2)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right]\}$$

$$\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2}.$$

By exchanging the integral and expectation, we have

$$\begin{aligned}
&\mathbb{E}\left\{H_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\} \\
=&\iint_{\mathbb{R}^2}\{\mathbb{E}[(\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\right]) \\
&(\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]-\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right])]\}^2 \\
&\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
=&\iint_{\mathbb{R}^2}\{\mathbb{E}\left(\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right) \\
&-\mathbb{E}\left[\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\right]\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right]\}^2 \\
&\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
=&\iint_{\mathbb{R}^2}\{\mathbb{E}\left(\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right)\}^2 \\
&\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
&-2\iint_{\mathbb{R}^2}\mathbb{E}\left(\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right) \\
&\mathbb{E}\left[\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\right]\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right] \\
&\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
&+\iint_{\mathbb{R}^2}\{\mathbb{E}\left[\mathcal{B}_n\left[s_{*1},t;(s_1,t_1)\right]\right]\mathbb{E}\left[\mathcal{B}_n\left[s_{*2},t;(s_1,t_1)\right]\right]\}^2
\end{aligned}$$

$$
\begin{aligned}
&\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
\doteq &\iint_{\mathbb{R}^2} I_{n31}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
&-2\iint_{\mathbb{R}^2} I_{n32}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
&+\iint_{\mathbb{R}^2} I_{n33}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2}. \qquad \text{(A.1.37)}
\end{aligned}
$$

We consider the three terms separately. For $I_{n31}(s_{*1},s_{*2},t)$, by (A.1.29), we could find

$$
I_{n31}(s_{*1},s_{*2},t)=\left[\sum_{k=1}^{12} I_{n1k}(s_{*1},s_{*2},t)\right]^2.
$$

By the proof of Lemma A.1.6, we could have

$$
\begin{aligned}
&\iint_{\mathbb{R}^2} I_{n31}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2} \\
=&\iint_{\mathbb{R}^2}\left\{\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\left[2h\int_{\mathbb{R}}\phi\left(z_s\right)\phi\left(z_s+\frac{s_{*2}-s_{*1}}{h}\right)\right.\right. \\
&\int_{\mathbb{R}}\phi\left(z_t\right)\Phi\left(z_t\right)g(s_{*1}-hz_s,t-hz_t)dz_tdz_s \\
&-h\iint_{\mathbb{R}^2}\int_{\frac{s_{*1}-s_{*2}}{h}+w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*1}-hz_s,t-hz_t\right) \\
&dz_tdz_s\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u \\
&-h\iint_{\mathbb{R}^2}\int_{\frac{s_{*2}-s_{*1}}{h}+w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*2}-hz_s,t-hz_t\right) \\
&dz_tdz_s\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u \\
&+\iiint_{\mathbb{R}^3}\int_{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}^{\infty}\Phi\left(s_{*1}-hw_{u1}\right) \\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
\end{aligned}
$$

$$\phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{u1}dw_{u2}dw_{v1}dw_{v2}$$

$$+\iiint_{\mathbb{R}^3}\int_{-\infty}^{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}\Phi(s_{*2}-hw_{u2})$$

$$\frac{\partial\mathcal{C}}{\partial u}\left(\Phi(s_{*1}-hw_{u1}),\Phi(t-hw_{v1})\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi(s_{*2}-hw_{u2}),\Phi(t-hw_{v2})\right)$$

$$\left.\phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{u1}dw_{u2}dw_{v1}dw_{v2}\right]$$

$$+\Bigg[-\frac{h}{\phi(s_{*1})\,\phi(s_{*2})}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi(z_s)\Big[\Phi(w_v)\,g(s_{*1}-hz_s,t-hw_v)$$

$$+\int_{w_v}^{\infty}\phi(z_t)\,g(s_{*1}-hz_s,t-hz_t)\,dz_t\Big]dz_s$$

$$\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s_{*2}-hw_u),\Phi(t-hw_v)\right)\phi'(w_u)\,\phi(w_v)\,dw_v dw_u$$

$$-\frac{h}{\phi(s_{*1})\,\phi(s_{*2})}\iint_{\mathbb{R}^2}\int_{\mathbb{R}}\phi(z_s)\Big[\Phi(w_v)\,g(s_{*2}-hz_s,t-hw_v)$$

$$+\int_{w_v}^{\infty}\phi(z_t)\,g(s_{*2}-hz_s,t-hz_t)\,dz_t\Big]dz_s$$

$$\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s_{*1}-hw_u),\Phi(t-hw_v)\right)\phi'(w_u)\,\phi(w_v)\,dw_v dw_u$$

$$+h^2\left[\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*1}))\right]\Bigg[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s_{*2})\right)+\Phi(s_{*2})\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi(s_{*2}),\Phi(t)\right)$$

$$+\Phi(t)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi(s_{*2}),\Phi(t)\right)\Bigg]+o(h^2)$$

$$+h^2\left[\mathcal{C}_{2|1}(\Phi(t)\mid\Phi(s_{*2}))\right]\Bigg[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s_{*1})\right)+\Phi(s_{*1})\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi(s_{*1}),\Phi(t)\right)$$

$$+\Phi(t)\frac{\partial^2\mathcal{C}}{\partial u\partial v}\left(\Phi(s_{*1}),\Phi(t)\right)\Bigg]+o(h^2)$$

$$+\frac{1}{\phi(s_{*1})\,\phi(s_{*2})}\iiiint_{\mathbb{R}^4}F_{S,T}(s_{*1}-hw_{u1},t-hw_{v2})$$

$$\frac{\partial\mathcal{C}}{\partial u}\left(\Phi(s_{*1}-hw_{u1}),\Phi(t-hw_{v1})\right)\frac{\partial\mathcal{C}}{\partial v}\left(\Phi(s_{*2}-hw_{u2}),\Phi(t-hw_{v2})\right)$$

$$\phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{v1}dw_{u1}dw_{v2}dw_{u2}$$

$$+\frac{1}{\phi(s_{*1})\,\phi(s_{*2})}\iiiint_{\mathbb{R}^4}F_{S,T}(s_{*2}-hw_{u1},t-hw_{v2})$$

$$
\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
$$

$$
\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v1}dw_{u1}dw_{v2}dw_{u2}
$$

$$
+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{w_{v1}}^{\infty}\Phi\left(t-hw_{v2}\right)
$$

$$
\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
$$

$$
\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}
$$

$$
+\frac{1}{\phi\left(s_{*1}\right)\phi\left(s_{*2}\right)}\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{v1}}\Phi\left(t-hw_{v1}\right)
$$

$$
\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial v}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
$$

$$
\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{v2}dw_{v1}dw_{u1}dw_{u2}
$$

$$
-h^2\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)+\Phi\left(s_{*1}\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)
$$

$$
+\Phi\left(t\right)\frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\Bigg]
$$

$$
\Bigg[\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*2}\right)\right)+\Phi\left(s_{*2}\right)\frac{\partial^2 \mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)
$$

$$
+\Phi\left(t\right)\frac{\partial^2 \mathcal{C}}{\partial u \partial v}\left(\Phi\left(s_{*2}\right),\Phi\left(t\right)\right)\Bigg]+o(h^2)\Bigg]\Bigg\}^2
$$

$$
\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*2}ds_{*1},
$$

where the last nine terms in the square are neglectable, since the weighted integrals of their squares are of order $\mathcal{O}(h^4)$, and the weighted integrals of the interaction terms between them and the other terms are also of order $\mathcal{O}(h^4)$. Then, by change of variable $u=\frac{s_{*2}-s_{*1}}{h}$, we have

$$
\iint_{\mathbb{R}^2} I_{n31}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2}
$$

$$
\begin{aligned}
=&\iint_{\mathbb{R}^2}\left[2h\int_{\mathbb{R}}\phi\left(z_s\right)\phi\left(z_s+\frac{s_{*2}-s_{*1}}{h}\right)\right.\\
&\int_{\mathbb{R}}\phi\left(z_t\right)\Phi\left(z_t\right)g(s_{*1}-hz_s,t-hz_t)dz_tdz_s\\
&-h\iint_{\mathbb{R}^2}\int_{\frac{s_{*1}-s_{*2}}{h}+w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*1}-hz_s,t-hz_t\right)\\
&dz_tdz_s\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u\\
&-h\iint_{\mathbb{R}^2}\int_{\frac{s_{*2}-s_{*1}}{h}+w_u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*2}-hz_s,t-hz_t\right)\\
&dz_tdz_s\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u\\
&+\iiint_{\mathbb{R}^3}\int_{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}^{\infty}\Phi\left(s_{*1}-hw_{u1}\right)\\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)\\
&\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}\\
&+\iiint_{\mathbb{R}^3}\int_{-\infty}^{\frac{s_{*1}-s_{*2}}{h}+w_{u2}}\Phi\left(s_{*2}-hw_{u2}\right)\\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*2}-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)\\
&\left.\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}\right]^2\\
&\frac{\omega\left(\Phi\left(s_{*1}\right)\right)}{\phi\left(s_{*1}\right)}\frac{\omega\left(\Phi\left(s_{*2}\right)\right)}{\phi\left(s_{*2}\right)}ds_{*2}ds_{*1}+\mathcal{O}(h^4)\\
=&h\iint_{\mathbb{R}^2}\left[2h\int_{\mathbb{R}}\phi\left(z_s\right)\phi\left(z_s+u\right)\int_{\mathbb{R}}\phi\left(z_t\right)\Phi\left(z_t\right)g(s_{*1}-hz_s,t-hz_t)dz_tdz_s\right.\\
&-h\iint_{\mathbb{R}^2}\int_{w_u-u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*1}-hz_s,t-hz_t\right)dz_tdz_s\\
&\frac{\partial\mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}+hu-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u\\
&-h\iint_{\mathbb{R}^2}\int_{w_u+u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)g\left(s_{*1}+hu-hz_s,t-hz_t\right)dz_tdz_s
\end{aligned}
$$

$$
\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_u\right),\Phi\left(t-hw_v\right)\right)\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u
$$

$$
+\iiint_{\mathbb{R}^3}\int_{w_{u2}-u}^{\infty}\Phi\left(s_{*1}-hw_{u1}\right)
$$

$$
\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}+hu-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
$$

$$
\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}
$$

$$
+\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{u2}-u}\Phi\left(s_{*1}+hu-hw_{u2}\right)
$$

$$
\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}-hw_{u1}\right),\Phi\left(t-hw_{v1}\right)\right)\frac{\partial \mathcal{C}}{\partial u}\left(\Phi\left(s_{*1}+hu-hw_{u2}\right),\Phi\left(t-hw_{v2}\right)\right)
$$

$$
\left.\phi'\left(w_{u1}\right)\phi\left(w_{v1}\right)\phi'\left(w_{u2}\right)\phi\left(w_{v2}\right)dw_{u1}dw_{u2}dw_{v1}dw_{v2}\right]^2
$$

$$
\frac{\omega\left(\Phi\left(s_{*1}\right)\right)}{\phi\left(s_{*1}\right)}\frac{\omega\left(\Phi\left(s_{*1}+hu\right)\right)}{\phi\left(s_{*1}+hu\right)}duds_{*1}+\mathcal{O}(h^4).
$$

Then we apply Taylor expansion at (s_{*1}, t) to every term in the square, and by the fact in (A.1.2), we have

$$
\iint_{\mathbb{R}^2} I_{n31}(s_{*1},s_{*2},t)\phi\left(s_{*1}\right)\omega\left(\Phi\left(s_{*1}\right)\right)\phi\left(s_{*2}\right)\omega\left(\Phi\left(s_{*2}\right)\right)ds_{*1}ds_{*2}
$$

$$
=h^3\iint_{\mathbb{R}^2}\left[2\phi\left(s_{*1}\right)\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s_{*1})\right)\int_{\mathbb{R}}\phi\left(z_s\right)\phi\left(z_s+u\right)\int_{\mathbb{R}}\phi\left(z_t\right)\Phi\left(z_t\right)dz_tdz_s\right.
$$

$$
-\phi\left(s_{*1}\right)\mathcal{C}_{2|1}^2\left(\Phi(t)\mid\Phi(s_{*1})\right)\iint_{\mathbb{R}^2}\int_{w_u-u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)dz_tdz_s
$$

$$
\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u
$$

$$
-\phi\left(s_{*1}\right)\mathcal{C}_{2|1}^2\left(\Phi(t)\mid\Phi(s_{*1})\right)\iint_{\mathbb{R}^2}\int_{w_u+u}^{\infty}\phi\left(z_s\right)\int_{\mathbb{R}}\phi\left(z_t\right)dz_tdz_s
$$

$$
\phi'\left(w_u\right)\phi\left(w_v\right)dw_vdw_u
$$

$$
+\iiint_{\mathbb{R}^3}\int_{w_{u2}-u}^{\infty}\Big[-w_{u1}\phi\left(s_{*1}\right)\mathcal{C}_{2|1}^2\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)
$$

$$
\left.+(u-w_{u1}-w_{u2})\phi\left(s_{*1}\right)\Phi\left(s_{*1}\right)\mathcal{C}_{2|1}\left(\Phi\left(t\right)\mid\Phi\left(s_{*1}\right)\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi\left(s_{*1}\right),\Phi\left(t\right)\right)\right]
$$

$$\phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{u1}dw_{u2}dw_{v1}dw_{v2}$$

$$+\iiint_{\mathbb{R}^3}\int_{-\infty}^{w_{u2}-u}\Bigg[(u-w_{u2})\phi(s_{*1})\,\mathcal{C}^2_{2|1}\left(\Phi(t)\mid\Phi(s_{*1})\right)$$

$$+(u-w_{u1}-w_{u2})\phi(s_{*1})\,\Phi(s_{*1})\,\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s_{*1})\right)\frac{\partial^2\mathcal{C}}{\partial u^2}\left(\Phi(s_{*1}),\Phi(t)\right)\Bigg]$$

$$\phi'(w_{u1})\,\phi(w_{v1})\,\phi'(w_{u2})\,\phi(w_{v2})\,dw_{u1}dw_{u2}dw_{v1}dw_{v2}\Bigg]^2$$

$$\frac{\omega^2\left(\Phi(s_{*1})\right)}{\phi^2(s_{*1})}du\,ds_{*1}+o(h^3)$$

$$=h^3\iint_{\mathbb{R}^2}\left[\phi(s)\,\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)\frac{\phi\left(\frac{u}{\sqrt{2}}\right)}{\sqrt{2}}\right.$$

$$\left.-\phi(s)\,\mathcal{C}^2_{2|1}\left(\Phi(t)\mid\Phi(s)\right)\frac{\phi\left(\frac{u}{\sqrt{2}}\right)}{\sqrt{2}}\right]^2\frac{\omega^2\left(\Phi(s)\right)}{\phi^2(s)}du\,ds+o(h^3)$$

$$=h^3\frac{1}{2\sqrt{2\pi}}\int_{\mathbb{R}}\left[\mathcal{C}_{2|1}\left(\Phi(t)\mid\Phi(s)\right)-\mathcal{C}^2_{2|1}\left(\Phi(t)\mid\Phi(s)\right)\right]^2\omega^2\left(\Phi(s)\right)ds+o(h^3), \tag{A.1.38}$$

as $n\to\infty$. By similar steps of change of variables and Taylor expansions above, and using the results in Lemma A.1.6, we could find

$$\iint_{\mathbb{R}^2}I_{n32}(s_{*1},s_{*2},t)\phi(s_{*1})\,\omega\left(\Phi(s_{*1})\right)\phi(s_{*2})\,\omega\left(\Phi(s_{*2})\right)ds_{*1}ds_{*2}=\mathcal{O}(h^4), \tag{A.1.39}$$

$$\iint_{\mathbb{R}^2}I_{n33}(s_{*1},s_{*2},t)\phi(s_{*1})\,\omega\left(\Phi(s_{*1})\right)\phi(s_{*2})\,\omega\left(\Phi(s_{*2})\right)ds_{*1}ds_{*2}=\mathcal{O}(h^4), \tag{A.1.40}$$

as $n\to\infty$.

Substituting (A.1.38), (A.1.39), and (A.1.40) into (A.1.37), and applying change of variable $u=\Phi(s)$, and back substitution $\alpha=\Phi(t)$, we have

$$\mathbb{E}\left\{H_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\}$$

$$=h^3 \frac{1}{2\sqrt{2\pi}} \int_0^1 \left[\mathcal{C}_{2|1}(\alpha \mid u) - \mathcal{C}_{2|1}^2(\alpha \mid u)\right]^2 \frac{\omega^2(u)}{\phi\left(\Phi^{-1}(u)\right)} du + o(h^3),$$

as $n \to \infty$. Further if $Y \perp_\alpha X$, by the truth $\mathcal{C}_{2|1}(\alpha \mid u) \equiv \alpha$, we have the corresponding result.

Then we prove the normality of U_n. We could have the following properties.

1. Function $H_n\left[(s_{*1}, t_{*1}), (s_{*2}, t_{*2})\right]$ is symmetric, i.e.

$$\begin{aligned} & H_n\left[(s_{*1}, t_{*1}), (s_{*2}, t_{*2})\right] \\ = & \int_{\mathbb{R}} V_n(s_{*1}, t_{*1}; s, t) V_n(s_{*2}, t_{*2}; s, t) \phi(s)\, \omega(\Phi(s))\, ds \\ = & H_n\left[(s_{*2}, t_{*2}), (s_{*1}, t_{*1})\right]. \end{aligned}$$

2. $\mathbb{E}\left\{ H_n\left[(s_1, t_1), (s_2, t_2)\right] \middle| (s_1, t_1)\right\} = 0$ almost surely. This could be proved by

$$\begin{aligned} & \mathbb{E}\left\{ H_n\left[(s_1, t_1), (s_2, t_2)\right] \middle| (s_1, t_1)\right\} \\ = & \mathbb{E}\left[\int_{\mathbb{R}} V_n(s_1, t_1; s, t) V_n(s_2, t_2; s, t) \phi(s)\, \omega(\Phi(s))\, ds \middle| (s_1, t_1)\right] \\ = & \int_{\mathbb{R}} V_n(s_1, t_1; s, t) \mathbb{E}\left[V_n(s_2, t_2; s, t) \middle| (s_1, t_1)\right] \phi(s)\, \omega(\Phi(s))\, ds \\ = & \int_{\mathbb{R}} V_n(s_1, t_1; s, t) \mathbb{E}\left[V_n(s_2, t_2; s, t)\right] \phi(s)\, \omega(\Phi(s))\, ds, \end{aligned}$$

and noticing $\mathbb{E}\left[V_n(s_2, t_2; s, t)\right] = 0$, where $V_n(\cdot, \cdot; \cdot, \cdot)$ is defined in (A.1.36).

3. We define

$$G_n\left[(s_{*1}, t_{*1}), (s_{*2}, t_{*2})\right]$$

$$\doteq \mathbb{E}\left\{H_n\left[(s_1,t_1),(s_{*1},t_{*1})\right]H_n\left[(s_1,t_1),(s_{*2},t_{*2})\right]\right\}.$$

By similar proof with that of $\mathbb{E}\left\{H_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\}$, and mimicking the proof of Lemma 3 in Hall, 1984, we could have,

$$\mathbb{E}\left\{H_n^4\left[(s_1,t_1),(s_2,t_2)\right]\right\} =\mathcal{O}(h^5),$$
$$\mathbb{E}\left\{G_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\} =\mathcal{O}(h^7),$$

as $n \to \infty$.

Finally, we check the condition of the Theorem 1 in Hall, 1984. When $nh \to \infty$ as $n \to \infty$, we have

$$\begin{aligned}&\frac{\mathbb{E}\left\{G_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\}+n^{-1}\mathbb{E}\left\{H_n^4\left[(s_1,t_1),(s_2,t_2)\right]\right\}}{\left[\mathbb{E}\left\{H_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\}\right]^2}\\ =&\frac{\mathcal{O}(h^7)+n^{-1}\mathcal{O}(h^5)}{\left[h^3\sigma_3^2(\omega)+o(h^3)\right]^2}\\ =&\mathcal{O}(h)+\mathcal{O}(n^{-1}h^{-1})\to 0,\end{aligned}$$

as $n \to \infty$. Hence, by applying Theorem 1 in Hall, 1984, we have

$$U_n \xrightarrow{d} N\left(0,\frac{1}{2}n^2\mathbb{E}\left\{H_n^2\left[(s_1,t_1),(s_2,t_2)\right]\right\}\right),$$

as $n \to \infty$. And the normality of I_{n3} follows directly from the above results. □

Lemma A.1.9. *Suppose Condition 3.4 holds. Let $h \to 0$ as $n \to \infty$. Consider I_{n4} defined in* (3.9.14). *For any quantile level $\alpha \in (0,1)$, we have the following results hold as $n \to \infty$.*

(a) *When Y is **quantile independent** with X at the quantile level α, we have*

$$I_{n4} = h^4 M_{\perp_\alpha,4}(\omega) + o(h^4),$$

where $M_{\perp_\alpha,4}(\omega)$ is defined in Section 3.8.

(b) *When Y is **quantile dependent** with X at the quantile level α, we have*

$$I_{n4} = \int_0^1 \left[\mathcal{C}_{2|1}\left(\alpha \mid u\right) - \alpha\right]^2 \omega(u)du + h^2 M_{\not\perp_\alpha,4}(\omega) + o(h^2),$$

where $M_{\not\perp_\alpha,4}(\omega)$ is defined in Section 3.8.

Proof. By definition, we have

$$I_{n4} = \int_0^1 \left\{\mathbb{E}\left[A_n\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)\right] - \alpha\right\}^2 \omega(u)du,$$

where $\mathbb{E}\left[A_n\left(\cdot, \cdot\right)\right]$ is given by Lemma A.1.3. Then the proof of this lemma follows directly from the truth $\mathcal{C}_{2|1}\left(\alpha \mid u\right) \equiv \alpha$, and result in (A.1.35), when $Y \perp_\alpha X$. □

Lemma A.1.10. *Suppose Condition 3.4 holds. Let $h \to 0$ as $n \to \infty$. Consider $R_n(\cdot, \cdot)$ defined in* (3.9.9). *We have the following results hold as $n \to \infty$.*

$$\int_0^1 R_n^2\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)\omega(u)du = \mathcal{O}_{a.s.}\left(n^{-1/2}h^{-1}\log n(\log\log n)^{1/2}\right).$$

Further when $h = \mathcal{O}(n^{-\beta})$, $\beta \in (0, \frac{1}{2})$, *we have*

$$\int_0^1 R_n^2\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)\omega(u)du = o_{a.s.}(1).$$

Proof. By definition of $R_n(\cdot,\cdot)$, change of variables $s = \Phi^{-1}(u)$, $s^* = \frac{s-\Phi^{-1}(u^*)}{h}$ and $t^* = \frac{t-\Phi^{-1}(v^*)}{h}$, and substitution $t = \Phi^{-1}(\alpha)$, we have

$$\begin{aligned}
&\int_0^1 R_n^2\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)\omega(u)du \\
=&\int_{\mathbb{R}} R_n^2\left(s,t\right)\omega\left(\Phi(s)\right)\phi(s)ds \\
=&\frac{1}{h}\int_{\mathbb{R}}\left\{\iint_{\mathcal{I}^2}\left[\mathbb{C}_n(u^*,v^*) - \mathbb{G}_n(u^*,v^*)\right]\phi'\left(\frac{s-\Phi^{-1}(u^*)}{h}\right)\phi\left(\frac{t-\Phi^{-1}(v^*)}{h}\right)\right. \\
&\left.\frac{du^*}{h\phi(\Phi^{-1}(u^*))}\frac{dv^*}{h\phi(\Phi^{-1}(v^*))}\right\}^2\frac{\omega\left(\Phi(s)\right)}{\phi(s)}ds \\
=&\frac{1}{h}\int_{\mathbb{R}}\left\{\iint_{\mathbb{R}^2}\left[\mathbb{C}_n(\Phi(s-hs^*),\Phi(t-ht^*)) - \mathbb{G}_n(\Phi(s-hs^*),\Phi(t-ht^*))\right]\right. \\
&\left.\phi'\left(s^*\right)\phi\left(t^*\right)ds^*dt^*\right\}^2\frac{\omega\left(\Phi(s)\right)}{\phi(s)}ds.
\end{aligned}$$

By Cauchy-Schwarz Inequality, we have

$$\begin{aligned}
&\int_0^1 R_n^2\left(\Phi^{-1}(u), \Phi^{-1}(\alpha)\right)\omega(u)du \\
\leqslant&\frac{1}{h}\int_{\mathbb{R}}\iint_{\mathbb{R}^2}\left[\mathbb{C}_n(\Phi(s-hs^*),\Phi(t-ht^*)) - \mathbb{G}_n(\Phi(s-hs^*),\Phi(t-ht^*))\right]^2 \\
&\left(s^*\right)^2\phi\left(s^*\right)\phi\left(t^*\right)ds^*dt^*\frac{\omega\left(\Phi(s)\right)}{\phi(s)}ds \\
\leqslant&\frac{1}{h_n}\int_{\mathbb{R}}\left[\sup_{(s^*,t^*)\in\mathbb{R}^2}\left|\mathbb{C}_n(\Phi(s-hs^*),\Phi(t-ht^*)) - \mathbb{G}_n(\Phi(s-hs^*),\Phi(t-ht^*))\right|\right]^2
\end{aligned}$$

$$\iint_{\mathbb{R}^2} (s^*)^2 \phi(s^*) \phi(t^*) ds^* dt^* \frac{\omega(\Phi(s))}{\phi(s)} ds$$

$$= \frac{1}{h} \left[\sup_{(u',v') \in \mathcal{I}^2} |\mathbb{C}_n(u',v') - \mathbb{G}_n(u',v')| \right]^2$$

$$\iiint_{\mathbb{R}^3} (s^*)^2 \phi(s^*) \phi(t^*) ds^* dt^* \frac{\omega(\Phi(s))}{\phi(s)} ds$$

$$= \frac{1}{h} \left[\sup_{(u',v') \in \mathcal{I}^2} |\mathbb{C}_n(u',v') - \mathbb{G}_n(u',v')| \right]^2 \int_0^1 \frac{\omega(u)}{\phi^2(\Phi^{-1}(u))} du.$$

The rest of the proof follows directly from (3.9.8) and Condition 3.4. □

A.2 Simulations for Bandwidth Selection in Chapter 3

We would showcase the bandwidth selection we proposed in Corollary 3.3.1 minimized the MISE(α, h, ω) asymptotically. Throughout this section, we use the weight function $\omega(u) = \phi^2(\Phi^{-1}(u))$. With 500 replicates, and $n = 100, 2000$, we consider the following three experiments:

1. $(X, Y) \sim \mathrm{MVT}_2(\mathbf{0}_2, \mathbf{I}_{2\times 2}, 5)$, which means (X, Y) follows a 2-dimensionl multivariate t distribution with mean $(0, 0)^\intercal$ and variance identity matrix. And we consider the quantile levels $\alpha = 0.5, 0.75$, which corresponds to subexperiments 1.a and 1.b.

2. $Y = X \cdot \varepsilon$, $X \sim \mathrm{Pareto}(5, 1)$. This means X follows Pareto distribution with shape parameter $k = 5$, and scale parameter $\alpha = 1$, and ε is from $N(0, 1)$ and independent with X. We consider quantile level $\alpha = 0.5$.

3. $X \sim \mathrm{Lognormal}(0, 1)$, $Y \sim \mathrm{Weibull}(0.5, 1)$, and $\mathcal{C} = \mathcal{C}_2^{\mathrm{Gu}}$. This means X follows Log-normal distribution with logarithm of location parameter $\mu = 0$, and logarithm of scale parameter $\sigma = 1$;

Y follows Weibull distribution with shape parameter $k = 0.5$, and scale parameter $\lambda = 1$; the copula of (X, Y) is a Gumbel copula with $\theta = 2$. We consider quantile level $\alpha = 0.25$.

It is worth noticing that in Experiments 1 and 2, we consider the case when Y is quantile independent from X at the quantile level α. Correspondingly, MISE $(\alpha, h, \omega) = \mathbb{E}\left[\widehat{D}_\alpha(X, Y; \omega, h)\right]$, and we have $h^*(\alpha, \omega) = \left[\frac{M_{\perp\alpha,2}(\omega)}{4M_{\perp\alpha,4}(\omega)}\right]^{\frac{1}{5}} \cdot n^{-\frac{1}{5}}$, where

$$
\begin{aligned}
M_{\perp_\alpha,2}(\omega) =& \frac{\alpha}{4\pi} - \frac{\alpha^2}{4\pi}, \\
M_{\perp_\alpha,4}(\omega) =& \frac{2\alpha^2 - 4\alpha\phi'(\Phi^{-1}(\alpha)) + 3\left[\phi'(\Phi^{-1}(\alpha))\right]^2}{24\sqrt{3}\pi}.
\end{aligned}
$$

Remark A.2.1. *In Experiment 3, since we have*

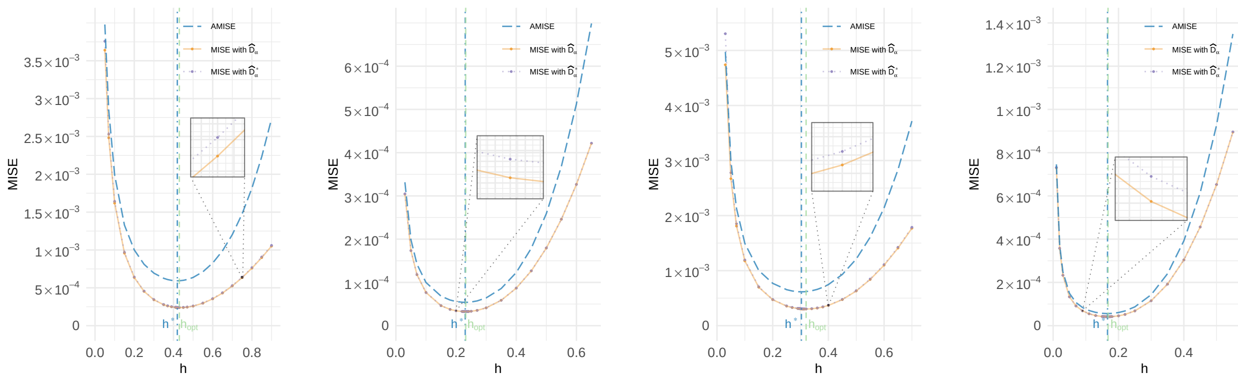
$$
\mathcal{C}(u, v) = \exp\left[-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{\frac{1}{\theta}}\right],
$$

with $\theta = 2$, it is easy to prove

$$
\mathcal{C}_{2|1}(v \mid u) = -\frac{\log(u)\exp\left\{-\left[(\log u)^2 + (\log v)^2\right]^{\frac{1}{2}}\right\}}{u\left[(\log u)^2 + (\log v)^2\right]^{\frac{1}{2}}}.
$$

Then numerically, we have

$$
\begin{aligned}
\textit{AMISE}\,(0.25, h, \omega) &= h^4 \int_0^1 B^2(u, v)\omega(u)du + \frac{1}{nh}\int_0^1 \sigma^2(u, v)\omega(u)du \\
&= 0.005585 \times h^4 + 0.01181 \times \frac{1}{nh}.
\end{aligned}
$$

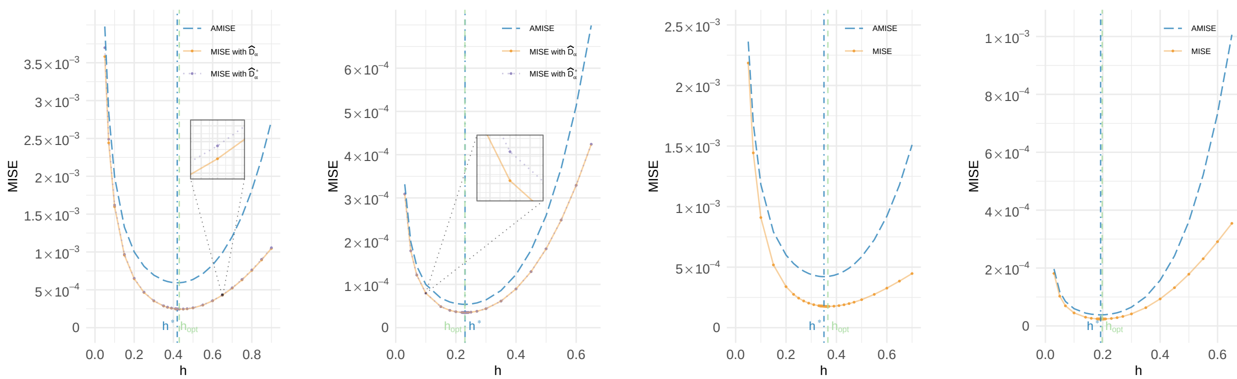


(a) Experiment 1.a with $\alpha = 0.5$, and $n = 100$.

(b) Experiment 1.a with $\alpha = 0.5$, and $n = 2000$.

(c) Experiment 1.b with $\alpha = 0.75$, and $n = 100$.

(d) Experiment 1.b with $\alpha = 0.75$, and $n = 2000$.



(e) Experiment 2 with $\alpha = 0.5$, and $n = 100$.

(f) Experiment 2 with $\alpha = 0.5$, and $n = 2000$.

(g) Experiment 3 with $\alpha = 0.25$, and $n = 100$.

(h) Experiment 3 with $\alpha = 0.25$, and $n = 2000$.

Figure A.5: Bandwidth selection for sample Copula Quantile Dependence.

Consequently, we have $h^*(0.25, \omega) = 0.8803 \times n^{-\frac{1}{5}}$*. Besides, for* $MISE\,(0.25, h, \omega)$*, we use approximation,*

$$MISE^*\,(0.25, h, \omega) \approx \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{\mathcal{C}}_{2|1}\,(0.25 \mid \widehat{u}_i) - \mathcal{C}_{2|1}\,(0.25 \mid \widehat{u}_i)\right]^2 \omega(\widehat{u}_i)$$

instead for easy computation.

The results of the 3 experiments are given in Figure A.5. Eight patterns are plotted for three experiments and $n = 100, 2000$. Over 500 replications, the sample MISE (α, h, ω) are plotted versus bandwidth h. Yellow solid lines represent the results using $\widehat{D}_\alpha(X, Y; \omega, h)$, purple dashed lines represent the results using $\widehat{D}^*_\alpha(X, Y; \omega, h)$, and blue dashed lines represent the results using AMISE (α, h, ω). The vertical blue dot dashed lines are references for the optimal bandwidth $h^*_\alpha(\omega)$ proposed in Corollary 3.3.1, and the vertical green dashed lines are references for the empirical optimal bandwidth by minimizing the empirical MISE. In each pattern of Experiments 1 and 2, we show a Zoom-in window to illustrate the closeness between $\widehat{D}_\alpha(X, Y; \omega, h)$ and $\widehat{D}^*_\alpha(X, Y; \omega, h)$. It shows not only the bandwidth is optimally selected, but also the closeness between $\widehat{D}_\alpha(X, Y; \omega, h)$ and $\widehat{D}^*_\alpha(X, Y; \omega, h)$.

A.3 Bias Adjustification Term in Chapter 3

In this subsection, we introduce the method for figuring out the bias adjustification term, δ_n, in Corollary 3.6.1 by simulations. The target of the bias adjustification term is to improve the empirical performance of the QC-Screen with FDR control.

By considering random variables, (X, Y), which are independent with each other, we may calculate the Copula Quantile Dependence, $\widehat{D}_\alpha(X, Y; \omega, h)$, under null case. With $\omega(u) = \phi^2(\Phi^{-1}(u))$, optimal bandwidth $h^*_\alpha(\omega)$, $\alpha = 0.1, 0.15, \cdots, 0.9$, and $n = 100 * 2^q$ with $q = 0, 1, \cdots, 8$, we calculate statistics,

$$\widehat{D}^\ddagger_\alpha(X, Y; \omega, h^*_\alpha(\omega)) \doteq \frac{n^{-1}\left[h^*_\alpha(\omega)\right]^{-1} M_{\perp\alpha,2}(\omega) + \left[h^*_\alpha(\omega)\right]^4 M_{\perp\alpha,4}(\omega)}{\sqrt{4n^{-1}\left[h^*_\alpha(\omega)\right]^4 \sigma^2_{\perp\alpha,1}(\omega) + 2n^{-2}\left[h^*_\alpha(\omega)\right]^{-1} \sigma^2_{\perp\alpha,3}(\omega)}} \cdot \log\left[\frac{\widehat{D}_\alpha(X, Y; \omega, [h^*_\alpha(\omega)])}{n^{-1}\left[h^*_\alpha(\omega)\right]^{-1} M_{\perp\alpha,2}(\omega) + \left[h^*_\alpha(\omega)\right]^4 M_{\perp\alpha,4}(\omega)}\right],$$

which is the $\widehat{D}^{\dagger}_{\alpha}(X, Y; \omega, h^*_{\alpha}(\omega))$ without bias adjustification. For each (α, n) combination, with 5000 replication, we find the $\delta^*_n(\alpha)$ as below

$$\delta^*_n(\alpha) \doteq \sup\left\{t \in \mathbb{R} : \text{Algorithm 3.2 selects no variables with } \widehat{D}^{\ddagger}_{\alpha}(X, Y; \omega, h^*_{\alpha}(\omega)) + t\right\}.$$

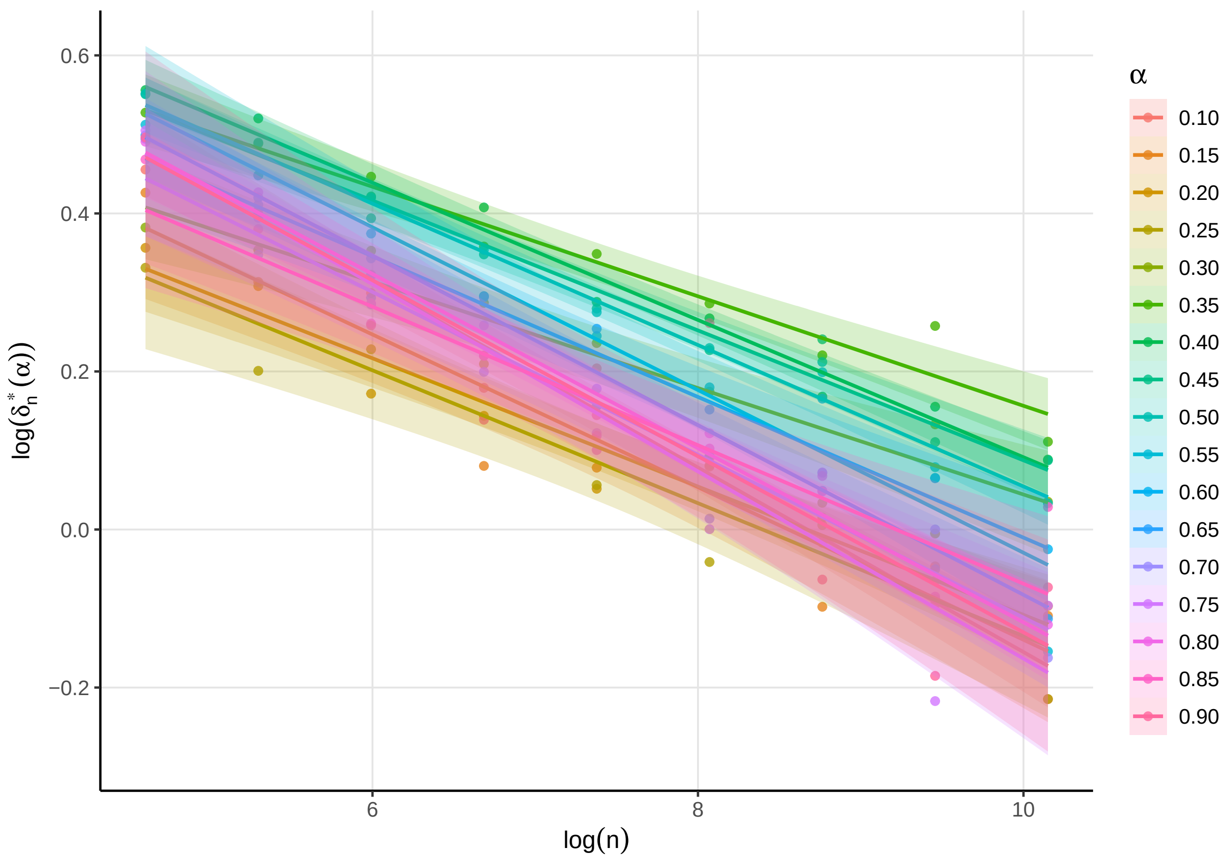


Figure A.6: Scatter plots, fitted lines and confidence intervals of linear regressions between $\log\left(\delta^*_n(\alpha)\right)$ and $\log(n)$.

In this set, we may apply Algorithm 3.2 with different γ values. Larger γ value would induce a smaller $\delta^*_n(\alpha)$, hence make the final FDR control algorithm be more conservative and select less variables. However, large γ value would usually guarantee the FDR being controlled well. Practically, people hardly select very high FDR level, hence, we consider $\gamma = 0.35$ in the above equation, which would be conservative enough. Correspondingly, $\forall\alpha$, with $\delta_n^*(\alpha)$'s, we construct a power decay converging sequence, $k_{\alpha 1} n^{-k_{\alpha 2}}$, with some constant $k_{\alpha 2} > 0$, to characterize the bias justification term. In Figure A.6, for each α, we plot the the graphs between $\log\left(\delta_n^*(\alpha)\right)$ and $\log(n)$. 17 patterns are plot for different values of α. The result clearly shows a linear pattern and verifies the correctness of power decay of the bias adjustification term. In addition, it is also noticeable that all the lines in this graph are nearly parallel with each other, which motivates us to select a uniform $k_{\alpha 2}$ over α. Upon the simulation results, we suggest to use $k_{\alpha 2} = 0.1$.

Then, we applied least square estimator to find the $k_{\alpha 1}$ for different α values, and the results are shown in Table A.1.

Table A.1: Results of bias adjustification term

α	$k_{\alpha 1}$
0.10	2.436
0.15	2.343
0.20	2.390
0.25	2.478
0.30	2.650
0.35	3.006
0.40	3.044
0.45	2.985
0.50	2.858
0.55	2.759
0.60	2.716
0.65	2.649
0.70	2.514
0.75	2.455
0.80	2.487
0.85	2.529
0.90	2.459

A.4 Variable Names of the U.S. 2020 Economic Data

In this section, we show the list of variables of the U.S. 2020 economic data.

Table A.2: Variable names of U.S. 2020 economic data

Variable	Names
X_1	Number of owner-occupied housing units with a mortgage
X_2	Median value of owner-occupied housing units with a mortgage (dollars)
X_3	Median household income in the past 12 months of owner-occupied housing units with a mortgage (dollars)
X_4	Median monthly housing costs of owner-occupied housing units with a mortgage (dollars)
X_5	Median real estate taxes of owner-occupied housing units with a mortgage (dollars)
X_6	Number of owner-occupied housing units
X_7	Number of renter-occupied housing units
X_8	Median household income in the past 12 months of owner-occupied housing units (dollars)
X_9	Median household income in the past 12 months of renter-occupied housing units (dollars)
X_{10}	Median monthly housing costs of occupied housing units (dollars)
X_{11}	Median monthly housing costs of owner-occupied housing units (dollars)

Continued on the next page

Variable	Rewritten Names
X_{12}	Median monthly housing costs of renter-occupied housing units (dollars)
X_{13}	Population 16 years and over
X_{14}	Proportion of population 16 years and over in labor force
X_{15}	Proportion of population 16 years and over in civilian labor force
X_{16}	Proportion of population 16 years and over in employed civilian labor force
X_{17}	Proportion of population 16 years and over in unemployed civilian labor force
X_{18}	Proportion of population 16 years and over in armed forces
X_{19}	Proportion of population 16 years and over not in labor force
X_{20}	Population of civilian labor force
X_{21}	Unemployment Rate in civilian labor force
X_{22}	Population of females 16 years and over
X_{23}	Proportion of population of females 16 years and over in labor force
X_{24}	Proportion of population of females 16 years and over in civilian labor force
X_{25}	Proportion of population of females 16 years and over in employed civilian labor force
X_{26}	Number of householders own children under 6 years
X_{27}	Proportion of householders own children under 6 years with all parents in family in labor force
X_{28}	Number of householders own children 6 to 17 years

Continued on the next page

Variable	Rewritten Names
X_{29}	Proportion of householders own children 6 to 17 years with all parents in family in labor force
X_{30}	Population of workers 16 years and over
X_{31}	Proportion of workers 16 years and over commuting to work with car, truck, or van (drove alone)
X_{32}	Proportion of workers 16 years and over commuting to work with car, truck, or van (carpooled)
X_{33}	Proportion of workers 16 years and over commuting to work with public transportation (excluding taxicab)
X_{34}	Proportion of workers 16 years and over commuting to work by walking
X_{35}	Proportion of workers 16 years and over commuting to work by other means
X_{36}	Proportion of workers 16 years and over working from home
X_{37}	Mean travel time to work of workers 16 years and over (minutes)
X_{38}	Civilian employed population 16 years and over
X_{39}	Proportion of civilian employed population 16 years and over in management, business, science, and arts occupations
X_{40}	Proportion of civilian employed population 16 years and over in service occupations
X_{41}	Proportion of civilian employed population 16 years and over in sales and office occupations

Continued on the next page

Variable	Rewritten Names
X_{42}	Proportion of civilian employed population 16 years and over in natural resources, construction, and maintenance occupations
X_{43}	Proportion of civilian employed population 16 years and over in production, transportation, and material moving occupations
X_{44}	Proportion of civilian employed population 16 years and over in agriculture, forestry, fishing and hunting, and mining industries
X_{45}	Proportion of civilian employed population 16 years and over in construction industries
X_{46}	Proportion of civilian employed population 16 years and over in manufacturing industries
X_{47}	Proportion of civilian employed population 16 years and over in wholesale trade industries
X_{48}	Proportion of civilian employed population 16 years and over in retail trade industries
X_{49}	Proportion of civilian employed population 16 years and over in transportation and warehousing, and utilities industries
X_{50}	Proportion of civilian employed population 16 years and over in information industries
X_{51}	Proportion of civilian employed population 16 years and over in finance and insurance, and real estate and rental and leasing industries

Continued on the next page

Variable	Rewritten Names
X_{52}	Proportion of civilian employed population 16 years and over in professional, scientific, and management, and administrative and waste management services industries
X_{53}	Proportion of civilian employed population 16 years and over in educational services, and health care and social assistance industries
X_{54}	Proportion of civilian employed population 16 years and over in arts, entertainment, and recreation, and accommodation and food services industries
X_{55}	Proportion of civilian employed population 16 years and over in other services, except public administration industries
X_{56}	Proportion of civilian employed population 16 years and over in public administration industries
X_{57}	Proportion of civilian employed population 16 years and over private wage and salary workers
X_{58}	Proportion of civilian employed population 16 years and over government workers
X_{59}	Proportion of civilian employed population 16 years and over self-employed in own not incorporated business workers
X_{60}	Proportion of total households with earnings
X_{61}	Mean earnings of households with earnings (dollars)
X_{62}	Proportion of total households with Social Security
X_{63}	Mean Social Security income of households with Social Security (dollars)

Continued on the next page

Variable	Rewritten Names
X_{64}	Proportion of total households with retirement income
X_{65}	Mean retirement income of households with retirement income (dollars)
X_{66}	Proportion of total households with Supplemental Security Income
X_{67}	Mean Supplemental Security Income of households with Supplemental Security Income (dollars)
X_{68}	Proportion of total households with cash public assistance income
X_{69}	Mean cash public assistance income of households with cash public assistance income (dollars)
X_{70}	Proportion of total households with Food Stamp/SNAP benefits in the past 12 months
X_{71}	Per capita income (dollars)
X_{72}	Median earnings for workers (dollars)
X_{73}	Median earnings for male full-time, year-round workers (dollars)
X_{74}	Median earnings for female full-time, year-round workers (dollars)
X_{75}	Civilian noninstitutionalized population
X_{76}	Proportion of civilian noninstitutionalized population with health insurance coverage
X_{77}	Proportion of civilian noninstitutionalized population with private health insurance
X_{78}	Proportion of civilian noninstitutionalized population with public coverage

Continued on the next page

Variable	Rewritten Names
X_{79}	Proportion of civilian noninstitutionalized population with no health insurance coverage
X_{80}	Civilian noninstitutionalized population under 19 years
X_{81}	Proportion of civilian noninstitutionalized population under 19 years with no health insurance coverage
X_{82}	Civilian noninstitutionalized population 19 to 64 years
X_{83}	Civilian noninstitutionalized population 19 to 64 years in labor force
X_{84}	Civilian noninstitutionalized population 19 to 64 years in employed labor force
X_{85}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with health insurance coverage
X_{86}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with private health insurance
X_{87}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with public coverage
X_{88}	Proportion of civilian noninstitutionalized population 19 to 64 years in employed labor force with no health insurance coverage
X_{89}	Civilian noninstitutionalized population 19 to 64 years in unemployed labor force
X_{90}	Proportion of civilian noninstitutionalized population 19 to 64 years in unemployed labor force with health insurance coverage

Continued on the next page

Variable	Rewritten Names
X_{91}	Proportion of civilian noninstitutionalized population 19 to 64 years in unemployed labor force with private health insurance
X_{92}	Proportion of civilian noninstitutionalized population 19 to 64 years in unemployed labor force with public coverage
X_{93}	Proportion of civilian noninstitutionalized population 19 to 64 years in unemployed labor force with no health insurance coverage
X_{94}	Civilian noninstitutionalized population 19 to 64 years not in labor force
X_{95}	Proportion of civilian noninstitutionalized population 19 to 64 years not in labor force with health insurance coverage
X_{96}	Proportion of civilian noninstitutionalized population 19 to 64 years not in labor force with private health insurance
X_{97}	Proportion of civilian noninstitutionalized population 19 to 64 years not in labor force with public coverage
X_{98}	Proportion of civilian noninstitutionalized population 19 to 64 years not in labor force with no health insurance coverage
X_{99}	Percentage of families and people whose income in the past 12 months is below the poverty level among all families

Continued on the next page

Variable	Rewritten Names
X_{100}	Percentage of families and people whose income in the past 12 months is below the poverty level among all families with related children of the householder under 18 years
X_{101}	Percentage of families and people whose income in the past 12 months is below the poverty level among all families with related children of the householder under 5 years only
X_{101}	Percentage of families and people whose income in the past 12 months is below the poverty level among all families with related children of the householder under 5 years only
X_{102}	Percentage of families and people whose income in the past 12 months is below the poverty level among married couple families
X_{103}	Percentage of families and people whose income in the past 12 months is below the poverty level among married couple families with related children of the householder under 18 years
X_{104}	Percentage of families and people whose income in the past 12 months is below the poverty level among married couple families with related children of the householder under 5 years only
X_{105}	Percentage of families and people whose income in the past 12 months is below the poverty level among families with female householder, no spouse present

Continued on the next page

Variable	Rewritten Names
X_{106}	Percentage of families and people whose income in the past 12 months is below the poverty level among families with female householder, no spouse present with related children of the householder under 18 years
X_{107}	Percentage of families and people whose income in the past 12 months is below the poverty level among families with female householder, no spouse present with related children of the householder under 5 years only
X_{108}	Percentage of families and people whose income in the past 12 months is below the poverty level among all people under 18 years
X_{109}	Percentage of families and people whose income in the past 12 months is below the poverty level among related children of the householder under 18 years
X_{110}	Percentage of families and people whose income in the past 12 months is below the poverty level among related children of the householder under 5 years
X_{111}	Percentage of families and people whose income in the past 12 months is below the poverty level among related children of the householder 5 to 17 years
X_{112}	Percentage of families and people whose income in the past 12 months is below the poverty level among all people 18 years and over
X_{113}	Percentage of families and people whose income in the past 12 months is below the poverty level among all people 18 to 64 years

Continued on the next page

Variable	Rewritten Names
X_{114}	Percentage of families and people whose income in the past 12 months is below the poverty level among all people 65 years and over
X_{115}	Percentage of families and people whose income in the past 12 months is below the poverty level among all people in families
X_{116}	Percentage of families and people whose income in the past 12 months is below the poverty level among all unrelated individuals 15 years and over
X_{117}	Below poverty level population for whom poverty status is determined
X_{118}	Below poverty level female population for whom poverty status is determined
X_{119}	Below poverty level population 25 years and over for whom poverty status is determined
X_{120}	Below poverty level population in civilian labor force of 16 years and over for whom poverty status is determined
X_{121}	Below poverty level population 16 years and over for whom poverty status is determined
X_{122}	Percent of below poverty level population for whom poverty status is determined
X_{123}	Percent of below poverty level female population for whom poverty status is determined
X_{124}	Percent of below poverty level population 25 years and over for whom poverty status is determined

Continued on the next page

Variable	Rewritten Names
X_{125}	Percent of below poverty level population in civilian labor force of 16 years and over for whom poverty status is determined
X_{126}	Percent of below poverty level population 16 years and over for whom poverty status is determined
X_{127}	Number of households
X_{128}	Median household income (dollars)
X_{129}	Number of families
X_{130}	Median family income (dollars)
X_{131}	Number of married-couple families
X_{132}	Median married-couple family income (dollars)
X_{133}	Number of nonfamily households
X_{134}	Median nonfamily household income (dollars)
X_{135}	Total population
X_{136}	Median age (years)
X_{137}	Population in households
X_{138}	Average household size
X_{139}	Average family size
X_{140}	Population 15 years and over
X_{141}	Male population 15 years and over

Continued on the next page

Variable	Rewritten Names
X_{142}	Female population 15 years and over
X_{143}	Population 3 years and over enrolled in school
X_{144}	Male population 3 years and over enrolled in school
X_{145}	Female population 3 years and over enrolled in school
X_{146}	Population 25 years and over
X_{147}	Female population 15 to 50 years
X_{148}	Female population 15 to 50 years who had a birth in the past 12 months
X_{149}	Unmarried female population 15 to 50 years who had a birth in the past 12 months
X_{150}	Population 30 years and over
X_{151}	Civilian population 18 years and over
X_{152}	Civilian noninstitutionalized population
X_{153}	Civilian noninstitutionalized population under 18 years
X_{154}	Civilian noninstitutionalized population 18 to 64 years
X_{155}	Civilian noninstitutionalized population 65 years and older
X_{156}	Population 1 year and over
X_{157}	Native born population
X_{158}	Foreign-born population
X_{159}	Foreign-born naturalized U.S. citizen population
X_{160}	Foreign-born not a U.S. citizen population

Continued on the next page

Variable	Rewritten Names
X_{161}	Population born outside the United States
X_{162}	Foreign-born population excluding population born at sea
X_{163}	Population 5 years and over

Bibliography

Abedi, V., Olulana, O., Avula, V., Chaudhary, D., Khan, A., Shahjouei, S., Li, J., & Zand, R. (2021). Racial, economic, and health inequality and covid-19 infection in the united states. *Journal of racial and ethnic health disparities*, *8*, 732–742.

Abramovich, F., Benjamini, Y., Donoho, D. L., & Johnstone, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *The Annals of Statistics*, *34*(2), 584–653. https://doi.org/10.1214/009053606000000074

Aharonov, D., Ambainis, A., Kempe, J., & Vazirani, U. (2001). Quantum walks on graphs. *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, 50–59.

Alfonsi, A., & Jourdain, B. (2014). A remark on the optimal transport between two probability measures sharing the same copula. *Statistics & Probability Letters*, *84*, 131–134.

Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., Davis, S. J., Leather, J., Meyer, B., Mihaylov, E., et al. (2020). Economic uncertainty before and during the covid-19 pandemic. *Journal of public economics*, *191*, 104274.

Attal, S., Guillotin-Plantard, N., & Sabot, C. (2015). Central limit theorems for open quantum random walks and quantum measurement records. *Annales Henri Poincaré*, *16*, 15–43.

Barber, R. F., & Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of statistics*, 2055–2085.

Bellman, R. (1966). Dynamic programming. *science*, *153*(3731), 34–37.

Carethers, J. M. (2021). Insights into disparities observed with covid-19. *Journal of internal medicine*, *289*(4), 463–473.

Chen, S. X., & Huang, T.-M. (2007). Nonparametric estimation of copula functions for dependence modelling. *Canadian Journal of Statistics*, *35*(2), 265–282.

Chisaki, K., Hamada, M., Konno, N., & Segawa, E. (2009). Limit theorems for discrete-time quantum walks on trees. *Interdisciplinary Information Sciences*, *15*(3), 423–429.

Cui, H., Li, R., & Zhong, W. (2015). Model-free feature screening for ultrahigh dimensional discriminant analysis. *Journal of the American Statistical Association*, *110*(510), 630–641.

Dai, C., Lin, B., Xing, X., & Liu, J. S. (2023). False discovery rate control via data splitting. *Journal of the American Statistical Association*, *118*(544), 2503–2520.

De Keyser, S., & Gijbels, I. (2025). High-dimensional copula-based wasserstein dependence. *Computational Statistics & Data Analysis*, *204*, 108096.

Del Barrio, E., González-Sanz, A., & Loubes, J.-M. (2024). Central limit theorems for general transportation costs. *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques*, *60*(2), 847–873.

Del Barrio, E., & Loubes, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *The Annals of Probability*, *47*(2), 926–951. https://doi.org/10.1214/18-AOP1275

Dieci, L., & Omarov, D. (2024). Solving semi-discrete optimal transport problems: Star shapedeness and newton's method. *Numerical Algorithms*, 1–56.

Donoho, D. L., et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, *1*(2000), 32.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*(420), 998–1004. Retrieved April 17, 2024, from http://www.jstor.org/stable/2290637

Fan, J. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66*. Routledge.

Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*.

Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *70*(5), 849–911.

Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, *20*(1), 101.

Fan, J., Samworth, R., & Wu, Y. (2009). Ultrahigh dimensional feature selection: Beyond the linear model. *The Journal of Machine Learning Research*, *10*, 2013–2038.

Fan, J., Shao, Q.-M., & Zhou, W.-X. (2018). Are discoveries spurious? distributions of maximum spurious correlations and their applications. *Annals of statistics*, *46*(3), 989.

Fan, J., & Zhou, W.-X. (2016). Guarding against spurious discoveries in high dimensions. *Journal of Machine Learning Research*, *17*(203), 1–34.

Fermanian, J.-D., Radulovic, D., & Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, *10*(5), 847–860.

Fournier, N., & Guillin, A. (2015). On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, *162*(3), 707–738.

Geenens, G. (2014). Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, *109*(505), 346–358.

Geenens, G., Charpentier, A., & Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, *23*(3), 1848–1873. https://doi.org/10.3150/15-BEJ798

Grover, L. (1996). A fast quantum mechanical algorithm for database search. *arXiv preprint quant-ph/9605043*.

Guo, X., Ren, H., Zou, C., & Li, R. (2023). Threshold selection in feature screening for error rate control. *Journal of the American Statistical Association*, *118*(543), 1773–1785.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157–1182.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, *14*(1), 1–16.

Hall, P., & Miller, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics*, *18*(3), 533–550.

Hallin, M., Mordant, G., & Segers, J. (2021). Multivariate goodness-of-fit tests based on wasserstein distance. *Electronic Journal of Statistics*, *15*, 1328–1371.

Hartmann, V., & Schuhmacher, D. (2020). Semi-discrete optimal transport: A solution procedure for the unsquared euclidean distance case. *Mathematical Methods of Operations Research*, *92*(1), 133–163.

Hastie, T. (2009). The elements of statistical learning: Data mining, inference, and prediction.

He, X., Wang, L., & Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, *41*(1), 342–369. https://doi.org/10.1214/13-AOS1087

Kempe, J. (2003). Quantum random walks: An introductory overview. *Contemporary Physics*, *44*(4), 307–327.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual review of economics*, *9*(1), 155–176.

Konno, N. (2005). Limit theorem for continuous-time quantum walk on the line. *Physical Review E*, *72*(2), 026113.

Lei, J. (2020). Convergence and concentration of empirical measures under wasserstein distance in unbounded functional spaces. *Bernoulli*, *26*(1), 767–798.

Li, G., Li, Y., & Tsai, C.-L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association*, *110*(509), 246–261.

Li, R., Zhong, W., & Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, *107*(499), 1129–1139.

Li, Z., & Van Leeuwen, M. (2023). Explainable contextual anomaly detection using quantile regression forests. *Data Mining and Knowledge Discovery*, *37*(6), 2517–2563.

Lin, Y., Liu, X., & Hao, M. (2018). Model-free feature screening for high-dimensional survival data. *Science China Mathematics*, *61*, 1617–1636.

Linton, O., & Xiao, Z. (2017). Quantile regression applications in finance. In *Handbook of quantile regression* (pp. 381–407). Chapman; Hall/CRC.

Liu, J., Zhong, W., & Li, R. (2015). A selective overview of feature screening for ultrahigh-dimensional data. *Science China Mathematics*, *58*, 1–22.

Liu, W., Ke, Y., Liu, J., & Li, R. (2022). Model-free feature screening and fdr control with knockoff features. *Journal of the American Statistical Association*, *117*(537), 428–443.

Lv, J., & Liu, J. S. (2014). Model selection principles in misspecified models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(1), 141–167.

Ma, X., & Zhang, J. (2016). Robust model-free feature screening via quantile correlation. *Journal of Multivariate Analysis*, *143*, 472–480.

Machida, T. (2013). Realization of the probability laws in the quantum central limit theorems by a quantum walk. *Quantum Info. Comput.*, *13*(5–6), 430–438.

Mai, Q., He, D., & Zou, H. (2023). Coordinatewise gaussianization: Theories and applications. *Journal of the American Statistical Association*, *118*(544), 2329–2343.

Mai, Q., & Zou, H. (2013). The kolmogorov filter for variable screening in high-dimensional binary classification. *Biometrika*, *100*(1), 229–234.

McDiarmid, C. (1989). On the method of bounded differences. In J. Siemons (Ed.), *Surveys in combinatorics, 1989: Invited papers at the twelfth british combinatorial conference* (pp. 148–188). Cambridge University Press.

McKennan, C., & Nicolae, D. (2019). Accounting for unobserved covariates with varying degrees of estimability in high-dimensional biological data. *Biometrika*, *106*(4), 823–840.

Montanaro, A., & Pallister, S. (2016). Quantum algorithms and the finite element method. *Physical Review A*, *93*(3), 032324.

Mordant, G., & Segers, J. (2022). Measuring dependence between random vectors via optimal transport. *Journal of Multivariate Analysis*, *189*, 104912.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, *9*(1), 141–142. https://doi.org/10.1137/1109020

Nayak, A., & Vishwanath, A. (2000). Quantum walk on the line.

Nguyen, L. H., Chevapatrakul, T., & Yao, K. (2020). Investigating tail-risk dependence in the cryptocurrency markets: A lasso quantile regression approach. *Journal of Empirical Finance*, *58*, 333–355.

Ni, L., & Fang, F. (2016). Entropy-based model-free feature screening for ultrahigh-dimensional multiclass classification. *Journal of Nonparametric Statistics*, *28*(3), 515–530.

Nies, T. G., Staudt, T., & Munk, A. (2022). Transport dependency: Optimal transport based dependency measures. *Contributions to the Theory of Statistical Optimal Transport*, 137.

Ozair, S., Lynch, C., Bengio, Y., Van den Oord, A., Levine, S., & Sermanet, P. (2019). Wasserstein dependency measure for representation learning. *Advances in Neural Information Processing Systems*, *32*.

Pan, W., Wang, X., Xiao, W., & Zhu, H. (2019). A generic sure independence screening procedure. *Journal of the American Statistical Association*.

Paul, A., Englert, P., & Varga, M. (2021). Socio-economic disparities and covid-19 in the usa. *Journal of Physics: Complexity*, *2*(3), 035017.

Polyakova, M., Kocks, G., Udalova, V., & Finkelstein, A. (2020). Initial economic damage from the covid-19 pandemic in the united states is more widespread across ages and geographies than initial mortality impacts. *Proceedings of the National Academy of Sciences*, *117*(45), 27934–27939.

Powell, D. (2020). Quantile treatment effects in the presence of covariates. *Review of Economics and Statistics*, *102*(5), 994–1005.

Rippl, T., Munk, A., & Sturm, A. (2016). Limit laws of the empirical wasserstein distance: Gaussian distributions. *Journal of Multivariate Analysis*, *151*, 90–109.

Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, *18*(3), 764–782. https://doi.org/10.3150/11-BEJ387

Shao, X., & Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association*, *109*(507), 1302–1318.

Shor, P. W. (1994). Algorithms for quantum computation: Discrete logarithms and factoring. *Proceedings 35th annual symposium on foundations of computer science*, 124–134.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP*, *8*(3), 229–231.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, *7*(4), 784–802.

Stigler, S. M. (1984). Studies in the history of probability and statistics xl boscovich, simpson and a 1760 manuscript note on fitting a linear relation. *Biometrika*, *71*(3), 615–620.

Székely, G. J., & Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *The Annals of Statistics*, *42*(6), 2382–2412. https://doi.org/10.1214/14-AOS1255

Thakur, N., Lovinsky-Desir, S., Bime, C., Wisnivesky, J. P., & Celedón, J. C. (2020). The structural and social determinants of the racial/ethnic disparities in the us covid-19 pandemic. what's our role? *American journal of respiratory and critical care medicine*, *202*(7), 943–949.

Tong, Z., Cai, Z., Yang, S., & Li, R. (2023). Model-free conditional feature screening with fdr control. *Journal of the American Statistical Association*, *118*(544), 2575–2587.

Venegas-Andraca, S. (2022). *Quantum walks for computer scientists*. Springer Nature.

Venegas-Andraca, S. E. (2012). Quantum walks: A comprehensive review. *Quantum Information Processing*, *11*(5), 1015–1106.

Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. CRC press.

Wang, L., Wu, Y., & Li, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association*, *107*(497), 214–222.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359–372.

Weed, J., & Bach, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, *25*(4A), 2620–2648. https://doi.org/10.3150/18-BEJ1065

Wei, Y., Kehm, R. D., Goldberg, M., & Terry, M. B. (2019). Applications for quantile regression in epidemiology. *Current Epidemiology Reports*, *6*, 191–199.

Xie, J., Lin, Y., Yan, X., & Tang, N. (2020). Category-adaptive variable screening for ultra-high dimensional heterogeneous categorical data. *Journal of the American Statistical Association*, *115*(530), 747–760.

Xu, Q., Fan, Z., Jia, W., & Jiang, C. (2019). Quantile regression neural network-based fault detection scheme for wind turbines with application to monitoring a bearing. *Wind Energy*, *22*(10), 1390–1401.

Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, *93*(441), 228–237.

Yu, K., Lu, Z., & Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician*, *52*(3), 331–350.

Zhang, T., & Ke, Y. (2024). Consistent sampling with smoothed quantum walk. *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, *01*, 25–30. https://doi.org/10.1109/QCE60285.2024.00013

Zhou, T., & Zhu, L. (2017). Model-free feature screening for ultrahigh dimensional censored regression. *Statistics and Computing*, *27*, 947–961.

Zhou, Y., & Zhu, L. (2018). Model-free feature screening for ultrahigh dimensional datathrough a modified blum-kiefer-rosenblatt correlation. *Statistica Sinica*, *28*(3), 1351–1370.

Zhu, L.-P., Li, L., Li, R., & Zhu, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, *106*(496), 1464–1475.