

Faculty of Sciences

Department of Mathematics and Statistical Sciences

Student E-mail Address: otlaadisam@biust.ac.bw

**ROBUST ESTIMATION AND INFERENCE IN SINGLE-INDEX VARYING
COEFFICIENT REGRESSION MODELS WITH RESPONSES MISSING
AT RANDOM**

by

MASEGO OTLAADISA

Student ID number: 13100022

BSc (Hons) Statistics (UNAM), MSc (Statistics) (BIUST)

A Dissertation/Thesis Submitted to the Faculty of Sciences in Partial Fulfilment of the
Requirements for the Award of the Degree of Doctor of Philosophy in Statistics of BIUST

Supervisor(s): Title, Names

Prof. Huybrechts F. Bindele	Prof. Asheber Abebe	Prof. Boikanyo Makubate
Dept. of Maths and Statistics	Dept. of Maths and Statistics	Dept. of Maths and Statistical Sciences
University of South Alabama	Auburn University	BIUST
hbindele@southalabama.edu	abebeas@auburn.edu	makubateb@biust.ac.bw

March 2022

DECLARATION REGARDING THE WORK AND COPYRIGHT

Candidate (please write in caps or type): MASEGO OTLAADISA

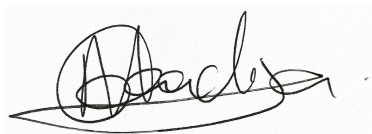
Student ID: 13100022

Thesis Titled: Robust Estimation and Inference in Single-Index Varying Coefficient Regression Models with Responses Missing at Random

I, MASEGO OTLAADISA, certify that the Thesis is all my own original work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

(If the thesis is based on a group project, then the student must indicate the extent of her / his contribution, with reference to any other theses submitted or published by each collaborator in the project, and a declaration to this effect must be included in the thesis)

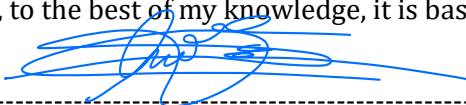
This dissertation/thesis is copyright material protected under the Berne Convention, the Copyright and Neighbouring Rights Act, Act. No. 8 of 2000 and other international and national enactments, in that behalf, on intellectual property. It must not be reproduced by any means, in full or in part, except for short extracts in fair dealing; for researcher private study, critical scholarly review or discourse with an acknowledgement, without the written permission of the office of the Postgraduate School, on behalf of both the author and the BIUST.



07/03/2022

Signed: ----- Date: -----

Primary Supervisor (please write in caps or type): HUYBRECHTS FRAZIER ACHARD BINDELE
I, HUYBRECHTS FRAZIER ACHARD BINDELE, hereby confirm that I have inspected the above titled thesis and, to the best of my knowledge, it is based on the original work of the candidate.



Signed: ----- Date: 08/03/2022

Acknowledgements

The work presented in this thesis was jointly done at Auburn University, United States of America and BIUST, funded by Simons Foundations, BIUST International Linkages Department and BIUST. My deepest gratitude goes to Professor Asheber Abebe, for giving me an opportunity of getting an exposure in the United States of America. I would like to gratefully acknowledge BIUST International Linkages Department for sponsoring my visit to Auburn University. I had a great exposure and gained invaluable knowledge while at Auburn University. Therefore, I would like to gratefully acknowledge the Office of International Programs and Graduate School at Auburn for recognizing me and making my stay and studies comfortable. My grateful acknowledgments also goes to Simon's Foundations and BIUST for financing my tuition fees while studying at BIUST.

I would like to express my deepest sincere gratitude to my advisors, Professors Huybrechts F. Bindele and Asheber Abebe, who expertly offered their priceless assistance, support and guidance. Their unswerving zeal for statistics kept me constantly engaged with my research and their personal kindness helped make my time at Auburn enjoyable. They have taught me to have a tradition of working independently, smart and hard. I would also like to acknowledge Professor Boikanyo Makubate for being my supervisor in BIUST.

I would also like to thank Professors; Oganeditse Boikanyo, Broderick Oluyede and Alphonse Amey for their valuable support in the Department. I highly appreciate my family for always believing and supporting my dreams. This work is dedicated to my mother who always wished for me to get the highest educational level possible.

Above all, I thank God who gave me good health, strength and wisdom to complete this thesis.

Abstract

Nowadays, when collecting data, it has become almost unavoidable to end up with missing data due to various reasons. Very often, these reasons are out of the control of the investigator. For regression settings, data can be missing in the response space, covariate space, or in both the response and covariate spaces. Such missingness can be related to different missing data mechanisms. The most commonly encountered missingness mechanisms in the literature include missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In data analytics, the handling of missing data has captured much attention in the statistical community. The classical approach for handling missing data involves the complete case analysis, which ignores observations with missing information in the study. This approach has been demonstrated to result in biased and/or less efficient estimates, especially when the proportion of missing data is significant. Hence, it is of utmost importance to develop methodologies for handling missing data for better statistical inference.

The main objective of this thesis is to derive robust and efficient estimates and make inferences in a single-index varying coefficient regression model (SIVCM) and its special case when some responses are assumed to be missing at random. The SIVCM has gained popularity in recent years due to its flexibility, interpretability and ability to overcome the curse of dimensionality. It has been used in many areas such as ecology, medical science, epidemiology, economics, finance, politics, and so on, to capture and model changing patterns.

This thesis considers two approaches of handling point and interval estimations of parameters in SIVCM with responses missing at random: the rank-based normal approximation approach and the rank-based empirical likelihood approach. The consistency and asymptotic normality of the rank-based normal approximation estimators are established under certain mild regularity conditions. On the other hand, under the same mild regularity conditions, asymptotic chi-square distributions of the rank-based empirical likelihood functions are established. Furthermore, robust confidence regions/intervals of the true model parameters are derived. Monte Carlo simulation studies are carried out and show that the proposed methods result in robust and more efficient estimators for the rank-based normal approximation approach when compared to the least squares

and least absolute deviations methods, when dealing with heavy tailed, contaminated model error distributions and/or when data contain gross outliers in the response space. Generally, the same experiments show that the proposed empirical likelihood approaches for interval estimation perform better when compared to their normal approximations counterpart.

Research Outputs

The contents of Chapter 2 of this thesis are already published in *Statistical Papers* and contents of Chapter 3 are to appear in the *Journal of Nonparametric Statistics*.

Contents

1	Introduction	1
1.1	Introduction	1
1.2	The SIVCM and some existing estimation methods	2
1.3	Proposed methodology	5
1.4	Robust inference via empirical likelihood	5
1.5	Aim and Objective	6
1.6	Contribution	7
1.7	Thesis outline	7
2	Estimation in Single Index Model with MAR Responses	9
2.1	Introduction	9
2.2	3-Step estimation under MAR responses	9
	Step 1: Complete case analysis	9
	Step 2: Imputation of y -missing values	11
	Step 3: Re-estimation	12
2.3	Computational Algorithm	12
2.4	Asymptotic properties	14
	Consistency	15
	Asymptotic Normality	15
2.5	Simulation & Real Data Applications	16
	Simulation Study	16
	Real Data Application: CD4 and CD8 Data	28
2.6	Conclusion	31
3	Estimation in SIVCM with MAR Responses	32
3.1	Introduction	32
3.2	3-Step estimation with MAR responses	32
	Step 1: Complete case analysis	32
	Step 2: Imputation of y -missing values	33
	Step 3: Re-estimation	34
3.3	Computational Algorithm	34

3.4	Asymptotic properties	36
	Consistency of $\hat{\beta}_{jn}$	37
	Asymptotic Normality of $\hat{\beta}_{jn}$	38
3.5	Simulation & Real Data Application	41
	Simulation Study	41
	Real Application to Catch per Unit Effort (CPUE) of Groundfish Data .	44
3.6	Discussion	48
4	Robust Empirical likelihood inference for the SIVCM	50
4.1	Introduction	50
4.2	Complete-case rank-based empirical likelihoods for β_0 and G	51
4.3	Empirical likelihood of the index parameter under MAR responses	55
4.4	Empirical likelihood of the varying coefficients under MAR responses . .	58
4.5	Simulation & Real Data	61
	Simulation Study	61
4.6	Discussion	66
5	General Conclusion	67
	References	68
6	Appendices	79
6.1	Appendix chapter 2	79
6.2	Appendix chapter 3	90
6.3	Appendix chapter 4	104

List of Figures

1	Estimated functions for Model 1 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN	24
2	Estimated functions for Model 2 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN	24
3	Estimated functions for Model 3 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN	25

4	Mean angle between full data and imputed data estimated single-index directions	25
5	Ten-fold cross-validation errors and estimated functions for the CD4 and CD8 data	30
6	Estimated functions for the SIVCM fit on fish catch data using three estimation methods for either complete case data (CC) or imputing missing responses using simple imputation (SI) or inverse probability weighting (IP).	47
7	Final smooths for rank-estimated SIVCM fit on fish catch data with missing response imputed using both SI and IP imputations.	48
8	Empirical likelihood coverage probabilities (EL-CP) and normal approximation coverage probabilities (NA-CP) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from SI under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$	63
9	Empirical likelihood coverage probabilities (EL-CP) and normal approximation coverage probabilities (NA-CP) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from IPW under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$	64
10	Empirical likelihood average lengths (EL-AL) and normal approximation average lengths (NA-AL) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from SI under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$	65
11	Empirical likelihood average lengths (EL-AL) and normal approximation average lengths (NA-AL) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from IPW under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$	66

List of Tables

1	The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the $MSEs \times 100$ of $\hat{g}$ for Model 1	20
2	The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the $MSEs \times 100$ of $\hat{g}$ for Model 2	21
3	The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSE of $\hat{g}$ for Model 3	22

4	P-values of the Multivariate Distribution Normality test of $\sqrt{n}(\hat{\beta} - \beta_0)$ for Model 3 under $N(0, 1)$ and t_3	23
5	The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSEs $\times 100$ of $\hat{g}$ for Model 1	27
6	Comparison of the complete case and imputed analyses	28
7	The mean and SD of $A(\hat{\beta}, \beta_0)$ and the MSE $\times 10^2$ of $\hat{G}$	43
8	Comparison of the complete case and imputed analysis	44
9	Cross Validation for fish data	46
10	CP (AL) of a 95% confidence interval for β_{02} under $\varepsilon \sim N(0, 1)$, t_3 , $CN(0.1, 10)$ with y imputed under random simple imputation (SI).	62
11	CP (AL) of a 95% confidence interval for β_{02} under $\varepsilon \sim N(0, 1)$, t_3 , $CN(0.1, 10)$ with y imputed under inverse probability weighting imputation (IPW).	63

1 Introduction

1.1 Introduction

Nowadays, when collecting data, it has become almost inevitable to end up with missing data due to various reasons. These can include equipment malfunction, contamination of samples, manufacturing defects, drop out in clinical trials, inhospitable conditions for data collection, and incorrect data entry. For regression modeling, missing data can occur in the response space, covariate space, or both in the response and covariate spaces. Such missingness can be related to different missing mechanisms. The most frequently encountered missing data mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), which are well discussed in Rubin (1976). Mathematically, considering the random sample $(y_i, \mathbf{x}_i, \delta_i)$, $i = 1, \dots, n$, where $\delta_i = 1$, if y_i is observed and $\delta_i = 0$, if y_i is missing, the above mechanisms can be expressed as follows:

1. MCAR mechanism is equivalent to $P(\delta = 1|y, \mathbf{x}) = P(\delta = 1)$,
2. MAR mechanism is equivalent to $P(\delta = 1|y, \mathbf{x}) = P(\delta = 1|\mathbf{x})$, and
3. MNAR mechanism is equivalent to y and δ being always dependent, even after controlling $\mathbf{x}$.

Missing data problems have been studied by many authors that include Rubin (1976), Little & Rubin (1983), Little (1988a), Little (1988b), Little & Rubin (1989), Allison (2000), Allison (2001), Scheffer (2002), Switzer III & Roth (2002), Wayman (2003), Fox-Wasylyshyn & El-Masri (2005), Enders (2010), Critical Data (2016), and Little & Rubin (2019), just to cite a few. How to deal with missing data in regression modeling has captured great attention in the statistical community due to the shortcomings brought by the classical way of handling this problem. The complete case analysis approach,

which has been the classical way of handling this problem, omits missing values in the analysis leading to reduction of statistical power and potential bias in parameter estimates (Little, 1992; Little & Rubin, 2002). To mitigate this problem, various methods of imputing missing values have been developed and studied by various authors. These methods include mean/median/mode imputation, deterministic and stochastic regression imputations, multiple imputation, linear interpolation, k-nearest neighbors (KNN) and inverse probability weighting procedures. For more information about these methods readers can refer to Little & Rubin (2002), Enders (2010) and Critical Data (2016). In regression modeling, missing data problem has not been an exception, hence, it is of utmost importance to develop methodologies for handling missing data in a regression set up for better and accurate inferences.

1.2 The SIVCM and some existing estimation methods

The single-index varying coefficient regression model (SIVCM) is an important regression model that has been used in exploring and modeling non-linearity and changing patterns in many research areas such as ecology, medical science, epidemiology, economics, finance, politics, and others. This model has gained popularity due to its flexibility and interpretability. Its ability to overcome the "curse of dimensionality" has made it a powerful tool in multivariate estimation. Our interest in the SIVCM is given to a robust and efficient estimation of the model parameters in the presence of missing responses. To be precise, suppose $y_i \in \mathbb{R}$ is the response variable, $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})^\tau$ and $\mathbf{z}_i = (z_{0i}, \dots, z_{qi})^\tau$ with $z_{0i} = 1$, are explanatory variables defined in $\mathbb{R}^{p+q+1}$. The SIVCM has the form

$$y_i = \sum_{k=0}^q g_k(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \mathbf{z}_{ki} + \varepsilon_i, \quad 1 \leq i \leq n, \quad (1.1)$$

where $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters defined in a compact space $\mathcal{B} \in \mathbb{R}^p$, $g_k(\cdot) : \mathcal{F} \rightarrow \mathbb{R}$, with $\mathcal{F} := \{u : u = \mathbf{x}_i^\tau \boldsymbol{\beta}_0\}$, $k = 0, \dots, q$ are unknown regression coefficient functions. The unspecified link functions $g_k(\cdot)$ allow for model flexibility and thus model (1.1) mitigates the risk of model misspecification. The model errors ε_i are conditionally independent and identically distributed with $E(\varepsilon_i | \mathbf{x}_i) = 0$ and $E(\varepsilon_i^2 | \mathbf{x}_i) > 0$. For model (1.1) to be identifiable, we assume that $\|\boldsymbol{\beta}_0\| = 1$ and the first component of

β_0 is positive through out the study. Because of this restriction, β_0 is known as the single-index direction.

Model (1.1) includes a class of important semiparametric models. When setting $q = 0$ in model (1.1), we obtain the single-index model (SIM) given by

$$y_i = g_0(\mathbf{x}_i^\tau \beta_0) + \varepsilon_i, \quad 1 \leq i \leq n. \quad (1.2)$$

In the absence of missing data, this model has been extensively studied by Ichimura (1987), Ichimura (1993), Hardle *et al.* (1993), Horowitz & Hardle (1996), Zhu & Fang (1996), Chiou & Müller (1998), Xia & Li (1999), Naik & Tsai (2000), Hristache *et al.* (2001), Yu & Ruppert (2002), Xue & Zhu (2006), Zhu & Xue (2006), Wang & Yang (2009), Chang *et al.* (2010), Wang *et al.* (2010), Zeng *et al.* (2012), Wang & Wang (2015), Bindele *et al.* (2018a) and many others. Other special cases for the SIVCM include the varying-coefficient model (VCM) (Chen & Qin, 1993; Hastie & Tibshirani, 1993; Wu *et al.*, 1998; Fan *et al.*, 1999; Cai *et al.*, 2000a; Wang & Yang, 2009; Cai *et al.*, 2000b; Xue & Zhu, 2007b), and the adaptive varying-coefficient linear model (AVCM) (Fan *et al.*, 2003; Lu & Cheng, 2007). These models have been widely used in application. Just as for model (1.2), in the absence of missing data, model (1.1) has also been tackled by many authors such as Xia *et al.* (2007), Wong *et al.* (2008), Xue & Pang (2013), Xue *et al.* (2012), Feng & Xue (2015), Luo *et al.* (2016) and Sun *et al.* (2019). From these contributions, the estimation of β_0 was done using many approaches that include spline estimation, least squares estimation, M-estimation, average derivative estimation, sliced inverse regression, maximum likelihood, L_2 -cross validation, empirical likelihood, and rank-based approaches. To the best of our knowledge, there has not been much literature for the SIVCM in the presence of missing data. The available literature on this model for missing response or missing covariates include recent contributions by Wang *et al.* (2010), Dong & Peng (2013), Guo *et al.* (2015), Li & Yang (2016) and Qin *et al.* (2019). Most of these contributions are based on the LS method. While the LS approach is usually preferred because it is computationally simple, its resulting estimators are known to be non-robust and less efficient in the presence of outliers, model contaminations and/or heavy-tail model error distributions. As alternative to the LS,

the least absolute deviations (LAD) method, which is robust was proposed, but is also known to result in inefficient estimators. Another alternative is the maximum likelihood (ML) method, which is robust and efficient but requires the specification of the model error distribution. This supports the fact that the ML method is vulnerable to model misspecification, which is likely to happen, as $g_k(\cdot)$, $k = 0, \dots, q$ in (1.1) are unspecified. Moreover, the ML approach assumes a joint parametric distribution for covariates and response, and it ignores the missing data mechanism, Wang *et al.* (1998). Therefore, with incomplete data, the ML approach can be very sensitive to model misspecification, since missing data are imputed from their conditional distribution given the covariates, Robins *et al.* (1995). In the same spirit to derive robust estimators of the model parameters, Yang & Yang (2014) and Song *et al.* (2016) developed robust estimation approaches based on the penalized robust regression and modal regression. Unfortunately, just as the LAD, these approaches lack efficiency. Recently, Sun *et al.* (2019) proposed a rank-based approach for model (1.1) that resulted in robust and efficient estimators of the model parameters. It is worth pointing out that their method was developed for the complete case analysis. In this thesis, we follow the approach proposed by Sun *et al.* (2019) for model (1.1), but now assuming that some responses are missing at random. There have been no studies on model (1.1) with regards to missing data until last year when Zhao *et al.* (2020) considered model (1.1) with covariates missing at random. Therefore, this study will contribute to the missing data literature for model (1.1) when some responses are missing at random. Since introduced by Jaeckel (1972), it is well known that as long as the outliers are on the response space, or the data comes from a heavy-tailed distribution, rank-based estimators outperform the LS, LAD and ML estimators (Hettmansperger & McKean, 2011). The motivation on considering the rank-based method comes from the fact that it does not require the specification of the model error distribution and hence, it is free from the risk of model misspecification. Its estimator has simple geometric interpretability as for the LS approach, Wang & Yang (2009). However, in regression modeling, the rank-based approach can be affected by outliers in the covariate space. In such cases, a weighted rank-based objective function as developed by Krasker & Welsch (1982), Sievers (1983) and Bindele & Abebe (2012) can be considered.

1.3 Proposed methodology

From model (1.1) with $q = 0$, setting the residuals as $z_i(\boldsymbol{\beta}) = y_i - g_0(\mathbf{x}_i^T \boldsymbol{\beta})$, as introduced by Jaeckel (1972), the rank-based objective function is defined by

$$D_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{R(z_i(\boldsymbol{\beta}))}{n+1}\right) z_i(\boldsymbol{\beta}),$$

where

- φ is a nondecreasing continuous function defined on $(0, 1)$, and satisfies $\varphi(u) = -\varphi(1 - u)$.
- φ can be standardized as $\int_0^1 \varphi(u) du = 0$ and $\int_0^1 \varphi^2(u) du = 1$.
- $R(t) = \sum_{j=1}^n I\{z_j(\boldsymbol{\beta}) \leq t\}$.

Theorem 1. *Suppose that $a(1) \leq \dots \leq a(n)$, $\sum a(i) = 0$ and $a(i) = \varphi(i/(n+1)) = -a(n+1-i)$, Jaeckel (1972). Then the function defined by*

$$\|\mathbf{u}\|_a = \sum_{i=1}^n a(R(u_i)) u_i,$$

for some vector $\mathbf{u} = (u_1, \dots, u_n)^T$, is a pseudo-norm.

The proof of this theorem can be found in Hettmansperger & McKean (2011), hence readers seeking for details are referred to the aforementioned reference.

In practice, the scores $a(1), \dots, a(n)$ are generated using the score generating function φ defined above as $a(i) = \varphi(i/(n+1))$. Minimizing the rank-based objective function above is known to result in robust and efficient estimators when dealing with contaminated, heavy-tailed model error distributions and/or in the presence of gross outliers in the response space (Hettmansperger & McKean, 2011).

1.4 Robust inference via empirical likelihood

Since introduced by Owen (1988), the empirical likelihood (EL) has gained popularity as a way of constructing robust confidence regions/intervals by avoiding estimating the estimator's covariance matrix that is necessary in the normal approximation (NA) approach

for possible statistical inference about the true regression parameters. The EL method has become a powerful tool in nonparametric statistical inference and has been used by many authors to make inference about the parameters in different regression models. In regard to making inference about parameters of models in (1.1), authors that have used empirical likelihood approach include Gao & Liang (1997), Zhu & Xue (2006), Xue & Zhu (2006), You & Zhou (2006), Xue & Zhu (2007b), Li *et al.* (2010), Xue *et al.* (2012), and Yang *et al.* (2017) in the absence of missing data. When it comes to missing data, Xue & Pang (2013) employed the empirical likelihood approach for model (1.2) with missing covariates. Once again, the advantage of this method is that, it does not require the estimation of the estimator's covariance matrix unlike the normal approximation approach. Moreover, the EL method is transformation invariant and range preserving (Hall and La Scala (1990)). Generally, the method is also known for providing more accurate confidence intervals/regions compared to its normal approximation counterpart. It is worth pointing out most of the contributions cited above are based on the least squares (LS) estimating equation, which is known to result in non-robust and inefficient estimates, except for normal model error distribution.

1.5 Aim and Objective

The main objective of this thesis is to derive robust and efficient estimates and make robust inference in a single-index varying coefficient regression model (SIVCM) and its special case when some responses are assumed to be missing at random.

The specific objectives are to:

1. use the rank-based estimation approach to estimate the parameters of the single index varying coefficient model in general and the single index model in particular;
2. establish asymptotic properties of the proposed estimators;
3. asses the performance of the proposed estimators using Monte Carlo simulations;
4. demonstrate via Monte Carlo simulations, the gain in estimation efficiency by imputing missing responses based on either the simple random imputation or the inverse probability weighting imputation;

5. demonstrate the performance of the proposed method through application to real data; and,
6. make robust inference by employing the empirical likelihood approach in a single-index varying coefficient regression model.

1.6 Contribution

To the best of our knowledge, this is the first work that addresses the estimation (based on the rank-based approach) and the statistical inference (via the EL method) for the index direction and the link functions in model (1.1) when some responses are missing at random. This work is motivated by the importance of robust and efficient statistical inference for decision making.

1.7 Thesis outline

The remainder of this thesis is organized as follows: in Chapter 2 and Chapter 3 we propose a rank-based estimation procedure for models (1.1) and (1.2) with MAR responses starting with model (1.2). Asymptotic properties of the proposed estimators are established under mild regularity conditions. In the same chapters, we demonstrate the gain in estimation efficiency by imputing missing responses based on either the simple random imputation or the inverse probability weighting imputation. Furthermore, the proposed method results in estimators that are robust to model error contamination, to heavy-tail model error distributions and in the presence of outliers compared to the LS and the LAD approaches. When it comes to the estimation of the link functions, a backfitting type algorithm is proposed by iterating between the link functions through local linear estimation procedure and estimating β_0 by a one-step rank estimation. The local linear estimation of $g_k(\cdot)$, $k = 0, \dots, q$, involves bandwidth selection, which is done via a leave-one-out cross-validation. In Chapter 4, a rank-based robust empirical likelihood (EL) approach is developed in an effort to derive robust confidence regions/intervals of model parameters. Monte Carlo simulations show that this approach provides more accurate confidence regions/intervals compared its normal approximation (NA) counterpart

and those obtained from the LS and LAD analogous. Finally, technical details of the theoretical development in the thesis are given in Chapter 5.

2 Estimation in Single Index Model with MAR Responses

2.1 Introduction

In this chapter, we consider model (1.2) with MAR responses, where we propose the rank-based estimation procedure for the estimation of β_0 and $g_0(\cdot)$. Under some mild regularity conditions, we show that the proposed estimators are strongly consistent and have asymptotic normal distributions. Monte Carlo simulation study show that we gain estimation efficiency by imputing the missing responses. When samples are drawn from heavy-tailed or contaminated model error distributions, the proposed estimators are more efficient than their LS and LAD counterparts. As expected, when the model error is derived from a normal distribution, the LS index direction estimator tends to be more efficient than the rank-based index direction estimator. However, the LS link function estimator remains less efficient than the rank-based link function estimator. A real data example is analyzed and cross-validation studies show that the proposed procedure provides better prediction than the LS and LAD methods when the responses contain outliers and are missing at random.

2.2 3-Step estimation under MAR responses

Step 1: Complete case analysis

For model identifiability, assume that $\|\beta_0\| = 1$ and the first component of β_0 is positive. Let $\delta_i = 1$, if y_i is observed and $\delta_i = 0$, if y_i is missing. Considering the random sample $\{(\mathbf{x}_i, y_i, \delta_i)\}_{i=1}^n$ and premultiplying Equation (1.2) by δ_i , we obtain $\delta_i y_i = \delta_i g_0(\mathbf{x}_i^T \beta_0) + \delta_i \varepsilon_i$. Under the MAR assumption, we have $E(\delta Y | \mathbf{x}) = E(\delta g_0(\mathbf{x}^T \beta_0) | \mathbf{x}) = g_0(\mathbf{x}^T \beta_0) E(\delta | \mathbf{x})$. With the assumption that $E(\delta Y | \mathbf{x}) = E(\delta Y | \mathbf{x}^T \beta_0)$, for $u = \mathbf{x}^T \beta_0$, obtain

$$g_0(u) = \frac{E(\delta Y | \mathbf{x}^T \beta_0 = u)}{E(\delta | \mathbf{x}^T \beta_0 = u)}. \quad (2.1)$$

From this, the complete case analysis rank-based estimator of β_0 , denoted by $\tilde{\beta}_n^c$ is

obtained by minimizing

$$L_n(\boldsymbol{\beta}) = \sum_{i=1}^n \varphi \left(\frac{R(e_i(\boldsymbol{\beta}))}{n+1} \right) e_i(\boldsymbol{\beta}), \quad (2.2)$$

where $R(e_i(\boldsymbol{\beta}))$ is the rank of $e_i(\boldsymbol{\beta}) = \delta_i(y_i - g_0(\mathbf{x}_i^T \boldsymbol{\beta}_0))$ among $\delta_1(y_1 - g_0(\mathbf{x}_1^T \boldsymbol{\beta}_0)), \dots, \delta_n(y_n - g_0(\mathbf{x}_n^T \boldsymbol{\beta}_0))$, and φ is a general non-decreasing, continuous score function defined on $(0, 1)$. It is worth pointing out that $\tilde{\boldsymbol{\beta}}_n^c$ is $\sqrt{n}$ -consistent; see Bindele (2017). It should be noted that in Equation (2.2), $\boldsymbol{\beta}_0$ and $g_0(\cdot)$ are not known. A Nadaraya-Watson estimator of $g_0(u)$ is given by

$$\tilde{g}(u) = \frac{\sum_{i=1}^n \delta_i y_i K_h(u - \mathbf{x}_i^T \boldsymbol{\beta})}{\sum_{i=1}^n \delta_i K_h(u - \mathbf{x}_i^T \boldsymbol{\beta})}, \quad (2.3)$$

where $K_h(t) = K(t/h)$ with K being some kernel function and h the corresponding bandwidth. Following Einmahl & Mason (2005), under some mild regularity conditions, it can be shown that the estimator in Equation (2.3) is a uniformly strongly consistent estimator of $g_0(u)$. From this, the complete case rank-based objective function is defined as

$$\tilde{L}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \varphi \left(\frac{R(\tilde{e}_i(\boldsymbol{\beta}))}{n+1} \right) \tilde{e}_i(\boldsymbol{\beta}), \quad (2.4)$$

where $\tilde{e}_i(\boldsymbol{\beta}) = \delta_i(y_i - \tilde{g}(\mathbf{x}_i^T \boldsymbol{\beta}))$. From the continuity and boundedness of φ together with the uniform strong consistency of $\tilde{g}(u|\boldsymbol{\beta})$, it can be shown that $L_n(\boldsymbol{\beta})$ and $\tilde{L}_n(\boldsymbol{\beta})$ are asymptotically equivalent. Therefore, the complete case analysis rank-based estimator ($\tilde{\boldsymbol{\beta}}_n^c$) of $\boldsymbol{\beta}_0$ can be obtained by minimizing Equation (2.4) instead. Note that inference about $\boldsymbol{\beta}_0$ in Equations (2.2) or (2.4) is based on ignoring observations with missing responses. As pointed out earlier, complete case analysis is known to result in biased estimators. Substantial efforts have been made to develop methods that can curb this problem. For instance, Little & Rubin (2002) have shown that, under MAR, when the link function $g_0(\cdot)$ is known the ML approach may lead to an efficient estimator of $\boldsymbol{\beta}_0$. However, when the link function $g_0(\cdot)$ is unknown, the ML approach is known for its potential high risk of model misspecification. In the same manner, in order to explore the entire information, we consider imputing for the missing y -values following similar

ideas of Bindele (2015) and Bindele & Abebe (2015).

Step 2: Imputation of y -missing values

Let $j = 1$ and $j = 2$ correspond to the simple random imputation and the inverse marginal probability weighting imputation, respectively. Then, putting $\pi(\mathbf{x}_i) = P(\delta_i = 1|\mathbf{x}_i, y_i) = P(\delta_i = 1|\mathbf{x}_i)$ under MAR, the imputed responses, z_{ij} , are given as follows

$$z_{ij} = \begin{cases} \delta_i y_i + (1 - \delta_i) \tilde{g}(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c), & \text{if } j = 1; \\ \frac{\delta_i}{\pi(\mathbf{x}_i)} y_i + \left(1 - \frac{\delta_i}{\pi(\mathbf{x}_i)}\right) \tilde{g}(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c), & \text{if } j = 2. \end{cases} \quad (2.5)$$

If the functional form of $\pi(\mathbf{x})$ in the above equation is unknown, it can be estimated non-parametrically as discussed in Bindele (2015) by $\hat{\pi}(\mathbf{x}) = \sum_{i=1}^n \omega_{ni}(\mathbf{x}) \delta_i$, with $\omega_{ni}(\mathbf{t})$ defined by

$$\omega_{ni}(\mathbf{t}) = \frac{\Omega_b(\mathbf{t} - \mathbf{x}_i)}{\sum_{k=1}^n \Omega_b(\mathbf{t} - \mathbf{x}_k)},$$

and $\Omega_b(\mathbf{t}) = \Omega(\mathbf{t}/b)$, where Ω is a kernel function with bandwidth b . Following the ideas in Einmahl & Mason (2005), it can be shown that $\hat{\pi}(\mathbf{x}) \rightarrow \pi(\mathbf{x})$ *a.s.* Henceforth, the imputed responses can be defined as

$$\tilde{z}_{ij} = \begin{cases} \delta_i y_i + (1 - \delta_i) \tilde{g}(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c), & \text{if } j = 1; \\ \frac{\delta_i}{\hat{\pi}(\mathbf{x}_i)} y_i + \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{x}_i)}\right) \tilde{g}(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c), & \text{if } j = 2. \end{cases} \quad (2.6)$$

Remark 2.1. *For high dimensional $\mathbf{x}$, it is worth pointing out that one may face the curse of dimensionality issue when estimating $\pi(\mathbf{x})$ that often occurs in multivariate kernel density estimation. To overcome this issue, lately substantial efforts have been made to derive ways of evading the curse of dimensionality issue. One way, is through the use of vine copulas as proposed by Nagler & Czado (2016) or by considering the approach proposed by Guo et al. (2018).*

Step 3: Re-estimation

Now, consider the following rank-based objective function given by

$$D_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \varphi\left(\frac{R(\xi_{ij}(\boldsymbol{\beta}))}{n+1}\right) \xi_{ij}(\boldsymbol{\beta}), \quad (2.7)$$

where $\xi_{ij}(\boldsymbol{\beta})$ are defined as $\xi_{ij}(\boldsymbol{\beta}) = \tilde{z}_{ij} - g_0(\mathbf{x}_i^\tau \boldsymbol{\beta})$. Our interest is placed in the estimation of $\boldsymbol{\beta}_0$ and $g_0(\cdot)$. It is worth pointing out that while the Nadaraya-Watson estimator applied in Step 1 provides quick starting values, it has been proven that local linear estimators tend to perform better than local constant estimators for estimating the function $g_0(\cdot)$. From now on, we propose estimating $g_0(\cdot)$ based on a local linear estimator (Fan & Gijbels, 1996). Under the smoothness assumption on $g_0(\cdot)$ and applying the mean value theorem to $g_0(u)$, for $u \in \mathcal{F}$, one can approximate $g_0(\mathbf{x}_i^\tau \boldsymbol{\beta})$ as $g_0(\mathbf{x}_i^\tau \boldsymbol{\beta}) \approx g_0(\mathbf{v}^\tau \boldsymbol{\beta}) + g'_0(\mathbf{v}^\tau \boldsymbol{\beta})(\mathbf{x}_i - \mathbf{v})^\tau \boldsymbol{\beta}$, for any $\mathbf{v}$ in the neighborhood of $\mathbf{x}$. From this, rewrite $\xi_{ij}(\boldsymbol{\beta})$ as $\xi_{ij}(\boldsymbol{\beta}) = \tilde{z}_{ij} - a - b\mathbf{x}_{0i}^\tau \boldsymbol{\beta}$, where $a = g_0(\mathbf{v}^\tau \boldsymbol{\beta})$, $b = g'_0(\mathbf{v}^\tau \boldsymbol{\beta})$ and $\mathbf{x}_{0i} = \mathbf{x}_i - \mathbf{v}$. Since the estimation of $g_0(\cdot)$ and $g'_0(\cdot)$ involve kernel smoothing, consider minimizing

$$\tilde{D}_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_i \varphi\left(\frac{R(\xi_{ij}(\boldsymbol{\beta}))}{n+1}\right) \xi_{ij}(\boldsymbol{\beta}), \quad (2.8)$$

instead of $D_{jn}(\boldsymbol{\beta})$. In Equation(2.8), $w_i \equiv w_i(\tilde{\boldsymbol{\beta}}_n^c)$ is the weight introduced to control boundary effects and is defined as

$$w_i = K_h(\mathbf{x}_{0i}^\tau \tilde{\boldsymbol{\beta}}_n^c) / \sum_{i=1}^n K_h(\mathbf{x}_{0i}^\tau \tilde{\boldsymbol{\beta}}_n^c).$$

Finally, we can obtain the rank-based estimator of $\boldsymbol{\beta}_0$ as

$$\hat{\boldsymbol{\beta}}_n^j = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\text{Argmin}} \tilde{D}_{jn}(\boldsymbol{\beta}). \quad (2.9)$$

2.3 Computational Algorithm

For computational purposes, $\mathbf{v}$ needs to be specified. Suppose that $\mathbf{v} = \mathbf{x}_k$ and set $a_k = g_0(\mathbf{x}_k^\tau \boldsymbol{\beta})$, $b_k = g'_0(\mathbf{x}_k^\tau \boldsymbol{\beta})$, then $g_0(\mathbf{x}_i^\tau \boldsymbol{\beta}) \approx a_k + b_k(\mathbf{x}_i - \mathbf{x}_k)^\tau \boldsymbol{\beta}$. For each fixed k with

$\xi_{ij}^k(\boldsymbol{\beta}) = \tilde{z}_{ij} - a_k - b_k \mathbf{x}_{ik}^\tau \boldsymbol{\beta}$ and $\mathbf{x}_{ik} = \mathbf{x}_i - \mathbf{x}_k$, let

$$\tilde{D}_{jn}^k(\boldsymbol{\beta}) = \tilde{D}_{jn}(a_k, b_k, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_{ik} \varphi \left(\frac{R(\xi_{ij}^k(\boldsymbol{\beta}))}{n+1} \right) \xi_{ij}^k(\boldsymbol{\beta}), \quad (2.10)$$

where the weight function $w_{ik} \equiv w_{ik}(\tilde{\boldsymbol{\beta}}_n^c) = K_h(\mathbf{x}_{ik}^\tau \tilde{\boldsymbol{\beta}}_n^c) / \sum_{k=1}^n \sum_{i=1}^n K_h(\mathbf{x}_{ik}^\tau \tilde{\boldsymbol{\beta}}_n^c)$.

Remark 2.2. When minimizing $\tilde{D}_{jn}^k(\boldsymbol{\beta})$, one has to face the estimation of a_k , b_k and $\boldsymbol{\beta}$. Following the MAVE approach described in Xia et al. (2002) with $\delta_i = 10^{-4}$ for $i = 1, 2, 3$, the alternating estimation procedure can be used as follows:

Step 1: For $j = 1, 2$, consider $\tilde{\boldsymbol{\beta}}_n^{j(0)}$ the initial estimator of $\boldsymbol{\beta}_0$ obtained from model (1.2) when the function g_0 can be, for example, chosen as a linear function. Set $r = 1$.

Step 2: For each k , compute $(\hat{a}_k^{j(0)}, \hat{b}_k^{j(0)})$ as $(\hat{a}_k^{j(0)}, \hat{b}_k^{j(0)}) = \underset{a_k, b_k}{\text{Argmin}} \tilde{D}_{jn}^k(\tilde{\boldsymbol{\beta}}_n^{j(0)})$.

Step 3: Compute $\hat{\boldsymbol{\beta}}_n^{j(r)}$ as $\hat{\boldsymbol{\beta}}_n^{j(r)} = \underset{\boldsymbol{\beta}}{\text{Argmin}} \tilde{D}_{jn}(\hat{a}_k^{j(r-1)}, \hat{b}_k^{j(r-1)}, \boldsymbol{\beta})$.

Step 4: For each k , compute $(\hat{a}_k^{j(r)}, \hat{b}_k^{j(r)})$ as $(\hat{a}_k^{j(r)}, \hat{b}_k^{j(r)}) = \underset{a_k, b_k}{\text{Argmin}} \tilde{D}_{jn}^k(\hat{\boldsymbol{\beta}}_n^{j(r)})$.

Step 5: If

$$\frac{|\hat{b}_k^{j(r)} - \hat{b}_k^{j(r-1)}|}{|\hat{b}_k^{j(r-1)}|} \leq \delta_1, \quad \frac{|\hat{a}_k^{j(r)} - \hat{a}_k^{j(r-1)}|}{|\hat{a}_k^{j(r-1)}|} \leq \delta_2 \quad \text{and} \quad \frac{\|\hat{\boldsymbol{\beta}}_n^{j(r)} - \hat{\boldsymbol{\beta}}_n^{j(r-1)}\|}{\|\hat{\boldsymbol{\beta}}_n^{j(r-1)}\|} \leq \delta_3,$$

where $\delta_1, \delta_2, \delta_3$ are small positive constants giving relative tolerance limits, then take

$\hat{\boldsymbol{\beta}}_n^j = \hat{\boldsymbol{\beta}}_n^{j(r)}$, $\hat{a}_k^j = \hat{a}_k^{j(r)}$, $\hat{b}_k^j = \hat{b}_k^{j(r)}$ and STOP. Otherwise, let $r = r + 1$ and go back to

Step 3.

Thus, for fixed n , we get $\hat{\boldsymbol{\beta}}_n^j = \lim_{r \rightarrow \infty} \hat{\boldsymbol{\beta}}_n^{j(r)}$. It is worth pointing out that at each step of the algorithm described above, the bandwidth h has to be selected and this is done by selecting h as $\hat{h}^{j(r)} = \underset{h}{\text{Argmin}} \tilde{D}_{jn}(\hat{a}_k^{j(r)}, \hat{b}_k^{j(r)}, \hat{\boldsymbol{\beta}}_n^{j(r)}, h)$, for $r = 0, 1, 2, \dots$

Now, set $\hat{a} = \hat{g}(t)$, $\hat{b} = \hat{g}'(t)$ for any $t \in \mathcal{T}$ and, $\nu_{ij}(\boldsymbol{\beta}) = \tilde{z}_{ij} - \hat{a} - \hat{b} \mathbf{x}_{0i}^\tau \boldsymbol{\beta}_0$. Define $Q_{jn}(\boldsymbol{\beta})$

as

$$Q_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_i \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) \nu_{ij}(\boldsymbol{\beta}).$$

From the strong consistency of $\hat{g}(t)$ and $\hat{g}'(t)$, it is obtained in a straightforward manner that $\nu_{ij}(\boldsymbol{\beta}) - \xi_{ij}(\boldsymbol{\beta}) \rightarrow 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$. From this, we have

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} |Q_{jn}(\boldsymbol{\beta}) - \tilde{D}_{jn}(\boldsymbol{\beta})| = 0 \text{ a.s.}$$

The proof of this fact follows that of Theorem 1 in Bindele (2017). So, for the sake of brevity, it is omitted here. From this asymptotic equivalence result, the strong consistency of $\hat{\boldsymbol{\beta}}_n^j$ can be established using $\tilde{D}_{jn}(\boldsymbol{\beta})$.

2.4 Asymptotic properties

In this section, we derive the asymptotic properties (strong consistency and asymptotic distribution) of $\hat{\boldsymbol{\beta}}_n^j$. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. For $i = 1, \dots, n$, since $\varepsilon_i \equiv \varepsilon_i(\boldsymbol{\beta}_0)$ are assumed to be i.i.d., $\xi_{ij} \equiv \xi_{ij}(\boldsymbol{\beta}_0)$ are i.i.d. with common distribution F_j , and for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $\xi_{ij}(\boldsymbol{\beta})$ are independent with distribution $F_{ij\boldsymbol{\beta}}$. We also consider the following assumptions:

($\mathcal{J}_1$) φ is a nondecreasing, bounded and twice continuously differentiable score function with bounded derivatives, defined on $(0, 1)$, and assume that φ can be standardized as

$$\int_0^1 \varphi(u) du = 0 \quad \text{and} \quad \int_0^1 \varphi^2(u) du = 1.$$

($\mathcal{J}_2$) ε_i , $i = 1, \dots, n$, are i.i.d. continuous errors with common distribution F , finite Fisher information and almost surely bounded first derivative of the density function f' , where $f(t) = dF(t)/dt$.

($\mathcal{J}_3$) $g(\cdot)$ is a function defined on $\mathcal{F} = \{\mathbf{x}^T \boldsymbol{\beta} : \boldsymbol{\beta} \in \mathcal{B}, \mathbf{x} \in \mathbb{R}^p\}$. There exist an arbitrary function $J(\cdot)$ independent of $\boldsymbol{\beta}$ such that $\|\nabla_{\boldsymbol{\beta}}^r \kappa(\boldsymbol{\beta}, \mathbf{x})\| \leq J(\mathbf{x})$, for $r = 0, 1, 2, 3$, $\kappa(\boldsymbol{\beta}, \mathbf{x}) = g(\mathbf{x}^T \boldsymbol{\beta})$ is three times continuously differentiable with respect to $\boldsymbol{\beta}$, and $E[J^\alpha(\mathbf{x})] < \infty$, for some $\alpha \geq 1$. Also, the function $t \mapsto g(t)$ is three times continuously differentiable for any $t \in \mathcal{F}$.

($\mathcal{J}_4$) Letting $m_{\boldsymbol{\beta}}$ be the density function of $\mathbf{x}^T \boldsymbol{\beta}$, we assume that $m_{\boldsymbol{\beta}}$ have bounded continuous derivatives up to order 2. Also, $\inf_{t \in \mathcal{F}} m_{\boldsymbol{\beta}}(t) > \alpha$, with $\alpha > 0$, $\forall \boldsymbol{\beta} \in \mathcal{B}$.

Moreover, there exists c_0 such that $0 < c_0 \leq \pi(\mathbf{x}) < 1$, for all $\mathbf{u}$.

($\mathcal{J}_5$) $K(\cdot)$ is a symmetric density function with compact support. The bandwidth $h_n \sim n^{-\delta}$ with $1/6 \leq \delta \leq 1/4$ and satisfies $nh_n^p/\log n \rightarrow \infty$ as $n \rightarrow \infty$.

($\mathcal{J}_6$) $\sup_{\mathbf{x}} E[|\varepsilon|^r|\mathbf{x}] < \infty$, where r is the order of smoothness of $g(\cdot)$. Also, $E[\exp\{\lambda\|\mathbf{x}\|\}] < \infty$, for some λ .

($\mathcal{J}_7$) $\beta_0 \in \text{Int}(\mathcal{B})$ with $\|\beta_0\| = 1$, and for $j = 1, 2$, $\beta_0 = \underset{\beta \in \mathcal{B}}{\text{Argmin}} \lim_{n \rightarrow \infty} E[\tilde{D}_{jn}(\beta)]$.

Remark 2.3. ($\mathcal{J}_1$) and ($\mathcal{J}_2$) are regularity assumptions commonly used in the framework of rank-based estimation (Hettmansperger & McKean, 2011; Bindele & Abebe, 2015). On the other hand, ($\mathcal{J}_3$) – ($\mathcal{J}_6$) are necessary and sufficient in ensuring the uniform strong consistency of either the local constant or the local linear estimators of $g(t)$ and their derivatives, as given in Lemma 1; see Delecroix et al. (2006). Also, these same assumptions are used in the application of the Strong Law of Large Numbers at various occasions in the proof of Theorems 2 and 3. Finally, ($\mathcal{J}_7$) is required for the identifiability of the true parameter, β_0 .

Consistency

The following theorem gives the stochastic equicontinuity of the sequence of rank objective functions $\{\tilde{D}_{jn}(\beta)\}_{n \geq 1}$ and the strong consistency of $\hat{\beta}_n^j$.

Theorem 2. Under ($\mathcal{J}_1$) – ($\mathcal{J}_7$), $\tilde{D}_{jn}(\beta) - E\{\tilde{D}_{jn}(\beta)\} \rightarrow 0$ a.s., as $n \rightarrow \infty$, $j = 1, 2$ and the sequence $\{\tilde{D}_{jn}(\beta)\}_{n \geq 1}$ is stochastically equicontinuous. Moreover, $\hat{\beta}_n^j \rightarrow \beta_0$ a.s., as $n \rightarrow \infty$.

Asymptotic Normality

To establish the asymptotic distribution of $\hat{\beta}_n^j$, set $\Sigma = E[w\{g'(\beta_0^T \mathbf{x})\}^2 \mathbf{x} \mathbf{x}^T]$ and for $j = 1, 2$, let

$$\begin{aligned} \nu_j(\mathbf{x}) &= E[\varphi(F_j(\xi_j))|\mathbf{x}], \quad \mu_j(\mathbf{x}) = E[f_j(\xi_j)\varphi'(F_j(\xi_j))|\mathbf{x}], \\ \mathbf{V}_j &= E[w\{g'(\mathbf{x}^T \beta_0)\}^2 \mathbf{x} \mathbf{x}^T \mu_j(\mathbf{x})] \quad \text{and} \quad \mathcal{V}_j = E[w\{g''(\mathbf{x}^T \beta_0)\} \mathbf{x} \mathbf{x}^T \nu_j(\mathbf{x})]. \end{aligned}$$

Also define $\Gamma := \{\mathbf{x} : m_{\beta_0}(\mathbf{x}^\tau \beta_0) \geq c\}$, with $c \geq c_0 > 0$ and let $\nabla_\beta = (\partial/\partial\beta_i)_i$ and $\nabla_\beta^2 = (\partial^2/\partial\beta_i\partial\beta_j)_{ij}$, for $\beta = (\beta_1, \dots, \beta_p)^\tau$, $i, j = 1, \dots, p$, denote the gradient and Hessian operators, respectively.

Under the smoothness assumption on φ and g , $\tilde{D}_{jn}(\beta)$ and $Q_{jn}(\beta)$ are weakly differentiable. Let $\tilde{S}_{jn}(\beta) = -\nabla_\beta \tilde{D}_{jn}(\beta)$ and $S_{jn}(\beta) = -\nabla_\beta Q_{jn}(\beta)$. Clearly, $\hat{\beta}_n^j$ being a minimizer of $\tilde{D}_{jn}(\beta)$, $\tilde{S}_{jn}(\hat{\beta}_n^j) = 0$. Precisely, $\tilde{S}_{jn}(\beta)$ is given by

$$\tilde{S}_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^n w_i \{g'(\mathbf{x}_i^\tau \beta)\} \mathbf{x}_i \varphi\left(\frac{R(\xi_{ij}(\beta))}{n+1}\right).$$

At the true parameter β_0 , $\tilde{S}_{jn}(\beta_0) = n^{-1} \sum_{i=1}^n w_i \{g'(\mathbf{x}_i^\tau \beta_0)\} \mathbf{x}_i \varphi(R(\xi_{ij}(\beta_0))/(n+1))$. Also, define $\tilde{T}_{jn}(\beta_0)$ as $\tilde{T}_{jn}(\beta_0) = n^{-1} \sum_{i=1}^n w_i \{g'(\mathbf{x}_i^\tau \beta_0)\} \mathbf{x}_i \varphi(F_{ij}(\xi_{ij}(\beta_0)))$. With this, the following theorem establishes the asymptotic distribution of $\hat{\beta}_n^j$.

Theorem 3. *Under assumptions $(\mathcal{J}_1)$ – $(\mathcal{J}_7)$, we have $\sqrt{n}(\hat{\beta}_n^j - \beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-)$, where $\mathbf{W}_j^-$ is the Moore Penrose inverse of $\mathbf{W}_j$, where $\mathbf{W}_j = -\mathbf{V}_j + \mathcal{V}_j$.*

Remark 2.4. *Note that if we assume that ε is independent of $\mathbf{x}$, ξ_j is also independent of $\mathbf{x}$ and $\mathbf{W}_j$ can be expressed using the rank scale parameter as $\mathbf{W}_j = \gamma_{j\varphi}^{-1} \Sigma$, where*

$$\gamma_{j\varphi}^{-1} = \int_0^1 \varphi(u) \varphi_{f_j}(u) du \quad \text{with} \quad \varphi_{f_j}(u) = \frac{f_j'(F_j^{-1}(u))}{f_j(F_j^{-1}(u))}.$$

Also, in practice and for further statistical inference about β_0 , the estimator's covariance matrix $\mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-$ has to be estimated. Sandwich type estimators of $\mathbf{V}_j$ and $\mathbf{W}_j$ can be obtained following ideas in Brunner & Denker (1994).

2.5 Simulation & Real Data Applications

Simulation Study

To study the finite samples performance of $\hat{\beta}_n^j$, we present simulation studies under two scenarios where in scenario 1, the proportion of missingness is between 10% – 20%, and between 50% – 70% in scenario 2. These allow us to investigate the impact of different proportions of missingness in estimation. We considered the following single index models

$$\begin{aligned}
\text{Model 1: } Y &= 1.5 \sin((\beta_0^T \mathbf{X})\pi) + \varepsilon, & \beta_0 &= (1, 2, 0, 2)/3, & \mathbf{X} &\sim (U[-1, 1])^{\otimes 4}, \\
\text{Model 2: } Y &= 4\sqrt{|(\beta_0^T \mathbf{X}) + 1|} + (\beta_0^T \mathbf{X}) + \varepsilon, & \beta_0 &= (2, -2, 4, -1)/5, & \mathbf{X} &\sim (U[-2, 2])^{\otimes 4}, \\
\text{Model 3: } Y &= 2(\beta_0^T \mathbf{X}) + 10 \exp(-(\beta_0^T \mathbf{X})^2/5) + \varepsilon, & \beta_0 &= (2, -2, 4, -1)/5, & \mathbf{X} &\sim (U[-2, 2])^{\otimes 4}.
\end{aligned}$$

Model 1 is based on an oscillating function often used in studying single-index models (Liu *et al.*, 2013). Models 2 and 3 were used in a simulation study by Zeng *et al.* (2012). The model error ε is generated from three different distributions: the standard normal distribution ($\varepsilon \sim N(0, 1)$), the contaminated normal distribution with contamination rate of 10% ($\varepsilon \sim CN \equiv 0.9N(0, 1) + 0.1N(0, 10^2)$), and the t distribution with 3 degrees of freedom ($\varepsilon \sim t_3$). These choices allow us to study the effect of tail thickness and contamination on the performance of the proposed estimator. In the rank-based objective function, we used the linear score function $\varphi(u) = \sqrt{12}(u - 1/2)$ with the local linear rank regression estimated using the iteratively reweighted least squares method of Sievers & Abebe (2004). In the local linear estimation of $g_0(\cdot)$, the kernel $K(\cdot)$ used is the Epanechnikov Kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$. The missing indicator, δ , is generated as $\delta \sim \text{Bernoulli}(\pi(\mathbf{x}))$, where $\pi(\mathbf{x})$ is obtained through two different scenarios:

Scenario 1 We considered three cases with the proportion of missingness between 10% – 20%, where

Case 1: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.8$,

Case 2: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.8 + 0.2|u(\mathbf{x}) - 0.5|$ if $|u(\mathbf{x}) - 0.5| \leq 0.25$, and 0.95 elsewhere, and

Case 3: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.9 - 0.1|u(\mathbf{x}) - 0.5|$ if $|u(\mathbf{x}) - 0.5| \leq 0.25$, and 0.8 elsewhere,

where $u(\mathbf{x}) = (4a)^{-1}\|\mathbf{x}\|_1$ with $\mathbf{x}$ generated from $U[-a, a]$, where $a = 1$ for Model 1 and $a = 2$ for Models 2 and 3. Since $u(\mathbf{X})$ follows a Bates distribution (Johnson *et al.*, 1995), it is obtained in a straightforward manner that Cases 1, 2 and 3 give on average 20%, 17% and 12% of missing responses, respectively. From samples of size of $n = 100$, estimates of β_0 and $g_0(\cdot)$ were obtained for the full data as well as for missing responses, where these were imputed using either the simple random imputation or the inverse probability weighted imputation methods. From 500 replications, the mean and standard deviation

of $A(\hat{\beta}, \beta_0)$, and the MSE of the estimated function $\hat{g}$ are calculated for the least squares (LS), the least absolute deviation (LAD), and the proposed rank (Rank) estimators. The angle between the estimated $\hat{\beta}$ and the true β_0 is calculated as

$$A(\hat{\beta}, \beta_0) = (180/\pi) \arccos(|\hat{\beta}^\tau \beta_0|),$$

with small values indicating good performance (Zeng *et al.*, 2012).

We checked the fidelity of Theorem 3 by running Royston’s multivariate normality test (Royston, 1992) for Model 3 under $N(0, 1)$ and t_3 . The simulation results are presented in Tables 1 – 4. We also report plots of $\hat{g}(\hat{\beta}^\tau \mathbf{x})$ versus $\hat{\beta}^\tau \mathbf{x}$ for missingness mechanism Case 2 in Figures 1 – 3.

For Model 1, the rank estimator performed better than the LS and LAD estimators in terms of estimating g_0 for all the cases considered including the full data (Table 1). A substantial difference is observed among the three methods when considering the contaminated normal error distribution. For the estimation of the index direction β_0 , as expected, the LS method gave better performance than the LAD and the rank methods for normal errors, while the reverse was observed for t_3 and contaminated normal errors, with the rank method showing best performance among the three methods. This was true for all missingness cases, both methods of imputation, and the full data. The performance of the estimated functions given in Figure 1 demonstrates this observation for Case 2 missingness. The LS estimate of the function is highly affected by aberrant observations in the t_3 and contaminated normal error cases often missing the underlying shape of the true function. As expected, Case 3 gave mean angles for the index direction as well as MSE for $g_0(\cdot)$ closest to the ones for the full data. This case was also where the performances of simple imputation and inverse probability weighted imputation were comparable. Simple imputation generally outperformed inverse probability weighted imputation except for the normal errors and t_3 (LS only).

For Models 2 and 3, the results of the simulation experiment are given in Table/Figure 2 and Table/Figure 3, respectively. Our observations regarding the performance of our proposed estimator remain similar to those for Model 1. In particular, the rank estimator gave much better performance than the LS and LAD estimators for estimating the index

direction β_0 as well as the function $g_0(\cdot)$ for the heavier tailed t_3 and the contaminated normal distributions. It can be observed in Figure 2 that the three methods have difficulty with dealing with the singularity in Model 2.

In Tables 1 – 3 we can see that the angle between the true direction and the estimated direction from the imputed data $A(\hat{\beta}_{\text{imp}}, \beta_0)$ is always larger than the angle between the true direction and the estimated direction from full data $A(\hat{\beta}_{\text{full}}, \beta_0)$. The quantity $\hat{\beta}_{\text{full}}$ can be regarded to be the best achievable for each method. Thus to evaluate the quality of our imputation strategy among the three methods, it is of interest to calculate the angle between the estimated direction from the imputed data and the estimated direction from full data $A(\hat{\beta}_{\text{imp}}, \hat{\beta}_{\text{full}})$. The average angles from LS, LAD and rank methods are plotted against each other (on the log scale for better visibility) in Figure 4. If the three methods provide similar performance in terms of imputation, then we expect the points to be on the 45° line. It is clear that the LS method does a slightly better job than the rank method and a much better job than the LAD for normal errors. The rank method provides better performance than the LS for a majority of the t_3 cases and an even a greater advantage for contaminated normal errors, especially for Model 1. The difference in how the rank and the LAD imputation methods perform with respect to LS is much more pronounced for t_3 errors with rank giving slightly better performance than LS but LAD giving slightly worse performance than LS. While loss of information is unavoidable when there are missing data, the spread away from the lower left corner in Figure 4 of the log angles indicates that there is room to improve the model fits by improving the imputation procedures. Finally, Table 4 confirms the asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ as described in Theorem 3. Specifically, we observe that while the proposed method gives p -values that are consistently greater than the significance level $\alpha = 0.05$ for almost all missingness cases under the considered model error distributions, the LS only passes the normality test for $\varepsilon \sim N(0, 1)$ as would be expected but fails to pass the normality test when $\varepsilon \sim t_3$; see Table 2. On the other hand, the LAD gives inconsistent p -values for all missingness cases under the two considered model error distributions; see Table 4.

Tab. 1: The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSEs $\times 100$ of $\hat{g}$ for Model 1

Dist	Imp	Case	LS			Rank			LAD		
			Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)
$N(0, 1)$	SI	1	7.239	3.593	17.248	9.227	5.622	15.304	13.318	11.073	35.728
		2	6.588	3.574	16.683	8.282	4.328	14.068	11.156	8.241	25.004
		3	6.761	3.567	16.063	8.841	5.727	14.304	11.391	7.168	28.319
	IP	1	7.376	3.659	15.592	9.822	7.001	17.708	13.978	11.939	36.436
		2	6.517	3.480	15.772	8.339	4.431	14.006	11.166	7.343	24.711
		3	6.934	3.706	14.546	9.013	6.348	16.053	12.649	11.050	31.788
	Full		5.562	2.634	12.027	7.113	3.180	9.789	8.916	5.770	20.402
	SI	1	18.212	17.726	62.719	15.079	14.356	42.354	18.183	15.673	62.806
		2	15.884	15.603	50.861	12.039	10.250	29.524	14.074	11.322	38.070
		3	16.689	16.848	59.727	13.653	13.170	36.336	17.187	16.053	53.025
t_3	IP	1	21.065	20.321	69.297	16.535	15.540	50.318	20.157	16.843	66.458
		2	17.163	17.831	54.515	13.140	12.269	28.530	15.430	13.111	44.659
		3	17.990	17.913	57.012	14.759	14.192	40.359	19.144	17.326	60.084
	Full		14.344	15.027	36.100	10.570	9.526	19.337	12.537	10.276	31.545
	SI	1	43.886	25.154	206.554	23.360	19.356	77.229	23.941	18.852	91.229
		2	40.443	25.520	186.119	17.448	14.033	58.560	21.166	17.503	70.294
		3	41.912	24.825	191.438	20.797	17.396	70.284	22.501	17.920	81.951
	IP	1	43.944	25.257	229.670	27.842	20.243	110.466	29.485	20.853	115.662
		2	42.158	25.946	199.163	20.159	15.478	77.480	22.005	15.945	79.456
		3	42.979	24.622	201.860	25.596	20.748	83.889	27.198	18.704	98.070
	Full		39.573	25.239	158.296	14.824	13.374	47.129	17.540	14.496	62.010

Tab. 2: The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSEs $\times 100$ of $\hat{g}$ for Model 2

Dist	Imp	Case	LS			Rank			LAD		
			Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)
$N(0, 1)$	SI	1	4.085	1.997	25.906	4.360	1.828	24.334	5.267	2.302	33.790
		2	3.933	1.854	24.792	4.238	1.779	23.674	5.131	2.191	32.537
		3	3.725	1.652	23.636	4.091	1.772	22.606	4.859	2.232	30.792
	IP	1	4.143	2.010	23.879	4.582	1.956	24.996	5.412	2.437	34.579
		2	3.977	1.883	23.205	4.442	1.869	24.197	5.226	2.311	32.788
		3	3.744	1.660	22.352	4.170	1.802	22.683	4.923	2.304	30.921
	Full		3.302	1.446	20.913	3.665	1.591	20.381	4.174	1.831	26.648
	SI	1	6.652	3.583	38.763	5.736	2.678	29.984	5.949	2.740	39.337
		2	6.540	3.444	37.173	5.542	2.574	28.263	5.893	2.607	37.278
		3	6.237	3.259	35.490	5.210	2.447	27.676	5.408	2.422	35.610
t_3	IP	1	6.963	3.722	37.708	6.217	2.928	31.941	6.368	2.889	42.022
		2	6.796	3.577	35.793	5.920	2.825	29.832	6.233	2.690	39.333
		3	6.322	3.367	34.430	5.430	2.599	28.732	5.552	2.542	37.101
	Full		5.478	2.768	29.833	4.443	1.968	23.697	4.789	2.166	30.788
SI	1	13.426	8.481	137.978	5.527	2.752	33.259	6.062	3.000	44.083	
	2	12.633	7.953	119.032	5.357	2.666	31.311	5.771	2.727	41.398	
	3	12.384	6.787	105.554	5.077	2.481	30.630	5.425	2.439	40.148	
CN	IP	1	14.046	8.691	167.394	6.915	3.361	44.764	7.660	3.724	53.416
		2	13.255	8.417	150.705	6.445	3.147	39.823	7.089	3.552	49.434
		3	12.702	7.144	107.177	5.703	2.748	35.024	6.045	3.112	43.198
	Full		10.909	5.515	70.910	4.314	1.902	25.515	4.742	2.023	32.585

Tab. 3: The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSE of $\hat{g}$ for Model 3

Dist	Imp	Case	LS			Rank			LAD		
			Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)
$N(0, 1)$	SI	1	3.464	1.488	24.012	3.780	1.660	11.591	4.149	1.855	11.715
		2	3.349	1.479	22.980	3.637	1.644	10.578	4.069	1.808	11.167
		3	3.216	1.375	21.325	3.501	1.569	9.711	3.838	1.656	9.889
	IP	1	3.568	1.561	19.724	3.865	1.752	11.313	4.400	1.978	12.243
		2	3.446	1.541	19.381	3.710	1.702	10.236	4.231	1.968	11.463
		3	3.231	1.386	18.896	3.539	1.586	9.426	3.896	1.684	10.068
	Full		2.752	1.229	17.277	3.011	1.372	8.074	3.415	1.592	8.555
	SI	1	5.016	2.652	34.110	4.122	1.931	15.299	4.574	2.088	15.753
		2	4.843	2.396	32.248	4.003	1.852	13.588	4.452	2.139	14.243
		3	4.808	2.356	30.194	3.860	1.758	13.107	4.285	1.913	13.404
	t_3	1	5.189	2.892	31.500	4.402	2.014	16.201	4.740	2.320	17.436
		2	5.020	2.683	29.459	4.167	1.970	14.244	4.560	2.301	15.628
		3	4.871	2.530	29.629	3.943	1.768	13.819	4.433	2.022	13.987
CN	Full		4.052	2.039	23.806	3.320	1.446	9.825	3.627	1.646	10.037
	SI	1	10.642	7.228	99.531	4.840	2.491	21.103	5.254	2.606	20.719
		2	10.196	6.898	91.059	4.682	2.416	19.832	4.981	2.542	18.58
		3	9.853	6.249	96.369	4.316	2.283	17.417	4.680	2.269	19.034
	IP	1	11.491	8.113	121.514	5.939	2.932	31.740	6.385	3.735	34.802
		2	10.811	7.330	88.125	5.596	2.766	26.920	5.863	3.110	28.044
		3	10.170	6.715	111.940	4.781	2.413	23.512	5.224	2.748	28.645
	Full		8.499	4.961	61.556	3.589	1.762	17.666	3.978	1.985	19.075

Tab. 4: P-values of the Multivariate Distribution Normality test of $\sqrt{n}(\hat{\beta} - \beta_0)$ for Model 3 under $N(0, 1)$ and t_3

			Model 3		
Dist	Imp	Case	LS	Rank	LAD
$N(0, 1)$	SI	1	0.049	0.210	0.004
		2	0.209	0.099	0.004
		3	0.173	0.307	0.109
	IP	1	0.007	0.199	0.007
		2	0.090	0.033	0.002
		3	0.072	0.446	0.459
	SI	1	0.000	0.112	0.082
		2	0.000	0.115	0.017
		3	0.000	0.395	0.891
t_3	IP	1	0.000	0.059	0.001
		2	0.000	0.014	0.000
		3	0.000	0.299	0.659

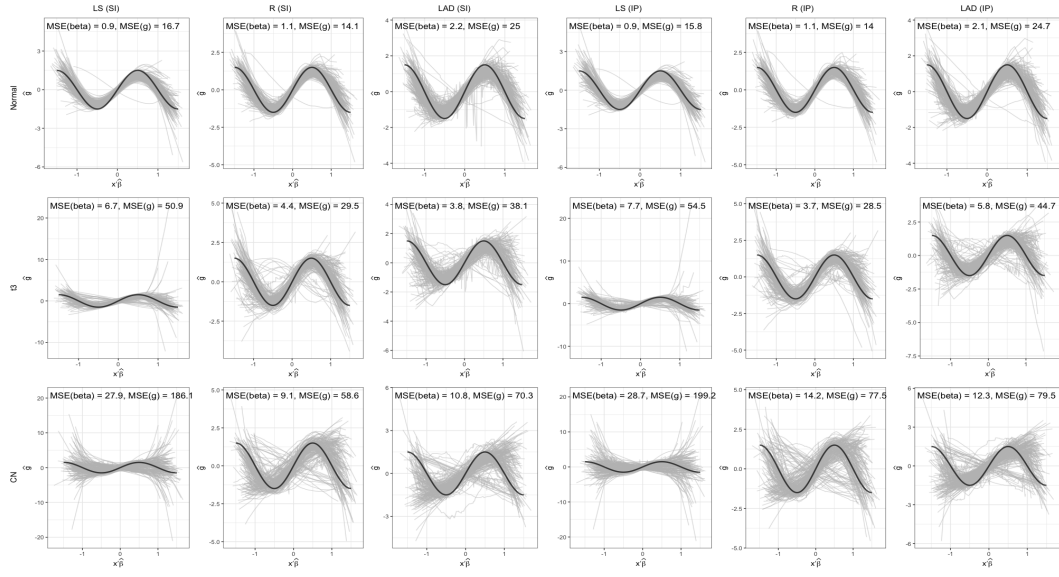


Fig. 1: Estimated functions for Model 1 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN

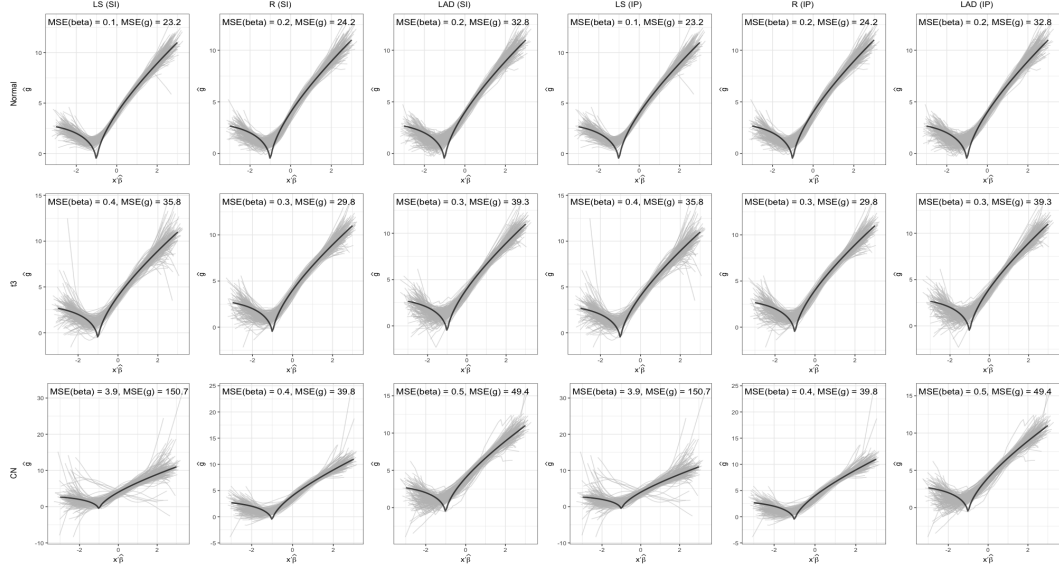


Fig. 2: Estimated functions for Model 2 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN

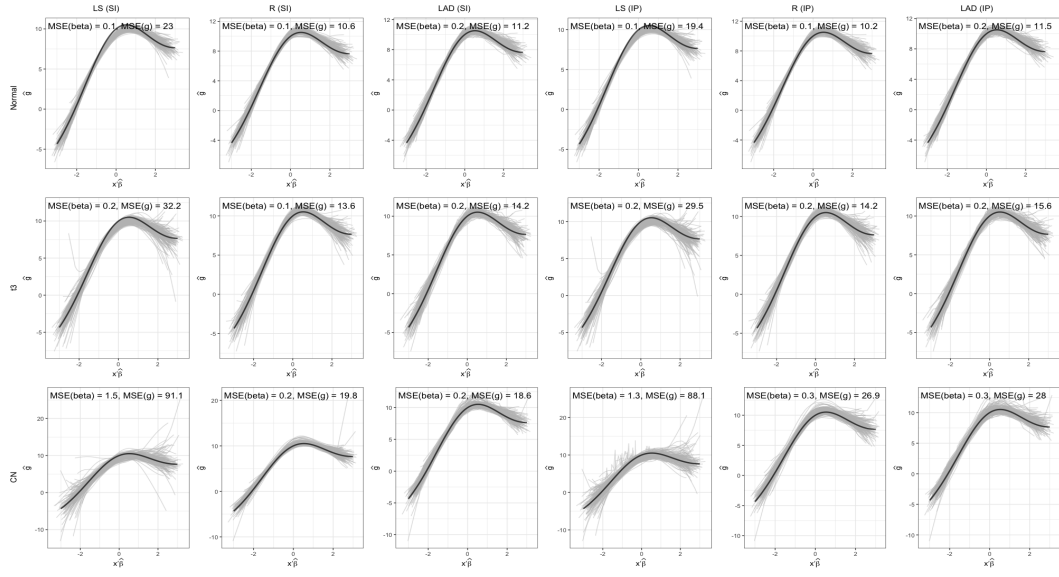


Fig. 3: Estimated functions for Model 3 under Case 2 and error distributions $N(0, 1)$, t_3 , and CN

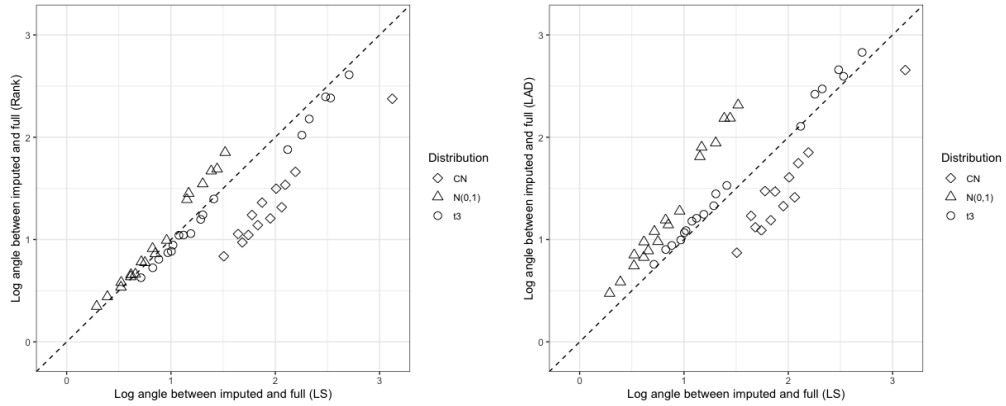


Fig. 4: Mean angle between full data and imputed data estimated single-index directions

Scenario 2 To investigate the impact of large proportions of missingness (50% – 70%), we consider Model 1 of (Liu *et al.*, 2013) under the following cases:

Case 4: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.3$.

Case 5: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.5 + 0.5|u(\mathbf{x}) - 0.5|$ if $|u(\mathbf{x}) - 0.5| \leq 0.2$, and 0.3 elsewhere.

Case 6: $\pi(\mathbf{x}) = P(\delta = 1|\mathbf{x}) = 0.5 - 0.5|u(\mathbf{x}) - 0.5|$ if $|u(\mathbf{x}) - 0.5| \leq 0.25$, and 0.4 elsewhere.

These cases give on average 70%, 50% and 60% of missing responses, respectively. As in scenario 1, using samples of size $n = 100$, $\hat{\beta}$ and $\hat{g}$ were estimated, where the missing responses were imputed under simple and inverse probability weighted imputations. Also, means and standard deviations of the angle between $\hat{\beta}$ and β_0 were calculated as well as mean square errors of $\hat{g}$. Moreover, to evaluate the gain in efficiency, the angle between the imputed estimate, $\hat{\beta}_I$, and the complete case estimate, $\hat{\beta}_{CC}$, as well as the L^2 -distance between the imputed estimated function $\hat{g}_I$ and the complete case estimated function $\hat{g}_{CC}$ were calculated. The results are displayed in Tables 5 and 6.

From these tables, we observe a similar trend as in our simulation scenario 1 with respect to the performance of the proposed method compared to the LS and LAD. Generally, the larger the proportion of missingness, the larger the angle between $\hat{\beta}$ and β_0 as well as the MSEs of $\hat{g}$; see Tables 5 and 6. As advocated by many authors, the complete case analysis (ignoring observations with missing responses) can reduce estimation efficiency and, more importantly, can lead to biased estimates of the model parameters (Little, 1992; Little & Rubin, 2002). From the observation that the angle between $\hat{\beta}_I$ and $\hat{\beta}_{CC}$ as well as the L^2 -distance between $\hat{g}_I$ and $\hat{g}_{CC}$ are not close to 0 (see Table 6), we surmise that the estimation using imputed responses and the estimation based on the complete case analysis lead to different estimators. This confirms that the imputation procedure has improved the efficiency of the considered estimators.

Tab. 5: The mean and SD of the angle $A(\hat{\beta}, \beta_0)$ and the MSEs $\times 100$ of $\hat{g}$ for Model 1

Imp	Dist	Case	LS			Rank			LAD		
			Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)	Mean	SD	MSE($\hat{g}$)
SI	$N(0, 1)$	4	26.614	21.734	86.194	29.316	21.445	102.183	31.78	20.469	120.506
		5	21.693	20.74	65.079	25.441	22	85.967	30.285	21.477	91.531
		6	22.933	20.146	75.04	24.959	18.775	88.986	29.055	19.053	103.902
	t_3	4	40.541	26.871	170.775	38.595	25.217	176.846	42.316	24.689	180.004
		5	41.12	27.695	151.107	39.315	26.17	151.931	41.944	25	147.344
		6	35.372	24.65	140.852	32.932	22.006	143.506	36.782	22.584	147.063
	CN	4	49.628	24.571	750.891	39.19	24.207	264.384	41.245	22.149	254.656
		5	52.676	25.321	477.526	44.261	25.619	209.809	46.084	26.048	197.133
		6	48.391	24.997	672.319	38.177	22.236	224.325	39.047	21.718	207.63
IP	$N(0, 1)$	4	33.278	22.383	124.711	31.979	21.505	128.471	34.132	20.592	137.671
		5	28.233	23.925	98.448	29.32	23.326	106.148	35.28	22.789	113.551
		6	28.312	22.837	93.746	28.15	20.517	104.968	31.632	19.329	113.23
	t_3	4	45.248	25.958	241.412	40.969	23.849	208.412	45.688	23.533	215.936
		5	47.224	26.446	203.869	42.963	25.222	194.83	45.833	23.488	192.277
		6	41.696	25.253	196.835	38.501	22.333	188.226	41.722	22.184	189.254
	CN	4	51.052	23.747	878.931	46.55	23.781	409.606	48.29	21.952	386.2
		5	52.948	23.46	607.509	50.272	23.61	352.584	51.727	22.152	318.987
		6	49.866	22.962	797.752	46.644	22.812	343.172	46.413	22.633	304.155

Tab. 6: Comparison of the complete case and imputed analyses

Imp	Dist	Case	LS		Rank		LAD	
			$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{g}_I - \hat{g}_{CC}\ ^2$	$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{g}_I - \hat{g}_{CC}\ ^2$	$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{g}_I - \hat{g}_{CC}\ ^2$
SI	$N(0, 1)$	4	7.065	142.993	8.754	206.603	11.603	188.217
		5	7.688	119.971	10.261	180.556	12.906	146.046
		6	8.322	145.933	8.616	206.174	11.989	182.737
	t_3	4	10.817	216.502	14.189	250.774	18.72	218.478
		5	12.918	182.667	15.692	199.452	19.18	168.14
		6	10.688	198.221	13.629	229.675	18.125	200.084
	CN	4	13.38	751.434	22.456	347.567	22.977	285.474
		5	15.705	593.799	25.423	272.638	26.684	239.279
		6	15.017	784.476	21.852	324.807	23.726	275.352
IP	$N(0, 1)$	4	22.148	168.953	15.404	219.034	19.011	202.011
		5	22.738	149.225	17.377	191.096	23.467	160.838
		6	17.991	172.675	14.902	216.154	18.529	189.768
	t_3	4	29.37	283.145	22.009	277.601	27.885	242.901
		5	33.081	221.852	26.263	227.491	31.093	198.439
		6	28.451	235.755	23.242	254.665	27.34	227.224
	CN	4	30.195	855.494	34.53	454.683	35.704	407.437
		5	31.581	728.358	38.706	412.247	42.486	359.665
		6	30.539	903.646	33.618	426.429	38.007	364.015

Real Data Application: CD4 and CD8 Data

Patients who test positive for the HIV antibody develop AIDS when HIV destroys their immune system. Immune system function is measured by the CD4 cell count and a count of less than 200 is said to indicate a case of AIDS. On the other hand, results from mathematical modeling and experimental evaluations have indicated that CD8 responses are important in the control of HIV infection. In this section, we present an analysis of the AIDS clinical trial group (ACTG 315) study Wu & Wu (2002) where CD4 and CD8 cell counts were measured after initiation of an anti-viral therapy. We will use the number

of days after initiation of treatment (x_1), the log of viral load (x_2), and its baseline value (x_3) as covariates. The dataset contains a total of 317 observations with about 20% of CD4 and CD8 cell counts missing. While the CD4 variable did not appear to have any outliers, CD8 had numerous suspected outliers (values over 2000). We will consider both CD4 and CD8 cell counts as response in two single-index regression models. We made the MAR assumption that the missingness of the cell counts depends only on the predictors in the data. Given the superior performance of the rank estimator as evaluated in the simulation studies, we decided to focus on the performance of the rank estimator as compared to the commonly used LS estimator. We performed 10-fold cross validation studies to evaluate the predictive performance of the regression modeling based on the least squares (LS) and the rank (Rank) methods.

The results of this cross-validation study as well as the estimated functions for the CD4 and CD8 responses are given in Figure 5. For the CD4 response, we found that $\hat{\beta}_{LS} = (0.209, -0.123, -0.970)$ and $\hat{\beta}_{Rank} = (0.138, -0.990, 0.016)$. These were close to those found from the complete case analysis:

$$\hat{\beta}_{LS}^{CC} = (0.209, -0.141, -0.967) \quad \text{and} \quad \hat{\beta}_{Rank}^{CC} = (0.152, -0.986, 0.074).$$

The estimated functions for LS and Rank were quite similar. Based on these estimated values and the estimated function given in Figure 5, we note that increasing days after treatment was associated with increase in CD4 counts. Increasing viral load was associated with decreasing CD4 counts for the rank estimate while the reduction in CD4 counts was associated with increases in baseline viral load for the LS estimate. The 10-fold CV results indicate that the LS method provided better prediction than the rank method for the CD4 count data. For the CD8 data, we found $\hat{\beta}_{LS} = (0.005, 0.715, -0.699)$ and $\hat{\beta}_{Rank} = (0.059, -0.845, 0.531)$. These were close to those found from the complete case analysis $\hat{\beta}_{LS}^{CC} = (0.004, 0.720, -0.694)$ and $\hat{\beta}_{Rank}^{CC} = (0.144, -0.835, 0.531)$. The estimated functions were vastly different for the two cases. Both the LS and rank estimates indicated that decrease in CD8 counts was associated with increasing viral load and decreasing baseline viral load. The results suggest that the contribution of the two variables is similar in the LS case; however, the rank method indicates greater association of viral

load to CD8 counts. The difference is likely due to the influence of the outliers in CD8 count. This was corroborated by the 10-fold CV results which indicated that the rank method provided vastly superior prediction to the LS method for the CD8 count data.

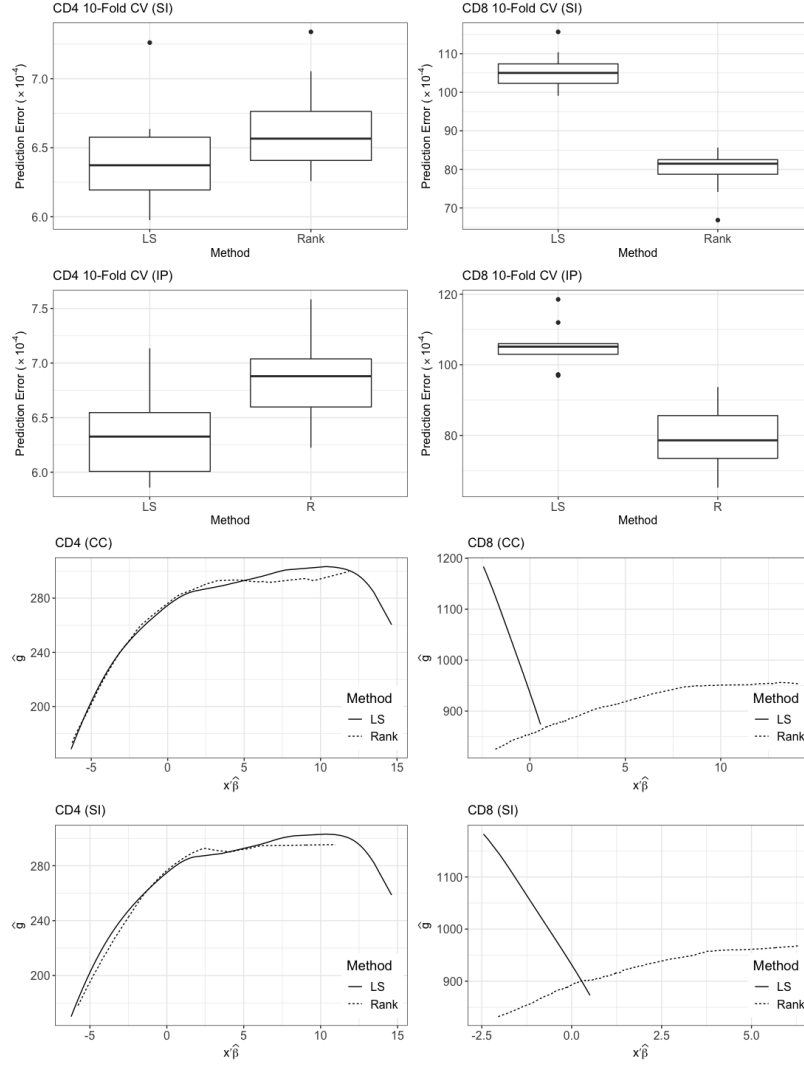


Fig. 5: Ten-fold cross-validation errors and estimated functions for the CD4 and CD8 data

2.6 Conclusion

This chapter provides a robust and efficient method for making inference about β_0 in a single-index regression model when some responses are missing at random. The proposed rank-based procedure can be used as an alternative tool when one is dealing with contaminated heavy-tailed error distributions, or when there are outliers in the response space. Asymptotic properties of the proposed estimators (consistency and asymptotic normal distribution) were established under mild regularity conditions. Monte Carlo simulation studies and a real data application have shown that the proposed estimators are robust and more efficient compared with the LS and LAD estimators when dealing with contaminated and heavy-tailed model error distributions, and/or when there are gross outliers in the response space. It is worth pointing out that there is gain in efficiency when imputing for missing values. However, there is still room for improving the quality of imputation methods used.

3 Estimation in SIVCM with MAR Responses

3.1 Introduction

In this chapter, we extend the approach discussed in the previous chapter to model (1.1) with MAR responses. We derive rank-based estimators of the model parameters. Under some mild regularity conditions, asymptotic properties of such estimators are established. To demonstrate the robustness and efficiency of the proposed estimators, Monte Carlo simulation experiments are carried out and show that under heavy-tailed and contaminated model error distributions, the rank-based method performs better in estimating the model parameters compared to its LS and LAD counterparts. An illustrative real data application supports the simulation outcomes.

3.2 3-Step estimation with MAR responses

Recall that setting $G(u) = (g_0(u), g_1(u), \dots, g_q(u))^T$, the SIVCM in (1.1) can be rewritten as

$$\mathbf{y} = \{G(\mathbf{x}^T \boldsymbol{\beta}_0)\}^T \mathbf{z} + \boldsymbol{\varepsilon},$$

where $\mathbf{x} = (x_1, \dots, x_p)^T$, $\mathbf{z} = (z_0, z_1, \dots, z_q)^T$ with $z_0 = 1$, $\mathbf{y} = (y_1, \dots, y_n)^T$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$. Considering the incomplete random sample given by $\{(y_i, \mathbf{x}_i, \mathbf{z}_i, \delta_i), i = 1, \dots, n\}$ and premultiply Equation (1.1) by δ_i , we get

$$\delta_i y_i = \delta_i \{G(\mathbf{x}_i^T \boldsymbol{\beta}_0)\}^T \mathbf{z}_i + \delta_i \varepsilon_i.$$

Under MAR, since δ is independent of $\mathbf{y}$, δ is also independent of $\boldsymbol{\varepsilon}$. Hence, $E(\delta \boldsymbol{\varepsilon} | (\mathbf{x}, \mathbf{z})) = E(\delta | (\mathbf{x}, \mathbf{z})) E(\boldsymbol{\varepsilon} | (\mathbf{x}, \mathbf{z})) = 0$.

Step 1: Complete case analysis

From the above model, G and $\boldsymbol{\beta}_0$ being unknown, they need to be estimated. As in the previous Chapter, to estimate G , consider the local linear approximation approach. For

fixed k , under the smoothness assumption on g_k , $g_k(u) \approx g_k(t) + g'_k(t)(u - t)$, $t \in \mathcal{F}$. Now, approximate $g_k(\mathbf{x}^\tau \boldsymbol{\beta})$ as $g_k(\mathbf{x}^\tau \boldsymbol{\beta}) \approx a_k + b_k \mathbf{x}_0^\tau \boldsymbol{\beta}$, where $a_k = g_k(\mathbf{x}^\tau \boldsymbol{\beta})$, $b_k = g'_k(\mathbf{x}^\tau \boldsymbol{\beta})$ and $\mathbf{x}_0 = \mathbf{x} - \mathbf{v}$, for any $\mathbf{v}$ in the neighborhood of $\mathbf{x}$. From this, as in Sun *et al.* (2019), the complete case rank-based estimators of $\boldsymbol{\beta}_0$, a_k and b_k are obtained as $(\tilde{\boldsymbol{\beta}}_n^c, \tilde{a}_k, \tilde{b}_k) = \underset{\boldsymbol{\beta}, a_k, b_k}{\text{Argmin}} L_n(\boldsymbol{\beta}, a_k, b_k)$, with

$$L_n(\boldsymbol{\beta}, a_k, b_k) = \frac{1}{n} \sum_{i=1}^n w_i \varphi \left(\frac{R(e_i(\boldsymbol{\beta}))}{n+1} \right) e_i(\boldsymbol{\beta}),$$

where $e_i(\boldsymbol{\beta}) = \delta_i \varepsilon_i = \delta_i y_i - \sum_{k=0}^q \delta_i (a_k + b_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki}$, $\varphi : (0, 1) \rightarrow \mathbb{R}$ is a bounded nondecreasing score function, $R(t) = \sum_{s=1}^n I(e_s(\boldsymbol{\beta}) \leq t)$, $\mathbf{x}_{0i} = \mathbf{x}_i - \mathbf{v}$ and $w_i = K_b(\mathbf{x}_{0i}^\tau \tilde{\boldsymbol{\beta}}) / \sum_{s=1}^n K_b(\mathbf{x}_{0s}^\tau \tilde{\boldsymbol{\beta}})$, with $\tilde{\boldsymbol{\beta}}$, some initial estimator and $K_b(t) = K(t/b)$, where K is a univariate kernel function with bandwidth b .

Step 2: Imputation of y -missing values

Once again, to take into account the missing information, we consider imputing the missing responses as before using the simple random imputation ($j = 1$) and the marginal inverse weighting probability ($j = 2$). To do so, setting $\tilde{e}_i = \sum_{k=0}^q (\tilde{a}_k + \tilde{b}_k \mathbf{x}_{0i}^\tau \tilde{\boldsymbol{\beta}}_n^c) z_{ki}$, the imputed responses are defined as

$$\mu_{ij} = \begin{cases} \delta_i y_i + (1 - \delta_i) \tilde{e}_i, & \text{if } j = 1; \\ \frac{\delta_i}{\pi(\mathbf{x}_i, \mathbf{z}_i)} y_i + \left(1 - \frac{\delta_i}{\pi(\mathbf{x}_i, \mathbf{z}_i)}\right) \tilde{e}_i, & \text{if } j = 2. \end{cases}$$

In this definition, if the functional of $\pi(\mathbf{x}, \mathbf{z}) = P(\delta = 1 | \mathbf{x}, \mathbf{z})$ is unknown, it can be estimated nonparametrically as follows: set $\mathbf{u} = (\mathbf{x}, \mathbf{z})$, and consider $\hat{\pi}(\mathbf{u}) = \sum_{j=1}^n \omega_{nj}(\mathbf{u}) \delta_j$, where

$$\omega_{nj}(\mathbf{u}) = \frac{\Omega_b(\mathbf{u} - \mathbf{u}_j)}{\sum_{r=1}^n \Omega_b(\mathbf{u} - \mathbf{u}_r)} \quad \text{and} \quad \Omega_b(\mathbf{t}) = \Omega(\mathbf{t}/b^{p+q}),$$

with $\Omega : \mathbb{R}^{p+q} \rightarrow \mathbb{R}$, a kernel function with bandwidth b . Following Einmahl & Mason (2005), it is obtained that $\hat{\pi}(\mathbf{u}) \rightarrow \pi(\mathbf{u})$ *a.s.* as $n \rightarrow \infty$. From this, the imputed responses

become

$$\tilde{\mu}_{ij} = \begin{cases} \delta_i y_i + (1 - \delta_i) \tilde{e}_i, & \text{if } j = 1; \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} y_i + \left(1 - \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)}\right) \tilde{e}_i, & \text{if } j = 2. \end{cases}$$

As pointed out in the previous chapter, in the estimation of $\pi(\mathbf{u})$, one may face the “curse of dimensionality” issue that occurs in multivariate nonparametric kernel estimation. One way to avoid this issue is by assuming that for $\mathbf{u} = (u_1, \dots, u_n)$, the u_i , $i = 1, \dots, n$ are independent and considering $\Omega_b(\mathbf{u}) = \prod_{i=1}^n K_b(u_i)$. In case the u_i are correlated, one may first consider eliminating multicollinearity by applying usual techniques and then, consider the suggested approach. Other ways are those proposed by Nagler & Czado (2016) and Guo *et al.* (2018) as discussed earlier.

Step 3: Re-estimation

Now, setting $\nu_{ij}(\boldsymbol{\beta}) = \tilde{\mu}_{ij} - \sum_{k=0}^q (a_k + b_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki}$, the updated rank-based objective function is given by

$$L_{jn}(\boldsymbol{\beta}, a_k, b_k) = \frac{1}{n} \sum_{i=1}^n w_i \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) \nu_{ij}(\boldsymbol{\beta}), \quad (3.1)$$

where $w_i \equiv w_i(\tilde{\boldsymbol{\beta}}) = K_b(\mathbf{x}_{0i}^\tau \tilde{\boldsymbol{\beta}}) / \sum_{r=1}^n K_b(\mathbf{x}_{0r}^\tau \tilde{\boldsymbol{\beta}})$, $\tilde{\boldsymbol{\beta}}$ could be taken to be $\tilde{\boldsymbol{\beta}}_n^c$. To this end, the rank-based estimators of $\boldsymbol{\beta}_0$, a_k and b_k are obtained as

$$(\hat{\boldsymbol{\beta}}_{jn}, \hat{a}_{jk}, \hat{b}_{jk}) = \underset{\boldsymbol{\beta}, a_k, b_k}{\text{Argmin}} L_{jn}(\boldsymbol{\beta}, a_k, b_k). \quad (3.2)$$

3.3 Computational Algorithm

For computational purposes, in the definition of $\mathbf{x}_0$, $\mathbf{v}$ needs to be specified. Set $\mathbf{v} = \mathbf{x}_l$ and $\mathbf{x}_{il} = \mathbf{x}_i - \mathbf{x}_l$. For $j = 1, 2$, define

$$L_{jn}^l(\boldsymbol{\beta}, a_{jk}, b_{jk}) = \frac{1}{n} \sum_{i=1}^n w_{il} \varphi \left(\frac{R(\xi_{il}^j(\boldsymbol{\beta}))}{n+1} \right) \xi_{il}^j(\boldsymbol{\beta}), \quad (3.3)$$

with $\xi_{il}^j(\boldsymbol{\beta}) = \tilde{\mu}_{ij} - \sum_{k=0}^q (a_k + b_k \mathbf{x}_{il}^\tau \boldsymbol{\beta}) \mathbf{z}_{ki}$ and $w_{il} \equiv w_{il}(\tilde{\boldsymbol{\beta}}) = K_b(\mathbf{x}_{il}^\tau \tilde{\boldsymbol{\beta}}) / \sum_{t=1}^n \sum_{s=1}^n K_b(\mathbf{x}_{ts}^\tau \tilde{\boldsymbol{\beta}})$.

To this end, the computations go as follows

Step 1: For $j = 1, 2$, consider $\hat{\boldsymbol{\beta}}_{jn}^{(0)} = \tilde{\boldsymbol{\beta}}_{jn}^c$ the initial estimator of $\boldsymbol{\beta}_0$. Set $r = 1$

Step 2: For each fixed l , compute $(\tilde{a}_{jk}^{l(0)}, \tilde{b}_{jk}^{l(0)})$ as $(\tilde{a}_{jk}^{l(0)}, \tilde{b}_{jk}^{l(0)}) = \underset{a_k, b_k}{\text{Argmin}} L_{jn}^l(\hat{\boldsymbol{\beta}}_{jn}^{(0)}, a_k, b_k)$.

Step 3: Compute $\hat{\boldsymbol{\beta}}_{jn}^{(r)}$ as $\hat{\boldsymbol{\beta}}_{jn}^{(r)} = \underset{\boldsymbol{\beta}}{\text{Argmin}} \sum_{l=1}^n L_{jn}^l(\boldsymbol{\beta}, \tilde{a}_{jk}^{l(r-1)}, \tilde{b}_{jk}^{l(r-1)})$.

Step 4: Compute $(\tilde{a}_{jk}^{l(r)}, \tilde{b}_{jk}^{l(r)})$ as $(\tilde{a}_{jk}^{l(r)}, \tilde{b}_{jk}^{l(r)}) = \underset{a_k, b_k}{\text{Argmin}} L_{jn}^l(\hat{\boldsymbol{\beta}}_{jn}^{(r)}, a_k, b_k)$ and

$$\tilde{a}_{jk}^{(r)} = \frac{1}{n} \sum_{l=1}^n \tilde{a}_{jk}^{l(r)}, \quad \tilde{b}_{jk}^{(r)} = \frac{1}{n} \sum_{l=1}^n \tilde{b}_{jk}^{l(r)}.$$

Step 5: If

$$\frac{|\tilde{b}_{jk}^{(r)} - \tilde{b}_{jk}^{(r-1)}|}{|\tilde{b}_{jk}^{(r-1)}|} \leq \delta_1, \quad \frac{|\tilde{a}_{jk}^{(r)} - \tilde{a}_{jk}^{(r-1)}|}{|\tilde{a}_{jk}^{(r-1)}|} \leq \delta_2 \quad \text{and} \quad \frac{\|\hat{\boldsymbol{\beta}}_{jn}^{(r)} - \hat{\boldsymbol{\beta}}_{jn}^{(r-1)}\|}{\|\hat{\boldsymbol{\beta}}_{jn}^{(r-1)}\|} \leq \delta_3,$$

where $\delta_1, \delta_2, \delta_3$ are small positive constants giving relative tolerance limits, then take $\hat{\boldsymbol{\beta}}_{jn} = \hat{\boldsymbol{\beta}}_{jn}^{(r)}$, $\tilde{a}_{jk} = \tilde{a}_{jk}^{(r)}$, $\tilde{b}_{jk} = \tilde{b}_{jk}^{(r)}$ and STOP. Otherwise, let $r = r + 1$ and go back to **Step 3**.

In our computations, we took $\delta_1 = \delta_2 = \delta_3 = 10^{-4}$. Thus, for fixed n , we end up with $\hat{\boldsymbol{\beta}}_{jn} = \lim_{r \rightarrow \infty} \hat{\boldsymbol{\beta}}_{jn}^{(r)}$. It is worth pointing out that at each step of the algorithm described above, the bandwidth h has to be selected and this is done by selecting h as $\hat{h}_j^{(r)} = \underset{h}{\text{Argmin}} L_{jn}(\tilde{a}_{jk}^{(r)}, \tilde{b}_{jk}^{(r)}, \hat{\boldsymbol{\beta}}_{jn}^{(r)}, h)$, for $r = 0, 1, 2, \dots$

Now, set $\hat{a}_k = \hat{g}_k(t)$, $\hat{b}_k = \hat{g}'_k(t)$ for any $t \in \mathcal{T} := \{t : t = \mathbf{x}^\tau \boldsymbol{\beta}\}$ and, $\tilde{\nu}_{ij}(\boldsymbol{\beta}) = \tilde{\mu}_{ij} - \sum_{k=0}^q (\hat{a}_k - \hat{b}_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki}$. Define $Q_{jn}(\boldsymbol{\beta})$ as

$$Q_{jn}(\boldsymbol{\beta}) \equiv Q_{jn}(\boldsymbol{\beta}, \hat{a}_k, \hat{b}_k) = \frac{1}{n} \sum_{i=1}^n w_i \varphi \left(\frac{R(\tilde{\nu}_{ji}(\boldsymbol{\beta}))}{n+1} \right) \tilde{\nu}_{ij}(\boldsymbol{\beta}).$$

3.4 Asymptotic properties

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. To establish the asymptotic properties of the proposed rank based estimators, consider the following assumptions:

- (I₁) φ is a nondecreasing, bounded and twice continuously differentiable score function with bounded derivatives, defined on $(0, 1)$, and assume that φ can be standardized as

$$\int_0^1 \varphi(u) du = 0 \quad \text{and} \quad \int_0^1 \varphi^2(u) du = 1.$$

- (I₂) $\varepsilon_i, i = 1, \dots, n$, are i.i.d. continuous random errors with common distribution F , finite Fisher information. Its probability density function f , where $f(t) = dF(t)/dt$, has almost surely bounded derivatives.

- (I₃) For $k = 0, \dots, q$, $g_k(\cdot)$ is a three times continuously differentiable function defined on $\mathcal{F}$ with bounded derivatives. For fixed k , there exists an arbitrary positive function $J_k(\cdot)$, independent of β such that $\|g_k^{(r)}(\mathbf{x}^\tau \beta)\| \leq J_k(\mathbf{x})$, with $E[J_k^\alpha(\mathbf{x})] < \infty$, for some $\alpha \geq 1$ and $r = 0, 1, 2, 3$.

- (I₄) Letting $m(\cdot)$ be the density function of $\mathbf{x}^\tau \beta_0$, we assume that $m(\cdot)$ have bounded continuous derivatives up to order 2 and $\inf_{t \in \mathcal{F}} m(t) > \gamma$, with $\gamma > 0$. Also, there exists c_0 such that $0 < c_0 \leq \pi(\mathbf{u}) < 1$, for all $\mathbf{u}$.

- (I₅) $\Omega(\cdot)$ and $K(\cdot)$ are respectively multivariate and univariate kernel functions with compact support and bandwidth $b \equiv b_n$. The bandwidth $b_n \sim n^{-\delta}$ with $1/6 \leq \delta \leq 1/4$ and satisfies $nb_n^{p+q}/\log n \rightarrow \infty$ for Ω and $nb_n/\log n \rightarrow \infty$ for K , as $n \rightarrow \infty$.

- (I₆) $\sup_{\mathbf{u}} E[|\varepsilon|^r | \mathbf{u}] < \infty$, where r is the order of smoothness of $g_k(\cdot)$. Also, $E[\exp\{\lambda \|\mathbf{x}\|\}] < \infty$ and $E[\exp\{\lambda \|\mathbf{z}\|\}] < \infty$ for some $\lambda > 0$.

- (I₇) $\beta_0 \in \text{Int}(\mathcal{B})$ with $\|\beta_0\| = 1$, and for $j = 1, 2$, $(\beta_0, G) = \underset{\beta \in \mathcal{B}, a_k, b_k}{\text{Argmin}} \lim_{n \rightarrow \infty} E[L_{jn}(\beta, a_k, b_k)]$.

Remark 3.1. Assumptions (I₁) and (I₂) are common in the framework of rank-based estimation; see Hettmansperger & McKean (2011). The boundedness of the score function is necessary to ensure the consistency of the proposed estimator for the index coefficient

and, provides robustness at the expense of efficiency. For example, for normal data, the asymptotically efficient score function is the normal quantile function, which is unbounded. However, using just linear scores (bounded), the rank method still maintains a fairly high asymptotic relative efficiency (ARE) compared to maximum likelihood estimation in the normal case. For linear models, this ARE is approximately 95.5% for normal data. Assumptions $(I_3) - (I_6)$ on the other hand, are common assumptions for estimation problems involving single-index models and ensure the strong consistency of the estimator of $G(\cdot)$; see Xia (2006), Xia et al. (2007), Hansen (2008) and Gu & Yang (2015), just to cite a few. (I_7) is the identifiability assumption, which ensures the distinction of the proposed estimators to the true parameters. This assumption together with the previous assumptions, ensure the strong consistency of the index estimator (Bindele, 2017) that is established in Theorem 5.

Moreover, under $(I_1) - (I_7)$, and following Delecroix et al. (2006), it can be obtained that for each k , $\hat{a}_k \rightarrow a_k$ and $\hat{b}_k \rightarrow b_k$ as $n \rightarrow \infty$, uniformly. From this, it can be shown in a straightforward manner that $\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta}) \rightarrow 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$, which leads to the following theorem:

Theorem 4. Under $(I_1) - (I_6)$ defined below, we have $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} |Q_{jn}(\boldsymbol{\beta}) - L_{jn}(\boldsymbol{\beta})| = 0$ a.s.

From this asymptotic equivalence result, the strong consistency of $\hat{\boldsymbol{\beta}}_{jn}$ can be established using $L_{jn}(\boldsymbol{\beta}) \equiv L_{jn}(\boldsymbol{\beta}, a_k, b_k)$.

Consistency of $\hat{\boldsymbol{\beta}}_{jn}$

For $j = 1, 2$, since ε_i are i.i.d, $\nu_{ij}(\boldsymbol{\beta}_0)$ are also i.i.d. with common distribution F_j , and for $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $\nu_{ij}(\boldsymbol{\beta})$ are independent with distribution F_{ij} . From this, the following theorem gives the strong consistency of $\hat{\boldsymbol{\beta}}_{jn}$.

Theorem 5. Under $(I_1) - (I_7)$, for $\mathbf{x} \in \Gamma := \{\mathbf{x} : m(\mathbf{x}^\tau \boldsymbol{\beta}_0) \geq c\}$, with $c \geq c_0 > 0$ and $j = 1, 2$, $\hat{\boldsymbol{\beta}}_{jn} \rightarrow \boldsymbol{\beta}_0$ a.s., as $n \rightarrow \infty$.

The proof of this theorem can be constructed along the lines as that of Theorem 2. So, for sake of brevity, the proof is omitted here.

Asymptotic Normality of $\hat{\beta}_{jn}$

For $\beta = (\beta_1, \dots, \beta_p)$, let $\nabla_\beta = (\partial/\partial\beta_i)_{i=1,\dots,p}$ and $\nabla_\beta^2 = (\frac{\partial^2}{\partial\beta_i\partial\beta_k})_{ik}$ for $i, k = 1, \dots, p$, denote the gradient and Hessian operators, respectively. Also, consider the following notations

$$\begin{aligned}\kappa_j(\mathbf{u}) &= E[\varphi(F_j(\nu_j(\beta_0)))|\mathbf{u}], \quad \omega_j(\mathbf{u}) = E[f_j(\nu_j(\beta_0))\varphi'(F_j(\nu_j(\beta_0)))|\mathbf{u}], \\ \pi_j(\mathbf{u}) &= E\{\varphi^2(F_j(\nu_{ij}(\beta_0)))|\mathbf{u}\}, \quad \mathbf{V}_j = E[w^2(\{G'(\mathbf{x}^\tau\beta_0)\mathbf{z}\}^2\mathbf{x}\mathbf{x}^\tau\pi_j(\mathbf{u}))], \quad \text{and} \\ \Sigma &= E[w^2(\{\nabla_{\beta_0}G(\mathbf{x}^\tau\beta_0)\}^\tau\mathbf{z})^2\mathbf{x}\mathbf{x}^\tau].\end{aligned}$$

It should be noted that $R(\nu_{ij}(\beta))$ is a piecewise function of the residuals that has a finite number of non-differentiability points. Thus, under the smoothness assumption on φ and g_k ; $k = 1, \dots, p$, $L_{jn}(\beta)$ and $Q_{jn}(\beta)$ are weakly differentiable. Set $\tilde{S}_{jn}(\beta) = -\nabla_\beta Q_{jn}(\beta)$ and $S_{jn}(\beta) = -\nabla_\beta L_{jn}(\beta)$. By definition, $\hat{\beta}_{jn}$ is solution of the equation $\tilde{S}_{jn}(\beta) = 0$. More precisely, $\tilde{S}_{jn}(\beta)$ can be written as

$$\tilde{S}_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i (\{\nabla_\beta \hat{G}(\mathbf{x}_i^\tau \beta)\}^\tau \mathbf{z}_i) \varphi(R(\tilde{\nu}_{ij}(\beta)))/(n+1).$$

Similarly, $S_{jn}(\beta)$ is given by

$$S_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i (\{\nabla_\beta G(\mathbf{x}_i^\tau \beta)\}^\tau \mathbf{z}_i) \varphi(R(\nu_{ij}(\beta)))/(n+1).$$

From this, define $T_{jn}(\beta)$ as

$$T_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i (\{\nabla_\beta G(\mathbf{x}_i^\tau \beta)\}^\tau \mathbf{z}_i) \varphi(F_{ij}(\nu_{ij}(\beta))).$$

By Lemma 6.6-iii) given in the appendix, we have $\sqrt{n}(S_{jn}(\beta_0) - T_{jn}(\beta_0)) \xrightarrow{P} 0$, which implies that $\sqrt{n}S_{jn}(\beta_0)$ and $\sqrt{n}T_{jn}(\beta_0)$ have the same asymptotic distribution (Hettmansperger & McKean, 2011). Moreover, from the same lemma, with probability 1, $S_{jn}(\beta) = T_{jn}(\beta) + o(1/n)$. A Taylor expansion of order 2 of $T_{jn}(\beta)$ around β_0 gives

$$T_{jn}(\beta) = T_{jn}(\beta_0) + (\beta - \beta_0)^\tau \nabla_\beta T_{jn}(\beta_0) + \frac{1}{2}(\beta - \beta_0)^\tau \nabla_\beta^2 T_{jn}(\zeta)(\beta - \beta_0), \quad (3.4)$$

where ζ belongs to the line segment joining β_0 and β . From this, we have the following theorem:

Theorem 6. *Under assumptions $(I_1) - (I_7)$ and for any $\mathbf{x} \in \Gamma$, the following hold:*

a. $\nabla_{\beta} T_{jn}(\beta_0) \rightarrow \mathbf{W}_j$ a.s., where

$$\mathbf{W}_j = -E[w^2\{G'(\mathbf{x}^\tau \beta_0)\}^\tau \mathbf{z})^2 \mathbf{x} \mathbf{x}^\tau \omega_j(\mathbf{u})] + E[w^2(\{G''(\mathbf{x}^\tau \beta_0)\}^\tau \mathbf{z}) \mathbf{x} \mathbf{x}^\tau \kappa_j(\mathbf{u})].$$

b. $\nabla_{\beta}^2 T_{jn}(\zeta_{jn})$ is almost surely bounded, where ζ_{jn} lies on the line segment joining β_0 and $\hat{\beta}_{jn}$.

This theorem together with the strong consistency of $\hat{\beta}_{jn}$, give the following the asymptotic normality theorem:

Theorem 7. *Under $(I_1) - (I_7)$, we have $\sqrt{n}(\hat{\beta}_{jn} - \beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-)$, where $\mathbf{W}_j^-$ is the Moore Penrose inverse of $\mathbf{W}_j$.*

The proof of this theorem is obtained following that of Theorem 3, taking into account Theorem 6 and Lemma 6.6 given in the Appendix.

Remark 3.2. *Once again, note that for further statistical inference about β_0 , the estimation of the estimator's covariance matrix $\mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-$ is necessary. Following Brunner & Denker (1994), consistent sandwich type estimators of $\mathbf{V}_j$ and $\mathbf{W}_j$ can be obtained. Readers seeking for details on how to obtain these estimators, are referred to the aforementioned paper. Also, if we assume that ε is independent of $\mathbf{u}$, $\nu_{ij}(\beta_0)$ will be independent of $\mathbf{u}_i$. Thus, by assumption (I_1) , $\mathbf{V}_j = \Sigma$ and $\mathbf{W}_j$ can be expressed using the rank scale parameter as $\mathbf{W}_j = \gamma_{j\varphi}^{-1} \Sigma$ similar to the linear model case, where*

$$\gamma_{j\varphi}^{-1} = \int_0^1 \varphi(u) \varphi_{f_j}(u) du \quad \text{with} \quad \varphi_{f_j}(u) = \frac{f'_j(F_j^{-1}(u))}{f_j(F_j^{-1}(u))}$$

and f_j being the probability density function of $\nu_{ij}(\beta_0) \equiv \nu_j(\beta_0)$. To see the latter, note that under independence we have

$$E[w^2(\{G'(\mathbf{x}^\tau \beta_0)\}^\tau \mathbf{z})^2 \mathbf{x} \mathbf{x}^\tau E\{f_j(\nu_j(\beta_0))\varphi'(F_j(\nu_j(\beta_0)))|\mathbf{u}\}] = \Sigma \times E\{f_j(\nu_j(\beta_0))\varphi'(F_j(\nu_j(\beta_0)))\}.$$

But

$$E[f_j(\nu_j(\boldsymbol{\beta}_0))\varphi'(F_j(\nu_j(\boldsymbol{\beta}_0)))] = \int_{-\infty}^{\infty} f_j(\nu)\varphi'(F_j(\nu))dF_j(\nu) = - \int_{-\infty}^{\infty} f'_j(\nu)\varphi(F_j(\nu))d\nu,$$

from the integration by parts, since $f_j(\nu)\varphi(F_j(\nu)) \rightarrow 0$ as $\nu \rightarrow \pm\infty$. Now, putting $u = F_j(\nu)$, we have

$$\int_{-\infty}^{\infty} f'_j(\nu)\varphi(F_j(\nu))d\nu = - \int_0^1 \varphi(u)\varphi_{f_j}(u)du = -\gamma_{j\varphi}^{-1}.$$

Similarly,

$$E[w^2\mathbf{x}\mathbf{x}^\tau(\{G''(\mathbf{x}^\tau\boldsymbol{\beta}_0)\}^\tau\mathbf{z})\varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))|\mathbf{u}] = E\{w^2\mathbf{x}\mathbf{x}^\tau(\{G''(\mathbf{x}^\tau\boldsymbol{\beta}_0)\}^\tau\mathbf{z})E\{\varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))\}.$$

Note that by assumption (I_1) , $E[\varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))] = \int_0^1 \varphi(t)dt = 0$. Thus, $\mathbf{W}_j = \gamma_{j\varphi}^{-1}\boldsymbol{\Sigma}$.

On the other hand, $\tilde{S}_{jn}(\hat{\boldsymbol{\beta}}_{jn}) = 0$ and by Lemma 6.6, $T_{nj}(\boldsymbol{\beta}) = \tilde{S}_{jn}(\boldsymbol{\beta}) + o_p(1/n)$. This together with equation (3.4) and Theorem 6, give

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_0) = -\mathbf{W}_j^- \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \mathbf{x}_i (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i) \varphi(F_j(\nu_j(\boldsymbol{\beta}_0))) + o_p(1).$$

From Corollary 3.5.2 of Hettmansperger & McKean (2011), the influence function of $\hat{\boldsymbol{\beta}}_{jn}$ at a given data point $(\mathbf{x}_0, \mathbf{z}_0, y_0)$ is obtained as

$$\Omega(\mathbf{x}_0, \mathbf{z}_0, y_0; \hat{\boldsymbol{\beta}}_{jn}) = -\mathbf{W}_j^- w(\boldsymbol{\beta}_0) \mathbf{x}_0 (\{G'(\mathbf{x}_0^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_0) \varphi(F(y_0 - \{G(\mathbf{x}_0^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_0)),$$

where F is the distribution function of $y_0 - \{G(\mathbf{x}_0^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_0$. Since φ is bounded by assumption (I_1) , it is not hard to see that $\Omega(\mathbf{x}_0, \mathbf{z}_0, y_0; \hat{\boldsymbol{\beta}}_{jn})$ is bounded in the direction of y_0 but unbounded in the direction of $(\mathbf{x}_0, \mathbf{z}_0)$. This entails that, while the proposed method is robust in the response space, it is not robust in design space. If robustness in the design space is of importance to investigators, then weighting methods described in Bindele & Abebe (2012) may be used.

From now, define $\hat{G}_j(t) = (\hat{g}_{0j}(t), \dots, \hat{g}_{qj}(t))^T$ and $\hat{G}_j''(t) = (\hat{g}_{0j}''(t), \dots, \hat{g}_{qj}''(t))^T$, for $t \in \tilde{\mathcal{A}} = \{t = \mathbf{x}^\tau \hat{\boldsymbol{\beta}}_{jn}\}$. The following theorem gives the asymptotic properties of $\hat{G}_j(t)$.

Theorem 8. Let $\omega_n = \sqrt{\log n / nh_n^{q+1}}$, and suppose that the derivative with respect to t of the function

$$t \mapsto \mathcal{W}_{0j}(t) = -E\{\mathbf{z}\mathbf{z}^\tau f_j(\nu_j(\boldsymbol{\beta}_0))\varphi'(F_j(\nu_j(\boldsymbol{\beta}_0)))|\mathbf{x}^\tau \boldsymbol{\beta}_0 = t\} + E\{\mathbf{z}\varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))|\mathbf{x}^\tau \boldsymbol{\beta}_0 = t\}$$

exists. Then, under assumptions $(I_1) - (I_8)$, we have

$$\sup_{|t| \leq c} \|\hat{G}_j(t) - G(t)\| = O_p((\omega_n + h_n^{2(q+1)})/c). \quad (3.5)$$

Moreover,

$$(nh_n^{q+1})^{1/2} \{\hat{G}_j(t) - G(t) - \boldsymbol{\eta}(t)h_n^{2(q+1)}\} \xrightarrow{\mathcal{D}} N\left(0, m^{-1}(t)\mathcal{W}_{0j}^{-1}(t)\mathcal{W}_{1j}(t)\mathcal{W}_{0j}^{-1}(t) \int K^2(u)du\right), \quad (3.6)$$

where

$$\boldsymbol{\eta}(t) = \frac{1}{2}G''(t) \int u^2 K(u)du, \quad \text{and} \quad \mathcal{W}_{1j}(t) = E\{\mathbf{z}\mathbf{z}^\tau \varphi^2(F_j(\nu_j(\boldsymbol{\beta}_0)))|\mathbf{x}^\tau \boldsymbol{\beta}_0 = t\}$$

The proof can be obtained in a similar way as given in Xia *et al.* (2007), and is therefore omitted here. Readers seeking for details are referred to the aforementioned paper.

Remark 3.3. Equation (3.6) indicates that $\hat{G}_j(t)$ is asymptotically unbiased, since $h_n^{2(q+1)} \rightarrow 0$, as $n \rightarrow \infty$. So, for small n and finite samples, $\hat{G}_j(t)$ is biased. Thus, when it comes to constructing confidence intervals/regions, biases corrections should be considered as in Levin & Kong (1990).

3.5 Simulation & Real Data Application

Simulation Study

To evaluate the performance of the proposed method, an extensive simulation study was conducted and the results were compared to the traditional least squares (LS) method and the least absolute deviation (LAD) method. For model (1.1), we considered the

model

$$Y = 3 \exp\{(-\mathbf{X}^\tau \boldsymbol{\beta}_0)^2\} + 0.8\{\mathbf{X}^\tau \boldsymbol{\beta}_0\} \mathbf{z}_1 + 1.5 \sin[\pi(\mathbf{X}^\tau \boldsymbol{\beta}_0)] \mathbf{z}_3 + \varepsilon,$$

where $\boldsymbol{\beta}_0 = (1, 2, 0, 2)/3$, $\mathbf{X} \sim (U[-1, 1])^{\otimes 4}$, $\mathbf{Z} \sim (N[0, 1])^{\otimes 4}$ and $G(\cdot) = (g_0, g_1, g_2, g_3, g_4)$, with $g_0(u) = e^{-u^2}$, $g_1(u) = u$, $g_3(u) = \sin \pi u$, and $g_2(u) = g_4(u) = 0$. The model error, ε , was generated from the standard normal distribution ($N(0, 1)$), the t -distribution with 3 degrees of freedom (t_3) and the contaminated normal distribution with 10% contamination rate ($CN(0, 10) = 0.9N(0, 1) + 0.1N(0, 10^2)$). In the rank-based objective function, we considered the linear Wilcoxon score function, $\varphi(u) = \sqrt{12}(u - 0.5)$, to ensure consistency and robustness of the proposed estimator. Since the estimation of $G(\cdot)$ involves kernel smoothing, the Epanechnikov Kernel, $K(u) = 0.75(1 - u^2)(|u| \leq 1)$, was considered. It is important to mention that the choice of the kernel function has less importance, as long as assumption (I_5) is satisfied (Einmahl & Mason, 2005; Delecroix *et al.*, 2006). The missingness of y_i was generated as $\delta \sim \text{Bernoulli}(\pi(\mathbf{x}_i, \mathbf{z}_i))$ and the missing y_i 's were imputed based on the simple imputation (SI) method and the inverse probability weighted (IP) procedure. Setting $u(\mathbf{x}, \mathbf{z}) = (4)^{-1}\|\mathbf{x}\|_1 - \Phi(\mu(\mathbf{z}))$, with $\mu(\mathbf{z}_i)$ being the mean of the vector $\mathbf{z}_i$ and Φ the standard normal cumulative distribution function, the response probability $\pi(\mathbf{x}_i, \mathbf{z}_i)$ was calculated from three different cases as

Case 1: $\pi(\mathbf{x}, \mathbf{z}) = P(\delta = 1|\mathbf{x}, \mathbf{z}) = 0.3 + |u(\mathbf{x}, \mathbf{z}) - 0.25|$ if $|u(\mathbf{x}, \mathbf{z}) - 0.25| \leq 0.25$, and 0.3 elsewhere,

Case 2: $\pi(\mathbf{x}, \mathbf{z}) = P(\delta = 1|\mathbf{x}, \mathbf{z}) = 0.5 + 0.5(|u(\mathbf{x}, \mathbf{z}) - 1.5|)$ if $|u(\mathbf{x}, \mathbf{z}) - 1.5| \leq 0.2$, and 0.5 elsewhere,

Case 3: $\pi(\mathbf{x}, \mathbf{z}) = P(\delta = 1|\mathbf{x}, \mathbf{z}) = 0.5 - |u(\mathbf{x}, \mathbf{z}) - 1.5|$ if $|u(\mathbf{x}, \mathbf{z}) - 1.5| \leq 0.25$, and 0.4 elsewhere,

where cases 1, 2, 3 gives on average the response rate of 30%, 50%, 40% respectively. From 500 replications, with $n = 200$, means and standard deviations of the angle between the estimated $\hat{\boldsymbol{\beta}}$ and the true $\boldsymbol{\beta}_0$; $A(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0)$, as well as mean squared errors (MSE) of $\hat{G}$, were calculated, where $A(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0) = (180/\pi) \arccos(|\hat{\boldsymbol{\beta}}^\tau \boldsymbol{\beta}_0|)$, with small values indicating good

performance (Zeng *et al.* , 2012). Conversely, the MSE of $\hat{G}$ was calculated as

$$MSE(\hat{G}) = \frac{1}{n \times (q+1)} \sum_{j=0}^q \sum_{k=1}^n |\hat{g}_j(u_k) - g_j(u_{0k})|^2,$$

where $(u_1, \dots, u_n)^\tau = \mathbf{X}^\tau \hat{\boldsymbol{\beta}}$ and $(u_{01}, \dots, u_{0n})^\tau = \mathbf{X}^\tau \boldsymbol{\beta}_0$. Since it is known that complete case analysis can lead to biased and/or less efficient estimates (Little, 1992; Little & Rubin, 2002), we evaluated how imputations have improved the estimation efficiency by calculating the means of $A(\hat{\boldsymbol{\beta}}_I, \hat{\boldsymbol{\beta}}_{CC})$; the angle between the imputed index estimate, $\hat{\boldsymbol{\beta}}_I$, and the complete case estimate, $\hat{\boldsymbol{\beta}}_{CC}$. Moreover, the Euclidean distance between the imputed vector functional coefficient estimate; $\hat{G}_I$, and the complete case vector functional coefficient estimate; $\hat{G}_{CC}$, was also calculated. Results of the simulation study are reported in Tables 7 and 8.

Tab. 7: The mean and SD of $A(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}_0)$ and the $MSE \times 10^2$ of $\hat{G}$

			LS			Rank			LAD			
Dist	Imp	Case	Mean	SD	MSE($\hat{G}$)	Mean	SD	MSE($\hat{G}$)	Mean	SD	MSE($\hat{G}$)	
$N(0, 1)$	SI	1	17.871	18.353	9509.787	20.207	19.785	9605.903	20.822	18.290	9706.757	
		2	7.508	6.368	8830.230	8.572	6.358	8809.943	9.993	8.246	8924.193	
		3	10.497	11.046	8987.754	11.942	12.036	9029.478	13.223	12.010	9138.704	
	IP	1	19.781	19.151	9789.496	21.081	19.717	9682.004	22.211	18.352	9773.411	
		2	8.288	8.424	8973.904	9.412	6.808	8917.463	11.642	9.380	9029.392	
		3	11.572	12.363	9133.780	13.123	12.552	9127.675	15.311	13.514	9277.547	
	Full		3.732	1.636	8622.096	4.389	1.919	8574.277	4.868	2.180	8596.799	
	t_3	SI	1	34.207	24.541	12106.678	30.395	22.915	10512.141	33.941	23.781	10713.683
			2	19.113	18.211	9826.707	13.959	12.315	9119.332	16.836	14.609	9244.873
3			25.221	21.816	10303.976	20.514	17.940	9457.118	24.360	19.424	9697.159	
IP		1	36.160	24.219	13391.928	32.550	23.215	10689.840	35.587	23.210	11038.367	
		2	21.951	19.753	9941.509	17.619	14.257	9308.311	20.976	16.338	9611.868	
		3	28.151	22.424	10624.456	23.908	19.481	9682.896	27.824	19.989	10108.978	
Full			8.474	9.918	8751.294	5.703	3.274	8535.013	6.417	3.754	8581.254	
CN		SI	1	49.483	23.430	18383.398	38.129	24.593	12294.197	42.814	24.938	12301.818
			2	45.807	25.068	13496.885	26.391	21.627	9924.579	29.68	22.599	9951.464
	3		47.502	24.645	15424.742	31.571	23.619	10776.071	36.431	24.696	10707.84	
	IP	1	50.430	23.705	19090.284	42.427	24.451	13153.894	47.074	23.968	13126.086	
		2	47.547	24.942	13690.266	33.794	23.084	10732.402	41.204	24.093	11090.797	
		3	48.750	24.185	16097.628	37.481	23.459	11708.186	45.355	24.15	11756.177	
	Full		30.644	22.289	9978.671	8.314	7.324	8637.458	10.933	12.029	8738.121	

Tab. 8: Comparison of the complete case and imputed analysis

Imp	Dist	Case	LS		Rank		LAD	
			$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{G}_I - \hat{G}_{CC}\ _2 \times 10^2$	$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{G}_I - \hat{G}_{CC}\ _2 \times 10^2$	$\bar{A}(\hat{\beta}_I, \hat{\beta}_{CC})$	$\ \hat{G}_I - \hat{G}_{CC}\ _2 \times 10^2$
SI	$N(0, 1)$	1	3.510	1790.058	4.005	1878.374	4.895	1878.746
		2	2.850	1583.326	3.197	1650.048	4.227	1646.927
		3	3.304	1624.300	3.604	1708.639	4.779	1712.397
	t_3	1	5.547	2212.046	6.944	2005.541	7.736	1995.720
		2	5.447	1851.914	5.813	1756.137	7.368	1735.425
		3	5.817	1959.406	6.552	1871.293	7.965	1874.997
	CN	1	6.947	3026.103	11.445	2321.817	15.359	2253.337
		2	10.670	2644.239	12.990	1853.113	15.274	1836.253
		3	8.678	2797.741	12.947	2079.979	16.187	1962.206
IP	$N(0, 1)$	1	9.476	1848.158	6.164	1902.003	8.315	1894.266
		2	5.073	1631.156	4.930	1678.748	7.426	1667.798
		3	7.274	1677.152	5.958	1735.096	8.095	1733.386
	t_3	1	16.618	2276.855	12.439	2066.290	14.123	2033.758
		2	13.326	1907.727	10.957	1801.762	15.083	1776.331
		3	15.827	2032.508	11.870	1939.168	14.940	1936.865
	CN	1	19.024	3119.342	20.058	2545.231	27.571	2432.572
		2	24.265	2721.314	25.962	2030.337	32.756	2056.317
		3	22.829	2902.559	24.890	2283.516	31.956	2204.083

From Table 7, it is observed that when $\varepsilon \sim N(0, 1)$, the LS procedure performs slightly better than the LAD and Rank procedures in terms of means and standard deviations of $A(\hat{\beta}, \beta_0)$, as would be expected, the Rank method gives more consistent estimates in terms of MSEs of $\hat{G}$. Under t_3 and $CN(0, 10)$, it is clearly observed that the proposed method dominates the LS and LAD methods. It is not surprising to see that for the full data, while the LS procedure performs better under $N(0, 1)$ as expected, the LAD and Rank procedures perform better under t_3 and $CN(0, 10)$, with the best performance by the Rank procedure. Generally, it is observed that the larger the rate of missingness, the larger the means of $A(\hat{\beta}, \beta_0)$ and MSEs of $\hat{G}$. It is also worth pointing out the SI approach tends to have a better performance than the IP approach of imputation. From Table 8 on the other hand, it is observed that means of $A(\hat{\beta}_I, \hat{\beta}_{CC})$ and $\|\hat{G}_I - \hat{G}_{CC}\|_2$ are not tending to zero, which indicates that imputation has improved the estimation efficiency.

Real Application to Catch per Unit Effort (CPUE) of Groundfish Data

The proposed method was applied to a complex food web system in the northern Pacific Ocean data. Four marine groundfish species, Pacific cod, Pacific halibut, sablefish, and

walleye pollock, interact regularly throughout their lifetimes in the Gulf of Alaska and Bering Sea. Pacific halibut serving as the main predator for adult sablefish (Center, 2019a), and Pacific cod and walleye pollock are important prey species for Pacific halibut (Aydin *et al.*, 2015; Hollowed *et al.*, 2000; Livingston, 1993). We were interested in understanding the effect of the three prey species on Pacific halibut in relation to changes in a suite of environmental variables.

Yearly catch per unit effort (CPUE) of these four groundfish at multiple locations across the Gulf of Alaska (Center, 2020) were matched with the annual means of March, April, and May averages for environmental covariates, including ocean chemical components (Center, 2018), zooplankton biomass concentrations (National Marine Fisheries Service, 2014; O'Brien, 2007), and temperature at the sea floor (Center, 2019b). The groundfish surveys were conducted by Japan in the early years, but transitioned to being conducted entirely by the United States National Marine Fisheries Service by 1994 (Center, 2019a). Data were therefore restricted to only surveys conducted by Japan with stations where at least 10 years of observations were recorded to reduce noise from differences between sampling methodologies and capture consistent spatiotemporal trends in environment and fish interactions. This resulted in a final data set of 188 observations from 1985 to 1994.

The SIVCM in (1.1) was employed, where y is the CPUE of Pacific halibut; $\mathbf{z} = (z_1, z_2, z_3)^\tau$ is the CPUEs of Pacific cod, sablefish and walleye pollock, respectively; and the matrix $\mathbf{x} = (x_1, \dots, x_{10})^\tau$ contains the environmental variables with depth = x_1 , plankton = x_2 , bottom temperature = x_3 , alkalinity = x_4 , chlorophyll = x_5 , nitrate = x_6 , oxygen = x_7 , phosphate = x_8 , salinity = x_9 , and silicate = x_{10} . In the response space, there were about 18.1% of missing values, which were imputed using both the simple imputation (SI) and the inverse probability weighting (IP) procedure. Since ranges of the environmental factors varied widely, covariates in the $\mathbf{x}$ matrix and groundfish CPUEs were centered and scaled to have mean zero and variance one.

We fit model (1.1) using the proposed method (Rank) and compared the results to those obtained using the LS and LAD approaches. A 10-fold cross-validation was performed, where fitting was done on the complete case training data, the index parameter β_0 and the vector functional coefficient $G = \{g_0, g_1, g_2, g_3\}$ were estimated, then the missing y

values for the training data were imputed. Finally, re-estimation of the index coefficient and the vector functional coefficient was performed on the training data with the imputed y values. Finally, MSE of the estimated coefficients for the test data were calculated. Functions estimated by the LS approach for the complete case data were highly affected by outliers, and imputation via inverse probability weighting produced inconsistent predictions for both LS and LAD approaches (Figure 6). The estimated functions from the proposed method remained consistently unaffected by outliers, regardless of whether complete case or imputed data were used. This bore out in the cross-validation, where the LS approach had the highest mean and median MSEs for both the simple and the inverse probability weighting imputations (Table 9).

Tab. 9: Means, standard deviations (SD), Medians and interquartile ranges (IQR) of mean squared error (MSE) from 10-fold cross-validation for each estimation method when using either simple imputation (SI) or inverse probability weighting (IP).

Imputation	Estimation	Mean MSE (SD)	Median MSE (IQR)
IP	LAD	1.141(1.440)	0.461 (0.990)
	LS	2.227(1.930)	1.620 (2.025)
	R	1.108(1.335)	0.495 (0.784)
SI	LAD	1.138(1.461)	0.588 (0.788)
	LS	1.744(1.532)	1.388 (2.140)
	R	1.109(1.356)	0.417 (0.871)

The estimated index coefficient for the proposed method was

$$\hat{\beta}_0 = (0.79, -0.01, -0.27, 0.47, -0.005, 0.035, -0.0001, -0.0003, -0.201, -0.21)^\tau,$$

and the estimated coefficient functions are shown in Figure 7. While $\hat{g}_2$ is monotonic, $\hat{g}_0$ and $\hat{g}_1$ exhibit functional changes around $\mathbf{x}^\tau \hat{\beta}_0 = 0$, and $\hat{g}_3$ has a peak at $\mathbf{x}^\tau \hat{\beta}_0 = 1.5$. The suite of environmental variables in $\mathbf{z}$ negatively affected Pacific halibut CPUE in the absence of prey interactions when $\mathbf{x}^\tau \hat{\beta}_0 < 0$, as indicated by the intercept function $\hat{g}_0$. Pacific halibut CPUE was also negatively affected by Pacific cod CPUE in the presence of these environmental conditions when $\mathbf{x}^\tau \hat{\beta}_0 < 0$, as illustrated by the g_1 function. On the other hand and as shown by the g_2 function, sablefish CPUE positively influenced the CPUE of Pacific halibut across the entire range of environmental factors.

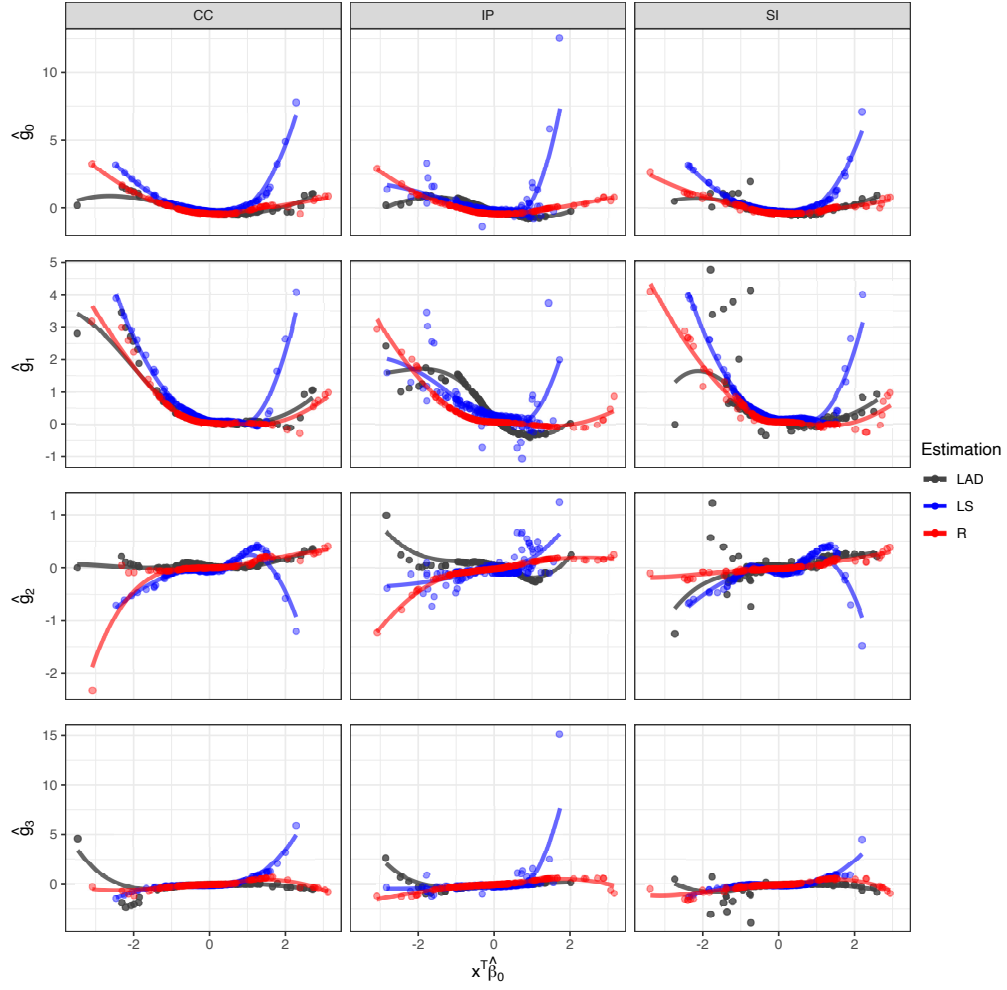


Fig. 6: Estimated functions for the SIVCM fit on fish catch data using three estimation methods for either complete case data (CC) or imputing missing responses using simple imputation (SI) or inverse probability weighting (IP).

Similarly, Pacific halibut CPUE was positively impacted by the CPUE of walleye pollock under the suite of environmental variables until $\mathbf{x}^T \hat{\beta}_0 > 1.5$, at which point the effect was negative. It is clear that the environmental thresholds for the three prey species differ substantially, therefore these species with analogous habitats respond in opposing ways to the same environmental variables. Changes in trophic interactions and species compositions, such as a decrease in foragers with simultaneous increase in predators, can signal strong regime shifts in ocean ecosystems (Anderson & Piatt, 1999).

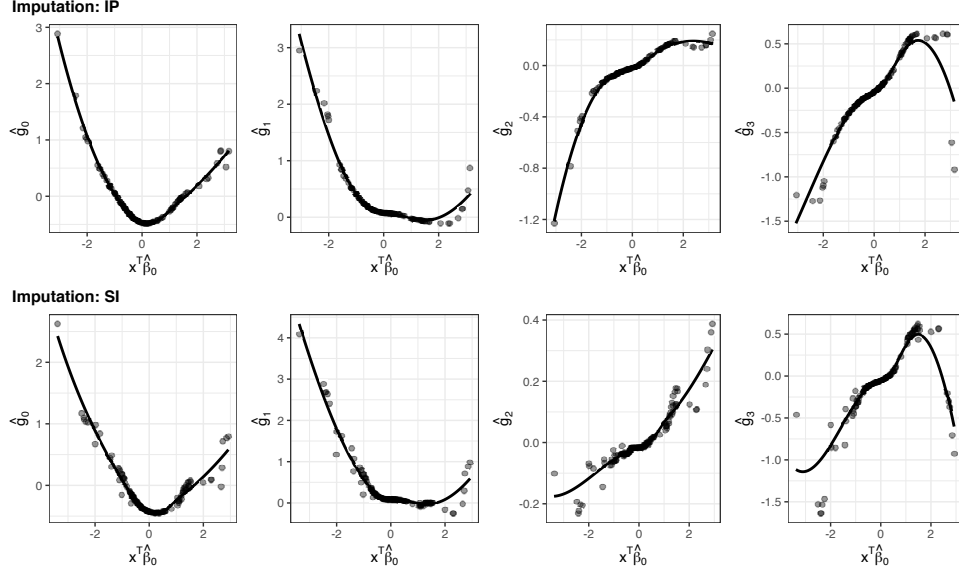


Fig. 7: Final smooths for rank-estimated SIVCM fit on fish catch data with missing response imputed using both SI and IP imputations.

Modeling groundfish food web systems in the North Pacific can therefore provide important insights for such trophic relationships with potential contributions to systemic ecological shifts. The application of the proposed rank-based method to the groundfish data illustrates the effectiveness of robust methods to handle outliers common in ecological data and allows for modeling of complex deep-water interactions in the presence of unobserved responses. In terms of estimating the functional coefficients, it is worth pointing out that from this data, the IP tends to provide better fits than the SI; see Figure 7.

3.6 Discussion

This chapter proposes a rank-based approach for estimating the index coefficient and the functional coefficients in model (1.1) whenever some responses are missing at random. A computational algorithm was developed to obtain the proposed estimators. Once again, under mild regularity conditions, these estimators are shown to be strongly consistent and possess asymptotic normal distributions. In the simulation study conducted, the estimators proved to be the best when compared to the LS and LAD estimators when

dealing with heavy-tailed and contaminated model error distributions. This was also proven in real data application, where the LS and LAD methods gave inconsistent results due to outliers in the data, while the proposed method was not affected. Lastly, as in the previous chapter, it was established that imputing for missing responses improves estimation efficiency.

4 Robust Empirical likelihood inference for the SIVCM

4.1 Introduction

Inference about the true regression coefficients based on the normal approximation (NA) method, requires the estimation of the estimator's covariance matrix, which, depending on the model structure, may be a difficult task. A fully nonparametric approach that avoids estimating the estimator's covariance matrix is known as the empirical likelihood (EL) method introduced by Owen (1988). This approach has been proven to possess the advantage over the NA method when it comes to providing robust confidence regions/intervals for the true regression parameter. Since it was introduced, the EL method has gained considerable attention as a way of avoiding estimating the estimator's covariance matrix in the normal approximation method. Based on a general estimating equations, an EL approach was derived by Qin & Lawless (1994). For nonlinear functional operators such as general U-statistic, a jackknife EL method is developed in Jing *et al.* (2009). Using rank-based estimating equations, EL approaches are developed in Bindele & Zhao (2016) and Bindele *et al.* (2018b). Many other contributions involving EL approaches can be found in Hall & La Scala (1990), Chen & Hall (1993), Kitamura (1997), Peng (2004), Xue & Zhu (2006), Xue & Zhu (2007b) and references therein. Most of these works demonstrate the advantages of EL approaches compared to their normal approximations or bootstrap approaches counterpart. In the context of missing data under the missing at random mechanism, EL approaches are developed in Wang *et al.* (2004), Wang & Rao (2002), Sun *et al.* (2009), and Xue (2009) just to cite a few. Most of these papers use LS estimating equations, while it is well understood that the LS objective function results in non-robust parameter estimators when dealing with heavy-tailed, contaminated and skewed model error distributions, and/or in the presence of unusual observations (outliers and leverages). Although a number of improvements have been made when considering the EL method based on the LS estimating equation, the burden that comes with the LS objective function, also affects the resulting confidence regions/intervals. Motivated

by the interest to provide robust and efficient inference about the model parameters, we propose a robust EL approach. The proposed approach is derived by considering the rank-based estimating equation from the rank-based dispersion function (Jaeckel, 1972). As for the LS approach, the rank-based method is known to possess not only a simple geometric interpretability but also, to result in robust and more efficient estimators compared to many existing estimation methods (Hettmansperger & McKean, 2011). Under MAR responses for regression models with simple structure, rank-based estimation method with inference based on normal approximation was proposed by Bindele (2015) and Bindele & Abebe (2015). As pointed out earlier, in the absence of missing data, the normal approximation method for model (1.1) is studied by Sun *et al.* (2019). While these contributions have shown advantages compared to estimation methods such as the LS, LAD and other methods of moments, they are all based on normal approximation inference, known to result in less accurate confidence regions/intervals. Most of the papers that have dealt with the EL approach for model (1.1), address the construction of confidence regions/intervals of the index parameter by using the LS estimating equation. Few papers that have considered the EL approach for the functional coefficients include Xue & Zhu (2007a), Yang & Yang (2014) and references therein. In an effort to derive more accurate confidence regions/intervals for both the index and functional coefficients in model (1.1) with MAR responses, in this chapter, a rank-based EL approach is proposed.

4.2 Complete-case rank-based empirical likelihoods for β_0 and G

Let

$$e_{ij}(\boldsymbol{\beta}) = \begin{cases} \delta_i \left(y_i - \sum_{k=0}^q (\tilde{a}_k + \tilde{b}_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki} \right), & j = 1 \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} \left(y_i - \sum_{k=0}^q (\tilde{a}_k + \tilde{b}_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki} \right), & j = 2, \end{cases}$$

where $\tilde{a}_k$ and $\tilde{b}_k$ are complete case estimates of a_k and b_k , as defined in the previous chapter. This can be rewritten as

$$e_{ij}(\boldsymbol{\beta}) = \begin{cases} \delta_i \left(y_i - \{\tilde{G}(\mathbf{x}_i^\tau \boldsymbol{\beta})\}^\tau \mathbf{z}_i \right), & j = 1 \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} \left(y_i - \{\tilde{G}(\mathbf{x}_i^\tau \boldsymbol{\beta})\}^\tau \mathbf{z}_i \right), & j = 2, \end{cases}$$

where $\{\tilde{G}(\mathbf{x}_i^\tau \boldsymbol{\beta})\}^\tau \mathbf{z}_i = \sum_{k=0}^q (\tilde{a}_k + \tilde{b}_k \mathbf{x}_{0i}^\tau \boldsymbol{\beta}) z_{ki}$. From this, the complete case objective function of $\boldsymbol{\beta}$ is defined as

$$\mathcal{N}_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \varphi \left(\frac{R(e_{ij}(\boldsymbol{\beta}))}{n+1} \right) e_{ij}(\boldsymbol{\beta}). \quad (4.1)$$

Similarly, setting

$$v_{ij}(G) = \begin{cases} \delta_i(y_i - \{G(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c)\}^\tau \mathbf{z}_i), & j = 1 \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} \left(y_i - \{G(\mathbf{x}_i^\tau \tilde{\boldsymbol{\beta}}_n^c)\}^\tau \mathbf{z}_i \right), & j = 2, \end{cases}$$

where $\tilde{\boldsymbol{\beta}}_n^c$ is the complete case estimator of $\boldsymbol{\beta}_0$ as also discussed in the previous chapter, the complete case objective function of G is given by

$$K_{jn}(G) = \frac{1}{n} \sum_{i=1}^n w_i \varphi \left(\frac{R(v_{ij}(G))}{n+1} \right) v_{ij}(G), \quad (4.2)$$

where w_i is defined in the previous chapter. From equations (4.1) and (4.2), it is observed in a straightforward manner that $\tilde{\boldsymbol{\beta}}_{jn}^c = \text{Argmin}_{\boldsymbol{\beta} \in \mathcal{B}} \mathcal{N}_{jn}(\boldsymbol{\beta})$ and $\tilde{G}_{jn} = \text{Argmin}_G K_{jn}(G)$. Thus, the corresponding estimating equations are given by $\nabla_{\boldsymbol{\beta}} \mathcal{N}_{jn}(\boldsymbol{\beta}) = 0$ and $\nabla_G K_{jn}(G) = 0$, respectively. Now, define the following random variables

$$\ell_{ij}(\boldsymbol{\beta}_0) = \begin{cases} \delta_i \mathbf{x}_i \{ \tilde{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi \left(\frac{R(e_{ij}(\boldsymbol{\beta}_0))}{n+1} \right) & j = 1 \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} \mathbf{x}_i \{ \tilde{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi \left(\frac{R(e_{ij}(\boldsymbol{\beta}_0))}{n+1} \right) & j = 2 \end{cases}$$

and

$$k_{ij}(G) = \begin{cases} \delta_i w_i \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G))}{n+1} \right) & j = 1 \\ \frac{\delta_i}{\hat{\pi}(\mathbf{u}_i)} w_i \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G))}{n+1} \right) & j = 2. \end{cases}$$

For $j = 1, 2$, let κ_{ij} and ζ_{ij} , $i = 1, \dots, n$ be probability masses attached to $\ell_{ij}(\boldsymbol{\beta}_0)$ and $k_{ij}(G)$, respectively. Define the following empirical log likelihood functions:

$$\mathcal{L}_j(\boldsymbol{\beta}_0) = -2 \sup_{\kappa_{ij}} \left\{ \sum_{i=1}^n \log(n \kappa_{ij}) : \kappa_{ij} \geq 0, \sum_{i=1}^n \kappa_{ij} = 1 \text{ and } \sum_{i=1}^n \kappa_{ij} \ell_{ij}(\boldsymbol{\beta}_0) = 0 \right\}, \quad j = 1, 2$$

and

$$\mathcal{K}_j(G) = -2 \sup_{\zeta_{ij}} \left\{ \sum_{i=1}^n \log(n\zeta_{ij}) : \zeta_{ij} \geq 0, \sum_{i=1}^n \zeta_{ij} = 1 \text{ and } \sum_{i=1}^n \zeta_{ij} k_{ij}(G) = 0 \right\}, \quad j = 1, 2.$$

As long as 0 is inside convex hulls of $\ell_{ij}(\beta_0)$ and $k_{ij}(G)$, maximizers of $\mathcal{L}_j(\beta_0)$ and $\mathcal{K}_j(G)$ exist. Using the Lagrange multiplier approach, $\mathcal{L}_j(\beta_0)$ is maximized at $\kappa_{ij} = 1/n(1 + \boldsymbol{\mu}^\tau \ell_{ij}(\beta_0))$ and $\mathcal{K}_j(G)$ is maximized at $\zeta_{ij} = 1/n(1 + \boldsymbol{\nu}^\tau k_{ij}(G))$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\nu} \in \mathbb{R}^{q+1}$. To see this, let

$$Q(\kappa_{ij}, \lambda_0, \boldsymbol{\mu}) = \sum_{i=1}^n \log(n\kappa_{ij}) + \lambda_0 \left(1 - \sum_{i=1}^n \kappa_{ij}\right) - \boldsymbol{\mu}^\tau \sum_{i=1}^n \kappa_{ij} \ell_{ij}(\beta_0), \quad j = 1, 2.$$

Putting $\frac{\partial Q(\kappa_{ij}, \lambda_0, \boldsymbol{\mu})}{\partial \kappa_{ij}} = 0$, we have

$$\frac{\partial Q(\kappa_{ij}, \lambda_0, \boldsymbol{\mu})}{\partial \kappa_{ij}} = \frac{1}{n\kappa_{ij}} - \lambda_0 - \boldsymbol{\mu}^\tau \ell_{ij}(\beta_0) = 0, \quad j = 1, 2.$$

Consequently, the value of $\lambda_0 = 1$, obtained by solving

$$\sum_{i=1}^n \kappa_{ij} \frac{\partial Q(\kappa_{ij}, \lambda_0, \boldsymbol{\mu})}{\partial \kappa_{ij}} = 0.$$

From this, we get the optimal κ_{ij}' s as

$$\kappa_{ij} = \frac{1}{n(1 + \boldsymbol{\mu}^\tau \ell_{ij}(\beta_0))}, \quad j = 1, 2, \tag{4.3}$$

where, $\boldsymbol{\mu}$ satisfies

$$\begin{aligned} 0 &= \sum_{i=1}^n \kappa_{ij} \ell_{ij}(\beta_0) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\ell_{ij}(\beta_0)}{1 + \boldsymbol{\mu}^\tau \ell_{ij}(\beta_0)}, \quad j = 1, 2. \end{aligned}$$

Similarly,

$$\zeta_{ij} = \frac{1}{n(1 + \boldsymbol{\nu}^\tau k_{ij}(G))}, \quad j = 1, 2, \quad (4.4)$$

where, $\boldsymbol{\nu}$ satisfies

$$\begin{aligned} 0 &= \sum_{i=1}^n \zeta_{ij} k_{ij}(\boldsymbol{\beta}_0) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{k_{ij}(G)}{1 + \boldsymbol{\nu}^\tau k_{ij}(G)}, \quad j = 1, 2. \end{aligned}$$

Substituting (4.3) and (4.4) in the definitions of $\mathcal{L}_j(\boldsymbol{\beta}_0)$ and $\mathcal{K}_j(G)$, we have

$$\mathcal{L}_j(\boldsymbol{\beta}_0) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\mu}^\tau \ell_{ij}(\boldsymbol{\beta}_0)), \quad \text{and} \quad \mathcal{K}_j(G) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\nu}^\tau k_{ij}(G)), \quad j = 1, 2.$$

From this, we have the following theorem

Theorem 9. *Set*

$$\begin{aligned} \Xi_1 &= E \left(\pi(\mathbf{u}) \mathbf{x} \mathbf{x}^\tau [\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}]^2 \right), \quad \Xi_2 = E \left(\pi^{-1}(\mathbf{u}) \mathbf{x} \mathbf{x}^\tau [\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}]^2 \right) \\ \mathbf{A}_1 &= E \left(\pi(\mathbf{u}) \mathbf{x} \mathbf{x}^\tau [\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}]^2 \varphi^2(F(\varepsilon)) \right), \quad \mathbf{A}_2 = E \left(\pi^{-1}(\mathbf{u}) \mathbf{x} \mathbf{x}^\tau [\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}]^2 \varphi^2(F(\varepsilon)) \right) \\ \Delta_1 &= E \left[\pi(\mathbf{u}) w^2 \mathbf{z} \mathbf{z}^\tau \right], \quad \Delta_2 = E \left[\pi^{-1}(\mathbf{u}) w^2 \mathbf{z} \mathbf{z}^\tau \right] \\ \mathbf{B}_1 &= E \left(\pi(\mathbf{u}) w^2 \mathbf{z} \mathbf{z}^\tau \varphi^2(F(\varepsilon)) \right), \quad \text{and} \quad \mathbf{B}_2 = E \left(\pi^{-1}(\mathbf{u}) w^2 \mathbf{z} \mathbf{z}^\tau \varphi^2(F(\varepsilon)) \right). \end{aligned}$$

Under assumptions $(I_1) - (I_7)$, we have

- i) $\sqrt{n} \nabla_{\boldsymbol{\beta}} \mathcal{N}_{jn}(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Xi_j)$ and $\mathcal{L}_j(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \sum_{i=1}^p \lambda_{ij} \chi_{1,i}^2$, where the λ_{ij} for $j = 1, 2$, are the eigenvalues of $\Xi_j^{1/2} \mathbf{A}_j^{-1} \Xi_j^{1/2}$ and the $\chi_{1,i}^2$ are independent χ^2 distributions with one degree of freedom.
- ii) $\sqrt{nh_n^{q+1}} \nabla_G \mathcal{K}_{jn}(G) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Delta_j)$ and $\mathcal{K}_j(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \sum_{i=0}^q \rho_{ij} \chi_{1,i}^2$, where the ρ_{ij} for $j = 1, 2$, are the eigenvalues of $\Delta_j^{1/2} \mathbf{B}_j^{-1} \Delta_j^{1/2}$ and the $\chi_{1,i}^2$ are independent χ^2 distributions with one degree of freedom.

The proof of this theorem can be constructed along the lines as that of Theorems 1 & 2 of Bindele *et al.* (2018b). So, for sake of brevity it is omitted here. Readers seeking for

details are referred to the aforementioned reference. This theorem allows us to construct empirical likelihood confidence regions/intervals of the complete case and the weighted complete case of the true model parameters.

As pointed out by many authors, while the complete case analysis can result in a valid inference for a small proportion of missing responses, it can result in biased and inefficient estimates, mainly when dealing with large proportion of missingness (Little, 1992). Therefore, to take into account the missing information, imputation has been proposed in the previous chapter, from which estimates of the model parameters have been derived and their asymptotic properties established. However, further statistical inference about the true model parameters, will require the estimation of the proposed estimators covariance matrices, which is a difficult task because of the model structure and the considered objective function. To avoid such a burden, a rank-based empirical likelihood approach is considered, as discussed in the next sections.

4.3 Empirical likelihood of the index parameter under MAR responses

Recall that $\tilde{S}_{jn}(\boldsymbol{\beta}) = -\nabla_{\boldsymbol{\beta}} Q_{jn}(\boldsymbol{\beta})$ and $\hat{\boldsymbol{\beta}}_{jn}$ is solution of the estimating equation $\tilde{S}_{jn}(\boldsymbol{\beta}) = 0$. Under assumption (I_7) , $E[\tilde{S}_{jn}(\boldsymbol{\beta}_0)] \rightarrow 0$, which indicates that the estimating equation $\tilde{S}_{jn}(\boldsymbol{\beta}) = 0$ is asymptotically unbiased. From

$$\tilde{S}_{jn}(\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i (\{\nabla_{\boldsymbol{\beta}} \hat{G}(\mathbf{x}_i^T \boldsymbol{\beta}_0)\}^T \mathbf{z}_i) \varphi(R(\tilde{\nu}_{ij}(\boldsymbol{\beta}_0))/(n+1)),$$

set $z_{ij}(\boldsymbol{\beta}_0) = w_i \mathbf{x}_i (\{\nabla_{\boldsymbol{\beta}} \hat{G}(\mathbf{x}_i^T \boldsymbol{\beta}_0)\}^T \mathbf{z}_i) \varphi(R(\tilde{\nu}_{ij}(\boldsymbol{\beta}_0))/(n+1))$. Under the MAR assumption, we have $E(z_{ij}(\boldsymbol{\beta}_0)) = 0$. Let $(p_{1j}, \dots, p_{nj})^T$ denote a vector of probability values satisfying $\sum_i^n p_{ij} = 1$ and $p_{ij} \geq 0$ for all i and $j = 1, 2$. Then, the empirical log-likelihood ratio function for $\boldsymbol{\beta}_0$ is given by

$$R_j(\boldsymbol{\beta}_0) = -2 \sup_{p_{ij}} \left\{ \sum_{i=1}^n \log(np_{ij}) : p_{ij} \geq 0, \sum_{i=1}^n p_{ij} = 1 \text{ and } \sum_{i=1}^n p_{ij} z_{ij}(\boldsymbol{\beta}_0) = 0 \right\}, \quad j = 1, 2.$$

A direct application of the Lagrange multiplier, results on

$$p_{ij} = \frac{1}{n(1 + \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0))}, \quad (4.5)$$

where $\boldsymbol{\lambda} \in \mathbb{R}^p$, which can be estimated by solving the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{z_{ij}(\boldsymbol{\beta}_0)}{1 + \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0)} = \mathbf{0}. \quad (4.6)$$

Using (4.5), it is obtained in a straightforward manner that

$$R_j(\boldsymbol{\beta}_0) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0)). \quad (4.7)$$

From this, we have the following theorem

Theorem 10. *Under the assumptions $(I_1) - (I_7)$, we have*

$$R_j(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \chi_p^2 \quad \text{as } n \rightarrow \infty \quad \text{and } j = 1, 2.$$

From this theorem, the empirical likelihood confidence region for $\boldsymbol{\beta}_0$ with a nominal level of confidence $1 - \alpha$, is given by

$$\mathcal{C}_1^j = \{\boldsymbol{\beta} : R_j(\boldsymbol{\beta}) \leq \chi_p^2(\alpha)\}, \quad (4.8)$$

where $\chi_p^2(\alpha)$ is the $(1 - \alpha)^{th}$ percentile of the χ^2 -distribution with p degrees of freedom.

Remark 4.1. *Recall that under the MAR assumption, $E(z_{ij}(\boldsymbol{\beta}_0)) = 0$. Putting*

$$\boldsymbol{\Lambda}_{jn} = \frac{1}{n} \sum_{i=1}^n z_{ij}(\boldsymbol{\beta}_0) z_{ij}^\tau(\boldsymbol{\beta}_0),$$

it is obtained under the assumptions of Theorem 7 that $\boldsymbol{\Lambda}_{nj} - \text{Var}(\tilde{S}_{jn}(\boldsymbol{\beta}_0)) \rightarrow 0$ a.s. as $n \rightarrow \infty$. When it comes to statistical inference about $\boldsymbol{\beta}_0$ based on the normal approximation, the estimator's covariance matrix $\mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-$ has to be estimated. Letting $\hat{\mathbf{V}}_j$ and $\hat{\mathbf{W}}_j$ be sandwich type estimators of $\mathbf{V}_j$ and $\mathbf{W}_j$ that can be obtained as in Brunner & Denker (1994), it can be shown that $\hat{\mathbf{V}}_j \rightarrow \mathbf{V}_j$ a.s. and $\hat{\mathbf{W}}_j \rightarrow \mathbf{W}_j$ a.s. Also, putting

$\widehat{\mathbf{\Lambda}}_{jn} = \frac{1}{n} \sum_{i=1}^n z_{ij}(\widehat{\boldsymbol{\beta}}_n^j) z_{ij}^\tau(\widehat{\boldsymbol{\beta}}_n^j)$, we have $\widehat{\mathbf{\Lambda}}_{jn} - \widehat{\mathbf{V}}_j \rightarrow 0$ a.s. Hence, the estimator's covariance matrix can be estimated by $\widehat{\mathbf{M}}_j = \widehat{\mathbf{W}}_j^- \widehat{\mathbf{\Lambda}}_{jn} \widehat{\mathbf{W}}_j^-$. To this end, a $(1-\alpha) \times 100\%$ normal approximation confidence region of $\boldsymbol{\beta}_0$ is given by

$$\mathcal{C}_2^j = \left\{ \boldsymbol{\beta} : \left(\widehat{\boldsymbol{\beta}}_n^j - \boldsymbol{\beta} \right) \widehat{\mathbf{M}}_j^- \left(\widehat{\boldsymbol{\beta}}_n^j - \boldsymbol{\beta} \right) \leq \chi_p^2(\alpha) \right\}. \quad (4.9)$$

where $\chi_p^2(\alpha)$ is the $(1-\alpha)^{th}$ percentile of the χ^2 -distribution with p degrees of freedom.

In practice, one is mostly interested in constructing confidence intervals of the components of $\boldsymbol{\beta}_0$ instead of $\boldsymbol{\beta}_0$ itself. To do so, partition $\boldsymbol{\beta}_0$ as $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}, \boldsymbol{\beta}_{02})$, where $\boldsymbol{\beta}_{01}$ is r -dimensional and $\boldsymbol{\beta}_{02}$ is $(p-r)$ -dimensional. Similarly, for any $\boldsymbol{\beta} \in \mathcal{B}$, write $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, where $\boldsymbol{\beta}_1$ is r -dimensional and $\boldsymbol{\beta}_2$ is $(p-r)$ -dimensional. Following Qin & Lawless (1994) and Yang & Zhao (2012), the profile empirical likelihood is obtained as

$$\Gamma_j(\boldsymbol{\beta}_1) = \inf_{\boldsymbol{\beta}_2} R_j(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2).$$

Following similar arguments as in Yang & Zhao (2012), the following Wilks theorem is established.

Theorem 11. *Under the assumptions of Theorem 10, $\Gamma_j(\boldsymbol{\beta}_{01}) \xrightarrow{\mathcal{D}} \chi_r^2$, as $n \rightarrow \infty$, $j = 1, 2$.*

The $100(1-\alpha)\%$ profile EL confidence region for $\boldsymbol{\beta}_{01}$ is obtained as

$$\mathcal{C}_3^j = \{ \boldsymbol{\beta}_{01} : \Gamma_j(\boldsymbol{\beta}_{01}) \leq \chi_r^2(\alpha) \},$$

where $\chi_r^2(\alpha)$ is the $(1-\alpha)^{th}$ percentile of the χ^2 -distribution with r degrees of freedom. Our simulation studies are based on this profile empirical likelihood confidence region.

4.4 Empirical likelihood of the varying coefficients under MAR responses

For any fixed $t_0 \in \mathcal{A} := \{\mathbf{x}^\tau \boldsymbol{\beta}_0\}$, consider the imputed residuals $v_{ij}(G(t_0)) = \tilde{\mu}_{ij} - G^\tau(t_0) \mathbf{z}_i$.

The rank-based objective function of the coefficient function $G(t_0)$ is given by

$$J_{jn}(G(t_0)) = \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \varphi \left(\frac{R(v_{ij}(G(t_{0i})))}{n+1} \right) v_{ij}(G(t_{0i})) w_i(t_{0i}),$$

where $G(t_0) = (g_0(t_0), \dots, g_q(t_0))^\tau$ and $w_i(t_{0i}) = K_h(t_{0i}) / \sum_{k=1}^n K_h(t_{0k})$, with $t_{0i} = \mathbf{x}_i^\tau \boldsymbol{\beta}_0$.

The rank-based estimator of $G(t_0)$ is given by

$$\hat{G}_j(t_0) = \underset{G(t_0)}{\operatorname{Argmin}} J_{jn}(G(t_0)).$$

Letting $\nabla G(t_0) = \left(\frac{\partial G(t_0)}{\partial g_0(t_0)}, \dots, \frac{\partial G(t_0)}{\partial g_q(t_0)} \right)^\tau$, we have

$$\nabla J_{jn}(G(t_0)) = -\frac{1}{nh_n^{q+1}} \sum_{i=1}^n [\nabla G(t_{0i})]^\tau \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G(t_{0i})))}{n+1} \right) w_i(t_{0i}),$$

and so, $\hat{G}(t_0)$ is solution of the estimating equation $\nabla J_{jn}(G(t_0)) = 0$. Now, setting

$$\phi_{ij}(G(t_0)) = [\nabla G(t_0)]^\tau \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G(t_0)))}{n+1} \right) w_i(t_0),$$

let η_{ij} , $i = 1, \dots, n$, $j = 1, 2$ be some probabilities satisfying $\sum_{i=1}^n \eta_{ij} = 1$ and $\eta_{ij} \geq 0$.

The empirical log-likelihood function of $G(t_0)$ is defined as

$$H_j(G(t_0)) = -2 \sup_{\eta_{ij}} \left\{ \sum_{i=1}^n \log(n\eta_{ij}) : \sum_{i=1}^n \eta_{ij} = 1, \eta_{ij} \geq 0, \sum_{i=1}^n \eta_{ij} \phi_{ij}(G(t_{0i})) = 0, \right\}$$

from which applying the Lagrange multiplier approach yield $\eta_{ij} = \frac{1}{n(1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_0)))}$, where $\boldsymbol{\xi} \in \mathbb{R}^{q+1}$. Note that $\boldsymbol{\xi}$ solve the equation

$$\sum_{i=1}^n \frac{\phi_{ij}(G(t_{0i}))}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = 0. \quad (4.10)$$

From this, it is obtained in a straightforward manner that

$$H_j(G(t_0)) = 2 \sum_{i=1}^n \log(1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))), \quad (4.11)$$

which entails the following theorem

Theorem 12. *Under assumptions $(I_1) - (I_8)$ and letting $\omega_n \rightarrow 0$ as $n \rightarrow \infty$, if $G(\cdot)$ is the true vector of coefficient functions, then*

$$H_j(G(t_0)) \xrightarrow{\mathcal{D}} \chi_{q+1}^2 \quad \text{as } n \rightarrow \infty \quad \text{and } j = 1, 2.$$

This theorem enable us to construct confidence regions for $G(\cdot)$. A $(1 - \alpha) \times 100\%$ confidence region for $G(t_0)$ is given by

$$\mathcal{C}_4^j := \{G(t_0) : H_j(G(t_0)) \leq \chi_{q+1}^2(\alpha)\},$$

where $\chi_{q+1}^2(\alpha)$ is the $(1 - \alpha)^{th}$ percentile of the χ^2 -distribution with $(q + 1)$ -degrees of freedom.

Remark 4.2. *Setting $\Omega_{nj}(t_0) = (nh_n^{q+1})^{-1} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i}))$, it can be shown that*

$$H_j(G(t_0)) = \left(\frac{1}{\sqrt{nh_n^{q+1}}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \right)^\tau \Omega_{nj}^-(t_0) \left(\frac{1}{\sqrt{nh_n^{q+1}}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \right) + o_p(1), \quad (4.12)$$

where $\Omega_{nj}^-(t_0)$ is the generalized inverse of $\Omega_{nj}(t_0)$. From Theorem 8, it can be shown in a straightforward manner that $\sqrt{nh_n^{q+1}} \nabla J_{jn}(G(t_0)) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathcal{W}_{1j}^{-1}(t_0))$, where $\mathcal{W}_{1j}(t_0) = \lim_{n \rightarrow \infty} \Omega_{nj}(t_0)$ with probability 1.

As in the discussion that follows Remark 4.1, it is often of interest to construct individual confidence intervals for $g_k(t_0)$, $k = 0, \dots, q$. This can be done by considering the marginal profile empirical for $g_k(t_0)$ as follows: as in Xue & Zhu (2007a), let $\mathbf{e}_k$ be the $(q + 1)$ -dimensional unit vector with k^{th} component equal to 1 and all the others equal

to 0. Set $\gamma_j(g_k(t_0)) = -h_n^q \mathbf{e}_k^\tau \nabla J_{jn}(G(t_0))$, $k = 0, \dots, q$. Specifically,

$$\gamma_j(g_k(t_0)) = \frac{1}{nh_n} \sum_{i=1}^n \mathbf{e}_k^\tau [\nabla G(t_{0i})]^\tau \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G(t_{0i})))}{n+1} \right) w_i(t_{0i}) = \frac{1}{nh_n} \sum_{i=1}^n \nu_{ij}(g_k(t_{0i})),$$

where $\nu_{ij}(g_k(t_{0i})) = \mathbf{e}_k^\tau [\nabla G(t_{0i})]^\tau \mathbf{z}_i \varphi \left(\frac{R(v_{ij}(G(t_{0i})))}{n+1} \right) w_i(t_{0i})$. Note that $\hat{g}_{jk}(t_0)$ solves $\gamma_j(g_k(t_0)) = 0$. So, considering some probability masses attached with each function, say ω_{ij} satisfying $\omega_{ij} \geq 0$ and $\sum_{i=1}^n \omega_{ij} = 1$, for $k = 0, \dots, q$, partial profil empirical log-likelihood functions can be defined as

$$S_j(g_k(t_0)) = -2 \sup_{\omega_{ij}} \left\{ \sum_{i=1}^n \log(n\omega_{ij}) : \sum_{i=1}^n \omega_{ij} = 1, \omega_{ij} \geq 0, \sum_{i=1}^n \omega_{ij} \nu_{ij}(g_k(t_{0i})) = 0 \right\}$$

Once again, a direct application of the Lagrange multiplier approach yields a maximum

$$\omega_{ij} = \frac{1}{n(1 + \mu \nu_{ij}(g_k(t_0)))},$$

where $\mu \in \mathbb{R}$ solves the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{\nu_{ij}(g_k(t_{0i}))}{1 + \mu \nu_{ij}(g_k(t_{0i}))} = 0. \quad (4.13)$$

Remark 4.3. From Theorem 12 together with Remark 4.2, a direct application of the Cr  mer Wold device (Serfling, 1980) entails that for $k = 0, \dots, q$,

$$\sqrt{nh_n} \mathcal{W}_{1j}^{-1/2}(t_0) \gamma_j(g_k(t_0)) = \mathbf{e}_k^\tau \mathcal{W}_{1j}^{-1/2}(t_0) \sqrt{nh_n^{q+1}} \nabla J_{jn}(G(t_0)) \xrightarrow{\mathcal{D}} N(0, 1), \quad \text{as } n \rightarrow \infty.$$

From this remark, we have the following theorem

Theorem 13. Under assumptions $(I_1) - (I_8)$ and $\omega_n \rightarrow 0$ as $n \rightarrow \infty$, if $g_k(t_0)$, $k = 0, \dots, q$ are the true coefficient functions, we have $S_j(g_k(t_0)) \xrightarrow{\mathcal{D}} \chi_1^2$ as $n \rightarrow \infty$, $j = 1, 2$.

The proof of this theorem follows directly from that of Theorem 12 by taking into account Remark 4.3. So, for sake of brevity, such a proof will not be included here.

To this end, $(1 - \alpha) \times 100\%$ partial confidence intervals for $g_k(t_0)$ are given by

$$\mathcal{C}_{5k}^j := \{g_k(t_0) : S_j(g_k(t_0)) \leq \chi_1^2(\alpha)\}, \quad k = 0, \dots, q,$$

where $\chi_1^2(\alpha)$ is the $(1 - \alpha)^{th}$ percentile of the χ^2 -distribution with 1 degree of freedom.

4.5 Simulation & Real Data

Simulation Study

We evaluate the performance of the proposed approach by considering a simple varying coefficient single index model given by

$$y = g_0(\mathbf{z}^T \boldsymbol{\beta}_0) + g_1(\mathbf{z}^T \boldsymbol{\beta}_0)x + \varepsilon,$$

where $g_0(t) = 3e^{-t^2}$, $g_1(t) = \frac{3}{2} \sin \pi t$, $\mathbf{z} = (z_1, z_2, z_3, z_4)^T$ is random vector uniformly distributed on $[-1, 1]^{\otimes 4}$, $x \sim N(0, 1)$ and $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^T = (1/3, 2/3, 0, 2/3)^T$. With samples of size $n = 100$, the model error is generated from three different distributions: the standard normal distribution ($\varepsilon \sim N(0, 1)$), the t -distribution with 3 degrees of freedom ($\varepsilon \sim t_3$) and the contaminated normal distribution with 10% contamination rate ($\varepsilon \sim CN(0.1, 10) = 0.9N(0, 1) + 0.1N(0, 10^2)$). The first model error is used to conduct a comparative study of the proposed method with respect to the LS method, while the two other model error distributions are used to investigate the effect of contamination and tail thickness on the proposed method. In the process of estimating the index and functional coefficients together with the estimation of the response probability $\pi(x, \mathbf{z})$, we consider the Epanechnikov Kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$. The missing indicator δ was generated as $\delta \sim \text{Bernoulli}(\pi(x, \mathbf{z}))$, where the response probability $\pi(x, \mathbf{z})$ was obtained from three different cases:

Case 1: $\pi(x, \mathbf{z}) = P(\delta = 1|x, \mathbf{z}) = 0.3 + |u(x, \mathbf{z}) - 0.25|$, if $|u(x, \mathbf{z}) - 0.25| \leq 0.25$, and 0.3 elsewhere,

Case 2: $\pi(x, \mathbf{z}) = P(\delta = 1|x, \mathbf{z}) = 0.5 + 0.5(|u(x, \mathbf{z}) - 1.5|)$, if $|u(x, \mathbf{z}) - 1.5| \leq 0.2$, and 0.5 elsewhere,

Case 3: $\pi(x, \mathbf{z}) = P(\delta = 1|x, \mathbf{z}) = 0.5 - |u(x, \mathbf{z}) - 1.5|$, if $|u(x, \mathbf{z}) - 1.5| \leq 0.25$, and 0.4 elsewhere,

where $u(x, \mathbf{z}) = (4)^{-1}|x| - \Phi(\mu(\mathbf{z}))$, with $\mu(\mathbf{z})$ being the mean of the vector $\mathbf{z}$ and Φ the standard normal cumulative distribution function. The three cases give on average 30%, 50%, and 40% response rates, respectively. In the considered objective function, the Wilcoxon score function, $\varphi(u) = \sqrt{12}(u-0.5)$, was considered. From 5000 replications and based on the normal approximation (NA) and the profile empirical likelihood (EL) with missing responses being imputed using either random simple imputation (SI) or inverse probability weighting imputation (IPW), coverage probabilities (CP) and average lengths (AL) of 95% confidence intervals for both the index coefficient β_{02} and the functional coefficient, $g_1(t_0)$, where $t_0 = \mathbf{z}^T \beta_0$, are calculated. These are compared with those obtained from two competing approaches: the LS and LAD. The simulation results are displayed in the following Tables 10-11 and Figures 8-11.

Tab. 10: CP (AL) of a 95% confidence interval for β_{02} under $\varepsilon \sim N(0, 1)$, t_3 , $CN(0.1, 10)$ with y imputed under random simple imputation (SI).

ε	cases	NALS-SI	ELLS-SI	NALAD-SI	ELLAD-SI	NAR-SI	ELR-SI	NAR-CC	ELR-CC
$N(0, 1)$	Case 1	92.88% (1.78)	94.65% (0.95)	90.69%(2.79)	93.13%(1.55)	92.71% (1.89)	94.56% (0.98)	92.25% (1.97)	94.27% (1.02)
	Case 2	93.37% (1.29)	94.87% (0.78)	91.89%(1.98)	93.97%(1.35)	93.29% (1.31)	94.83% (0.86)	93.15% (1.43)	94.78% (0.89)
	Case 3	93.08% (1.65)	94.81% (0.88)	90.91%(2.48)	93.36%(1.43)	92.96% (1.72)	94.77% (0.92)	92.93% (1.75)	94.48% (0.97)
t_3	Case 1	91.91% (2.59)	93.82% (1.76)	92.77%(1.72)	93.98%(1.38)	94.12% (1.09)	94.98% (0.74)	93.97% (1.28)	94.71% (0.97)
	Case 2	92.29% (2.16)	93.94% (1.44)	93.35%(1.36)	94.42%(1.02)	94.38% (0.93)	95.03% (0.54)	94.07% (1.08)	94.86% (0.79)
	Case 3	92.07% (2.37)	93.86% (1.63)	93.01%(1.68)	94.33%(1.18)	94.21% (1.03)	94.88% (0.68)	93.99% (1.19)	94.82% (0.86)
$CN(0.1, 10)$	Case 1	92.01% (2.48)	93.91% (1.65)	93.15%(1.57)	94.47%(1.13)	93.96% (1.01)	94.98% (0.62)	93.79% (1.13)	94.91% (0.91)
	Case 2	92.51% (2.04)	94.02% (1.33)	93.69%(1.21)	94.74%(0.98)	94.57% (0.84)	95.00% (0.48)	94.23% (0.97)	95.04% (0.76)
	Case 3	92.19% (2.34)	93.89% (1.55)	93.41%(1.35)	94.52%(1.08)	94.36% (0.91)	95.03% (0.57)	94.07% (1.09)	94.97% (0.83)

When it comes to the index coefficient β_{02} , under the standard normal distribution, as can be seen from the above tables, the LS under both the NA and EL approaches (NALS-SI, NALS-IPW, ELLS-SI and ELLS-IPW), performs better compared to the Rank (NAR-SI, NAR-IPW, ELR-SI and ELR-IPW) and LAD (NALAD-SI, NALAD-IPW, ELLAD-SI and ELLAD-IPW) approaches. It is worth noting that while both the LS and Rank have similar performance with a better performance by the LS, the LAD results in a worse performance. On the other hand, under t_3 and $CN(0.10)$, the Rank approach shows its dominance by providing CP that are closer to the nominal level and shorter

Tab. 11: CP (AL) of a 95% confidence interval for β_{02} under $\varepsilon \sim N(0, 1)$, t_3 , $CN(0.1, 10)$ with y imputed under inverse probability weighting imputation (IPW).

ε	cases	NALS-IPW	ELLS-IPW	NALAD-IPW	ELLAD-IPW	NAR-IPW	ELR-IPW	NAR-WCC	ELR-WCC
$N(0, 1)$	Case 1	93.28% (1.33)	94.94% (0.78)	91.95%(1.94)	93.72%(1.26)	93.06% (1.38)	94.81% (0.87)	93.17% (1.51)	94.67% (0.92)
	Case 2	93.78% (1.07)	95.02% (0.61)	92.53%(1.47)	94.19%(1.12)	93.57% (1.17)	94.97% (0.68)	93.48% (1.28)	94.89% (0.77)
	Case 3	93.35% (1.24)	94.99% (0.73)	92.27%(1.78)	93.88%(1.19)	93.27% (1.31)	94.88% (0.79)	93.21% (1.46)	94.76% (0.85)
t_3	Case 1	92.23% (2.42)	93.91% (1.64)	92.97%(1.56)	94.37%(1.17)	94.52% (0.93)	95.03% (0.65)	94.38% (1.09)	94.89% (0.83)
	Case 2	92.81% (1.99)	94.15% (1.25)	93.96%(1.18)	94.68%(0.91)	94.73% (0.82)	94.99% (0.49)	94.57% (0.94)	94.86% (0.67)
	Case 3	92.47% (2.28)	94.03% (1.43)	93.17%(1.51)	94.52%(0.87)	94.44% (0.86)	94.98% (0.54)	94.27% (1.01)	94.94% (0.78)
$CN(0.1, 10)$	Case 1	92.32% (2.31)	94.18% (1.52)	93.24%(1.43)	94.68%(1.09)	94.66% (0.87)	95.05% (0.59)	94.47% (1.06)	94.97% (0.76)
	Case 2	92.92% (1.87)	94.23% (1.13)	93.84%(0.82)	94.87%(0.79)	94.57% (0.84)	94.97% (0.41)	94.68% (0.90)	95.01% (0.63)
	Case 3	92.58% (2.19)	94.10% (1.33)	93.41%(1.27)	94.72%(0.81)	94.61% (0.85)	95.01% (0.53)	94.51% (0.97)	94.98% (0.71)

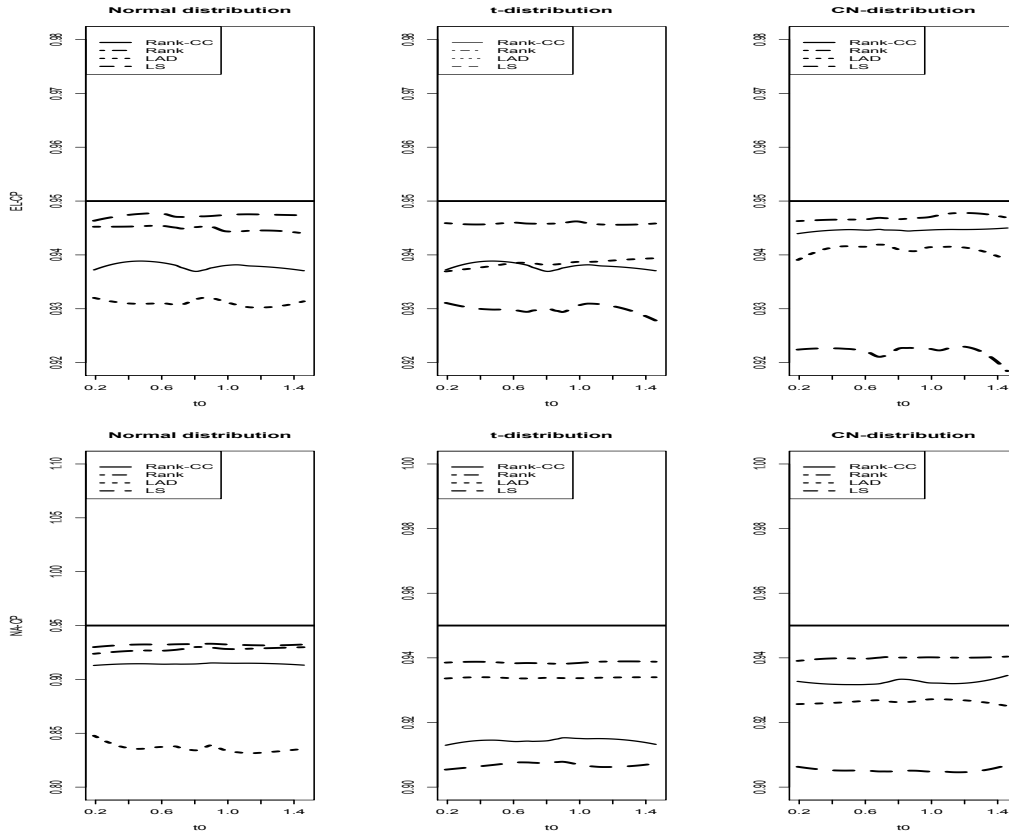


Fig. 8: Empirical likelihood coverage probabilities (EL-CP) and normal approximation coverage probabilities (NA-CP) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from SI under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$.

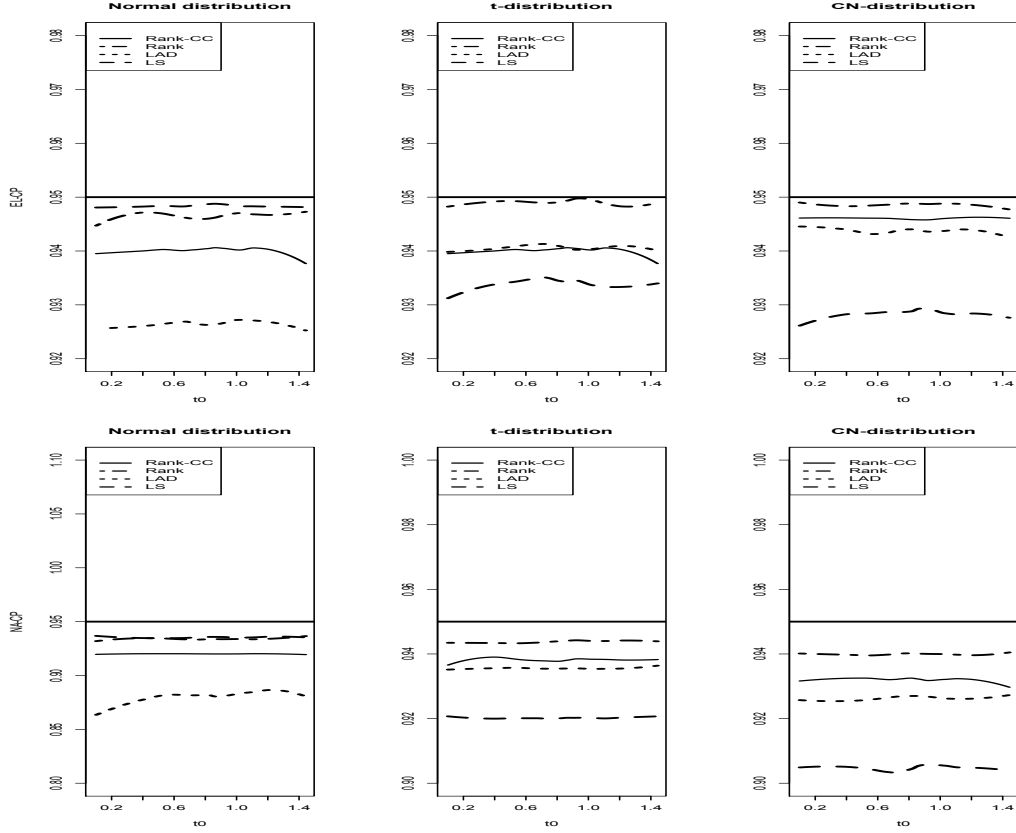


Fig. 9: Empirical likelihood coverage probabilities (EL-CP) and normal approximation coverage probabilities (NA-CP) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from IPW under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$.

AL compared to the LS and LAD methods. One also note that under these two model error distributions, the LS has a worse performance compared to the Rank and LAD. Generally, EL approaches perform better compared to their NA approaches counterpart. It is further observed that all the considered approaches perform better when missing responses are imputed using the inverse probability weighting imputation.

The proposed method is also compared with its complete case and weighted complete case counterparts. It is observed that the proposed approach with imputed missing responses performs better than the CC and WCC methods. This supports the importance of imputation, as imputation has improved the estimation inefficiency introduced by the

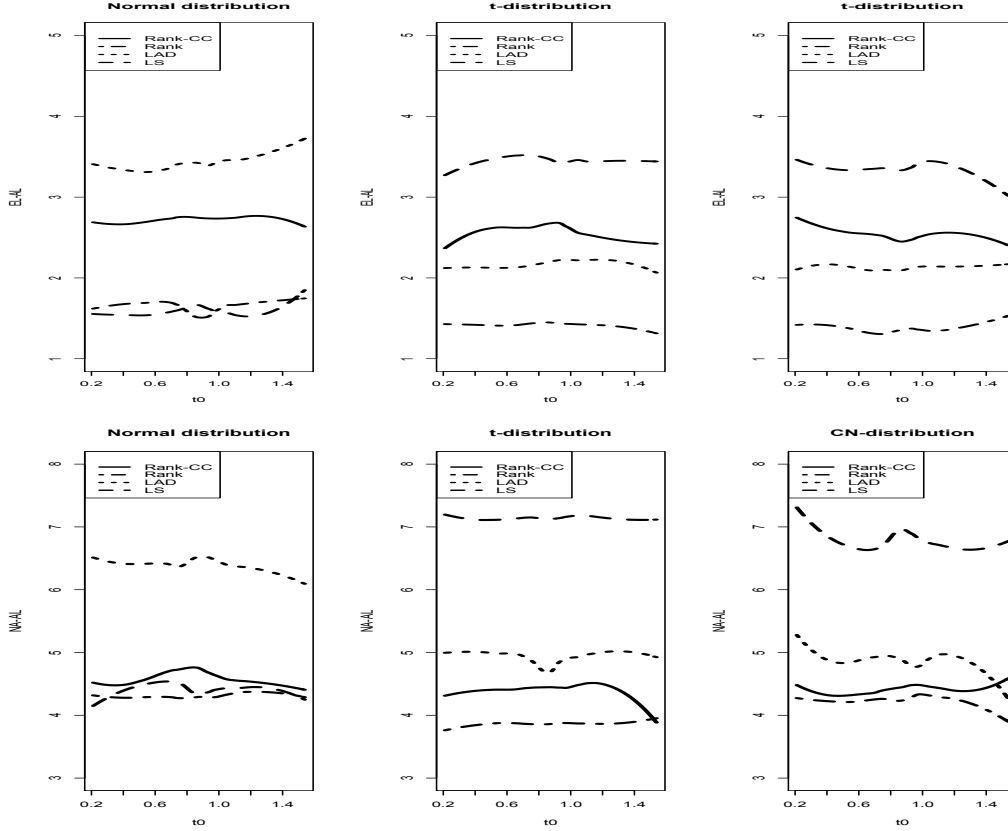


Fig. 10: Empirical likelihood average lengths (EL-AL) and normal approximation average lengths (NA-AL) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from SI under Case 2 and error distributions $N(0,1)$, t_3 , and $CN(0.1, 10)$.

complete case analysis through the NA approach.

For the functional coefficient, similar observations are made, as the Rank-EL approach provide better coverage probabilities and average lengths of a 95% confidence interval under t_3 and $CN(0.1, 10)$. It is also observed that for the standard normal model error distribution, the LS-EL slightly performs better compared to the Rank-EL approach, as would be expected, see Figures 8-11.

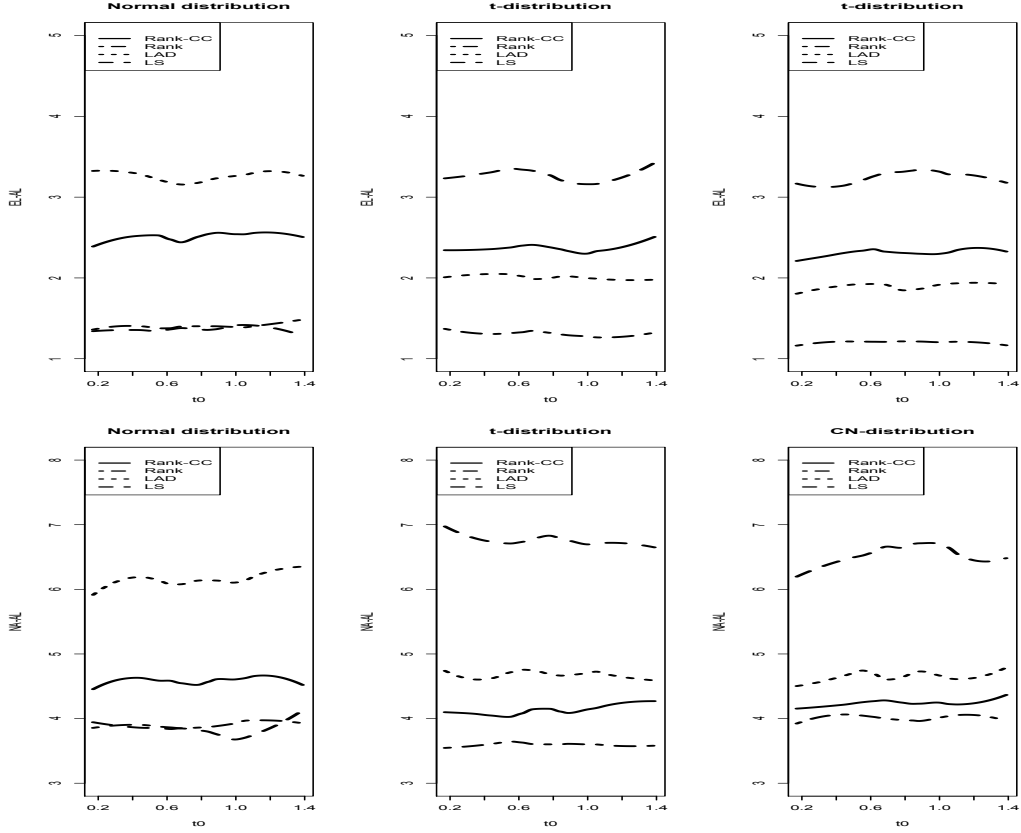


Fig. 11: Empirical likelihood average lengths (EL-AL) and normal approximation average lengths (NA-AL) of a 95% confidence interval for $g_1(t_0)$ with missing responses imputed from IPW under Case 2 and error distributions $N(0, 1)$, t_3 , and $CN(0.1, 10)$.

4.6 Discussion

In this chapter, we have proposed a rank-based empirical likelihood approach as alternative to the normal approximation method discussed in the previous chapter. As pointed out earlier, the proposed method does not require the estimation of rank-based estimators covariance matrices required in the normal approximation method for further possible statistical inference. Monte Carlo simulation has demonstrated that the rank-based EL approach is not only robust but provides more accurate confidence intervals/regions compared its normal approximation counterpart, it also dominates the LS and the LAD.

5 General Conclusion

This thesis has investigated two approaches of handling point and interval estimations of parameters in model (1.1) with responses missing at random. For the two approaches, one is based on the normal approximation and the other, on the empirical likelihood. The rank-based normal approximation introduced method, has discussed the parameters' estimation in model (1.1) with responses missing at random. Under some mild regularity conditions, asymptotic properties of the resulting estimators have been established. It is recognized that for further statistical inference on the true model parameters, the rank-based normal approximation method requires the estimation of the estimators' covariance matrices, which depending on the model structure can be a very difficult task. To avoid such a scenario, based on rank-based estimating equations, rank-based empirical likelihood functions have been defined. Under the same mild regularity conditions, these empirical likelihood functions have been proven to possess chi-square asymptotic distributions, from which robust confidence regions/intervals of the true model parameters have been derived. Based on extensive Monte Carlo simulation experiments, the proposed methods have been shown to be robust and more efficient compared to the least squares and least absolute deviations methods, when dealing with heavy tailed and contaminated model error distributions and/or when data contain gross outliers in the response space. The same simulation results have also shown that the proposed empirical likelihood approaches perform better compared to their normal approximations counterpart. It is worth pointing out that the proposed methods can be affected by outliers in the covariates space. In this case, weighted versions of the proposed objective functions should be considered; see Bindele & Abebe (2012). Generally, when dealing with model (1.1) under MAR responses for contaminated and heavy-tailed model error distributions, the work in this thesis has demonstrated that it would be preferable to use the proposed methods for better statistical inference. Finally, while this thesis has investigated model (1.1) with MAR responses, it may be of interest to extend the proposed work for model (1.1) with responses missing not at random and missing covariates.

References

- Allison, Paul D. 2000. Multiple imputation for missing data: A cautionary tale. *Sociological methods & research*, **28**(3), 301–309.
- Allison, Paul D. 2001. *Missing data*. Sage publications.
- Anderson, Paul J., & Piatt, John F. 1999. Community reorganization in the Gulf of Alaska following ocean climate regime shift. *Marine Ecology Progress Series*, **189**, 117–123.
- Andrews, Donald WK. 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 43–72.
- Aydin, K., Barbeaux, S.J., Barnard, D., Chilton, L., Clark, B., Conners, M.E., Conrath, C., Dalton, M., Echave, K., Fritz, L., Furuness, M., Hanselman, D., Haynie, A., Hoff, J., Honkalehto, T., Hulson, P.J., Ianelli, J., Kotwicki, S., Lauth, R., Lowe, S., Lunsford, C., McGilliard, C., McKelvey, D., Nichol, D., Norcross, B., Ormseth, O.A., Palsson, W., Rodgveller, C.J., Rooper, C.N., Spencer, P., Spies, I., Stockhausen, W., Stram, D., TenBrink, T., Thompson, G., Tribuzio, C., Wilderbuer, T., & Williamson, N. 2015 (November). *Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions*. Tech. rept. North Pacific Fishery Management Council.
- Bindele, Huybrechts F. 2015. The Signed-rank estimator for nonlinear regression with responses missing at random. *Electron. J. Statist.*, **9**(1), 1424–1448.
- Bindele, Huybrechts F. 2017. Strong consistency of the general rank estimator. *Communications in Statistics - Theory and Methods*, **46**(2), 532–539.
- Bindele, Huybrechts F., & Abebe, Ash. 2015. Semi-parametric rank regression with missing responses. *Journal of Multivariate Analysis*, **142**, 117 – 132.
- Bindele, Huybrechts F., & Abebe, Asheber. 2012. Bounded influence nonlinear signed-rank regression. *Canadian Journal of Statistics*, **40**(1), 172–189.

- Bindele, Huybrechts F., & Zhao, Yichuan. 2016. Signed-rank regression inference via empirical likelihood. *Journal of Statistical Computation and Simulation*, **86**(4), 729–739.
- Bindele, Huybrechts F, Abebe, Ash, & Meyer, Karlene N. 2018a. General rank-based estimation for regression single index models. *Annals of the Institute of Statistical Mathematics*, **70**(5), 1115–1146.
- Bindele, Huybrechts F., Abebe, Asheber, & Meyer, Nicole K. 2018b. Robust confidence regions for the semi-parametric regression model with responses missing at random. *Statistics*, **52**(4), 885–900.
- Brunner, Edgar, & Denker, Manfred. 1994. Rank statistics under dependent observations and applications to factorial designs. *Journal of Statistical Planning and Inference*, **42**(3), 353 – 378.
- Cai, Zongwu, Fan, Jianqing, & Li, Runze. 2000a. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, **95**(451), 888–902.
- Cai, Zongwu, Fan, Jianqing, & Yao, Qiwei. 2000b. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, **95**(451), 941–956.
- Center, Alaska Fisheries Science. 2019a. *AFSC/ABL: Longline Sablefish Survey*. <https://noaa-fisheries-afsc.data.socrata.com/dataset/AFSC-ABL-Longline-Sablefish-Survey/itxd-qjvg/data>. Accessed: 2014-04-14.
- Center, Alaska Fisheries Science. 2019b. *AFSC/RACE/GAP: RACEBASE Database*. http://www.afsc.noaa.gov/RACE/groundfish/survey_data/default.htm.
- Center, Alaska Fisheries Science. 2020 (July). *AFSC/ABL: Longline Sablefish Survey from 1979-05-24 to 2017-08-27*. <https://inport.nmfs.noaa.gov/inport/item/17218>. NOAA National Centers for Environmental Information.

- Center, National Data Buoy. 2018. *Meteorological and oceanographic data collected from the National Data Buoy Center Coastal-Marine Automated Network (C-MAN) and moored (weather) buoys*. <https://accession.nodc.noaa.gov/NDBC-CMANWx>.
- Chang, Ziqing, Xue, Liugen, & Zhu, Lixing. 2010. On an asymptotically more efficient estimation of the single-index model. *Journal of Multivariate Analysis*, **101**(8), 1898–1901.
- Chen, Jiahua, & Qin, Jing. 1993. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**(1), 107–116.
- Chen, Song Xi, & Hall, Peter. 1993. Smoothed Empirical Likelihood Confidence Intervals for Quantiles. *The Annals of Statistics*, **21**(3), pp. 1166–1181.
- Chiou, Jeng-Min, & Müller, Hans-Georg. 1998. Quasi-likelihood regression with unknown link and variance functions. *Journal of the American Statistical Association*, **93**(444), 1376–1387.
- Critical Data, MIT. 2016. *Secondary analysis of electronic health records*. Springer Nature.
- Delecroix, Michel, Hristache, Marian, & Patilea, Valentin. 2006. On semiparametric estimation in single-index regression. *Journal of Statistical Planning and Inference*, **136**(3), 730 – 769.
- Dong, Yiran, & Peng, Chao-Ying Joanne. 2013. Principled missing data methods for researchers. *SpringerPlus*, **2**(1), 1–17.
- Einmahl, Uwe, & Mason, David M. 2005. Uniform in Bandwidth Consistency of Kernel-Type Function Estimators. *The Annals of Statistics*, **33**(3), 1380–1403.
- Enders, Craig K. 2010. *Applied missing data analysis*. Guilford press.
- Fan, Jianqing, & Gijbels, Irene. 1996. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Vol. 66. CRC Press.
- Fan, Jianqing, Zhang, Wenyang, *et al.* . 1999. Statistical estimation in varying coefficient models. *Annals of Statistics*, **27**(5), 1491–1518.

- Fan, Jianqing, Yao, Qiwei, & Cai, Zongwu. 2003. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, **65**(1), 57–80.
- Feng, Sanying, & Xue, Liugen. 2015. Model detection and estimation for single-index varying coefficient model. *Journal of Multivariate Analysis*, **139**, 227–244.
- Fox-Wasylyshyn, Susan M, & El-Masri, Maher M. 2005. Handling missing data in self-report measures. *Research in nursing & health*, **28**(6), 488–495.
- Gao, Jiti, & Liang, Hua. 1997. Statistical inference in single-index and partially nonlinear models. *Annals of the Institute of Statistical Mathematics*, **49**(3), 493–517.
- Gu, Lijie, & Yang, Lijian. 2015. Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electronic Journal of Statistics*, **9**(1), 1540–1561.
- Guo, Xu, Niu, Cuizhen, Yang, Yiping, & Xu, Wangli. 2015. Empirical likelihood for single index model with missing covariates at random. *Statistics*, **49**(3), 588–601.
- Guo, Xu, Fang, Yun, Zhu, Xuehu, Xu, Wangli, & Zhu, Lixing. 2018. Semiparametric double robust and efficient estimation for mean functionals with response missing at random. *Computational Statistics & Data Analysis*, **128**, 325–339.
- Hájek, J., Šidák, Z., & Sen, P.K. 1999. *Theory of Rank Tests*. Probability and mathematical statistics. Academic Press.
- Hall, Peter, & La Scala, Barbara. 1990. Methodology and Algorithms of Empirical Likelihood. *International Statistical Review*, **58**(2), 109–127.
- Hansen, Bruce E. 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**(03), 726–748.
- Hardle, Wolfgang, Hall, Peter, & Ichimura, Hidehiko. 1993. Optimal Smoothing in Single-Index Models. *Ann. Statist.*, **21**(1), 157–178.
- Hastie, Trevor, & Tibshirani, Robert. 1993. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, **55**(4), 757–779.

- Hettmansperger, Thomas P., & McKean, Joseph W. 2011. *Robust Nonparametric Statistical Methods*. Second edn. Monographs on Statistics and Applied Probability, vol. 119. CRC Press, Boca Raton, FL.
- Hollowed, Anne Babcock, Ianelli, James N., & Livingston, Patricia A. 2000. Including predation mortality in stock assessments: a case study for Gulf of Alaska walleye pollock. *ICES Journal of Marine Science*, **57**(2), 279–293.
- Horowitz, Joel L., & Hardle, Wolfgang. 1996. Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates. *Journal of the American Statistical Association*, **91**(436), 1632–1640.
- Hristache, Marian, Juditsky, Anatoli, & Spokoiny, Vladimir. 2001. Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, 595–623.
- Ichimura, Hidehiko. 1987. *Estimation of single index models*. Ph.D. thesis, Massachusetts Institute of Technology.
- Ichimura, Hidehiko. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, **58**(1 - 2), 71 – 120.
- Jaekel, Louis A. 1972. Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, **43**, 1449–1458.
- Jing, Bing-Yi, Yuan, Junqing, & Zhou, Wang. 2009. Jackknife Empirical Likelihood. *Journal of the American Statistical Association*, **104**(487), 1224–1232.
- Johnson, N.L., Kotz, S., & Balakrishnan, N. 1995. *Continuous univariate distributions*. Wiley series in probability and mathematical statistics: Applied probability and statistics, no. v. 2. Wiley & Sons.
- Kitamura, Yuichi. 1997. Empirical Likelihood Methods with Weakly Dependent Processes. *The Annals of Statistics*, **25**(5), pp. 2084–2102.
- Krasker, William S., & Welsch, Roy E. 1982. Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.*, **77**(379), 595–604.

- Levin, Bruce, & Kong, Fanhui. 1990. Bartlett's Bias Correction to the Profile Score Function is a Saddlepoint Correction. *Biometrika*, **77**(1), 219–221.
- Li, Gaorong, Zhu, Lixing, Xue, Liugen, & Feng, Sanying. 2010. Empirical likelihood inference in partially linear single-index models for longitudinal data. *Journal of Multivariate Analysis*, **101**(3), 718–732.
- Li, Tingting, & Yang, Hu. 2016. Inverse probability weighted estimators for single-index models with missing covariates. *Communications in Statistics-theory and Methods*, **45**(5), 1199–1214.
- Little, Roderick J. A., & Rubin, Donald B. 2002. *Statistical Analysis with Missing Data*. Second edn. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- Little, Roderick JA. 1988a. Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, **6**(3), 287–296.
- Little, Roderick JA. 1988b. A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, **83**(404), 1198–1202.
- Little, Roderick JA. 1992. Regression with missing X's: a review. *Journal of the American Statistical Association*, **87**(420), 1227–1237.
- Little, Roderick JA, & Rubin, Donald B. 1983. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, **37**(3), 218–220.
- Little, Roderick JA, & Rubin, Donald B. 1989. The analysis of social science data with missing values. *Sociological Methods & Research*, **18**(2-3), 292–326.
- Little, Roderick JA, & Rubin, Donald B. 2019. *Statistical analysis with missing data*. Vol. 793. John Wiley & Sons.
- Liu, Jicai, Zhang, Riquan, Zhao, Weihua, & Lv, Yazhao. 2013. A robust and efficient estimation method for single index models. *J. Multivariate Anal.*, **122**, 226–238.

- Livingston, Patricia A. 1993. Importance of predation by groundfish, marine mammals and birds on walleye pollock *Theragra chalcogramma* and Pacific herring *Clupea pallasii* in the eastern Bering Sea. *Marine Ecology Progress Series*, **102**, 205–215.
- Lu, Xuewen, & Cheng, Tsung-Lin. 2007. Randomly censored partially linear single-index models. *Journal of Multivariate Analysis*, **98**(10), 1895–1922.
- Luo, Xinchao, Zhu, Lixing, & Zhu, Hongtu. 2016. Single-index varying coefficient model for functional responses. *Biometrics*, **72**(4), 1275–1284.
- Nagler, Thomas, & Czado, Claudia. 2016. Evading the curse of dimensionality in non-parametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, **151**, 69 – 89.
- Naik, Prasad, & Tsai, Chih-Ling. 2000. Partial least squares estimator for single-index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(4), 763–771.
- National Marine Fisheries Service, NOAA. 2014. *The Coastal & Oceanic Plankton Ecology, Production, & Observation Database*. <http://www.st.nmfs.noaa.gov/copepod>.
- Newey, Whitney K. 1991. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, 1161–1167.
- O’Brien, T.D. 2007. *COPEPOD: The Global Plankton Database. A review of the 2007 database contents and new quality control methodology*. Tech. rept. NOAA Tech. Memo. NMFS-F/ST-34. U.S. Dep. Commerce.
- Owen, Art B. 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**(2), 237–249.
- Owen, Art B. 1990. Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics*, **18**(1), 90–120.
- Peng, Liang. 2004. Empirical-Likelihood-Based Confidence Interval for the Mean with a Heavy-Tailed Distribution. *The Annals of Statistics*, **32**(3), pp. 1192–1214.

- Qin, Jin, & Lawless, Jerry. 1994. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, **22**(1), 300–325.
- Qin, Jing, Yu, Tao, Li, Pengfei, Liu, Hao, & Chen, Baojiang. 2019. Using a monotone single-index model to stabilize the propensity score in missing data problems and causal inference. *Statistics in medicine*, **38**(8), 1442–1458.
- Rao, BLS Prakasa. 2009. Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics*, **61**(2), 441–460.
- Robins, James M., Rotnitzky, Andrea, & Zhao, Lue Ping. 1995. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*, **90**(429), 106–121.
- Royston, Patrick. 1992. Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and computing*, **2**(3), 117–119.
- Rubin, Donald B. 1976. Inference and missing data. *Biometrika*, **63**(3), 581–592.
- Scheffer, Judi. 2002. Dealing with missing data. *Res. Lett. Inf. Math. Sci.*, **3**, 153–160.
- Serfling, Robert J. 1980. *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.
- Sievers, Gerald L. 1983. A weighted dispersion function for estimation in linear models. *Comm. Statist. A—Theory Methods*, **12**(10), 1161–1179.
- Sievers, Gerald L., & Abebe, Asheber. 2004. Rank estimation of regression coefficients using iterated reweighted least squares. *J. Stat. Comput. Simul.*, **74**(11), 821–831.
- Song, Yunquan, Jian, Ling, & Lin, Lu. 2016. Robust exponential squared loss-based variable selection for high-dimensional single-index varying-coefficient model. *Journal of Computational and Applied Mathematics*, **308**, 330–345.
- Sun, Wei, Bindele, Huybrechts F, Abebe, Ash, & Correia, Hannah. 2019. General local rank estimation for single-index varying coefficient models. *Journal of Statistical Planning and Inference*, **202**, 57–79.

- Sun, Zhihua, Wang, Qihua, & Dai, Pengjie. 2009. Model checking for partially linear models with missing responses at random. *J. Multivar. Anal.*, **100**(4), 636–651.
- Switzer III, Fred S, & Roth, Philip L. 2002. Coping with missing data. *Journal of Applied Psychology*.
- Wang, C. Y., Wang, Suojin, Gutierrez, Roberto G., & Carroll, R. J. 1998. Local linear regression for generalized linear models with missing data. *Ann. Statist.*, **26**(3), 1028–1050.
- Wang, Guannan, & Wang, Li. 2015. Spline estimation and variable selection for single-index prediction models with diverging number of index parameters. *Journal of Statistical Planning and Inference*, **162**(Supplement C), 1 – 19.
- Wang, Li, & Yang, Lijian. 2009. Spline estimation of single-index models. *Statistica Sinica*, 765–783.
- Wang, Qihua, & Rao, J. N. K. 2002. Empirical Likelihood-Based Inference under Imputation for Missing Response Data. *The Annals of Statistics*, **30**(3), pp. 896–924.
- Wang, Qihua, Linton, Oliver, & Härdle, Wolfgang. 2004. Semiparametric Regression Analysis with Missing Response at Random. *Journal of the American Statistical Association*, **99**(466), 334–345.
- Wang, Yanhua, Shen, Junshan, He, Shuyuan, & Wang, Qihua. 2010. Estimation of single index model with missing response at random. *Journal of Statistical Planning and Inference*, **140**(7), 1671 – 1690.
- Wayman, Jeffrey C. 2003. Multiple imputation for missing data: What is it and how can I use it. *Page 16 of: Annual Meeting of the American Educational Research Association, Chicago, IL*, vol. 2.
- Whitt, Ward. 2002. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer Science & Business Media.
- Wong, Heung, Ip, Wai-cheung, & Zhang, Riquan. 2008. Varying-coefficient single-index model. *Computational statistics & data analysis*, **52**(3), 1458–1476.

- Wu, Colin O, Chiang, Chin-Tsang, & Hoover, Donald R. 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American statistical Association*, **93**(444), 1388–1402.
- Wu, Lang, & Wu, Hulin. 2002. Missing time-dependent covariates in human immunodeficiency virus dynamic models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **51**(3), 297–318.
- Xia, Yingcun. 2006. Asymptotic distributions for two estimators of the single-index model. *Econometric Theory*, **22**(6), 1112–1137.
- Xia, Yingcun, & Li, Wai Keung. 1999. On single-index coefficient regression models. *Journal of the American Statistical Association*, **94**(448), 1275–1285.
- Xia, Yingcun, Tong, Howell, Li, W. K., & Zhu, Li-Xing. 2002. An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(3), 363–410.
- Xia, Yingcun, Li, W, & Tong, Howell. 2007. Threshold variable selection using nonparametric methods. *Statistica Sinica*, **17**(1), 265.
- Xue, Liu-Gen, & Zhu, Lixing. 2006. Empirical likelihood for single-index models. *Journal of Multivariate Analysis*, **97**(6), 1295 – 1312.
- Xue, Liugen. 2009. Empirical Likelihood Confidence Intervals for Response Mean with Data Missing at Random. *Scandinavian Journal of Statistics*, **36**(4), 671–685.
- Xue, Liugen, & Pang, Zhen. 2013. Statistical inference for a single-index varying-coefficient model. *Statistics and Computing*, **23**(5), 589–599.
- Xue, Liugen, & Zhu, Lixing. 2007a. Empirical Likelihood for a Varying Coefficient Model With Longitudinal Data. *Journal of the American Statistical Association*, **102**(478), 642–654.
- Xue, Liugen, & Zhu, Lixing. 2007b. Empirical Likelihood Semiparametric Regression Analysis for Longitudinal Data. *Biometrika*, **94**(4), 921–937.

- Xue, Liugen, Wang, Qihua, *et al.* . 2012. Empirical likelihood for single-index varying-coefficient models. *Bernoulli*, **18**(3), 836–856.
- Yang, Guangren, Cui, Xia, & Hou, Sumin. 2017. Empirical likelihood confidence regions in the single-index model with growing dimensions. *Communications in Statistics-Theory and Methods*, **46**(15), 7562–7579.
- Yang, Hanfang, & Zhao, Yichuan. 2012. New empirical likelihood inference for linear transformation models. *Journal of Statistical Planning and Inference*, **142**(7), 1659–1668.
- Yang, Hu, & Yang, Jing. 2014. A robust and efficient estimation and variable selection method for partially linear single-index models. *Journal of Multivariate Analysis*, **129**, 227–242.
- You, Jinhong, & Zhou, Yong. 2006. Empirical likelihood for semiparametric varying-coefficient partially linear regression models. *Statistics & Probability Letters*, **76**(4), 412–422.
- Yu, Yan, & Ruppert, David. 2002. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, **97**(460), 1042–1054.
- Zeng, Peng, He, Tianhong, & Zhu, Yu. 2012. A Lasso-type approach for estimation and variable selection in single index models. *Journal of Computational and Graphical Statistics*, **21**(1), 92–109.
- Zhao, Yang, Xue, Liugen, Zhang, Jinghua, & Liu, Juanfang. 2020. Single-index varying-coefficient models with missing covariates at random. *Communications in Statistics-Simulation and Computation*, 1–15.
- Zhu, Li-Xing, & Fang, Kai-Tai. 1996. Asymptotics for Kernel Estimate of Sliced Inverse Regression. *The Annals of Statistics*, **24**(3), pp. 1053–1068.
- Zhu, Lixing, & Xue, Liugen. 2006. Empirical likelihood confidence regions in a partially linear single-index model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**(3), 549–570.

6 Appendices

These appendices are introduced to elaborate on all technical details through the entire thesis. It contains necessary lemmas needed in the proofs of theoretical results stated in Chapters 2, 3 and 4.

6.1 Appendix chapter 2

The following lemmas are necessary and sometimes sufficient in the establishment of the proofs of some of the main results stated in this thesis.

Lemma 6.1. *Let $\{A_n(\alpha)\}_{n \leq 1}$ be a stochastically equicontinuous random objective function defined on a compact space Θ such that $\hat{\alpha}_n = \text{Argmin}_{\alpha} A_n(\alpha)$ and for fixed n , there is a unique $\alpha_{0,n} \in \Theta$ satisfying $\alpha_{0,n} = \text{Argmin}_{\alpha} E[A_n(\alpha)]$ with $E[A_n(\alpha)]$ being continuous with respect to α . For $\alpha_0 \in \Theta$, let $\alpha_0 = \lim_{n \rightarrow \infty} \alpha_{0,n}$. Then, if for every $\alpha_0 \in \Theta$, $A_n(\alpha) - E[A_n(\alpha)] \rightarrow 0$ a.s. as $n \rightarrow \infty$, we have $\hat{\alpha}_n \rightarrow \alpha_0$ a.s. as $n \rightarrow \infty$.*

The proof of this lemma can be found in Andrews (1994) and Newey (1991), and is sufficient for the proof of Theorem 2.

Lemma 6.2. *Under assumptions $(\mathcal{J}_2) - (\mathcal{J}_6)$, we have*

$$\lim_{n \rightarrow \infty} \sup_{u \in \mathcal{F}} |\hat{g}^{(r)}(u) - g^{(r)}(u)| \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty, \quad r = 0, 1, 2.$$

This lemma whose proof can be constructed along the lines as given in Delecroix *et al.* (2006) ensures the uniform strong consistency of $\hat{g}$ with its derivatives. For sake of brevity the proof is omitted here.

Lemma 6.3. *Under assumptions $(\mathcal{J}_1) - (\mathcal{J}_7)$ and for $\mathbf{x} \in \Gamma$, we have*

$$\begin{aligned} \text{a. } & \lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \left\| \tilde{S}_{jn}(\beta) - S_{jn}(\beta) \right\| = 0 \text{ a.s. and } \frac{1}{n} \sum_{i=1}^n w_i \{g'(\mathbf{x}_i^T \beta_0)\}^2 \mathbf{x}_i \mathbf{x}_i^T \varphi^2(F_{ij}(\xi_{ij})) \rightarrow \mathbf{V}_j \text{ a.s.} \\ \text{b. } & \lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \left\| \tilde{S}_{jn}(\beta) - \tilde{T}_{jn}(\beta) \right\| = 0 \text{ a.s. and } \sqrt{n} \tilde{T}_{jn}(\beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V}_j) \text{ as } n \rightarrow \infty. \end{aligned}$$

This lemma shows that $\tilde{S}_{jn}(\boldsymbol{\beta})$ and $\tilde{T}_{jn}(\boldsymbol{\beta})$ are asymptotically equivalent. It can further be shown that $\sqrt{n}(\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)) \xrightarrow{P} 0$, which implies that $\sqrt{n}\tilde{S}_{jn}(\boldsymbol{\beta}_0)$ and $\sqrt{n}\tilde{T}_{jn}(\boldsymbol{\beta}_0)$ have the same asymptotic distribution (Hettmansperger & McKean, 2011). As a consequence of Lemma 6.3, with probability 1, $\tilde{S}_n(\boldsymbol{\beta}) = \tilde{T}_{jn}(\boldsymbol{\beta}) + o(1/n)$. A Taylor expansion of $\tilde{T}_{jn}(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$ gives

$$\tilde{T}_{jn}(\boldsymbol{\beta}) = \tilde{T}_{jn}(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau \nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau \nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\zeta})(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

where $\boldsymbol{\zeta}$ belongs in the line segment joining $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}$. Thus, with probability 1,

$$\tilde{S}_{jn}(\boldsymbol{\beta}) = \tilde{T}_{jn}(\boldsymbol{\beta}_0) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau \nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\tau \nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\zeta})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(1/n).$$

$\hat{\boldsymbol{\beta}}_n^j$ being a solution of $\tilde{S}_{jn}(\boldsymbol{\beta}) = 0$, we have

$$0 = \tilde{S}_{jn}(\hat{\boldsymbol{\beta}}_n^j) = \tilde{T}_{jn}(\boldsymbol{\beta}_0) + \nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) \cdot (\hat{\boldsymbol{\beta}}_n^j - \boldsymbol{\beta}_0) + \frac{1}{2}(\hat{\boldsymbol{\beta}}_n^j - \boldsymbol{\beta}_0)^\tau \cdot \nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\zeta}_n^j) \cdot (\hat{\boldsymbol{\beta}}_n^j - \boldsymbol{\beta}_0) + o(1/n), \quad (6.1)$$

where $\boldsymbol{\zeta}_n^j = \lambda \boldsymbol{\beta}_0 + (1 - \lambda) \hat{\boldsymbol{\beta}}_n^j$.

Lemma 6.4. *Under assumptions $(I_1) - (I_7)$ and for any $\mathbf{x} \in \Gamma$, the following hold:*

- a. $\nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) \rightarrow \mathbf{W}_j$ a.s.
- b. $\nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\zeta}_n^j)$ is almost surely bounded.

Proof of Lemma 6.3. Let $a_{ijn} = R(\xi_{ij}(\boldsymbol{\beta})) / (n + 1)$ and let c be any arbitrary positive constant. Then, $S_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i^\tau \hat{g}'(\mathbf{x}_i^\tau \boldsymbol{\beta}) \varphi(a_{ijn})$.

a. For any $\mathbf{x} \in \Gamma$ and $\boldsymbol{\beta} \in \mathcal{B}$,

$$\|\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta})\| \leq \frac{1}{n} \sum_{i=1}^n w_i \|\mathbf{x}_i\| |\varphi(a_{ijn})| |\hat{g}'(\mathbf{x}_i^\tau \boldsymbol{\beta}) - g'(\mathbf{x}_i^\tau \boldsymbol{\beta})|$$

Since φ is bounded, there exists a positive constant c such that $|\varphi(t)| \leq c$, for any $t \in (0, 1)$. Also, by definition, $w_i \leq 1$, for any i . This together with Cauchy-Swartz

inequality imply that

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta})\| \leq c \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \right]^{1/2} \left[\sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\hat{g}'(\mathbf{x}_i^\tau \boldsymbol{\beta}) - g'(\mathbf{x}_i^\tau \boldsymbol{\beta})|^2 \right]^{1/2}.$$

By Lemma 6.2, $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\hat{g}'(\mathbf{x}_i^\tau \boldsymbol{\beta}) - g'(\mathbf{x}_i^\tau \boldsymbol{\beta})| = 0$ a.s. and by SLLN, $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2$ converges almost surely to $E[\|\mathbf{x}\|^2] < \infty$ a.s., by assumption (I_6) . Hence,

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta})\| = 0 \text{ a.s.}$$

b. Also,

$$\tilde{S}_{jn}(\boldsymbol{\beta}) - \tilde{T}_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \boldsymbol{\beta}) [\varphi(a_{ijn}) - \varphi(F_{ij}(\xi_{ij}(\boldsymbol{\beta})))] .$$

Recall that following Hájek *et al.* (1999), $a_{ijn} - F_{ij}(\xi_{ij}(\boldsymbol{\beta})) \rightarrow 0$ a.s. as $n \rightarrow \infty$, for all $\boldsymbol{\beta} \in \mathcal{B}$ and for all $i, j = 1, 2$. So, by the generalized continuous mapping theorem (Whitt, 2002),

$$\max_{1 \leq i \leq n} |\varphi(a_{ijn}) - \varphi(F_{ij}(\xi_{ij}(\boldsymbol{\beta})))| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

From the boundedness of $g'(t)$ and the fact that $w_i \leq 1$, for all i , we have

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{S}_{jn}(\boldsymbol{\beta}) - \tilde{T}_{jn}(\boldsymbol{\beta})\| \leq \left[\frac{1}{n} \sum_{i=1}^n \{\|\mathbf{x}_i\| J(\mathbf{x}_i)\}^2 \right]^{1/2} \left[\sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\varphi(a_{ijn}) - \varphi(F_{ij}(\xi_{ij}(\boldsymbol{\beta})))|^2 \right]^{1/2},$$

by assumption $(\mathcal{J}_3)$ and Cauchy-Schwarz inequality. Thus,

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{S}_{jn}(\boldsymbol{\beta}) - \tilde{T}_{jn}(\boldsymbol{\beta})\| = 0 \text{ a.s.},$$

since the SLLN ensures that $\frac{1}{n} \sum_{i=1}^n \{\|\mathbf{x}_i\| J(\mathbf{x}_i)\}^2 \rightarrow E(\{\|\mathbf{x}\| J(\mathbf{x})\}^2) < \infty$ a.s., by assumption $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$ and,

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\varphi(a_{ijn}) - \varphi(F_{ij}(\xi_{ij}(\boldsymbol{\beta})))|^2 \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

Next,

$$E[\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)] = \frac{1}{n} \sum_{i=1}^n E \left[w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0) \left(\varphi \left(\frac{R(\xi_j(\beta_0))}{n+1} \right) - \varphi(F_j(\xi_j(\beta_0))) \right) \right],$$

from which Cauchy-Shwarz inequality gives

$$\begin{aligned} \|E[\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)]\| &\leq \\ \left[\frac{1}{n} \sum_{i=1}^n E\{(w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0))^2\} \right]^{1/2} &\left[E \left(\left| \varphi \left(\frac{R(\xi_j(\beta_0))}{n+1} \right) - \varphi(F_j(\xi_j(\beta_0))) \right|^2 \right) \right]^{1/2}. \end{aligned}$$

From the discussion above,

$$\left| \varphi \left(\frac{R(\xi_j(\beta_0))}{n+1} \right) - \varphi(F_j(\xi_j(\beta_0))) \right|^2 \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

This together with the generalized dominated convergence theorem gives

$$E \left(\left| \varphi \left(\frac{R(\xi_j(\beta_0))}{n+1} \right) - \varphi(F_j(\xi_j(\beta_0))) \right|^2 \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus, $E[\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)] \rightarrow 0$ as $n \rightarrow \infty$. By Chebychev's inequality, for any $\epsilon > 0$, we have

$$P \left(\sqrt{n}(\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)) > \epsilon \right) \leq \epsilon^{-2} E \left[n(\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0))^2 \right].$$

To complete the proof, it suffices to show that $E[n(S_n(\beta_0) - T_n(\beta_0))^2] \rightarrow 0$ as $n \rightarrow \infty$.

Following Hájek *et al.* (1999) page 192,

$$\begin{aligned} E \left[n(\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0))^2 \right] &= \frac{1}{n} E \left[\left\{ \sum_{i=1}^n w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0) \left(\varphi \left(\frac{R(\xi_{ij}(\beta_0))}{n+1} \right) - \varphi(F_j(\xi_{ij}(\beta_0))) \right) \right\}^2 \right] \\ &\leq \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^n E[(w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0))^2] \right\} E \left[\left(\varphi \left(\frac{R(\xi_j)}{n+1} \right) - \varphi(F_j(\xi_j)) \right)^2 \right] \end{aligned}$$

From the SLLN, it is not hard to see that

$$\frac{1}{n} \sum_{i=1}^n (w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0))^2 \leq \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 J^2(\mathbf{x}_i) \rightarrow E[\|\mathbf{x}\|^2 J^2(\mathbf{x})] < \infty \text{ a.s.},$$

by assumption $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$. This implies that the right hand side the above inequality converges to zero. Thus,

$$E \left[n (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0))^2 \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

and therefore,

$$\lim_{n \rightarrow \infty} P \left(\sqrt{n} (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)) > \epsilon \right) = 0.$$

Now, by assumption $(\mathcal{J}_7)$, we have $E\{\tilde{S}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$ as $n \rightarrow \infty$. Since

$$E\{\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0,$$

we have $E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$ as $n \rightarrow \infty$. Recall that $\tilde{T}_{jn}(\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \varphi(F_j(\xi_j)) = \frac{1}{n} \sum_{i=1}^n V_{ij}$, where $V_{ij} = w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \varphi[F_j(\xi_j)]$. To obtain the asymptotic distribution of $\tilde{T}_{jn}(\boldsymbol{\beta}_0)$, we employ the Cramér-Wold device (Serfling, 1980). To this end, set

$$U_{jn} = n^{-1/2} \sum_{i=1}^n \mathbf{a}^T V_{ij} = \mathbf{a}^T \tilde{T}_n(\boldsymbol{\beta}_0),$$

where $\mathbf{a} \in \mathbb{R}^q$. Since $E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$ and $E(U_{jn}) = \mathbf{a}^T E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\}$, we have $E(U_{jn}) \rightarrow 0$. From the fact that $(\mathbf{x}_i, y_i, \varepsilon_i)$, $i = 1, \dots, n$, is a random sample, $(\mathbf{x}_i, z_{ij}, \xi_{ij})$ is also a random sample. Conditional on $\mathbf{x}_i$, $\{V_{nj}\}_{n \geq 1}$ is a sequence of conditional independent random variables with identical distribution (Rao, 2009).

$$\text{Var}(U_{jn}) = \frac{1}{n} \sum_{i=1}^n E \left[(w_i \mathbf{a}^T \mathbf{x}_i g'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0))^2 E \{ \varphi^2(F_j(\xi_j)) | \mathbf{x}_i \} \right] \rightarrow \mathbf{a}^T \mathbf{V}_j \mathbf{a} \quad \text{as } n \rightarrow \infty.$$

Note that $\tilde{T}_{jn}(\boldsymbol{\beta}_0)$ is the sum of identically distributed conditionally independent functions of random variables. Hence, the limiting distribution is established by verifying the conditional Lindeberg-Feller condition for the applicability of the conditional Central

Limit Theorem. To do this, let $\zeta_{jn} = \sqrt{n}\tilde{T}_{jn}(\beta_0)$ and set $\sigma_{jn}^2 = \text{Var}(\zeta_{jn})$. We need to show that for any $\mu > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{nj}^2} \sum_{i=1}^n E \left[\frac{1}{n} (w_i g'(\mathbf{x}_i^\tau \beta_0))^2 E\{\varphi^2[F_j(\xi_j)|\mathbf{x}_i]\} \right] \times I \left(\left| \frac{1}{\sqrt{n}} (w_i g'(\mathbf{x}_i^\tau \beta_0)) \varphi[F_j(\xi_j)] \right| > \mu \sigma_{nj} \right) = 0.$$

To see this, note that since $w_i \leq 1$, for all i , we have

$$\frac{1}{\sqrt{n}} |w_i \mathbf{a}^\tau \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0)| \leq n^{-1/2} \|\mathbf{a}\| \|\mathbf{x}_i\| J(\mathbf{x}_i) \rightarrow 0 \text{ a.s. as } n \rightarrow \infty,$$

since $\|\mathbf{x}\|J(\mathbf{x})$ is almost surely finite, as $E(\|\mathbf{x}\|J(\mathbf{x})) < \infty$, by $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$. Set

$$\lambda_n = \left[\max_{1 \leq i \leq n} n^{-1} |w_i \mathbf{a}^\tau \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0)| \right]^{1/2}.$$

Then $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$, and, is independent of i . Since σ_{jn}^2 converges to a positive quantity, the ratio $\sigma_{jn}/\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$. Now, it is easy to see that

$$\sum_{i=1}^n E \left[\frac{1}{n} (w_i \mathbf{a}^\tau \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0))^2 E\{\varphi^2[F_j(\xi_{ij})|\mathbf{x}_i]\} I \left(\left| \frac{1}{\sqrt{n}} (w_i \mathbf{a}^\tau \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0)) \varphi[F_j(\xi_{ij})] \right| > \mu \sigma_{jn} \right) \right]$$

is less than or equal to

$$E \left\{ E \left[\varphi^2[F_j(\xi_j)] I \left(|\varphi[F_j(\xi_j)]| > \mu \sigma_{jn}/\lambda_n \right) \middle| \mathbf{x} \right] \right\} \times \frac{1}{n} \sum_{i=1}^n E [(w_i \mathbf{a}^\tau \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0))^2]. \quad (6.2)$$

In Equation (6.2), the limit of the second term, that is,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E [(w_i \mathbf{x}_i g'(\mathbf{x}_i^\tau \beta_0))^2] \leq E(\|\mathbf{x}\|^2 J^2(\mathbf{x})) < \infty.$$

From the boundedness of φ and applying the Dominated Convergence Theorem to the first term in Equation (6.2), we have

$$E \left\{ E \left[\varphi^2[F_j(\xi_j)] I \left(|\varphi[F_j(\xi_j)]| > \mu \sigma_{jn}/\lambda_n \right) \middle| \mathbf{x} \right] \right\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This established the conditional Lindeberg-Feller condition and thus, the conditional

Central Limit Theorem (Rao, 2009) gives $\sqrt{n}\tilde{T}_{jn}(\beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V}_j)$ as $n \rightarrow \infty$. $\square$

Proof of Lemma 6.4. Differentiating $\tilde{T}_{jn}(\beta)$ with respect to β , we have

$$\begin{aligned}\nabla_{\beta}\tilde{T}_{jn}(\beta) &= -\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} [g'(\mathbf{x}_i^{\tau} \beta)]^2 f_{ij}(\xi_{ij}(\beta)) \varphi'(F_{ij}(\xi_{ij}(\beta))) \\ &\quad + \frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} g''(\mathbf{x}_i^{\tau} \beta) \varphi(F_{ij}(\xi_{ij}(\beta))).\end{aligned}$$

At $\beta = \beta_0$, we have

$$\nabla_{\beta}\tilde{T}_{jn}(\beta_0) = -\frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} [g'(\mathbf{x}_i^{\tau} \beta_0)]^2 f_j(\xi_j) \varphi'(F_j(\xi_j)) + \frac{1}{n}\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} g''(\mathbf{x}_i^{\tau} \beta_0) \varphi(F_j(\xi_j)).$$

Note that

$$E[w_i \mathbf{x}_i \mathbf{x}_i^{\tau} [g'(\mathbf{x}_i^{\tau} \beta_0)]^2 E\{f_j(\xi_j) \varphi'(F_j(\xi_j)) | \mathbf{x}_i\}] = E[w \mathbf{x} \mathbf{x}^{\tau} \{g'(\mathbf{x}^{\tau} \beta_0)\}^2 \mu_j(\mathbf{x})] < \infty,$$

by assumptions $(\mathcal{J}_1)$, $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$. Also,

$$E[w_i \mathbf{x}_i \mathbf{x}_i^{\tau} g''(\mathbf{x}_i^{\tau} \beta_0) E\{\varphi(F_j(\xi_j)) | \mathbf{x}_i\}] = E[w_i \mathbf{x}_i \mathbf{x}_i^{\tau} g''(\mathbf{x}_i^{\tau} \beta_0) \nu_j(\mathbf{x}_i)],$$

which is also finite by assumptions $(\mathcal{J}_1)$, $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$. Now, a direct application of the SLLN, gives $\nabla_{\beta}\tilde{T}_{jn}(\beta_0) \rightarrow \mathbf{W}_j$ a.s. If ε is independent of $\mathbf{x}$, ξ_j is independent of $\mathbf{x}$. Hence,

$$\mu_j(\mathbf{x}) = E[f_j(\xi_j) \varphi'(F_j(\xi_j)) | \mathbf{x}] = E[f_j(\xi_j) \varphi'(F_j(\xi_j))] = \int f_j(\xi_j) \varphi'(F_j(\xi_j)) dF_j(\xi_j).$$

From integration by parts, $\int f_j(\xi_j) \varphi'(F_j(\xi_j)) dF_j(\xi_j) = -\int_{-\infty}^{\infty} f'_j(\xi_j) \varphi(F_j(\xi_j)) d\xi_j$ since $f_j(\xi_j) \varphi(F_j(\xi_j)) \rightarrow 0$ as $\xi_j \rightarrow \pm\infty$. Putting $x^j = F_j(\xi_j)$, we have

$$\int_{-\infty}^{\infty} f'_j(\xi_j) \varphi(F_j(\xi_j)) d\xi_j = -\int_0^1 \varphi(x^j) \varphi_{f_j}(x^j) dx^j = -\gamma_{j\varphi}^{-1}. \quad (6.3)$$

Also, $\nu_j(\mathbf{x}) = E[\varphi(F_j(\xi_j))|\mathbf{x}]$, from which putting $x^j = F_j(\xi_j)$, $dx^j = dF_j(\xi_j)$ and

$$\nu_j(\mathbf{x}) = E[\varphi(F_j(\xi_j))|\mathbf{x}] = E[\varphi(F_j(\xi_j))] = \int_{-\infty}^{\infty} \varphi(F_j(\xi_j)) dF_j(\xi_j) = \int_0^1 \varphi(x^j) dx^j = 0, \quad (6.4)$$

by assumption $(\mathcal{J}_1)$. From equations (6.3) and (6.4),

$$\begin{aligned} \mathbf{W}_j &= -E[w\mathbf{x}\mathbf{x}^\tau \{g'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^2 \mu_j(\mathbf{x})] + E[w\mathbf{x}\mathbf{x}^\tau \{g''(\mathbf{x}^\tau \boldsymbol{\beta}_0)\} \nu_j(\mathbf{x})] \\ &= -E[w\mathbf{x}\mathbf{x}^\tau \{g'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^2 \mu_j(\mathbf{x})] \\ &= \gamma_{j\varphi}^{-1} E[w\mathbf{x}\mathbf{x}^\tau \{g'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^2] \\ &= \gamma_{j\varphi}^{-1} \boldsymbol{\Sigma} \end{aligned}$$

Finally,

$$\begin{aligned} \nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\xi_n) &= -\frac{3}{n} \sum_{i=1}^n w_i(\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i g''(\xi_n^\tau \mathbf{x}_i) g'(\xi_n^\tau \mathbf{x}_i) \tilde{f}_i(\eta_i(\xi_n)) \varphi'(\tilde{F}_i(\tilde{\eta}_i(\xi_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i(\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i (g'(\mathbf{x}_i^\tau \xi_n))^3 \tilde{f}_i^2(\tilde{\eta}_i(\xi_n)) \varphi''(\tilde{F}_i(\tilde{\eta}_i(\xi_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i(\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i (g'(\mathbf{x}_i^\tau \xi_n))^3 \tilde{f}_i'(\tilde{\eta}_i(\xi_n)) \varphi'(\tilde{F}_i(\tilde{\eta}_i(\xi_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i(\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i g'''(\mathbf{x}_i^\tau \xi_n) \varphi'(\tilde{F}_i(\tilde{\eta}_i(\xi_n))) \end{aligned}$$

Assumptions $(\mathcal{J}_1) - (\mathcal{J}_3)$ and $(\mathcal{J}_6)$ together with a direct application of the SLLN, imply that each term to the right hand side of the above equality is bounded in probability, which in turn implies that $\nabla_{\boldsymbol{\beta}}^2 \tilde{T}_n(\xi_n)$ is bounded in probability. $\square$

Proof of Theorem 2. From the boundedness of φ , and by assumptions $(\mathcal{J}_3)$ and $(\mathcal{J}_6)$, for $j = 1, 2$, we have $Var(\xi_{ij}(\boldsymbol{\beta})) < \infty$, for all i and $\boldsymbol{\beta} \in \mathcal{B}$. Then,

$$\sum_{i=1}^n \frac{Var(\xi_{ij}(\boldsymbol{\beta}))}{n^2} \leq \frac{\max\{Var(\xi_{1j}(\boldsymbol{\beta})), \dots, Var(\xi_{nj}(\boldsymbol{\beta}))\}}{n} = O_p(1/n).$$

Following Bindele *et al.* (2018a), it is obtained that $\tilde{D}_{jn}(\boldsymbol{\beta}) - E\{\tilde{D}_{jn}(\boldsymbol{\beta})\} \rightarrow 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$ as $n \rightarrow \infty$, $j = 1, 2$. To complete the proof, we show that $\{\tilde{D}_{jn}(\boldsymbol{\beta})\}_{n \geq 1}$

is stochastically equicontinuous. For $j = 1, 2$, set $a_{ijn}(\boldsymbol{\beta}) = \frac{R(\xi_{ij}(\boldsymbol{\beta}))}{n+1}$. Then, for all $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{B}$, we have

$$\begin{aligned}
\tilde{D}_{jn}(\boldsymbol{\beta}_1) - \tilde{D}_{jn}(\boldsymbol{\beta}_2) &= \frac{1}{n} \sum_{i=1}^n w_i [\varphi(a_{ijn}(\boldsymbol{\beta}_1)) \xi_{ij}(\boldsymbol{\beta}_1) - \varphi(a_{ijn}(\boldsymbol{\beta}_2)) \xi_{ij}(\boldsymbol{\beta}_2)] \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ w_i [\varphi(a_{ijn}(\boldsymbol{\beta}_2)) + (\varphi(a_{ijn}(\boldsymbol{\beta}_1)) - \varphi(a_{ijn}(\boldsymbol{\beta}_2)))] \xi_{ij}(\boldsymbol{\beta}_1) \right. \\
&\quad \left. - \varphi(a_{ijn}(\boldsymbol{\beta}_2)) \xi_{ij}(\boldsymbol{\beta}_2) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n w_i \varphi(a_{ijn}(\boldsymbol{\beta}_2)) [g(\mathbf{x}_i^\tau \boldsymbol{\beta}_1) - g(\mathbf{x}_i^\tau \boldsymbol{\beta}_2)] \\
&\quad + \frac{1}{n} \sum_{i=1}^n w_i [\varphi(a_{ijn}(\boldsymbol{\beta}_1)) - \varphi(F_{ij\boldsymbol{\beta}_1}(\xi_{ij}(\boldsymbol{\beta}_1)))] \xi_{ij}(\boldsymbol{\beta}_1) \\
&\quad + \frac{1}{n} \sum_{i=1}^n w_i [\varphi(F_{ij\boldsymbol{\beta}_1}(\xi_{ij}(\boldsymbol{\beta}_1)) - \varphi(F_{ij\boldsymbol{\beta}_2}(\xi_{ij}(\boldsymbol{\beta}_2)))] \xi_{ij}(\boldsymbol{\beta}_1) \\
&\quad + \frac{1}{n} \sum_{i=1}^n w_i [\varphi(F_{ij\boldsymbol{\beta}_2}(\xi_{ij}(\boldsymbol{\beta}_2))) - \varphi(a_{ijn}(\boldsymbol{\beta}_2))] \xi_{ij}(\boldsymbol{\beta}_1),
\end{aligned}$$

with $\varphi(F_{ij\boldsymbol{\beta}}(\xi_{ij}(\boldsymbol{\beta}))) = \varphi(F_{ij\boldsymbol{\beta}}(\tilde{z}_{ij} - g(\mathbf{x}_i^\tau \boldsymbol{\beta}))) = h_{ij}(\boldsymbol{\beta})$. For $i = 1, \dots, n$, $F_{ij\boldsymbol{\beta}}$ is a cumulative distribution function of $\xi_{ij}(\boldsymbol{\beta})$, which is almost surely differentiable. So, $F_{ij\boldsymbol{\beta}}(\xi_{ij}(\boldsymbol{\beta}))$ are independently uniformly distributed in the interval $(0, 1)$ for all $\boldsymbol{\beta} \in \mathcal{B}$ and for $j = 1, 2$. By differentiability of $g(\cdot)$ and $\varphi(\cdot)$ with respect to $\boldsymbol{\beta}$, $h_{ij}(\cdot)$ is almost surely differentiable. Thus, applying the mean value theorem, there exist $\eta = \lambda \boldsymbol{\beta}_1 + (1 - \lambda) \boldsymbol{\beta}_2$, for $\lambda \in (0, 1)$ such that $h_{ij}(\boldsymbol{\beta}_1) - h_{ij}(\boldsymbol{\beta}_2) = h'_{ij}(\zeta)(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$, where

$$h'_{ij}(\eta) = -\mathbf{x}_i f_{ij\eta}(\xi_{ij}(\eta)) \varphi'(F_{ij\eta}(\xi_{ij}(\eta))) \nabla_\eta g(\mathbf{x}_i^\tau \eta), \text{ and } f_{ij\eta}(t) = \frac{dF_{ij\eta}(t)}{dt}.$$

By assumptions $(\mathcal{J}_3)$ and $(\mathcal{J}_4)$ together with the boundedness of φ' , there exists a positive

constant M and positive function $J(\mathbf{x})$ such that $\|h'_{ij}\| \leq M\|\mathbf{x}_i\|J(\mathbf{x}_i)$ *a.s.* Hence,

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n w_i [\varphi(F_{ij\beta_1}(\xi_{ij}\beta_1)) - \varphi(F_{ij\beta_2}(\xi_{ij}(\beta_2)))] \xi_{ij}(\beta_1) \right| \\ & \leq \|\beta_1 - \beta_2\| \frac{M}{n} \sum_{i=1}^n \|\mathbf{x}_i\| J(\mathbf{x}_i) |\xi_{ij}(\beta_1)| \\ & \leq \|\beta_1 - \beta_2\| M \left(\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 J^2(\mathbf{x}_i) \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^n |\xi_{ij}(\beta_1)|^2 \right)^{1/2}. \end{aligned}$$

Now, recall that $\xi_{ij}(\beta_1) = \tilde{z}_{ij} - g(\mathbf{x}_i^\tau \beta_1)$ and define z_{ij}^* as

$$z_{ij}^* = \begin{cases} \delta_i y_i + (1 - \delta_i) g(\mathbf{x}_i^\tau \beta_0), & \text{if } j = 1; \\ \frac{\delta_i}{\pi(\mathbf{x}_i)} y_i + \left(1 - \frac{\delta_i}{\pi(\mathbf{x}_i)}\right) g(\mathbf{x}_i^\tau \beta_0), & \text{if } j = 2. \end{cases} \quad (6.5)$$

Set $\lambda_{i1} = \delta_i$ and $\lambda_{i2} = \delta_i/\pi(\mathbf{x}_i)$. Also, $\hat{\lambda}_{i1} = \delta_i$ and $\hat{\lambda}_{i2} = \delta_i/\hat{\pi}(\mathbf{x}_i)$. Thus,

$$\tilde{z}_{ij} - z_{ij}^* = (\hat{\lambda}_{ij} - \lambda_{ij})(y_i - g(\mathbf{x}_i^\tau \beta_0)) + (1 - \hat{\lambda}_{ij})(\tilde{g}(\mathbf{x}_i^\tau \tilde{\beta}_n^c) - g(\mathbf{x}_i^\tau \beta_0)).$$

From this, under assumptions $(\mathcal{J}_3) - (\mathcal{J}_5)$,

$$|\tilde{z}_{ij} - z_{ij}^*| \leq \left| \frac{1}{\hat{\pi}(\mathbf{x}_i)} - \frac{1}{\pi(\mathbf{x}_i)} \right| (|y_i| + J(\mathbf{x}_i)) + |1 - \hat{\lambda}_{ij}| |\tilde{g}(\mathbf{x}_i^\tau \tilde{\beta}_n^c) - g(\mathbf{x}_i^\tau \beta_0)|.$$

The strong consistency of $\hat{\pi}(\mathbf{x})$, $\tilde{g}(\cdot)$ and $\tilde{\beta}_n^c$ together with the generalized continuous mapping theorem (Whitt, 2002), it is obtained that $|\tilde{z}_{ij} - z_{ij}^*| \rightarrow 0$ *a.s.*, for all i and $j=1,2$. Thus, by assumption $(\mathcal{J}_3)$,

$$\begin{aligned} |\xi_{ij}(\beta_1)| & \leq |\tilde{z}_{ij}| + J(\mathbf{x}_i) \leq |\tilde{z}_{ij} - z_{ij}^*| + |z_{ij}^*| + J(\mathbf{x}_i) \\ & \leq |\lambda_{ij}| |y_i| + (2 - \lambda_{ij}) J(\mathbf{x}_i) + o_p(1), \end{aligned}$$

from which

$$\frac{1}{n} \sum_{i=1}^n |\xi(\beta_1)|^2 \leq \frac{1}{n} \sum_{i=1}^n [|\lambda_{ij}| |y_i| + (2 - \lambda_{ij}) J(\mathbf{x}_i)]^2 := K_{jn} + o_p(1).$$

Clearly, by $\mathcal{J}_3$ and $\mathcal{J}_4$ together with the SLL, K_{jn} converges almost surely to a finite

quantity. Thus,

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n w_i [\varphi(F_{ij\beta_1}(\xi_{ij}\beta_1)) - \varphi(F_{ij\beta_2}(\xi_{ij}(\beta_2)))] \xi_{ij}(\beta_1) \right| \\
& \leq \| \beta_1 - \beta_2 \| \frac{M}{n} \sum_{i=1}^n \| \mathbf{x}_i \| J(\mathbf{x}_i) |\xi_{ij}(\beta_1)| \\
& \leq \| \beta_1 - \beta_2 \| M \left(\frac{1}{n} \sum_{i=1}^n \| \mathbf{x}_i \|^2 J^2(\mathbf{x}_i) \right)^{1/2} K_{jn}^{1/2} + o_p(1).
\end{aligned}$$

Also, once again, applying the mean value theorem, we have

$$g(\mathbf{x}_i^\tau \beta_1) - g(\mathbf{x}_i^\tau \beta_2) = \mathbf{x}_i \nabla_\zeta g(\mathbf{x}_i^\tau \zeta)(\beta_1 - \beta_2),$$

where $\zeta = \lambda \beta_1 + (1 - \lambda) \beta_2$ with $\lambda \in (0, 1)$. By assumption $(\mathcal{J}_1)$, there exists a positive constant A such that $|\varphi(u)| \leq A$, for all $u \in (0, 1)$. Thus, under assumptions $(\mathcal{J}_1) - (\mathcal{J}_3)$, we have

$$\left| \frac{1}{n} \sum_{i=1}^n w_i \varphi(a_{ijn}(\beta_2)) [g(\mathbf{x}_i^\tau \beta_1) - g(\mathbf{x}_i^\tau \beta_2)] \right| \leq \| \beta_1 - \beta_2 \| \frac{A}{n} \sum_{i=1}^n \mathbf{x}_i J(\mathbf{x}_i).$$

Moreover, a direct application of Cauchy-shwarz inequality gives

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n w_i [\varphi(a_{ijn}(\beta_1)) - \varphi(F_{ij}(\xi_{ij}(\beta_1)))] \xi_{ij}(\beta_1) \right| \\
& \leq \frac{1}{n} \sum_{i=1}^n |\varphi(a_{ijn}(\beta_1)) - \varphi(F_{ij\beta_1}(\xi_{ij}(\beta_1)))| |\xi_{ij}(\beta_1)| \\
& \leq \left(\max_{1 \leq i \leq n} |\varphi(a_{ijn}(\beta_1)) - \varphi(F_{ij\beta_1}(\xi_{ij}(\beta_1)))|^2 \right)^{1/2} K_{jn}^{1/2} + o_p(1).
\end{aligned}$$

Now, following Hájek *et al.* (1999), $a_{ijn}(\beta) - F_{ij\beta}(\xi_{ij}(\beta)) \rightarrow 0$ a.s., for all $\beta \in \mathcal{B}$ and for all i . By continuity of φ , we have $\varphi(a_{ijn}(\beta)) - \varphi(F_{ij\beta}(\xi_{ij}(\beta))) \rightarrow 0$ a.s., for all $\beta \in \mathcal{B}$ and for all i . Thus, $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}} |\varphi(a_{ijn}(\beta)) - \varphi(F_{ij\beta}(\xi_{ij}(\beta)))| \rightarrow 0$ a.s. Thus, as $n \rightarrow \infty$,

$$\left| \frac{1}{n} \sum_{i=1}^n w_i [\varphi(a_{ijn}(\beta_1)) - \varphi(F_{ij}(\xi_{ij}(\beta_1)))] \xi_{ij}(\beta_1) \right| \rightarrow 0 \text{ a.s.}$$

Similarly,

$$\left| \frac{1}{n} \sum_{i=1}^n w_i [\varphi(F_{ij}(\xi_{ij}(\beta_2)) - \varphi(a_{ijn}(\beta_2)))] \xi_{ij}(\beta_1) \right| \rightarrow 0 \text{ a.s.}$$

To this end, setting

$$B_{jn} = \frac{A}{n} \sum_{i=1}^n \mathbf{x}_i J(\mathbf{x}_i) + M \left(\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 J^2(\mathbf{x}_i) \right)^{1/2} K_{jn}^{1/2} + o_p(1),$$

$|\tilde{D}_{jn}(\beta_1) - \tilde{D}_{jn}(\beta_2)| \leq B_{jn} \|\beta_1 - \beta_2\|$, from which B_{jn} is almost surely bounded. Therefore, $\{\tilde{D}_{jn}(\beta)\}_{n \geq 1}$ is stochastically equicontinuous. By Lemma 6.1, $\hat{\beta}_n^j \rightarrow \beta_0$ a.s., as $n \rightarrow \infty$. $\square$

Proof of Theorem 3. From the consistency of $\hat{\beta}_n^j$ and the boundedness of $\nabla_{\beta}^2 \tilde{T}_{jn}(\xi_n)$, we have

$$(\hat{\beta}_n^j - \beta_0) \nabla_{\beta}^2 \tilde{T}_{jn}(\xi_n) \rightarrow 0 \text{ a.s.}$$

Then, equation (6.1) is reduced to $0 = \tilde{T}_{jn}(\beta_0) + \nabla_{\beta} \tilde{T}_{jn}(\beta_0) \cdot (\hat{\beta}_n^j - \beta_0) + o_p(1/n)$. Thus, taking into account Lemma 6.4-a., we have $\sqrt{n}(\hat{\beta}_n^j - \beta_0) = -\mathbf{W}_j^- \sqrt{n} \tilde{T}_{jn}(\beta_0) + o_p(1)$. Since $\sqrt{n} \tilde{T}_{jn}(\beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V}_j)$ as $n \rightarrow \infty$, we have $\sqrt{n}(\hat{\beta}_n^j - \beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{W}_j^- \mathbf{V}_j \mathbf{W}_j^-)$. $\square$

6.2 Appendix chapter 3

The following lemmas are key to the establishment of the asymptotic properties of the index coefficient estimator (Bindele, 2017; Sun *et al.*, 2019) and the uniform strong consistency of the functional coefficients estimators (Delecroix *et al.*, 2006; Xia *et al.*, 2007).

Lemma 6.5. *Let $\Gamma_n = \{\mathbf{x} : \hat{m}_{\tilde{\beta}}(\mathbf{x}^{\tau} \tilde{\beta}) \geq c\}$, where $\hat{m}_{\tilde{\beta}}(t) = (nb_n)^{-1} \sum_{k=1}^n K((t - \mathbf{x}^{\tau} \tilde{\beta})/b_n)$, for some constants $c \geq c_0 > 0$. Set $\mathcal{H}_n := \{h : c_1 n^{-\alpha_1} < h < c_2 n^{-\alpha_2}\}$, for some constants $c_1, c_2 > 0$ and $1/6 < \alpha_1 < \alpha_2 < 1/4$. Then,*

(i) *if $\delta > 0$, we have, $\sup_{\beta \in \mathcal{B}_n, h \in \mathcal{H}_n} |I_{\{\mathbf{x} : \hat{m}_{\beta}(\mathbf{x}^{\tau} \beta) \geq c\}}(\mathbf{x}) - I_{\Gamma}(\mathbf{x})| \leq I_{\Gamma^{\delta}}(\mathbf{x}) + I_{(\delta, \infty)}(M_n)$, where*

$$\Gamma^{\delta} = \{\mathbf{x}^{\tau} : |m_{\beta_0}(\mathbf{x}^{\tau} \beta_0) - c| \leq \delta\} \quad \text{and} \quad M_n = \max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}_n, h \in \mathcal{H}_n} |\hat{m}_{\beta}(\mathbf{x}_i^{\tau} \beta) - m_{\beta_0}(\mathbf{x}_i^{\tau} \beta_0)|$$

(ii) There exists a sequence $\delta_n \rightarrow 0$ such that $\delta_n/n^{-a\epsilon} \rightarrow \infty$, for some $a > 0$ and $0 < \epsilon < 1/2$, then $I_{(\delta_n, \infty)}(M_n) = o_p(n^{-a})$. Moreover, together with assumptions $(I_3) - (I_6)$, we have

$$\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}, h \in \mathcal{H}_n} |\hat{g}_k(\mathbf{x}^\tau \beta) - g_k(\mathbf{x}^\tau \beta)| I_\Gamma(\mathbf{x}) \rightarrow 0 \quad a.s. \text{ as } n \rightarrow \infty, k = 0, \dots, q$$

and

$$\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}, h \in \mathcal{H}_n} [\hat{g}_k^{(r)}(\mathbf{x}^\tau \beta) - g_k^{(r)}(\mathbf{x}^\tau \beta)] I_\Gamma(\mathbf{x}) \rightarrow 0 \quad a.s. \text{ as } n \rightarrow \infty \text{ and } k = 0, \dots, q,$$

for $r = 1, 2, 3$.

This lemma ensures the uniform strong consistency of $\hat{g}_k(t)$ and $\hat{g}_k^{(r)}(t)$, $r = 1, 2, 3$, for $t \in \mathcal{A}$, and $k = 0, \dots, q$. From this lemma, it is obtained that $\hat{G}'(t) \rightarrow G'(t)$ a.s. and $\hat{G}''(t) \rightarrow G''(t)$ a.s., for all $t \in \mathcal{A}$.

Lemma 6.6. Under assumptions $(I_1) - (I_7)$ and for any $\mathbf{x} \in \Gamma$, we have

$$(i) \quad \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i)^2 \varphi^2(F_j(\nu_j(\beta_0))) \rightarrow \mathbf{V}_j \text{ a.s.}$$

$$(ii) \quad \text{for all } \mu > 0, \lim_{n \rightarrow \infty} P\left(\sqrt{n} \|\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)\| > \mu\right) = 0 \text{ and } \sqrt{n} \tilde{T}_{jn}(\beta_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V}_j),$$

as $n \rightarrow \infty$;

$$(iii) \quad \lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \|S_{jn}(\beta) - T_{jn}(\beta)\| = 0, \text{ a.s.}, \lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \|\tilde{S}_{jn}(\beta) - S_{jn}(\beta)\| = 0 \text{ a.s.},$$

$$\lim_{n \rightarrow \infty} \sup_{\beta \in \mathcal{B}} \|\tilde{T}_{jn}(\beta) - T_{jn}(\beta)\| = 0 \text{ a.s.}; \text{ and}$$

$$(iv) \quad \lim_{n \rightarrow \infty} \|\nabla_\beta \tilde{T}_{jn}(\beta_0) - \nabla_\beta T_{jn}(\beta_0)\| = 0 \text{ a.s.}, \text{ where}$$

$$\tilde{T}_{jn}(\beta) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i (\{\nabla_\beta \hat{G}(\mathbf{x}_i^\tau \beta)\}^\tau \mathbf{z}_i) \varphi(F_{ij}(\tilde{\nu}_{ij}(\beta))),$$

$$\begin{aligned} \nabla_\beta \tilde{T}_{jn}(\beta_0) &= -\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \left(\{\hat{G}'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i \right)^2 \tilde{f}(\tilde{\nu}_j(\beta_0)) \varphi'(\tilde{F}(\tilde{\nu}_j(\beta_0))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{\hat{G}''(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i \varphi(\tilde{F}(\tilde{\nu}_j(\beta_0))), \end{aligned}$$

and

$$\begin{aligned}\nabla_{\beta} T_n(\beta_0) &= -\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} (\{G'(\mathbf{x}_i^{\tau} \beta_0)\}^{\tau} \mathbf{z}_i)^2 f(\nu_j(\beta_0)) \varphi'(F(\nu_j(\beta_0))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} \{G''(\mathbf{x}_i^{\tau} \beta_0)\}^{\tau} \mathbf{z}_i \varphi(F(\nu_j(\beta_0))).\end{aligned}$$

As in Lemma 6.3, from Lemma 6.6-(iii), we have $\sqrt{n}(S_n(\beta_0) - T_n(\beta_0)) \xrightarrow{P} 0$, which implies that $\sqrt{n}S_n(\beta_0)$ and $\sqrt{n}T_n(\beta_0)$ have the same asymptotic distribution (Hettmansperger & McKean, 2011).

Proof of Lemma 6.6.

(i) By assumptions $(I_1) - (I_6)$ and the fact that $w_i \leq 1$ for all i , we have

$$\|E[w_i \mathbf{x}_i \mathbf{x}_i^{\tau} (\{G'(\mathbf{x}_i^{\tau} \beta_0)\}^{\tau} \mathbf{z}_i)^2 \varphi^2(F_j(\nu_j(\beta_0)))]\| \leq c^2 E \left[\|\mathbf{x}\|^2 \left\{ \|z\| \sum_{k=0}^q J_k(\mathbf{x}) \right\}^2 \right] < \infty.$$

So, conditional on $\mathbf{u}$, a direct application of the conditional SLLN Rao (2009) gives

$$\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^{\tau} (\{G'(\mathbf{x}_i^{\tau} \beta_0)\}^{\tau} \mathbf{z}_i)^2 \varphi^2(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \rightarrow \mathbf{V}_j \text{ a.s.}$$

Note that if $\mathbf{u}$ is independent of ε , $\mathbf{u}$ is also independent of $\nu_j(\beta_0)$. Thus, by assumption (I_1) , $E\{\varphi^2(F_j(\nu_j(\beta_0)))\} = 1$. Hence, $\mathbf{V}_j = \Sigma$, $j = 1, 2$.

(ii) Put $\Phi_i(\beta_0) = \hat{G}'(\mathbf{x}_i^{\tau} \beta_0) - G'(\mathbf{x}_i^{\tau} \beta_0)$. From their definitions,

$$\begin{aligned}\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0) &= \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \Phi_i^{\tau}(\beta_0) \mathbf{z}_i \varphi \left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1} \right) \\ &\quad - \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \{\hat{G}'(\mathbf{x}_i^{\tau} \beta_0)\}^{\tau} \mathbf{z}_i \left[\varphi \left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1} \right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right]\end{aligned}$$

From the boundedness of φ and the fact that $w_i \leq 1$, for all i , we have

$$\begin{aligned} \|\tilde{S}_{jn}(\beta_0) - \tilde{T}_{jn}(\beta_0)\| &\leq \frac{c}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\Phi_i(\beta_0)\| \|\mathbf{z}_i\| \\ &\quad + \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\hat{G}'(\mathbf{x}_i^\tau \beta_0)\| \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right| \end{aligned}$$

Note that by the Cauchy-Swartz inequality,

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\Phi_i(\beta_0)\| \|\mathbf{z}_i\| \leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \right]^{1/2} \left[\max_{1 \leq i \leq n} |\Phi_i(\beta_0)|^2 \right]^{1/2}.$$

Assumption (I_5) together with the SLLN give $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \rightarrow E(\|\mathbf{x}\|^2 \|\mathbf{z}\|^2) < \infty$ a.s. Also, by Lemma 6.3, $\max_{1 \leq i \leq n} |\Phi_i(\beta_0)|^2 \rightarrow 0$ a.s. as $n \rightarrow \infty$. Moreover, once again, applying Cauchy-Swartz, we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\hat{G}'(\mathbf{x}_i^\tau \beta_0)\| \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\Phi_i(\beta_0)\| \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right| \\ &\quad + \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|G'(\mathbf{x}_i^\tau \beta_0)\| \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right| \\ &\leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\Phi_i(\beta_0)\|^2 \|\mathbf{z}_i\|^2 \right]^{1/2} \left[\left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right|^2 \right]^{1/2} \\ &\quad + \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \left(\sum_{j=0}^p J_j(\mathbf{x}_i) \right)^2 \|\mathbf{z}_i\|^2 \right]^{1/2} \left[\left| \varphi\left(\frac{R(\tilde{\nu}_j(\beta_0))}{n+1}\right) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right|^2 \right]^{1/2}. \end{aligned}$$

Note that $R(\tilde{\nu}_j(\beta_0))/(n+1) - \tilde{F}_j(\tilde{\nu}_j(\beta_0)) \rightarrow 0$ a.s. (Hájek *et al.*, 1999). By the generalized continuous mapping theorem (Whitt, 2002), we have

$$\left| \varphi(R(\tilde{\nu}_j(\beta_0))/(n+1)) - \varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) \right|^2 \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

Also, assumption (I_5) together with the SLLN, give

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \left(\sum_{k=0}^q J_k(\mathbf{x}_i) \right)^2 \|\mathbf{z}_i\|^2 \rightarrow E \left[\|\mathbf{x}\|^2 \left(\sum_{k=0}^q J_k(\mathbf{x}) \right)^2 \|\mathbf{z}\|^2 \right] < \infty \text{ a.s. as } n \rightarrow \infty.$$

Moreover,

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\Phi_i(\boldsymbol{\beta}_0)\|^2 \|\mathbf{z}_i\|^2 \leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^4 \right]^{1/2} \left[\max_{1 \leq i \leq n} \|\Phi_i(\boldsymbol{\beta}_0)\|^4 \right]^{1/2}.$$

Since $\max_{1 \leq i \leq n} \|\Phi_i(\boldsymbol{\beta}_0)\|^4 \rightarrow 0$; a.s. by Lemma 6.3 and $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^4 \rightarrow E(\|\mathbf{x}\|^4 \|\mathbf{z}\|^4)$ a.s.

as $n \rightarrow \infty$, by the SLLN, we have $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\Phi_i(\boldsymbol{\beta}_0)\|^2 \|\mathbf{z}_i\|^2 \rightarrow 0$; a.s. Hence,

$$\|\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}\| \rightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

Now, by Chebyshev's inequality, for any $\epsilon > 0$, we have

$$P \left(\sqrt{n} (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)) > \epsilon \right) \leq \epsilon^{-2} E \left[n (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0))^2 \right].$$

To complete the proof, it suffices to show that

$$E \left[n (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0))^2 \right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Following Hájek *et al.* (1999) page 192,

$$\begin{aligned} & E \left[n (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0))^2 \right] \\ &= \frac{1}{n} E \left[\left\{ \sum_{i=1}^n w_i \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \left(\varphi \left(\frac{R(\tilde{\nu}_j(\boldsymbol{\beta}_0))}{n+1} \right) - \varphi \left(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0)) \right) \right) \right\}^2 \right] \\ &\leq \frac{n}{n-1} \left\{ \frac{1}{n} \sum_{i=1}^n E[\|\mathbf{x}_i\|^2 (\{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i)^2] \right\} E \left[\left(\varphi \left(\frac{R(\tilde{\nu}_j(\boldsymbol{\beta}_0))}{n+1} \right) - \varphi \left(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0)) \right) \right)^2 \right] \end{aligned}$$

Following similar arguments as above, the right hand side of this inequality converges to zero as $n \rightarrow \infty$. Thus, $E \left[n (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0))^2 \right] \rightarrow 0$, as $n \rightarrow \infty$ and

therefore,

$$\lim_{n \rightarrow \infty} P \left(\sqrt{n} (\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)) > \epsilon \right) = 0.$$

Now, by assumption (I_7) , we have $E\{\tilde{S}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$, as $n \rightarrow \infty$. Since $E\{\tilde{S}_{jn}(\boldsymbol{\beta}_0) - \tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$, we have $E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$, as $n \rightarrow \infty$. To obtain the asymptotic distribution of $\tilde{T}_{jn}(\boldsymbol{\beta}_0)$, we employ the Cramér-Wold device (Serfling, 1980). To this end, set

$$U_{jn} = n^{-1/2} \sum_{i=1}^n w_i \mathbf{a}^\tau \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))] = \mathbf{a}^\tau \tilde{T}_{jn}(\boldsymbol{\beta}_0),$$

where $\mathbf{a} \in \mathbb{R}^q$. Since $E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\} \rightarrow 0$ and $E(U_{jn}) = \mathbf{a}^\tau E\{\tilde{T}_{jn}(\boldsymbol{\beta}_0)\}$, we have $E(U_{jn}) \rightarrow 0$. Letting

$$V_{ij} = w_i \mathbf{a}^\tau \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))],$$

from the fact that $(\mathbf{u}_i, y_i, \varepsilon_i)$, $i = 1, \dots, n$, is a random sample, conditional on $\mathbf{u}_i$, $\{V_{nj}\}_{n \geq 1}$ is a sequence of conditional independent random variables with identical distribution (Rao, 2009).

$$Var(U_{jn}) = \frac{1}{n} \sum_{i=1}^n E \left[\left(w_i \mathbf{a}^\tau \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \right)^2 E \left\{ \varphi^2(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) | \mathbf{u}_i \right\} \right]$$

converges almost surely to $\mathbf{a}^\tau \mathbf{V}_j \mathbf{a}$, as $n \rightarrow \infty$. Note that U_{jn} is the sum of identically distributed conditionally independent functions of random variables. Hence, the limiting distribution is established by verifying the conditional Lindeberg-Feller condition for the applicability of the conditional Central Limit Theorem. To this end, set $\sigma_{jn}^2 = Var(U_{jn})$. We need to show that for any $\mu > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{\sigma_{jn}^2} \sum_{i=1}^n E \left[\frac{1}{n} \left(w_i \mathbf{a}^\tau \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \right)^2 E \{ \varphi^2[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))] | \mathbf{u}_i \} \right] \quad (6.6)$$

To this end, we have

$$\frac{1}{\sqrt{n}} w_i |\mathbf{a}^\tau \mathbf{x}_i \{ \hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i| \leq n^{-1/2} \|\mathbf{a}\| \|\mathbf{x}_i\| \|\mathbf{z}_i\| \sum_{k=0}^q J_k(\mathbf{x}_i) + o(1/\sqrt{n}).$$

By assumptions (I_3) and (I_6) , $E[\|\mathbf{a}\|\|\mathbf{x}_i\|\|\mathbf{z}_i\|\sum_{k=0}^q J_k(\mathbf{z}_i)] < \infty$, $\|\mathbf{a}\|\|\mathbf{x}_i\|\|\mathbf{z}_i\|\sum_{k=0}^q J_k(\mathbf{x}_i)$ is bounded in probability, for all i . Thus, with probability 1, we have

$$\frac{1}{\sqrt{n}}|\mathbf{a}^\tau \mathbf{x}_i \{\widehat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Set $\lambda_n = [\max_{1 \leq i \leq n} n^{-1} \|\mathbf{x}_i\|\|\mathbf{z}_i\|\sum_{k=0}^q J_k(\mathbf{x}_i)]^{1/2} \|\mathbf{a}\|$. Then $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$, and, is independent of i . Since σ_{jn}^2 converges to a positive quantity, the ratio $\sigma_{jn}/\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$. Now conditioning on $\mathbf{u}_i$, it is easy to see that

$$\sum_{i=1}^n E\left[\frac{1}{n} V_{ij}^2 | \mathbf{u}_i\right] I\left(\left|\frac{1}{\sqrt{n}} V_{ij}\right| > \mu \sigma_{jn}\right)$$

is less than or equal to

$$E\left\{E\left[\varphi^2[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))]I\left(\left|\varphi[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))]\right| > \mu \sigma_{jn}/\lambda_n\right)\middle|\mathbf{u}\right]\right\} \times \\ \frac{1}{n} \sum_{i=1}^n E\left[(w_i \mathbf{a}^\tau \mathbf{x}_i \{\widehat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2\right].$$

By assumptions (I_3) and (I_6) , $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[(w_i \mathbf{a}^\tau \mathbf{x}_i \{\widehat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2] < \infty$. From the boundedness of φ , applying the Dominated Convergence Theorem to the first term, we have

$$E\left\{E\left[\varphi^2[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))]I\left(\left|\varphi[\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))]\right| > \mu \sigma_{jn}/\lambda_n\right)\middle|\mathbf{u}\right]\right\} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

This establishes the Lindeberg-Feller condition. Thus, by the conditional Central Limit Theorem (Rao, 2009), we have $\sqrt{n} \tilde{T}_{jn}(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{V}_j)$, as $n \rightarrow \infty$.

$$(iii) \text{ As before, } S_{jn}(\boldsymbol{\beta}) - T_{jn}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \{G'(\mathbf{x}_i^\tau \boldsymbol{\beta})\}^\tau \mathbf{z}_i \left[\varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi(F_{ij}(\nu_{ij}(\boldsymbol{\beta}))) \right].$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& \|S_{jn}(\boldsymbol{\beta}) - T_{jn}(\boldsymbol{\beta})\| \\
& \leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \|G'(\mathbf{x}_i^\tau \boldsymbol{\beta})\|^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n \left| \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) - \varphi(F_{ij}(\nu_{ij}(\boldsymbol{\beta}))) \right|^2 \right]^{1/2} \\
& \leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \right]^{1/2} \left[\max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \left| \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) - \varphi(F_{ij}(\nu_{ij}(\boldsymbol{\beta}))) \right|^2 \right]^{1/2}.
\end{aligned}$$

By the continuity of φ and the fact that $R(\nu_{ij}(\boldsymbol{\beta}))/n - F_{ij}(\nu_{ij}(\boldsymbol{\beta})) \rightarrow 0$ a.s., for all i and all $\boldsymbol{\beta} \in \mathcal{B}$, we have

$$\max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \left| \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) - \varphi(F_{ij}(\nu_{ij}(\boldsymbol{\beta}))) \right|^2 \rightarrow 0 \text{ a.s.}$$

On the other hand, the strong law of large numbers (SLLN) gives

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \rightarrow E \left[\|\mathbf{x}\|^2 \|\mathbf{z}\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}) \right\}^2 \right] < \infty \text{ a.s.},$$

by assumption (I_3) and (I_6) . Thus, $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \|S_{jn}(\boldsymbol{\beta}) - T_{jn}(\boldsymbol{\beta})\| = 0$ a.s.

Next, $\Psi_i(\boldsymbol{\beta}) \rightarrow 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$. we have

$$\begin{aligned}
\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \Psi_i^\tau(\boldsymbol{\beta}) \mathbf{z}_i \varphi \left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1} \right) \\
&\quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \{G'(\mathbf{x}_i^\tau \boldsymbol{\beta})\}^\tau \mathbf{z}_i \left[\varphi \left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1} \right) - \varphi \left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1} \right) \right].
\end{aligned}$$

Note that $\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta}) = \sum_{k=0}^q \left[\left(\hat{g}_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g_k(\mathbf{v}^\tau \boldsymbol{\beta}) \right) + \left(\hat{g}'_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g'_k(\mathbf{v}^\tau \boldsymbol{\beta}) \right) \boldsymbol{\beta}^\tau \mathbf{x}_{i0} \right] z_{ji}$.

Then,

$$\begin{aligned}
|\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})| &\leq (q+1) \max_{0 \leq k \leq q} |\hat{g}_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g_k(\mathbf{v}^\tau \boldsymbol{\beta})| \|b z_i\| \\
&\quad + (q+1) \max_{0 \leq k \leq q} |\hat{g}'_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g'_k(\mathbf{v}^\tau \boldsymbol{\beta})| \|\boldsymbol{\beta}\| \|\mathbf{x}_{i0}\| \|\mathbf{z}_i\|.
\end{aligned}$$

By assumption (I_6) , with probability 1, there exist constants c_1 and c_2 , $\|\mathbf{z}_i\| \leq c_1$

and $\|\mathbf{x}_{i0}\| \leq c_2$, for all i . Thus,

$$\begin{aligned} \max_{1 \leq i \leq n} |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})| &\leq (q+1)c_1 \max_{0 \leq k \leq q} |\hat{g}_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g_k(\mathbf{v}^\tau \boldsymbol{\beta})| \\ &\quad + (q+1)c_1c_2 \max_{0 \leq k \leq q} |\hat{g}'_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g'_k(\mathbf{v}^\tau \boldsymbol{\beta})| \|\boldsymbol{\beta}\|. \end{aligned}$$

By Lemma 6.5 and for all $\boldsymbol{\beta} \in \mathcal{B}$,

$$\max_{0 \leq k \leq q} |\hat{g}_k(\mathbf{v}^\tau \boldsymbol{\beta}) - g_k(\mathbf{v}^\tau \boldsymbol{\beta})| \rightarrow 0 \text{ a.s.} \quad \text{and} \quad \max_{0 \leq k \leq q} |\hat{g}'_k(\boldsymbol{\beta}^T z) - g'_k(\boldsymbol{\beta}^T z)| \rightarrow 0 \text{ a.s.}$$

Hence, $\max_{1 \leq i \leq n} |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})| \rightarrow 0 \text{ a.s.}$, for all $\boldsymbol{\beta} \in \mathcal{B}$. Also, following Hájek *et al.* (1999), letting $\tilde{F}_{ij}(t) = P(\tilde{\nu}_{ij}(\boldsymbol{\beta}) \leq t)$ and $F_{ij}(t) = P(\nu_{ij}(\boldsymbol{\beta}) \leq t)$, we have

$$R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))/((n+1)-\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\beta}))) \rightarrow 0 \text{ a.s.} \quad \text{and} \quad R(\nu_{ij}(\boldsymbol{\beta}))/((n+1)-F_{ij}(\nu_{ij}(\boldsymbol{\beta}))) \rightarrow 0 \text{ a.s.}, \quad (6.7)$$

for all $\boldsymbol{\beta} \in \mathcal{B}$ and for all i . Furthermore, since convergence almost sure implies convergence in probability, and convergence in probability implies convergence in distribution, we have $\tilde{F}_{ij}(t) \rightarrow F_{ij}(t)$, for all i . Now, taking into account the continuity of $\tilde{F}_{ij}(t)$ and $F_{ij}(t)$, and applying the generalized continuous mapping theorem (Whitt, 2002), we have $\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\beta})) - F_{ij}(\nu_{ij}(\boldsymbol{\beta})) \rightarrow 0 \text{ a.s.}$, for all $\boldsymbol{\beta} \in \mathcal{B}$ and for all i . Combining this with equation (6.7) together with the continuity of φ , yields

$$\varphi\left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) \rightarrow 0 \text{ a.s.} \quad \text{for all } \boldsymbol{\beta} \in \mathcal{B} \text{ and for all } i. \quad (6.8)$$

Then,

$$\begin{aligned} &\|\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta})\| \\ &\leq \frac{c}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\mathbf{z}_i\| \|\Phi_i(\boldsymbol{\beta})\| + \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\mathbf{z}_i\| \|G'(\mathbf{x}_i^\tau \boldsymbol{\beta})\| \left| \varphi\left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) \right|. \end{aligned}$$

From this equation, by the Cauchy-Schwarz inequality, we have

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\mathbf{z}_i\| \|\Phi_i(\boldsymbol{\beta})\| \leq \left(\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \right)^{1/2} \max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \Phi_i^2(\boldsymbol{\beta}) \rightarrow 0 \text{ a.s.},$$

as $\max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \Phi_i^2(\boldsymbol{\beta}) \rightarrow 0$ a.s. and $\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \rightarrow E(\|\mathbf{x}\|^2 \|\mathbf{z}\|^2) < \infty$ a.s., by the SLLN under assumption (I_6) . On the other hand, by the Cauchy-Schwarz inequality,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\| \|\mathbf{z}_i\| \|G'(\mathbf{x}_i^\tau \boldsymbol{\beta})\| \left| \varphi\left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) \right| \\ & \leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \right]^{1/2} \left[\max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \left| \varphi\left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) \right|^2 \right]^{1/2}. \end{aligned}$$

By equation (6.8), $\max_{1 \leq i \leq n} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \left| \varphi\left(\frac{R(\tilde{\nu}_{ij}(\boldsymbol{\beta}))}{n+1}\right) - \varphi\left(\frac{R(\nu_{ij}(\boldsymbol{\beta}))}{n+1}\right) \right|^2 \rightarrow 0$ a.s. and once again, by the SLLN, as before,

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \rightarrow E \left[\|\mathbf{x}\|^2 \|\mathbf{z}\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}) \right\}^2 \right] < \infty \text{ a.s.},$$

by assumption (I_3) and (I_5) . Thus, $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{S}_{jn}(\boldsymbol{\beta}) - S_{jn}(\boldsymbol{\beta})\| = 0$ a.s. Similarly,

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \|\tilde{T}_{jn}(\boldsymbol{\beta}) - T_{jn}(\boldsymbol{\beta})\| = 0 \text{ a.s.}$$

(iv) Putting $\tilde{\Psi}_{ij}(\boldsymbol{\beta}_0) = \tilde{f}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))\varphi'(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0)))$ and $\Psi_{ij}(\boldsymbol{\beta}_0) = f_j(\nu_j(\boldsymbol{\beta}_0))\varphi'(F_j(\nu_j(\boldsymbol{\beta}_0)))$,

$$\begin{aligned} & \nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) - \nabla_{\boldsymbol{\beta}} T_{jn}(\boldsymbol{\beta}_0) \\ & = -\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \left[\left(\{\hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i \right)^2 - (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2 \right] \tilde{\Psi}_{ij}(\boldsymbol{\beta}_0) \\ & \quad - \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2 \left[\tilde{\Psi}_{ij}(\boldsymbol{\beta}_0) - \Psi_{ij}(\boldsymbol{\beta}_0) \right] \\ & \quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{ \hat{G}'''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) - G'''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{x}_i \varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) \\ & \quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{ G'''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \left[\varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) - \varphi(F_j(\nu_j(\boldsymbol{\beta}_0))) \right]. \end{aligned}$$

By the strong consistency of $\hat{G}(t)$, $\tilde{\nu}_{ij}(\beta_0) - \nu_{ij}(\beta_0) \rightarrow 0$ a.s., for all i . Once again, by the continuity of $\tilde{F}_j$ and F_j , and the fact that $\tilde{F}_j(\tilde{\nu}_j(\beta_0)) - F_j(\nu_j(\beta_0)) \rightarrow 0$ a.s., for all i , the generalized continuous mapping theorem (Whitt, 2002), gives $\max_{1 \leq i \leq n} |\tilde{\Psi}_{ij}(\beta_0) - \Psi_{ij}(\beta_0)| \rightarrow 0$ a.s. Moreover, by the continuity of φ , we have $\varphi(\tilde{F}_j(\tilde{\nu}_j(\beta_0))) - \varphi(\nu_j(\beta_0)) \rightarrow 0$ a.s., for all i . Now,

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \left[\left(\{\hat{G}'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i \right)^2 - (\{G'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i)^2 \right] \tilde{\Psi}_{ij}(\beta_0) \right\| \\ &= \left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\Phi_i(\beta_0)^\tau \mathbf{z}_i) \left\{ \hat{G}'(\mathbf{x}_i^\tau \beta_0) + G'(\mathbf{x}_i^\tau \beta_0) \right\}^\tau \mathbf{z}_i \tilde{\Psi}_{ij}(\beta_0) \right\| \\ &\leq \left[\max_{1 \leq i \leq n} \|\Phi_i(\beta_0)\|^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n w_i^2 \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \hat{G}'(\mathbf{x}_i^\tau \beta_0) + G'(\mathbf{x}_i^\tau \beta_0) \right\}^2 \|\tilde{\Psi}_{ij}(\beta_0)\|^2 \right]^{1/2}, \end{aligned}$$

by Cauchy-Schwarz inequality. Clearly, $\max_{1 \leq i \leq n} \|\Phi_i(\beta_0)\|^2 \rightarrow 0$ a.s. Also, from the consistency of $\hat{G}(t)$, a direct application of the SLLN entails that

$$\frac{1}{n} \sum_{i=1}^n w_i^2 \|\mathbf{x}_i\|^2 \|\mathbf{z}_i\|^2 \left\{ \hat{G}'(\mathbf{x}_i^\tau \beta_0) + G'(\mathbf{x}_i^\tau \beta_0) \right\}^2 \|\tilde{\Psi}_{ij}(\beta_0)\|^2$$

converges almost surely to

$$4E[w^2 \|\mathbf{x}\|^2 \|\mathbf{z}\|^2 \|G'(\mathbf{x}^\tau \beta_0)\|^2 E\{|\Psi_j(\beta_0)| |\mathbf{u}|\}] < \infty \text{ a.s.}$$

Thus,

$$\left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \left[\left(\{\hat{G}'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i \right)^2 - (\{G'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i)^2 \right] \tilde{\Psi}_{ij}(\beta_0) \right\| \rightarrow 0 \text{ a.s.}$$

Also,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \beta_0)\}^\tau \mathbf{z}_i)^2 [\tilde{\Psi}_{ij}(\beta_0) - \Psi_{ij}(\beta_0)] \right\| &\leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^4 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^4 \right]^{1/2} \\ &\quad \times \left[\max_{1 \leq i \leq n} |\tilde{\Psi}_{ij}(\beta_0) - \Psi_{ij}(\beta_0)|^2 \right]^{1/2}, \end{aligned}$$

from which, $\max_{1 \leq i \leq n} |\tilde{\Psi}_{ij}(\beta_0) - \Psi_{ij}(\beta_0)|^2 \rightarrow 0$ a.s. and by the SLLN under assumptions

(I_3) and (I_6), we have

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^4 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^4 \rightarrow E \left[\|\mathbf{x}\|^4 \|\mathbf{z}\|^4 \left\{ \sum_{k=0}^q J_k(\mathbf{x}) \right\}^4 \right] < \infty \text{ a.s.}$$

Thus, $\left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2 \left[\tilde{\Psi}_{ij}(\boldsymbol{\beta}_0) - \Psi_{ij}(\boldsymbol{\beta}_0) \right] \right\| \rightarrow 0 \text{ a.s.}$ Next,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{ \hat{G}''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) - G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) \right\| &\leq M \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^2 \right]^{1/2} \\ &\times \left[\max_{1 \leq i \leq n} \|\hat{G}''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) - G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\|^2 \right], \end{aligned}$$

from which, by Lemma 6.6, $\max_{1 \leq i \leq n} \|\hat{G}''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) - G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\|^2 \rightarrow 0 \text{ a.s.}$ and by the SLLN under (I_6),

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^2 \rightarrow E [\|\mathbf{x}\|^4 \|\mathbf{z}\|^2] < \infty \text{ a.s.}$$

Hence, $\left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{ \hat{G}''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) - G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) \right\| \rightarrow 0 \text{ s.s.}$ Finally, by (I_3),

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{ G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0) \}^\tau \mathbf{z}_i \left[\varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) - \varphi(F_j(\nu_j(\boldsymbol{\beta}_0))) \right] \right\| \\ &\leq \left[\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \right]^{1/2} \times \left[\max_{1 \leq i \leq n} |\varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) - \varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))|^2 \right], \end{aligned}$$

from which, $\max_{1 \leq i \leq n} |\varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) - \varphi(F_j(\nu_j(\boldsymbol{\beta}_0)))|^2 \rightarrow 0 \text{ a.s.}$, by the continuity of φ and the fact that $\tilde{\nu}_j(\boldsymbol{\beta}_0) - \nu_j(\boldsymbol{\beta}_0) \rightarrow 0 \text{ a.s.}$, for all i . Furthermore, by the SLLN under (I_3) and (I_6),

$$\frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|^4 \|\mathbf{z}_i\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \rightarrow E \left[\|\mathbf{x}\|^4 \|\mathbf{z}\|^2 \left\{ \sum_{k=0}^q J_k(\mathbf{x}) \right\}^2 \right] < \infty \text{ a.s.}$$

Thus, $\left\| \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i \left[\varphi(\tilde{F}_j(\tilde{\nu}_j(\boldsymbol{\beta}_0))) - \varphi(F_j(\nu_j(\boldsymbol{\beta}_0))) \right] \right\| \rightarrow 0$ a.s. To this end,

$$\lim_{n \rightarrow \infty} \|\nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) - \nabla_{\boldsymbol{\beta}} T_{jn}(\boldsymbol{\beta}_0)\| = 0 \text{ a.s.},$$

which completes the proof. $\square$

Proof of Theorem 4. In this proof, we set c to be an arbitrary positive constant. Define $\hat{a}_{ijn}(\boldsymbol{\beta}) = R(\tilde{\nu}_{ij}(\boldsymbol{\beta})) / (n+1)$ and $a_{ijn}(\boldsymbol{\beta}) = R(\nu_{ij}(\boldsymbol{\beta})) / (n+1)$. By definition of $Q_{jn}(\boldsymbol{\beta})$ and $L_{jn}(\boldsymbol{\beta})$, we have

$$\begin{aligned} Q_{jn}(\boldsymbol{\beta}) - L_{jn}(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n w_i [\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) \tilde{\nu}_{ij}(\boldsymbol{\beta}) - \varphi(a_{ijn}(\boldsymbol{\beta})) \nu_{ij}(\boldsymbol{\beta})] \\ &= \frac{1}{n} \sum_{i=1}^n w_i [\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) - \varphi(a_{ijn}(\boldsymbol{\beta}))] \nu_{ij}(\boldsymbol{\beta}) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \varphi(a_{ijn}(\boldsymbol{\beta})) [\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})] \end{aligned}$$

Letting $\tilde{F}_{ij}$ be the cumulative distribution function of $\tilde{\nu}_{ij}(\boldsymbol{\beta})$, from Hájek *et al.* (1999), we have $\hat{a}_{ijn} - \tilde{F}_{ij} \rightarrow 0$ a.s. and $a_{ijn} - F_{ij} \rightarrow 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$, for all $i = 1, \dots, n$, and $j = 1, 2$. On the other hand, by the uniform strong consistency of $\hat{g}_k^{(r)}$ for $g_k^{(r)}$, $r = 0, 1$, we have $\limsup_{n \rightarrow \infty} \max_{\boldsymbol{\beta} \in \mathcal{B}} |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})| = 0$ a.s., which in turn implies that $\tilde{F}_{ij} - F_{ij} \rightarrow 0$ a.s., as $n \rightarrow \infty$, for all i and $j = 1, 2$. From the fact that $w_i \leq 1$, for all i , we have

$$\begin{aligned} |Q_{jn}(\boldsymbol{\beta}) - L_{jn}(\boldsymbol{\beta})| &\leq \frac{1}{n} \sum_{i=1}^n |\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) - \varphi(a_{ijn}(\boldsymbol{\beta}))| |\nu_{ij}(\boldsymbol{\beta})| \\ &\quad + \frac{1}{n} \sum_{i=1}^n |\varphi(a_{ijn}(\boldsymbol{\beta}))| |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})|. \end{aligned}$$

From the boundedness of φ , there exists $c > 0$ such that $|\varphi(u)| \leq c$, for all $u \in (0, 1)$. Also, since $\lim_{n \rightarrow \infty} (\hat{a}_{ijn}(\boldsymbol{\beta}) - a_{ijn}(\boldsymbol{\beta})) = 0$ a.s., for all $\boldsymbol{\beta} \in \mathcal{B}$, for all $i = 1, \dots, n$ and $j = 1, 2$, by the continuity of φ , applying the generalized continuous mapping theorem (Whitt, 2002), we have $\limsup_{n \rightarrow \infty} \max_{\boldsymbol{\beta} \in \mathcal{B}} |\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) - \varphi(a_{ijn}(\boldsymbol{\beta}))| = 0$ a.s. Moreover, by assumption

(I_3) , $|g_k(\mathbf{x}_i^\tau \boldsymbol{\beta})| \leq J_k(\mathbf{x}_i)$. Then,

$$|\nu_{ij}(\boldsymbol{\beta})| = \left| y_i - \sum_{k=0}^q g_k(\mathbf{x}_i^\tau \boldsymbol{\beta}) z_{ji} \right| \leq |y_i| + \sum_{k=0}^q J_k(\mathbf{x}_i) \|\mathbf{z}_i\| := \psi_i.$$

Thus, by Cauchy-Schwarz inequality

$$\begin{aligned} \sup_{\boldsymbol{\beta} \in \mathcal{B}} |Q_{jn}(\boldsymbol{\beta}) - L_{jn}(\boldsymbol{\beta})| &\leq \left[\sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) - \varphi(a_{ijn}(\boldsymbol{\beta}))|^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=1}^n |\psi_i|^2 \right]^{1/2} \\ &\quad + c \sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})|. \end{aligned}$$

Note that $E \left[\sup_{\boldsymbol{\beta} \in \mathcal{B}} |\nu_{ij}(\boldsymbol{\beta})|^2 \right] \leq E\{\psi_i^2\} < \infty$, by assumption (I_3) and (I_6) . So, by the SLLN,

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=1}^n \sup_{\boldsymbol{\beta} \in \mathcal{B}} |\nu_{ij}(\boldsymbol{\beta})|^2 - \frac{1}{n} \sum_{i=1}^n E \left[\sup_{\boldsymbol{\beta} \in \mathcal{B}} |\nu_{ij}(\boldsymbol{\beta})|^2 \right] \right\} = 0 \text{ a.s.}$$

Since

$$\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\varphi(\hat{a}_{ijn}(\boldsymbol{\beta})) - \varphi(a_{ijn}(\boldsymbol{\beta}))| = 0 \text{ a.s.} \quad \text{and} \quad \lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \max_{1 \leq i \leq n} |\tilde{\nu}_{ij}(\boldsymbol{\beta}) - \nu_{ij}(\boldsymbol{\beta})| = 0 \text{ a.s.},$$

we have $\lim_{n \rightarrow \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}} |Q_{jn}(\boldsymbol{\beta}) - L_{jn}(\boldsymbol{\beta})| = 0 \text{ a.s.}$ $\square$

Proof of Theorem 6.

(i) As shown above, since $\nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) - \nabla_{\boldsymbol{\beta}} T_{jn}(\boldsymbol{\beta}_0) \rightarrow 0 \text{ a.s.}$, considering the definition of $T_{jn}(\boldsymbol{\beta})$, we have

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} T_{jn}(\boldsymbol{\beta}_0) &= -\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2 f_j(\nu_j(\boldsymbol{\beta}_0)) \varphi'(F_j(\nu_j(\boldsymbol{\beta}_0))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i \varphi(F_j(\nu_i(\boldsymbol{\beta}_0))). \end{aligned}$$

By Lemma 6.6(i), conditional on $\mathbf{u}$, the conditional strong law of large numbers (Rao, 2009) gives

$$\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau (\{G'(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i)^2 f_j(\nu_j(\boldsymbol{\beta}_0)) \varphi'(F_j(\nu_i(\boldsymbol{\beta}_0))) \rightarrow E[\mathbf{w} \mathbf{x} \mathbf{x}^\tau (\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z})^2 \omega_j(\mathbf{u})],$$

and $\frac{1}{n} \sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i^\tau \{G''(\mathbf{x}_i^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z}_i \varphi(F_j(\nu_j(\boldsymbol{\beta}_0))) \rightarrow E[w \mathbf{x} \mathbf{x}^\tau \{G''(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z} \kappa_j(\mathbf{u})] \text{ a.s.}$
Thus,

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} \tilde{T}_{jn}(\boldsymbol{\beta}_0) \rightarrow \mathbf{W}_j &= -E[w \mathbf{x} \mathbf{x}^\tau (\{G'(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z})^2 \omega_j(\mathbf{u})] \\ &\quad + E[w \mathbf{x} \mathbf{x}^\tau \{G''(\mathbf{x}^\tau \boldsymbol{\beta}_0)\}^\tau \mathbf{z} \kappa_j(\mathbf{u})] \text{ a.s.} \end{aligned}$$

(ii) In a straightforward manner,

$$\begin{aligned} \nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\xi}_n) &= -\frac{3}{n} \sum_{i=1}^n w_i (\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i \{\hat{G}'''(\mathbf{x}_i^\tau \boldsymbol{\xi}_n)\}^\tau \mathbf{z}_i \tilde{f}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n)) \varphi'(\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \left(\mathbf{x}_i \{\hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\xi}_n)\}^\tau \mathbf{z}_i \right)^{\otimes 3} \tilde{f}_{ij}'(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n)) \varphi'(\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i \left(\mathbf{x}_i \{\hat{G}'(\mathbf{x}_i^\tau \boldsymbol{\xi}_n)\}^\tau \mathbf{z}_i \right)^{\otimes 3} \tilde{f}_{ij}^2(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n)) \varphi''(\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n))) \\ &\quad + \frac{1}{n} \sum_{i=1}^n w_i (\mathbf{x}_i \mathbf{x}_i^\tau) \mathbf{x}_i \{\hat{G}'''(\mathbf{x}_i^\tau \boldsymbol{\xi}_n)\}^\tau \mathbf{z}_i \varphi(\tilde{F}_{ij}(\tilde{\nu}_{ij}(\boldsymbol{\xi}_n))), \end{aligned}$$

where $\otimes$ stands for the tensor product. From this, taking into account assumptions $(I_3) - (I_6)$, by routine simple computations, it can be shown in a straightforward manner that each term to the right hand side of this equation is bounded by a quantity that converges almost surely to a finite quantity. Therefore, each term is almost surely bounded. Thus, $\nabla_{\boldsymbol{\beta}}^2 \tilde{T}_{jn}(\boldsymbol{\xi}_n)$ is almost surely bounded.

□

6.3 Appendix chapter 4

The following lemma is necessary for the proof of Theorem 12

Lemma 6.7. *Set $P_{jn}(G(t_0)) = \frac{1}{n} \sum_{i=1}^n [\nabla G(t_{0i})]^\tau \mathbf{z}_i \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i})))) v_{ij}(G(t_{0i})) w_i(t_{0i})$, where $\tilde{F}_{ijn}$ is the cumulative distribution function of $v_{ij}(G(t_0))$. Under $(I_1) - (I_8)$, we have*

$$\lim_{n \rightarrow \infty} \sup_{G(t_0)} \|J_{jn}(G(t_0)) - P_{jn}(G(t_0))\| = 0 \text{ a.s.}$$

Proof. Following Hájek *et al.* (1999), it is not hard to see that

$$R(v_{ij}(G(t_0)))/(n+1) - \tilde{F}_{ijn}(v_{ij}(G(t_0))) \rightarrow 0 \text{ a.s.}, \quad \text{for all } i, j = 1, 2.$$

By continuity of φ together with the generalized continuous mapping theorem (Whitt, 2002), we have $\max_{1 \leq i \leq n} |\varphi(R(v_{ij}(G(t_{0i})))/(n+1)) - \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i}))))| \rightarrow 0 \text{ a.s.}$ On the other hand,

$$\begin{aligned} & \|J_{jn}(G(t_0)) - P_{jn}(G(t_0))\| \\ & \leq \frac{1}{n} \sum_{i=1}^n \|\nabla G(t_{0i})\| \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(v_{ij}(G(t_{0i})))}{n+1}\right) - \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i})))) \right| |w_i(t_{0i})|. \end{aligned}$$

Note that by definition, $|w_i(t_0)| \leq 1$ and by assumption (J_2) , $\|\nabla G(t_0)\| \leq \sum_{k=0}^q J_k(\mathbf{x})$. Then, by Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \|J_{jn}(G(t_0)) - P_{jn}(G(t_0))\| \\ & \leq \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\} \|\mathbf{z}_i\| \left| \varphi\left(\frac{R(v_{ij}(G(t_{0i})))}{n+1}\right) - \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i})))) \right| \\ & \leq \left[\frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \|\mathbf{z}_i\|^2 \right]^{1/2} \left[\max_{1 \leq i \leq n} \left| \varphi\left(\frac{R(v_{ij}(G(t_{0i})))}{n+1}\right) - \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i})))) \right|^2 \right]^{1/2}. \end{aligned}$$

By (J_2) and (J_6) , $E \left[\left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \|\mathbf{z}_i\|^2 \right] < \infty$. So, by the Strong Law of Large Numbers, $\frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=0}^q J_k(\mathbf{x}_i) \right\}^2 \|\mathbf{z}_i\|^2 \rightarrow E \left[\left\{ \sum_{k=0}^q J_k(\mathbf{x}) \right\}^2 \|\mathbf{z}\|^2 \right] < \infty \text{ a.s.}$ Moreover, as $n \rightarrow \infty$,

$$\max_{1 \leq i \leq n} \left| \varphi\left(\frac{R(v_{ij}(G(t_{0i})))}{n+1}\right) - \varphi(\tilde{F}_{ijn}(v_{ij}(G(t_{0i})))) \right|^2 \rightarrow 0 \text{ a.s.}$$

Thus, $\lim_{n \rightarrow \infty} \sup_{G(t_0)} \|J_{jn}(G(t_0)) - P_{jn}(G(t_0))\| = 0 \text{ a.s.}$ □

Proof of Theorem 10. By definition, $\boldsymbol{\lambda}$ satisfies

$$\frac{1}{n} \sum_{i=1}^n \frac{z_{ij}(\boldsymbol{\beta}_0)}{1 + \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0)} = 0.$$

Now, by assumption (I_1) and (I_3) , one can obtain in a straightforward manner that $\max_{1 \leq i \leq n} \|z_{ij}(\beta_0)\| = o_p(n^{1/2})$. Also, $n^{-1} \sum_{i=1}^n \|z_{ij}(\beta_0)\|^3 = o_p(n^{1/2})$. Following the same argument as in Owen (1990), we obtain that $\|\lambda\| = O_p(n^{-1/2})$, and by Theorem 10, $\|S_{jn}(\beta_0)\| = O_p(n^{1/2})$. From this, a Taylor expansion to the right hand side of equation (4.6) gives

$$R_j(\beta_0) = 2 \sum_{i=1}^n \left[\lambda^\tau z_{ij}(\beta_0) - \frac{1}{2} (\lambda^\tau z_{ij}(\beta_0))^2 \right] + \gamma_n,$$

where $\|\gamma_n\| = O_p(1) \sum_{i=1}^n |\lambda^\tau z_{ij}(\beta_0)|^3 = O_p(1) \|\lambda\|^3 \sum_{i=1}^n \|z_{ij}(\beta_0)\|^3 = o_p(1)$. On the other hand, note that solving equation (4.5) is equivalent to solving

$$\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) (1 - \lambda^\tau z_{ij}(\beta_0)) + \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) \frac{(\lambda^\tau z_{ij}(\beta_0))^2}{1 + \lambda^\tau z_{ij}(\beta_0)} = 0,$$

which in turn can be rewritten as

$$\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) - \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) z_{ij}^\tau(\beta_0) \right) \lambda + \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) \frac{(\lambda^\tau z_{ij}(\beta_0))^2}{1 + \lambda^\tau z_{ij}(\beta_0)} = 0. \quad (6.9)$$

This implies that

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) z_{ij}^\tau(\beta_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) + \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) z_{ij}^\tau(\beta_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) \frac{(\lambda^\tau z_{ij}(\beta_0))^2}{1 + \lambda^\tau z_{ij}(\beta_0)}$$

from which, it follows that

$$\lambda = \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) z_{ij}^\tau(\beta_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) + o_p(1),$$

since $\left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) z_{ij}^\tau(\beta_0) \right)^{-1} \frac{1}{n} \sum_{i=1}^n z_{ij}(\beta_0) \frac{(\lambda^\tau z_{ij}(\beta_0))^2}{1 + \lambda^\tau z_{ij}(\beta_0)} = o_p(1)$. Also, premultiplying equation (6.9) by λ , yields

$$\frac{1}{n} \sum_{i=1}^n \lambda^\tau z_{ij}(\beta_0) - \frac{1}{n} \sum_{i=1}^n (\lambda^\tau z_{ij}(\beta_0))^2 + \frac{1}{n} \sum_{i=1}^n \frac{(\lambda^\tau z_{ij}(\beta_0))^3}{1 + \lambda^\tau z_{ij}(\beta_0)} = 0,$$

which results in

$$\sum_{i=1}^n (\lambda^\tau z_{ij}(\beta_0))^2 = \sum_{i=1}^n \lambda^\tau z_{ij}(\beta_0) + o_p(1)$$

since $\sum_{i=1}^n \frac{(\boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0))^3}{1 + \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0)} = o_p(1)$. Hence,

$$\begin{aligned} R_j(\boldsymbol{\beta}_0) &= \sum_{i=1}^n \boldsymbol{\lambda}^\tau z_{ij}(\boldsymbol{\beta}_0) + o_p(1) \\ &= \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\boldsymbol{\beta}_0) \right)^\tau \left(\sum_{i=1}^n z_{ij}(\boldsymbol{\beta}_0) z_{ij}^\tau(\boldsymbol{\beta}_0) \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n z_{ij}(\boldsymbol{\beta}_0) \right) + o_p(1) \end{aligned}$$

From this, note that by Theorem 10, we have $\frac{1}{n} \sum_{i=1}^n z_{ij}(\boldsymbol{\beta}_0) z_{ij}^\tau(\boldsymbol{\beta}_0) \rightarrow \mathbf{V}_j$ a.s. Thus,

$$R_j(\boldsymbol{\beta}_0) \approx \left(\mathbf{V}_j^{-1/2} \sqrt{n} S_{jn}(\boldsymbol{\beta}_0) \right)^\tau \left(\mathbf{V}_j^{-1/2} \sqrt{n} S_{jn}(\boldsymbol{\beta}_0) \right) + o_p(1).$$

On the other hand, noting that $\mathbf{V}_j^{-1/2} \sqrt{n} S_{jn}(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \mathbf{I})$, where $\mathbf{I}$ is a $p \times p$ identity matrix, $R_j(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \chi_p^2$ and the proof is complete. $\square$

Proof of Theorem 12. From equation (4.10), $\boldsymbol{\xi}$ satisfies

$$\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \frac{\phi_{ij}(G(t_{0i}))}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = 0. \quad (6.10)$$

Now, by assumption (I_1) and (I_3) , one can obtain in a straightforward manner that $\max_{1 \leq i \leq n} \|\phi_{ij}(t_{0i})\| = o_p(\sqrt{nh_n^{q+1}})$. Also, $\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \|\phi_{ij}(t_{0i})\|^3 = o_p(\sqrt{nh_n^{q+1}})$. Following the same argument as in the proof of Theorem 10, we have $\|\boldsymbol{\xi}\| = O_p(1/\sqrt{nh_n^{q+1}})$, and by Theorem 10, $\|\nabla J_{jn}(G(t_0))\| = O_p(\sqrt{nh_n^{q+1}})$. Once again, a direct Taylor expansion up to order 2 of the right hand side of equation (4.11) gives

$$H_j(G(t_0)) = 2 \sum_{i=1}^n \left[\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})) - \frac{1}{2} (\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2 \right] + \boldsymbol{\xi}_n,$$

where $\|\boldsymbol{\xi}_n\| = O_p(1) \sum_{i=1}^n |\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))|^3 = O_p(1) \|\boldsymbol{\xi}\|^3 \sum_{i=1}^n \|\phi_{ij}(G(t_{0i}))\|^3 = o_p(1)$. Following the same argument as in the proof of Theorem 10, solving equation (6.10) is equivalent to solving

$$\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) (1 - \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))) + \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \frac{(\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = 0,$$

which in turn can be rewritten as

$$\begin{aligned} & \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) - \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right) \boldsymbol{\xi} \\ & + \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \frac{(\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = 0. \end{aligned} \quad (6.11)$$

Solving this equation for $\boldsymbol{\xi}$ gives

$$\begin{aligned} \boldsymbol{\xi} = & \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right)^{-1} \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \\ & + \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right)^{-1} \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \frac{(\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} \end{aligned}$$

from which, it follows that

$$\boldsymbol{\xi} = \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right)^{-1} \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) + o_p(1),$$

as $\left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right)^{-1} \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \frac{(\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = o_p(1)$. Moreover, premultiplying equation (6.11) by $\boldsymbol{\xi}$, yields

$$\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})) - \frac{1}{nh_n^{q+1}} \sum_{i=1}^n (\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2 + \frac{1}{nh_n^{q+1}} \sum_{i=1}^n \frac{\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))^3}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = 0,$$

which in turn results in

$$\sum_{i=1}^n (\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^2 = \sum_{i=1}^n \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})) + o_p(1)$$

since $\sum_{i=1}^n \frac{(\boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})))^3}{1 + \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i}))} = o_p(1)$. Thus,

$$\begin{aligned} H_j(G(t_0)) &= \sum_{i=1}^n \boldsymbol{\xi}^\tau \phi_{ij}(G(t_{0i})) + o_p(1) \\ &= \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \right)^\tau \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) \right)^{-1} \\ &\quad \times \left(\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \right) + o_p(1) \end{aligned}$$

Noting that $\frac{1}{nh_n^{q+1}} \sum_{i=1}^n \phi_{ij}(G(t_{0i})) \phi_{ij}^\tau(G(t_{0i})) = \Omega_{nj}$ is symmetric, we have

$$H_j(G(t_0)) = \left(\Omega_{nj}^{-1/2}(t_0) \sqrt{nh_n^{q+1}} \nabla J_{jn}(G(t_0)) \right)^\tau \left(\Omega_{nj}^{-1/2}(t_0) \sqrt{nh_n^{q+1}} \nabla J_{jn}(G(t_0)) \right) + o_p(1).$$

To this end, from Remark 4.2, $\Omega_{nj}^{-1/2}(t_0) \sqrt{nh_n^{q+1}} \nabla J_{jn}(G(t_0)) \xrightarrow{\mathcal{D}} N_{q+1}(\mathbf{0}, \mathbf{I})$, where $\mathbf{I}$ is a $(q+1) \times (q+1)$ identity matrix. Thus, $H_j(G(t_0)) \xrightarrow{\mathcal{D}} \chi_{q+1}^2$, which completes the proof. $\square$