University of South Carolina
Scholar Commons

2018

Semiparametric Regression In The Presence Of Measurement Error

Xiang Li
University of South Carolina - Columbia

Follow this and additional works at: https://scholarcommons.sc.edu/etd

Part of the Statistics and Probability Commons

www.manaraa.com

Semiparametric Regression
In The Presence
Of Measurement Error

by

Xiang Li

Bachelor of Science
Anhui University 2012

Master of Science
University of Science and Technology of China 2014

_____

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2018

Accepted by:

Xianzheng Huang, Major Professor

John Grego, Committee Member

Dewei Wang, Committee Member

Jiajia Zhang, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

ACKNOWLEDGMENTS

This dissertation is the result of my past ten years' journey in Statistics. Thanks to Statistics, I met a lot of people and got a lot of help from them in this great trip.

First and foremost, I would like to thank my parents who always had faith in me for supporting my PhD study. Because of your support, I am able to broaden my eyes and concentrate on the thing I am interested in most.

I would also like to express my deepest gratitude to my advisor Dr. Xianzheng Huang, for her remarkable guidance. I am impressed by her great ability to pay attention to details and I am truly grateful to her for instilling in me that how to become a great researcher.

In the end, thanks my friends who ever help me during this long trip.

ABSTRACT

The error-in-covariates problem has received great attention among researchers who study semiparametric and nonparametric inference for regression models over the past two decades. Without correcting for the measurement error in covariates, estimators for covariate effect usually contain bias. To account for measurement error, much research have been done in mean regression (Liang et al., 1999; Fuller, 2009; Carroll et al., 2006) and quantile regression (He and Liang, 2000; Hardle et al., 2000; Wei and Carroll, 2009). In contrast, there is little research in mode regression and this motivates us to propose semiparametric methods to address this error-in-covariates problem in Chapters 1 and 3.

Chapter 1 considers estimating the mode of a response given an error-prone covariate X by assuming that the mode of Y given X is a linear function of X. It is first shown that ignoring measurement error typically leads to inconsistent inference for the mode of the response given the true covariate, as well as misleading inference for regression coefficients in the conditional mode model. To account for measurement error, the Monte Carlo corrected score method (Novick and Stefanski, 2002) is employed to numerically obtain an unbiased score function based on which the regression coefficients is estimated consistently. To relax the normality assumption on measurement error the first method requires, the corrected kernel method is proposed. In this method, an objective function constructed using deconvoluting kernels is maximized to obtain consistent estimators of the regression coefficients. Besides rigorous investigation on large sample properties of the new estimators, we study their finite sample performance via extensive simulation experiments, and find

that the proposed methods substantially outperform a naive inference method that ignores measurement error.

In Chapter 2, we assume that the mode of Y is a linear function of a covariate X and it also depends on another covariate T in an unspecified functional form. This leads to a partially linear model for the conditional mode. We employ B-splines to approximate the unspecified function that relates Y and T. To estimate the covariate effects explaining the association between Y and X, and at the same time, estimate the unspecified function linking Y and T, we develop two methods for inferring these two parts of the partially linear mode model. A simulation study is designed to show the performance of two proposed methods. Chapter 3 considers estimating the mode of a response in partially linear models when the aforementioned X is error-prone. To account for measurement error, we incorporate the corrected kernel method proposed in Chapter 1 and the proposed estimation methods in Chapter 2 to infer the parametric part and nonparametric part of the conditional mode accounting for measurement error in X. Results from simulation studies suggest that the proposed method substantially outperform a naive inference method that ignores measurement error. Instead of considering error-prone covariates, in Chapter 4, we consider a scenario where the response is contaminated by Berkson measurement error. In particular, we tackle the regression analysis for a pooled continuous response. Finally, Chapter 5 discusses future research in my dissertation.

TABLE OF CONTENTS

List of Tables

xi

LIST OF FIGURES

Chapter 1

Linear Mode Regression with Covariate Measurement Error

1.1 Introduction

Regression analysis has been a standard platform on which one studies the association between a response, Y, and covariates of interest, X. The majority of the literature on regression analysis is devoted to mean regression, where the mean of Y given X is the focal point of inference. There also exists a large body of work on quantile regression, where one infers quantiles of Y conditioning on X (Koenker, 2005). In contrast, there have been much less study on mode regression (Lee, 1989; Yao and Li, 2014; Chen et al., 2016), which aims to characterize the mode of Y given X. The mode of a distribution is an informative summary feature that is more of interest than the mean or quantiles in many applications (Parzen, 1962), such as biology (Hedges and Shah, 2003), economy (Huang and Yao, 2012), meteorology (Hyndman et al., 1996), astronomy (Bamford et al., 2008), and traffic engineering (Einbeck and Tutz, 2006), where the underlying distributions of Y given X are often skewed. In these referenced works, the most likely value of Y given a covariate value, as opposed to some average value of the response, is of scientific interest; and a location measure that is resistant to outliers, such as the mode, is more appealing. In these applications, some covariates cannot be measured directly or precisely, and only data for their error-contaminated surrogates are collected.

To address complications caused by error-prone covariates, a good collection of

1

methods for mean regression that account for covariate measurement error have been developed (Fuller, 2009; Carroll et al., 2006). There are also some approaches that take measurement error into consideration in quantile regression (He and Liang, 2000; Wei and Carroll, 2009; Wang et al., 2012). However, there is little research on mode regression in the presence of measurement error in covariates. The only work we are aware of that addresses measurement error in mode regression is by Zhou and Huang (2016), where the authors proposed nonparametric methods to estimate the mode of Y given X based on kernel density estimators. Different from the nonparametric route they took, here we consider a class of linear mode regression models, following the footsteps of existing works on mean regression (Fuller, 2009) and quantile regression (He and Liang, 2000; Wei and Carroll, 2009; Wang et al., 2012) with measurement error, where one starts by considering the conditional mean or quantiles as some linear function of covariates. This class of mode regression models has been mostly investigated by econometricians (Lee, 1989, 1993; Kemp and Silva, 2012), and all existing works assume error-free covariates. To the best of our knowledge, we are the first to investigate linear mode regression with covariate measurement error.

In this chapter, we first formulate the class of linear measurement error mode models in Section 1.2, and provide some preliminary analysis on the effect of measurement error on inference when one ignores measurement error. We propose two methods to estimate the regression coefficients in the model that account for measurement error in Section 1.3. Both methods depend on the choice of a bandwidth, for which we present a strategy in Section 1.4. Section 1.5 reports simulation studies where we compare the two proposed methods with a naive method that ignores measurement error, using estimates from the method proposed by Yao and Li (2014) applied to error-free data as benchmarks. Section 1.6 presents an application of the three methods to dietary data collected from the Women's Interview Survey of Health. We point out extension of the proposed methods under more general settings and discuss

follow-up research agendas in Section 1.7.

1.2 Preambles

Data and models

Suppose that the observed data consist of n independent data points, $\{(Y_j, W_j)\}_{j=1}^n$, where $\{W_j\}_{j=1}^n$ are surrogates of the unobserved covariate values, $\{X_j\}_{j=1}^n$, and Y_j given X_j follows a distribution specified by the probability density function $f_{Y|X}(y \mid x)$, for $j = 1, \ldots, n$. As in Grund and Hall (1995), we assume that $f_{Y|X}(y \mid x)$ has a unique largest mode; in particular, we assume a linear model for this conditional mode,

$$y_M(x) = \text{Mode}(Y_j \mid X_j = x) = \beta_0 + \beta_1 x \quad (j = 1, \ldots, n), \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$ is the regression coefficient vector containing parameters to be estimated.

A classical additive measurement error model is assumed in this study, according to which W_j relates to X_j via

$$W_j = X_j + U_j, \tag{1.2}$$

where U_j is the nondifferential measurement error (Carroll et al., 2006, Section 2.5), for $j = 1, \ldots, n$, following a distribution specified by the density function $f_U(u)$, of which the mean is zero and variance is σ^2. Measurement error in (1.2) being nondifferential essentially implies that, conditioning on X, Y and W are independent, where the index j is suppressed when we refer to a generic data point, X_j, Y_j, or W_j, for $j \in \{1, \ldots, n\}$. For model identifiability reasons, we assume $f_U(u)$ known. Considerations for cases where extra data are available to infer $f_U(u)$ are given in Section 1.7. Finally, we consider a univariate covariate for illustration purposes in the majority of the study, and discuss in Section 1.7 generalization to multivariate covariates that may include some error-free components.

3

NAIVE INFERENCE

Denote by $y_M^*(w)$ the mode of the conditional density of Y given $W = w$, $f_{Y|W}(y \mid w)$. In the context of linear mode regression, a naive inference method infers $y_M^*(w)$ assuming, as in (1.1), $y_M^*(w) = \beta_0^* + \beta_1^* w$. In what follows, we use an example to demonstrate that naive inference for the mode function can be misleading.

Suppose Y given $X = x$ follows a distribution with mean $m(x) = \alpha_0 + \alpha_1 x$ and standard deviation $\sigma(x) = \gamma_0 + \gamma_1 x$, where α_0, $\alpha_1 (\neq 0)$, γ_0 and γ_1 are constants free of x. In addition, suppose $X \sim N(\mu_X, \sigma_X^2)$ and $U \sim N(0, \sigma^2)$. Then, conditioning on $W = w$, Y follows a distribution with mean and standard deviation given by (Fuller, 2009)

$$
\begin{aligned}
m^*(w) &= \alpha_0 + (1-\lambda)\alpha_1\mu_X + \lambda\alpha_1 w, \\
\sigma^*(w) &= \sqrt{\{\gamma_0 + (1-\lambda)\gamma_1\mu_X + \lambda\gamma_1 w\}^2 + (1-\lambda)\alpha_1^2\sigma_X^2},
\end{aligned}
\tag{1.3}
$$

respectively, where $\lambda = \sigma_X^2/(\sigma_X^2 + \sigma^2)$ is the reliability ratio (Carroll et al., 2006, Section 3.2.1). Define two standardized mean residuals, $e = \{Y - m(X)\}/\sigma(X)$ and $e^* = \{Y - m^*(W)\}/\sigma^*(W)$. Denote by $e_M(x)$ the mode of e given $X = x$, and by $e_M^*(w)$ the mode of e^* given $W = w$. One can show that

$$
y_M(x) = m(x) + \sigma(x)e_M(x) = \alpha_0 + \alpha_1 x + (\gamma_0 + \gamma_1 x)e_M(x),
$$

and similarly

$$
\begin{aligned}
y_M^*(w) &= m^*(w) + \sigma^*(w)e_M^*(w) \\
&= \alpha_0 + (1-\lambda)\alpha_1\mu_X + \lambda\alpha_1 w + \\
&\sqrt{\{\gamma_0 + (1-\lambda)\gamma_1\mu_X + \lambda\gamma_1 w\}^2 + (1-\lambda)\alpha_1^2\sigma_X^2}\,e_M^*(w).
\end{aligned}
\tag{1.4}
$$

Comparing $y_M(x)$ and $y_M^*(w)$ above, one can see that, even if $e_M(x)$ and $e_M^*(w)$ are both constant functions, the naive mode $y_M^*(w)$ is not a linear function unless $\gamma_1 = 0$ or $\lambda = 1$, whereas the true mode $y_M(x)$ is linear if $e_M(x)$ does not depend on x.

This example suggests that naive linear mode regression usually involves an extra layer of model misspecification compared to naive linear mean regression. As one

4

sees in (1.3), when $m(x)$ is linear in x, $m^*(w)$ is also linear in w when X and U are independent normal random variables. Thus effects of measurement error on mode regression are in general far more complicated than those in the context of mean regression. In this example, if $\gamma_1 = 0$, β_1^* in the naive mode model revealed in (1.4) reduces to $\lambda\alpha_1$, which is attenuated compared to $\beta_1 = \alpha_1$ in (1.1) when $e_M(x)$ is free of x.

1.3 PROPOSED METHODS

INFERENCE IN THE ABSENCE OF MEASUREMENT ERROR

Given a fixed y in the support of Y, $Q_h(y) = n^{-1}\sum_{j=1}^{n}K_h(Y_j - y)$ is the local constant kernel density estimator (Silverman, 1986) of the density of Y evaluated at y, $f_Y(y)$, where $K(t)$ is a kernel, h is the bandwidth, and $K_h(t) = K(t/h)/h$. Since the mode of Y maximizes its density function $f_Y(y)$, a sensible estimator for the mode of Y is the maximizer of $Q_h(y)$. Motivated by this viewpoint, in the absence of covariate measurement error, Yao and Li (2014) proposed to estimate $\boldsymbol{\beta}$ by maximizing

$$Q_h(\beta) = n^{-1}\sum_{j=1}^{n}K_h(Y_j - \beta_0 - \beta_1 X_j). \tag{1.5}$$

Setting $K(t)$ as the standard normal density, Yao and Li (2014) developed an expectation maximization algorithm to compute their estimate of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$. In addition, they derived the order of the bias and variance of $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$ as $n \to \infty$, and established its asymptotic normality.

Naive implementation of Yao and Li's method using error-contaminated data is to substitute X_j with W_j in (1.5), resulting in a naive objective function one maximizes with respect to $\boldsymbol{\beta}$. Denote by $\hat{\boldsymbol{\beta}}_{\mathrm{NV}}$ the resultant naive estimator of $\boldsymbol{\beta}$. To account for measurement error, we revise this naive method from two perspectives.

MC

MONTE CARLO CORRECTED SCORE METHOD

Maximizing $Q_h(\boldsymbol{\beta})$ in (1.5) with respect to $\boldsymbol{\beta}$ is equivalent to solving the score equations for $\boldsymbol{\beta}$, $\sum_{j=1}^n \Psi(Y_j, X_j; \boldsymbol{\beta}) = \mathbf{0}$, where $\Psi(Y_j, X_j; \boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta}) K_h(Y_j - \beta_0 - \beta_1 X_j)$. In the presence of measurement error, naively applying Yao and Li's method amounts to using the naive score, $\Psi(Y, W; \boldsymbol{\beta})$, in place of the true score, $\Psi(Y, X; \boldsymbol{\beta})$. One way to correct this naive score-based estimation for measurement error is to construct a score function that depends on (Y, W), whose expectation conditioning on (Y, X) is equal to $\Psi(Y, X; \boldsymbol{\beta})$. This leads to the corrected score method (Nakamura, 1990), which has found its successes in linear mean regression, several nonlinear mean regression models (Carroll et al., 2006, Chapter 7), and some survival models (Song and Huang, 2005; Wang, 2006; Zucker and Spiegelman, 2008) with covariate measurement error.

Although the idea of correcting the naive score by using an unbiased estimator of the true score leads to a very general strategy to account for measurement error, such unbiased estimator, referred to as a corrected score, often does not exist in closed form. Novick and Stefanski (2002) developed a Monte Carlo procedure to numerically obtain a corrected score under the assumption that $U \sim N(0, \sigma^2)$ and $\Psi(Y, X; \beta)$ is an entire function with respect to its second argument (Boas, 2011). By using the standard normal kernel in (1.5), we have the true score $\Psi(Y, X, \boldsymbol{\beta})$ as an entire function in X, which allows us follow their Monte Carlo procedure to obtain an estimator of $\boldsymbol{\beta}$ via the following four-step algorithm.

- MC-1: For $b = 1, \ldots, B$, generate independent random errors, $\{U_{b,j}\}_{j=1}^n$, from $N(0, \sigma^2)$.

- MC-2: Form the complex-valued data, $\{\widetilde{W}_{b,j} = W_j + iU_{b,j}\}_{j=1}^n$, where i is the imaginary unit, for $b = 1, \ldots, B$.

6

- MC-3: Compute $\Psi_{\mathrm{MC},\,B}(Y_j, W_j; \boldsymbol{\beta}) = B^{-1} \sum_{b=1}^{B} \mathrm{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}$, where $\mathrm{Re}(t)$ denotes the real part of a complex-valued t.

- MC-4: Solve the following estimating equations for $\boldsymbol{\beta}$,

$$\sum_{j=1}^{n} \Psi_{\mathrm{MC},\,B}(Y_j, W_j; \boldsymbol{\beta}) = \mathbf{0}. \tag{1.6}$$

Denote the resultant estimator as $\hat{\beta}_{\mathrm{MC}}$.

By proving that $E[\mathrm{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\} \mid (Y_j, X_j)] = \Psi(Y_j, X_j; \boldsymbol{\beta})$, Novick and Stefanski (2002) showed that $\mathrm{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}$ is a corrected score that involves extra noise due to its dependence on $U_{b,j}$. A corrected score that is free of the extra noise is $E[\mathrm{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\} \mid (Y_j, W_j)]$, which usually cannot be derived analytically. This motivates MC-3 above, where one computes the average of $\{\mathrm{Re}\{\Psi(Y_j, \widetilde{W}_{b,j}; \boldsymbol{\beta})\}, b = 1, \ldots, B\}$ as an approximation of the aforementioned expectation. Clearly, this empirical mean, $\Psi_{\mathrm{MC},\,B}(Y_j, W_j; \boldsymbol{\beta})$, is also a corrected score, referred to as the Monte Carlo corrected score. Using the fact that $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ is an M-estimator that solves the estimating equations in (1.6) constructed from an unbiased score function, Novick and Stefanski (2002, Section 5) established the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$. Finally, they demonstrated that, even when the assumption of U being normally distributed or the true score function being entire is violated, $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ is still often less biased than the counterpart naive estimator.

Corrected kernel method

Even though the Monte Carlo corrected score method enjoys certain degree of robustness to the normality assumption on U, an alternative method that is well justified for more general error distributions is desirable. This motivates us to correct the naive method from a different angle. Instead of correcting the naive score function, we propose to correct the naive objective function for measurement error. This is accomplished by constructing an unbiased estimator of the summand in (1.5),

7

www.manaraa.com

$K_h(Y - \beta_0 - \beta_1 X)$, based on (Y, W). Since the objective function $Q_h(\beta)$ originates from a kernel density estimator, such unbiased estimator is readily available in Carroll and Hall (1988) and Stefanski and Carroll (1990), where the authors considered nonparametric density estimation in the presence of measurement error. Following their construction of a deconvoluting kernel, one can show that, conditioning on (Y, X), an unbiased estimator of $K_h(Y - \beta_0 - \beta_1 X)$ is $K_h^*(Y - \beta_0 - \beta_1 W)$, where $K_h^*(t) = K^*(t/h)/h$, and

$$K^*(t) = \frac{1}{2\pi} \int e^{-ist} \frac{\phi_K(s)}{\phi_U(-\beta_1 s/h)} ds, \tag{1.7}$$

in which $\phi_K(s)$ is the Fourier transform of $K(t)$, $\phi_U(s)$ is the characteristic function of U that does not vanish, both assumed to be even, and the integration is over the real line. Besides being used for density estimation in the works of Carroll, Hall, and Stefanski, Fan and Truong (1993) also used a deconvoluting kernel similar to that in (1.7) to construct a local constant estimator of $E(Y \mid X = x)$ in the presence of measurement error in X. Replacing the naive quantity, $K_h(Y - \beta_0 - \beta_1 W)$, with $K_h^*(Y - \beta_0 - \beta_1 W)$ gives the corrected objective function to be maximized with respect to $\boldsymbol{\beta}$,

$$Q_h^*(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^{n} K_h^*(Y_j - \beta_0 - \beta_1 W_j). \tag{1.8}$$

We call this method the corrected kernel method and denote the resultant estimator as $\hat{\boldsymbol{\beta}}_{\text{CK}}$. One existing work that also corrects an objective function for measurement error is Wang et al. (2012) in the context of linear quantile regression. In this work, the authors derived a smooth function depending on (Y, W), of which the conditional expectation given (Y, X) approaches to the true objective function as the smoothing parameter involved in the smooth function shrinks to zero.

Stefanski and Carroll (1990) studied the validity of the construction of (1.7) and its properties for two types of measurement error distributions, namely ordinary smooth error distributions and super smooth error distributions (Fan, 1991b). Their definitions are given next.

8

Definition 1. *The distribution of U is ordinary smooth of order b if, as $|t| \to \infty$, $d_0|t|^{-b} \leq |\phi_U(t)| \leq d_1|t|^{-b}$ for some positive constants d_0, d_1 and b.*

Definition 2. *The distribution of U is super smooth of order b if, as $|t| \to \infty$, $d_0|t|^{b_0}\exp(-|t|^b/d_2) \leq |\phi_U(t)| \leq d_1|t|^{b_1}\exp(-|t|^b/d_2)$ for some positive constants d_0, d_1, d_2, b, b_0, and b_1.*

For example, Laplace distributions are ordinary smooth of order $b = 2$, and normal distributions are super smooth of order $b = 2$. We derive the asymptotic bias and variance of $\hat{\boldsymbol{\beta}}_{\text{CK}}$ under each type of measurement error distributions, and also establish its asymptotic normality. These findings are summarized in the following two theorems. Detailed proofs are provided in Appendices A and B. Lemmas referenced in the theorems along with their proofs are given in Appendix C.

Denote by $g(\epsilon \mid x)$ the density of the mode residual, $\epsilon = Y - \beta_0 - \beta_1 x$, and let $\tilde{X} = (1, X)^{\mathrm{T}}$. The following three conditions on $g(\epsilon \mid x)$ and the covariate are assumed for the theorems.

(C1) The ℓ-th partial derivative of $g(\epsilon \mid x)$ with respect to ϵ, $g^{(\ell)}(\epsilon \mid x)$, is continuously differentiable around $\epsilon = 0$, for $\ell = 0, 1, 2, 3$, and $g^{(1)}(0 \mid x) = 0$ for all x in the support of X.

(C2) As $n \to \infty$, $n^{-1}\sum_{j=1}^{n} g(0 \mid X_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}$ and $n^{-1}\sum_{j=1}^{n} g^{(3)}(0 \mid X_j)\tilde{X}_j$ converge in probability, and $n^{-1}\sum_{j=1}^{n} g^{(2)}(0 \mid X_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}$ converges in probability to a negative definite matrix.

(C3) As $n \to \infty$, $n^{-1}\sum_{j=1}^{n} \|\tilde{X}_j\|^4 = O_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm.

Condition (C1) implies certain smoothness of $g(\epsilon \mid x)$ around its mode $\epsilon = 0$, which are mild assumptions typically required for kernel-based estimation of a function. Condition (C2) indicates existence of several expectations relevant to the asymptotic

mean and variance of the proposed estimator. Condition (C3) guarantees limits $\lim_{n\to\infty} n^{-1} \sum_{j=1}^n X_j^\ell$, for $\ell \le 4$, exist in probability. These assumptions are also imposed in Yao and Li (2014) and are indeed mild assumptions satisfied in a wide range of applications. Additional conditions that are required for proving the lemma referenced in the following two theorems are provided in Appendix A. These include conditions concerning $K(t)$ and $\phi_U(t)$. Conditions on $K(t)$ are imposed mainly to guarantee integrability of functions of the forms $t^{\ell_1}\phi_K^2(t)$ and $t^{\ell_1}K^{(\ell_2)}(t)$ for some positive integers ℓ_1 and ℓ_2. Essentially, these conditions suggest that $\phi_K(t)$ and $K^{(\ell_2)}(t)$ tail off fast enough as $|t| \to \infty$, which can be easily satisfied by choosing an adequate kernel such as the one we use for the corrected kernel method in the simulation study reported in Section 1.5. Conditions imposed on $\phi_U(t)$ are also mainly about how fast $\phi_U^{(\ell)}(t)$ tail off as $|t| \to \infty$ for some nonnegative integer ℓ.

Theorem 1. *Under conditions (C1)–(C3) and conditions in Lemma C, there exists a maximizer of $Q_h^*(\boldsymbol{\beta})$, denoted by $\hat{\boldsymbol{\beta}}_{CK}$, such that, as $n \to \infty$ and $h \to 0$,*

(i) when U follows an ordinary smooth distribution of order b, if $nh^{7+2b} \to 0$, then

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left(\sqrt{\frac{1}{nh^{3+2b}}}\right); \qquad (1.9)$$

(ii) when U follows a super smooth distribution of order b,
if $\exp(2|\beta_1|^b h^{-b}/d_2))/(nh^{b_6}) \to 0$, where $b_6 = \max\{3 - 2\min(b_2, b_3),$
$5 - 2\min(b_2, b_3, b_4), 7 - 2\min(b_2, b_3, b_4, b_5)\}$, in which b_ℓ, for $\ell = 2, 3, 4, 5$, are defined in Lemma C, then

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left\{\exp\left(\frac{|\beta_1|^b}{d_2 h^b}\right)\sqrt{\frac{1}{nh^{3-2\min(b_2, b_3)}}}\right\}. \qquad (1.10)$$

The error rates presented in Theorem 1 combine the rate of bias, appearing in the big-O part of (1.9) and (1.10), and the rate of standard deviation, as in the big-O_p part of (1.9) and (1.10), of $\hat{\boldsymbol{\beta}}_{CK}$. Three observations are worth pointing out regarding these

rates. First, the bias rate is not affected by measurement error, and coincides with the bias rate of Yao and Li's estimator in the absence of measurement error (Yao and Li, 2014, Theorem 2.2). Second, compared to the variance rate of Yao and Li's estimator in the absence of measurement error (Yao and Li, 2014, Theorem 2.2), the variance rates here are inflated due to measurement error. By setting $b = 0$, the variance rates suggested by (1.9) and (1.10) reduce to $O_p\{1/(nh^3)\}$, which is the variance rate of Yao and Li's estimator. Setting $b = 0$ is equivalent to setting $\sigma^2 = 0$, which leads to error-free covariate. Third, comparing (1.9) and (1.10) reveals that the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$ in the presence of super smooth measurement error is much slower than that when U is ordinary smooth. This is in line with the findings in density estimation (Carroll and Hall, 1988; Stefanski and Carroll, 1990), local polynomial estimation in mean regression (Fan and Truong, 1993; Delaigle et al., 2009; Huang and Zhou, 2017), and nonparametric mode regression (Zhou and Huang, 2016) in the presence of different types of measurement error.

Moments of certain functions that involve Fourier transform are derived in Appendix C to show Theorem 1. Results regarding these moments, along with strategies for deriving them, are also useful for establishing the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$, although additional assumptions listed under Conditions N in Appendix A in the supplementary materials are needed as well.

Theorem 2. *Under conditions N, besides the same assumptions imposed in Theorem 1, for the maximizer of $Q_h^*(\boldsymbol{\beta})$, $\hat{\boldsymbol{\beta}}_{CK}$, that satisfies the properties in Theorem 1,*

(i) if U follows an ordinary smooth distribution of order b,

$$\sqrt{nh^{3+2b}} \left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2 \mu_2 J^{*-1} Q/4 \right) \xrightarrow{d} N(0, J^{*-1} K_L J^{*-1}); \qquad (1.11)$$

where K_L is a constant matrix, $Q = \lim_{n\to\infty} n^{-1} \sum_{j=1}^n E\{g^{(3)}(0|X_j)\tilde{X}_j\}$, and $J^* = \lim_{n\to\infty} n^{-1} \sum_{j=1}^n E\{g^{(2)}(0|X_j)\tilde{X}_j \tilde{X}_j^{\mathrm{T}}\}$;

11

(ii) if U follows a super smooth distribution of order b,

$$\left\{ Var(\hat{\boldsymbol{\beta}}_{CK}) \right\}^{-1/2} \left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2 \mu_2 J^{*-1} Q/4 \right) \xrightarrow{d} N(0, 1), \qquad (1.12)$$

where $Var(\hat{\boldsymbol{\beta}}_{CK}) = O[\exp\{2|\beta_1|^b/(d_2 h^b)\}/\{nh^{3-2\min(b_2,b_3)}\}]$, *and* $\boldsymbol{\Sigma}^{-1/2}$ *denotes the inverse of the positive definite square root of a positive definite matrix* $\boldsymbol{\Sigma}$.

1.4 BANDWIDTH SELECTION

Kernel-based methods are typically sensitive to the choice of bandwidths. To address the complication in bandwidth selection due to measurement error, Delaigle and Hall (2008) developed a strategy for smoothing parameter selection that combines simulation-extrapolation (SIMEX) (Cook and Stefanski, 1994; Stefanski and Cook, 1995) and cross validation. We apply this strategy to choose h following the algorithm described next, where we aim to choose an h that optimizes inference for $\boldsymbol{\beta}$ in some sense. Generically denote by $\hat{\boldsymbol{\beta}}_h$ an estimator of $\boldsymbol{\beta}$ under consideration with the bandwidth fixed at h based on observed data $\{(Y_j, W_j)\}_{j=1}^n$.

- SM-1: Generate M sets of further contaminated covariate data, $\{W_{m,j}^* = W_j + U_{m,j}^*\}_{j=1}^n$, for $m = 1, \ldots, M$, where $\{U_{m,j}^*, j = 1, \ldots, n\}_{m=1}^M$ are independent random errors generated from $f_U(u)$.

- SM-2: For $m = 1, \ldots, M$, denote by $\hat{\boldsymbol{\beta}}_{h,m}^*$ the estimate of $\boldsymbol{\beta}$ based on data $\{(Y_j, W_{m,j}^*)\}_{j=1}^n$ using the method under consideration. Find

$$h_1 = \underset{h>0}{\arg\min} \; M^{-1} \sum_{m=1}^M (\hat{\boldsymbol{\beta}}_{h,m}^* - \hat{\boldsymbol{\beta}}_h)^{\mathrm{T}} \boldsymbol{S}_{h,1}^{-1} (\hat{\boldsymbol{\beta}}_{h,m}^* - \hat{\boldsymbol{\beta}}_h),$$

where $\boldsymbol{S}_{h,1}$ is the sample variance-covariance matrix of $\{\hat{\boldsymbol{\beta}}_{h,m}^* - \hat{\boldsymbol{\beta}}_h\}_{m=1}^M$.

- SM-3: Generate M sets of even further contaminated covariate data, $\{W_{m,j}^{**} = W_{m,j}^* + U_{m,j}^{**}\}_{j=1}^n$, for $m = 1, \ldots, M$, where $\{U_{m,j}^{**}, j = 1, \ldots, n\}_{m=1}^M$ are independent random errors generated from $f_U(u)$, which are also independent of $\{U_{m,j}^*, j = 1, \ldots, n\}_{m=1}^M$.

- SM-4: For $m = 1, \ldots, M$, denote by $\hat{\boldsymbol{\beta}}_{h,m}^{**}$ the estimate of $\boldsymbol{\beta}$ based on data $\{(Y_j, W_{m,j}^{**})\}_{j=1}^n$ using the method under consideration. Find

$$h_2 = \underset{h>0}{\arg\min}\, M^{-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}_{h,m}^{**} - \hat{\boldsymbol{\beta}}_{h,m}^{*})^{\mathrm{T}} \boldsymbol{S}_{h,2}^{-1} (\hat{\boldsymbol{\beta}}_{h,m}^{**} - \hat{\boldsymbol{\beta}}_{h,m}^{*}),$$

where $\boldsymbol{S}_{h,2}$ is the sample variance-covariance matrix of $\{\hat{\boldsymbol{\beta}}_{h,m}^{**} - \hat{\boldsymbol{\beta}}_{h,m}^{*}\}_{m=1}^{M}$.

- SM-5: Set the selected bandwidth as $h = h_1^2/h_2$.

The criterion we minimize in SM-2 and SM-4 is motivated by an ideal, or theoretical optimal, bandwidth given by $h_{\text{ideal}} = \arg\min_{h>0} E\{(\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{\Sigma}_h^{-1} (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})\}$, where $\boldsymbol{\Sigma}_h$ is the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_h$. The rationale behind this SIMEX procedure is that, as shown in Delaigle and Hall (2008), $\log(h_{\text{ideal}}) - \log(h_1) \approx \log(h_1) - \log(h_2)$ when σ^2 is small. And thus the value of h from SM-5 is a sensible approximation of h_{ideal}. Besides Delaigle and Hall (2008), Wang et al. (2012) also used a similar strategy to select the smoothing parameter in their problem of linear quantile regression with covariate measurement error.

1.5   Empirical Evidence

Simulation design

To assess finite sample performance of the proposed estimators, we design comparative experiments where $\hat{\boldsymbol{\beta}}_{\text{NV}}$, $\hat{\boldsymbol{\beta}}_{\text{MC}}$ (with $B = 1000$), and $\hat{\boldsymbol{\beta}}_{\text{CK}}$ are obtained based on simulated error-prone data $\{(Y_j, W_j)\}_{j=1}^n$, as well as $\hat{\boldsymbol{\beta}}_{\text{YL}}$ based on the corresponding error-free data $\{(Y_j, X_j)\}_{j=1}^n$. The fourth estimator serves as a gold standard in the sense that estimators, naive or non-naive, based on error-prone data are expected to be inferior in some regard than this estimator. Comparing the first three estimators with this reference estimator can shed light on how measurement error compromise the naive estimator, and whether or not the two proposed non-naive estimators improve over the naive estimator.

13

The kernel $K(t)$ used for obtaining $\hat{\boldsymbol{\beta}}_{\text{NV}}$, $\hat{\boldsymbol{\beta}}_{\text{MC}}$, and $\hat{\boldsymbol{\beta}}_{\text{YL}}$ is the standard normal density; and we use the kernel of which the Fourier transform is $\phi_K(t) = (1-t^2)^3 I(-1 \leq t \leq 1)$ for $\hat{\boldsymbol{\beta}}_{\text{CK}}$. The choice of kernel for the corrected kernel method is in part dictated by the technical conditions on $\phi_K(t)$ that arise from deriving asymptotic properties of $\hat{\boldsymbol{\beta}}_{\text{CK}}$. To mitigate the effects of data-driven bandwidth selection on the proposed estimators, in the first half of simulation, we use an approximation of h_{ideal} given by $\hat{h}_{\text{ideal}} = \arg\min_{h>0}(\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})^{\text{T}}\hat{\boldsymbol{\Sigma}}_h^{-1}(\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\Sigma}}_h$ is a bootstrap estimate of $\boldsymbol{\Sigma}_h$ based on 100 bootstrap samples. Clearly, $\hat{h}_{\text{ideal}}$ cannot be computed in practice since $\boldsymbol{\beta}$ is unknown in reality. In the second half of the simulation, we implement the SIMEX method described in Section 1.4, with $M = 10$, to select h for the proposed estimators. To preserve the integrity of $\hat{\boldsymbol{\beta}}_{\text{YL}}$, we run the Matlab code kindly provided by Professor Yao to compute $\hat{\boldsymbol{\beta}}_{\text{YL}}$ and $\hat{\boldsymbol{\beta}}_{\text{NV}}$, including their choice of bandwidth based on minimizing an estimate of the asymptotic mean squared error of Yao and Li's estimator of $\boldsymbol{\beta}$.

For ease of comparison, we follow the model setting in the simulation study presented in Yao and Li (2014) to generate error-free data. More specifically, for each of the two sample sizes, $n = 200$ and 400, the true covariate values $\{X_j\}_{j=1}^n$ are independent realizations from uniform$(0, 1)$. Given X_j, the response is generated according to $Y_j = 1 + 3X_j + (1 + 2X_j)e_j$, for $j = 1, \ldots, n$, where $\{e_j\}_{j=1}^n$ are independent errors from $0.5N(-1, 2.5^2) + 0.5N(1, 0.5^2)$. For this error distribution, $e_M(x) \approx 1$ for all $x \in [0, 1]$, and thus $y_M(x) \approx 2 + 5x$. Ignoring rounding error, we have the true mode regression coefficients $\boldsymbol{\beta} = (2, 5)^{\text{T}}$. The error contaminated covariate measures $\{W_j\}_{j=1}^n$ are generated according to (1.2), with U following a Laplace distribution and a normal distribution, respectively, whose mean is zero and variance σ^2 is set at four levels to achieve reliability ratios λ ranging from 0.75 to 0.9 at increments of 0.05.

14

Simulation results

Under each of sixteen model settings resulting from the combinations of n-$f_U(u)$-λ, 300 Monte Carlo replicate data sets of the form $\{(Y_j, X_j, W_j)\}_{j=1}^n$ are generated, producing 300 sets of estimates, $\{\hat{\boldsymbol{\beta}}_{\mathrm{NV}}, \hat{\boldsymbol{\beta}}_{\mathrm{MC}}, \hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\beta}}_{\mathrm{YL}}\}$, among which $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$ is not affected by the change in $f_U(u)$ or λ. Figure 1.1 presents the boxplots of these estimates when $n = 200$ for the case with Laplace measurement error when the approximated ideal bandwidth is used for $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$. Figure 1.2 depicts the boxplots of the estimates when $n = 200$, U is normal, and the approximated ideal bandwidth is used for $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$. Figures 1.3 and 1.4 provide the counterpart boxplots when h is chosen by the SIMEX method for $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$.

Overall, results for the two proposed methods that account for measurement error with bandwidths selected via the SIMEX method are very similar to those when the approximated ideal bandwidths are used. Except for higher variability, the two proposed estimates are comparable with the estimates obtained in the absence of measurement error, $\hat{\boldsymbol{\beta}}_{\mathrm{YL}}$; and the naive estimate, $\hat{\boldsymbol{\beta}}_{\mathrm{NV}} = (\hat{\beta}_{\mathrm{NV},0}, \hat{\beta}_{\mathrm{NV},1})^{\mathrm{T}}$, is compromised by measurement error in contrast. Under the current model setting, $\hat{\beta}_{\mathrm{NV},1}$ attenuates more towards null as error contamination in the covariate is more severe, that is, as λ decreases; and $\hat{\beta}_{\mathrm{NV},0}$ deviates more from the truth from above.

Between the two proposed estimators, $\hat{\boldsymbol{\beta}}_{\mathrm{MC}}$ appears to be more variable than $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$, especially in the presence of Laplace measurement error. This is expected because the Monte Carlo corrected score involves simulated pseudo measurement error. This source of variability can be more prominent when a small B is used to construct the Monte Carlo corrected score, $\Psi_{\mathrm{MC},\,B}$. But increasing B after certain point, say, going beyond the current level (1000) in the presented simulation experiments, becomes less profitable in terms of efficiency gain, especially considering the added computational burden with a much larger B. Another reason for the observed higher variability when U follows a Laplace distribution can be due to applying the Monte Carlo cor-

15
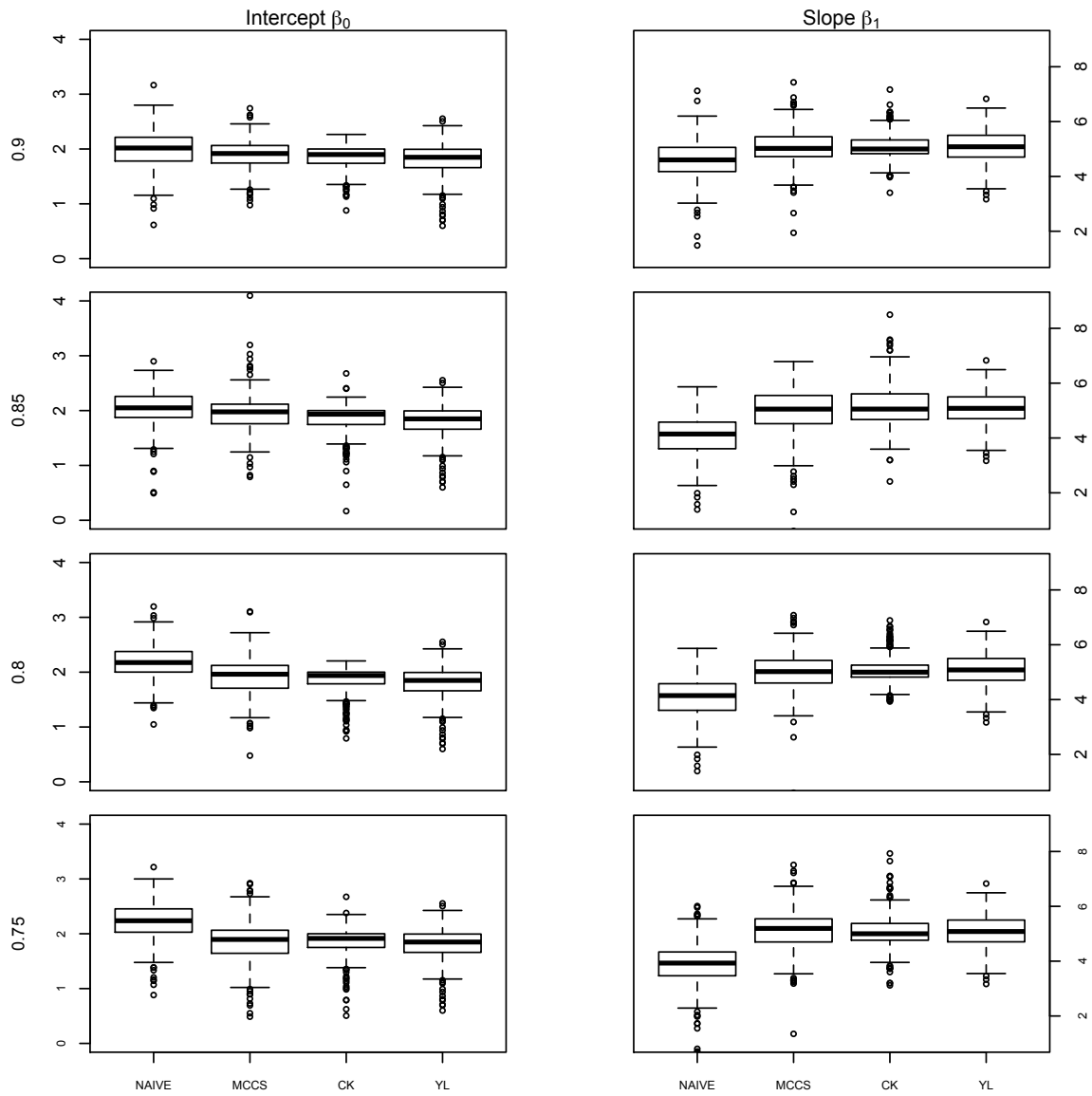
Figure 1.1: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is Laplace measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK), and Yao and Li's method (YL) in the absence of measurement error, respectively. The approximated theoretical optimal bandwidths are used for the Monte Carlo correcte score method and the corrected kernel method.

16

www.manaraa.com
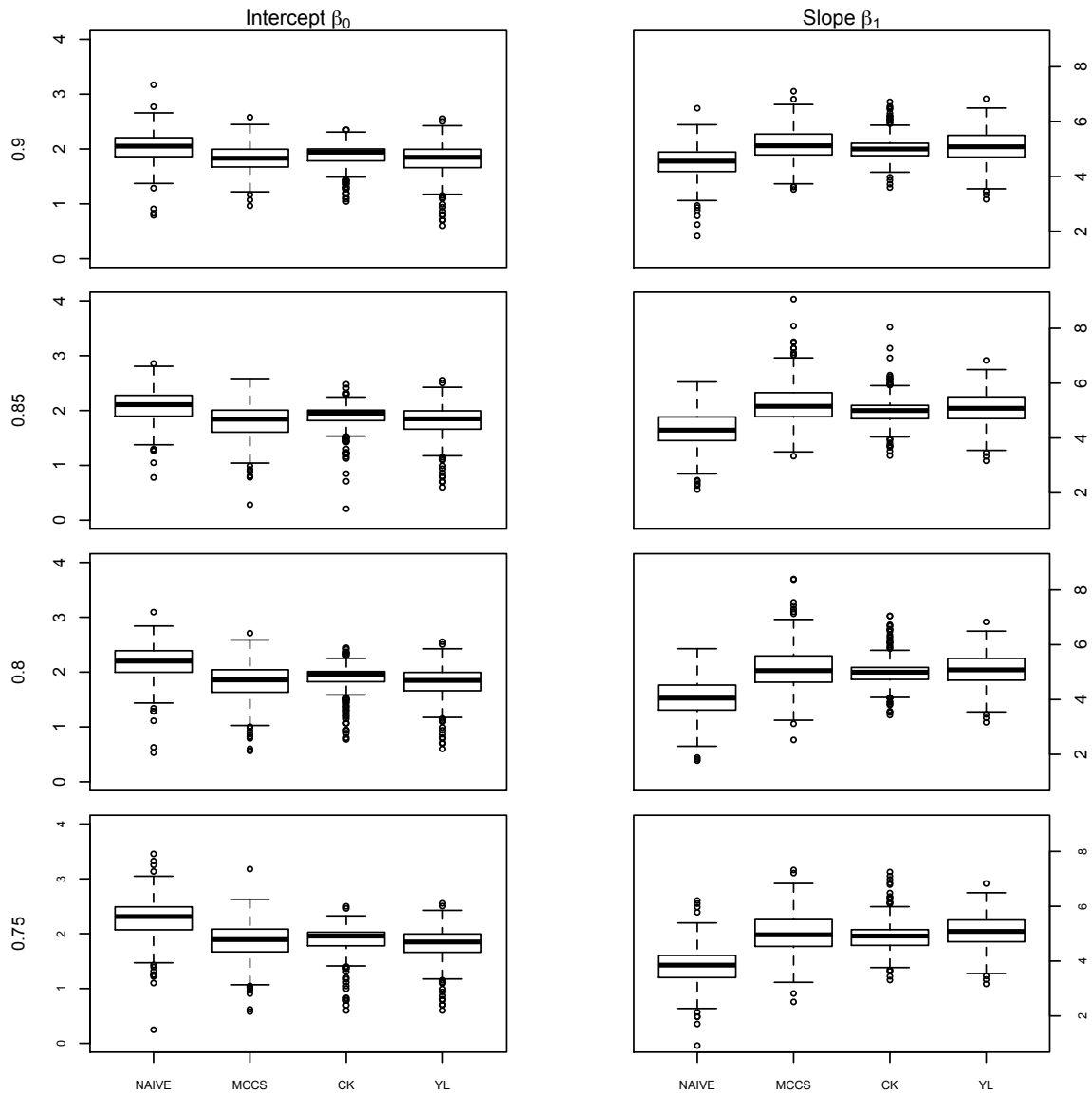
Figure 1.2: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is normal measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK), and Yao and Li's method (YL) in the absence of measurement error, respectively. The approximated theoretical optimal bandwidths are used for the Monte Carlo correcte score method and the corrected kernel method.

17
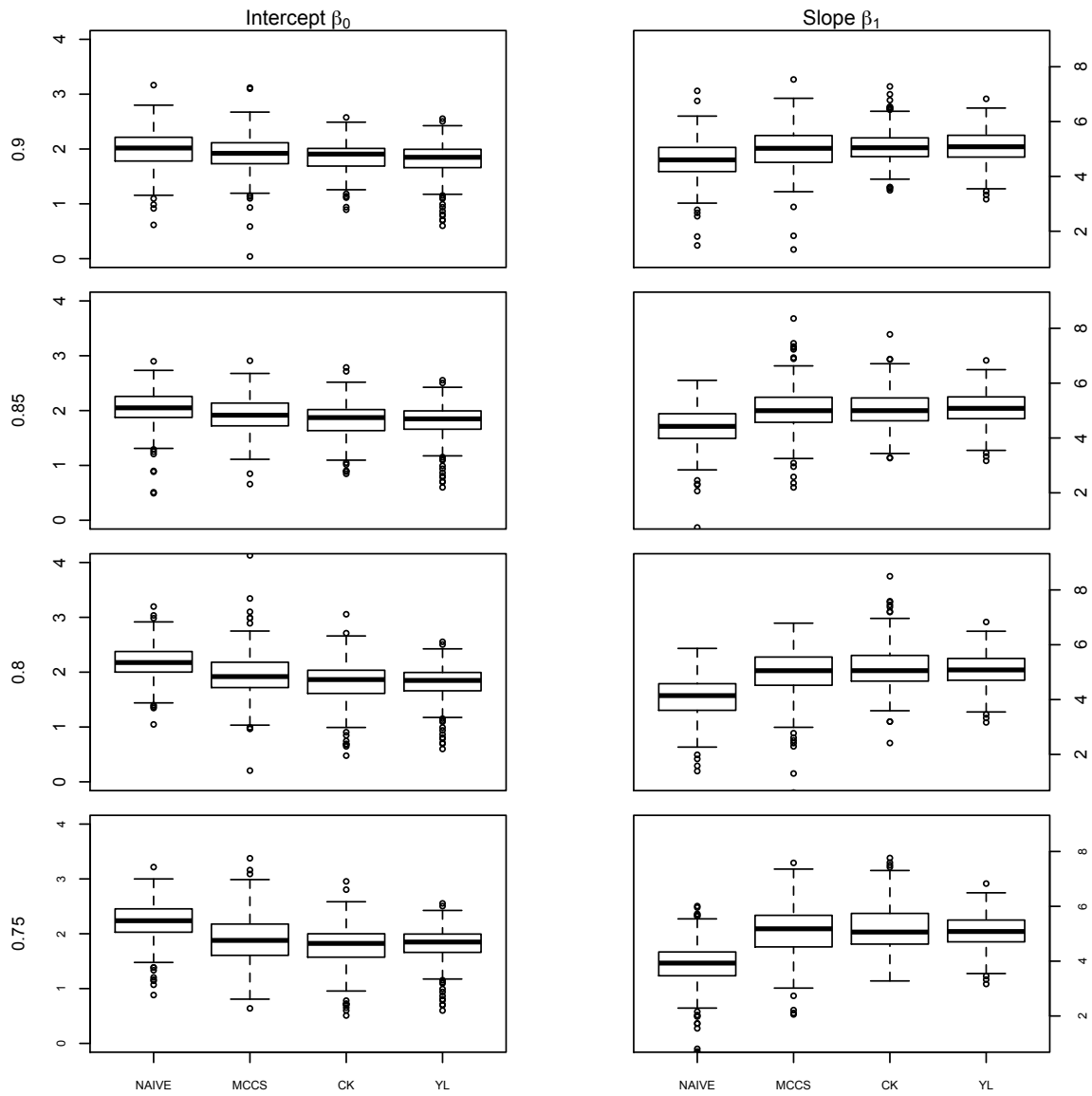
Figure 1.3: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is Laplace measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK), and Yao and Li's method (YL) in the absence of measurement error, respectively. Bandwidths chosen by the simulation-extrapolation method are used for the Monte Carlo corrected score method and the corrected kernel method.
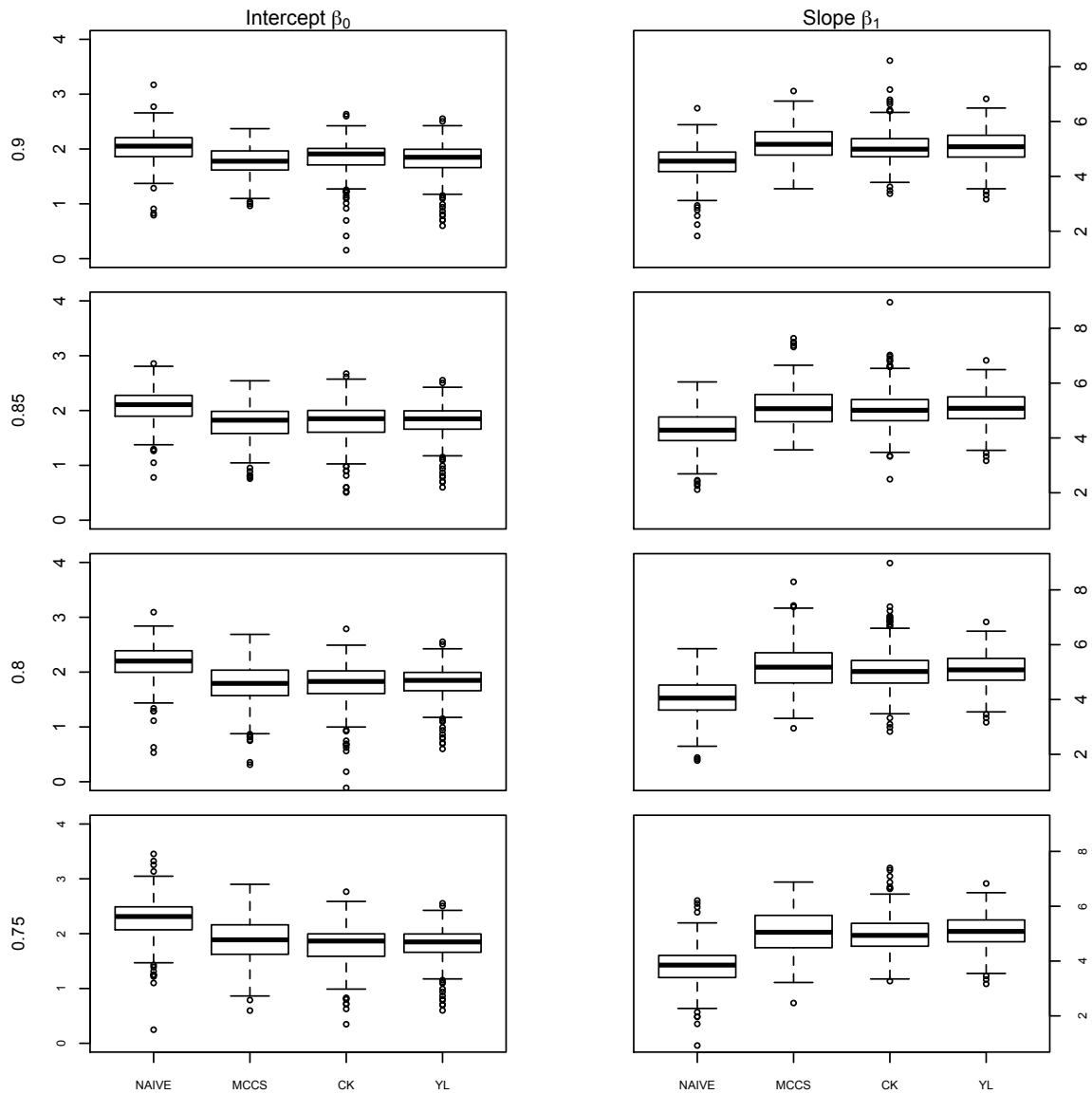
Figure 1.4: Boxplots of estimates of β_0 (on the left panels) and estimates of β_1 (on the right panels) when U is normal measurement error at four levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.9, 0.85, 0.8, 0.75$. Within each panel, the four estimates (from left to right) result from the naive method (NAIVE), the Monte Carlo corrected score method (MCCS), the corrected kernel method (CK), and Yao and Li's method (YL) in the absence of measurement error, respectively. Bandwidths chosen by the simulation-extrapolation method are used for the Monte Carlo corrected score method and the corrected kernel method.

rected score method when the normality assumption on U is violated. Although the corrected kernel method has neither aforementioned concern, computing the deconvoluting kernel requires some care as the integral that defines $K^*(t)$ in (1.7) can be computationally challenging, especially in the presence of normal measurement error (Delaigle and Gijbels, 2007). We use the fast Fourier transforms (Bailey and Swarztrauber, 1994) to compute these integrals, which can still be problematic at times when U is normal. To alleviate numerical inaccuracy in the numerical integration, we follow the suggestion in Meister (2004) and replace the normal characteristic function with the Laplace characteristic function in (1.7) even when U actually follows a normal distribution. The presented numerical results associated with $\hat{\boldsymbol{\beta}}_{\text{CK}}$ in this section are obtained using this treatment. We observe in our extensive numerical study that, when the numerical integration using fast Fourier transforms goes through smoothly with $\phi_U(s)$ as the normal characteristic function, using a Laplace characteristic function instead does not cause noticeable changes in $\hat{\boldsymbol{\beta}}_{\text{CK}}$; and using the latter often leads to smoother numerical implementation. The robustness to and the benefit of Laplace measurement error assumption was noted and investigated by Meister (2004) and Delaigle (2008). For instance, Delaigle (2008) showed that, if the assumed error distribution and the true error distribution match in regard to the first two moments, the bias due to misspecifying the error distribution is of order $O(h^2) + o(\sigma^2)$ when a second-order kernel is used in a kernel density estimator.

Tables 1.1 and 1.2 present Monte Carlo averages of the four considered estimates across 300 replicates along with their empirical standard errors for $\lambda \in \{0.75, 0.8\}$ and $\lambda \in \{0.85, 0.9\}$, respectively. Besides reinforcing the findings from Figures 1.1–1.4 that, compared to the naive estimator, the two proposed estimators are less compromised by measurement error and are closer to the benchmark estimator, these results also show that the performance of the proposed estimators improve in both accuracy and precision as the sample size increases. This is observed even for the

20

Monte Carlo corrected score method in the presence of Laplace measurement error, a case this method is not designed for.

Table 1.1: Monte Carlo averages of four sets of estimates over 300 Monte Carlo replicates when $\lambda = 0.75, 0.80$. Numbers in parentheses underneath the averages are empirical standard errors associated with the averages. The truth is $(\beta_0, \beta_1) = (2, 5)$. MCCS, Monte Carlo corrected score method; CK, corrected kernel method; YL, Yao and Li's method in the absence of measurement error

| | $\lambda = 0.75$ | | | | $\lambda = 0.80$ | | | |
| | $n = 200$ | | $n = 400$ | | $n = 200$ | | $n = 400$ | |
Method	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
	$U \sim \text{Laplace}(0, \sigma^2)$							
Naive	2.26	3.82	2.29	3.85	2.18	4.04	2.22	4.05
	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
MCCS	1.96	4.77	1.84	5.08	1.86	5.06	1.80	5.18
	(0.04)	(0.08)	(0.03)	(0.07)	(0.03)	(0.07)	(0.02)	(0.05)
CK	1.72	5.14	1.72	5.19	1.74	5.14	1.73	5.18
	(0.03)	(0.05)	(0.02)	(0.04)	(0.02)	(0.05)	(0.02)	(0.03)
	$U \sim N(0, \sigma^2)$							
Naive	2.21	3.88	2.27	3.92	2.17	4.07	2.18	4.16
	(0.02)	(0.05)	(0.01)	(0.03)	(0.02)	(0.05)	(0.01)	(0.03)
MCCS	2.02	4.62	1.92	4.88	1.75	5.18	1.78	5.17
	(0.04)	(0.09)	(0.02)	(0.04)	(0.02)	(0.05)	(0.03)	(0.06)
CK	1.77	5.02	1.79	4.97	1.80	4.99	1.84	4.99
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.04)
YL	1.83	5.08	1.87	5.05	1.83	5.08	1.87	5.05
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)

To this end, our focus has been estimating $\boldsymbol{\beta}$. Because modes can be used to predict the outcome Y, we also compare predictions using estimated modes from the above three linear mode regression methods and the local linear mode estimation using the nonparametric method developed by Zhou and Huang (2016). Table 1.3 provides such comparison in terms of the empirical coverage probability of a prediction interval (band) of width $c\sigma_e$ centered around an estimated mode line (or curve) from a considered method across 300 Monte Carlo replicates, for $c = 0.1, 0.2, 0.5$. Here, σ_e is the standard deviation of e_j, which is around 2 in the simulation. According to Table 1.3, all four considered methods applying to error-prone data yield prediction

Table 1.2: Monte Carlo averages of four sets of estimates over 300 Monte Carlo replicates when $\lambda = 0.85, 0.90$. Numbers in parentheses underneath the averages are empirical standard errors associated with the averages. The truth is $(\beta_0, \beta_1) = (2, 5)$. MCCS, Monte Carlo corrected score method; CK, corrected kernel method; YL, Yao and Li's method in the absence of measurement error

| | $\lambda = 0.85$ | | | | $\lambda = 0.9$ | | | |
| | $n = 200$ | | $n = 400$ | | $n = 200$ | | $n = 400$ | |
Method	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
	$U \sim \text{Laplace}(0, \sigma^2)$							
Naive	2.08	4.34	2.10	4.37	2.01	4.54	2.04	4.58
	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
MCCS	1.92	5.02	1.93	4.98	1.93	5.02	1.94	5.03
	(0.02)	(0.05)	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)	(0.04)
CK	1.83	5.05	1.91	5.05	1.86	5.09	1.91	5.05
	(0.02)	(0.04)	(0.01)	(0.04)	(0.02)	(0.03)	(0.01)	(0.03)
	$U \sim N(0, \sigma^2)$							
Naive	2.08	4.31	2.11	4.33	2.02	4.50	2.03	4.54
	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
MCCS	1.78	5.13	1.84	5.10	1.78	5.20	1.84	5.11
	(0.02)	(0.04)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
CK	1.88	5.05	1.89	4.95	1.90	5.06	1.94	4.97
	(0.02)	(0.05)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.03)
YL	1.83	5.08	1.87	5.05	1.83	5.08	1.87	5.05
	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)

intervals (bands) with similar empirical coverage probabilities as those from Yao and Li's linear mode regression method applying to error-free data. The observed similarity may not be surprising because prediction based on mean regression is also less affected by measurement error in covariates when compared to how covariate effects estimation is affected (Buonaccorsi, 1995). To have a more close-up comparison of estimated modes themselves, Table 1.4 presents the Monte Carlo averages of the point-wise error associated with each method, |the estimated $y_M(x) - y_M(x)$|, at $x = 0.5, 0.9$. From this more close-up comparison, one can see that using error-prone data for mode estimation tends to produce more bias than when one uses error-free data; but our two proposed methods substantially alleviate the bias seen in the naive mode estimates. The nonparametric method shows no advantage when

point-wise error of mode estimation is concerned, especially when the covariate value is near the boundary, e.g., $x = 0.9$. We acknowledge that the current simulation setting is designed for linear mode regression, with data simulated from models with a linear mode function. Nonparametric mode regression makes no assumption on the functional form of the conditional mode function, and thus it is expect to exhibit higher variability and less accuracy in estimating the mode than methods that take into account a simple (and true) functional form. Scenarios where the data generating process involves a nonlinear mode function are where one can benefit from employing the nonparametric method, which are scenarios beyond the scope of the current study.

Table 1.3: Monte Carlo averages of proportions of observed responses captured by a prediction interval (band) of width $c\sigma_e$, for $c = 0.1, 0.2, 0.5$, associated with each method across 300 Monte Carlo replicates. Numbers in parentheses underneath the averages are $100\times$(empirical standard error) associated with the averages. MCCS, Monte Carlo corrected score method; CK, corrected kernel method; NMR, Zhou and Huang's nonparametric mode regression; YL, Yao and Li's method in the absence of measurement error

	$\lambda = 0.85$						$\lambda = 0.9$					
	$n = 200$			$n = 400$			$n = 200$			$n = 400$		
Method	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$	$0.1\sigma_e$	$0.2\sigma_e$	$0.5\sigma_e$
	$U \sim \text{Laplace}(0, \sigma^2)$											
Naive	0.09	0.18	0.39	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.07)	(0.11)	(0.15)	(0.07)	(0.10)	(0.15)	(0.07)	(0.10)	(0.14)	(0.06)	(0.08)	(0.11)
MCCS	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.06)	(0.08)	(0.11)	(0.06)	(0.08)	(0.09)	(0.06)	(0.08)	(0.10)	(0.05)	(0.07)	(0.10)
CK	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.19	0.40
	(0.06)	(0.09)	(0.11)	(0.06)	(0.08)	(0.10)	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.10)
NMR	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38
	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)	(0.06)	(0.09)	(0.013)	(0.05)	(0.08)	(0.11)
	$U \sim N(0, \sigma^2)$											
Naive	0.09	0.18	0.39	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.07)	(0.11)	(0.17)	(0.06)	(0.10)	(0.15)	(0.06)	(0.09)	(0.12)	(0.06)	(0.09)	(0.12)
MCCS	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40
	(0.06)	(0.08)	(0.11)	(0.05)	(0.08)	(0.10)	(0.06)	(0.08)	(0.09)	(0.05)	(0.07)	(0.09)
CK	0.09	0.18	0.40	0.09	0.18	0.40	0.09	0.18	0.40	0.10	0.19	0.40
	(0.06)	(0.09)	(0.12)	(0.05)	(0.07)	(0.10)	(0.06)	(0.09)	(0.11)	(0.05)	(0.07)	(0.09)
NMR	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38	0.09	0.17	0.38
	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)	(0.06)	(0.09)	(0.13)	(0.06)	(0.08)	(0.11)
YL	0.09	0.18	0.40	0.09	0.19	0.41	0.09	0.18	0.40	0.09	0.19	0.41
	(0.06)	(0.10)	(0.13)	(0.05)	(0.07)	(0.10)	(0.06)	(0.10)	(0.13)	(0.05)	(0.07)	(0.10)

Table 1.4: Monte Carlo averages of point-wise errors, |estiamted mode−true mode|, associated with each method when $x = 0.5, 0.9$ across 300 Monte Carlo replicates. Numbers in parentheses are empirical standard error associated with the averages. MCCS, Monte Carlo corrected score method; CK, corrected kernel method; NMR, Zhou and Huang's nonparametric mode regression; YL, Yao and Li's method in the absence of measurement error

	$\lambda = 0.85$				$\lambda = 0.9$			
	$n = 200$		$n = 400$		$n = 200$		$n = 400$	
Method	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$	$x = 0.5$	$x = 0.9$
	$U \sim \text{Laplace}(0, \sigma^2)$							
Naive	0.28 (0.01)	0.53 (0.02)	0.25 (0.01)	0.51 (0.02)	0.27 (0.01)	0.47 (0.02)	0.20 (0.01)	0.38 (0.01)
MCCS	0.22 (0.01)	0.45 (0.03)	0.16 (0.02)	0.33 (0.03)	0.18 (0.01)	0.35 (0.03)	0.15 (0.01)	0.30 (0.02)
CK	0.20 (0.01)	0.35 (0.02)	0.17 (0.01)	0.26 (0.01)	0.21 (0.01)	0.32 (0.01)	0.18 (0.01)	0.28 (0.01)
NMR	0.66 (0.01)	1.10 (1.12)	0.65 (0.01)	1.24 (0.12)	0.68 (0.02)	1.31 (0.12)	0.66 (0.01)	0.58 (0.15)
	$U \sim N(0, \sigma^2)$							
Naive	0.28 (0.01)	0.59 (0.02)	0.25 (0.01)	0.54 (0.02)	0.25 (0.01)	0.46 (0.02)	0.21 (0.01)	0.39 (0.02)
MCCS	0.18 (0.01)	0.37 (0.02)	0.14 (0.01)	0.24 (0.01)	0.18 (0.01)	0.33 (0.02)	0.12 (0.01)	0.23 (0.01)
CK	0.21 (0.01)	0.33 (0.02)	0.21 (0.01)	0.26 (0.01)	0.20 (0.01)	0.40 (0.02)	0.18 (0.01)	0.25 (0.01)
NMR	0.64 (0.01)	0.95 (0.09)	0.34 (0.02)	0.52 (0.08)	0.63 (0.02)	1.28 (0.14)	0.68 (0.01)	1.05 (0.11)
YL	0.14 (0.01)	0.21 (0.01)	0.14 (0.01)	0.22 (0.01)	0.14 (0.01)	0.21 (0.01)	0.14 (0.01)	0.22 (0.01)

1.6 APPLICATION TO DIETARY DATA

In this section, we apply the proposed methods to a dietary data set from the Women's Interview Survey of Health. The data are from $n = 271$ subjects, each completing a food frequency questionnaire (FFQ) and six 24-hour food recalls on randomly selected days. We focus on studying the impact of the long-term usual intake (X) on the FFQ intake measured as the percent calories from fat (Y) (Carroll et al., 1997). Since the long-term intake cannot be measured directly, and the 24-hour recalls can be viewed as error-contaminated surrogates of it, we used the average of these recalls from each subject as a surrogate (W) of this subject's long-term intake. Figure 1.5 provides the histogram of FFQ intake and the scatter plot of it versus the 24-hour food recalls. The histogram indicates an underlying skewed distribution, and the scatter plot suggests existence of outliers in the observed data. These are both features that suggest mode regression can provide valuable information regarding the association between a response and a covariate that mean regression may not capture.

For illustration purposes, we consider a linear mode regression model for the mode of Y_j given X_j, where X_j is not observed but its error-contaminated surrogate
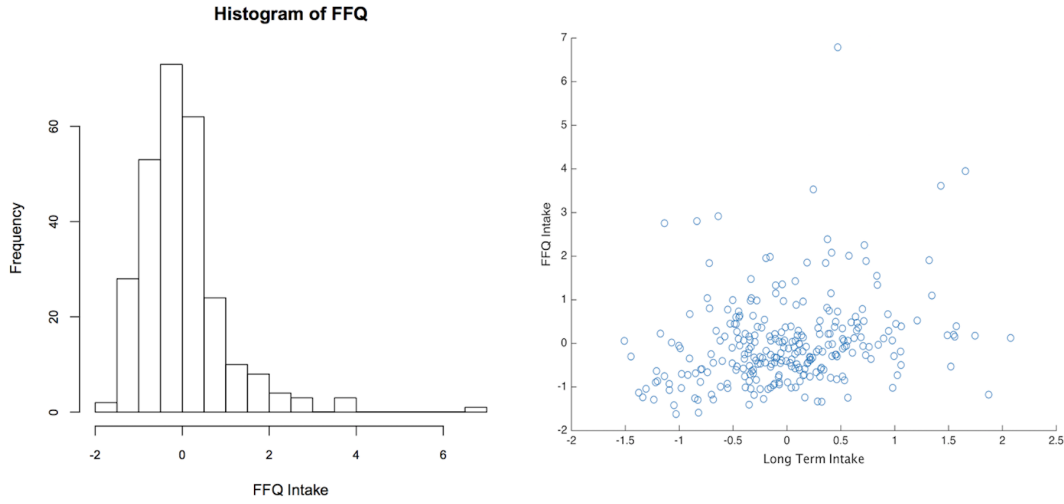
Figure 1.5: The histogram (on the left panel) of food frequency questionnaire intake and the scatter plot (on the right panel) of this quantity versus a surrogate of long-term intake for the dietary data.

W_j is, where $W_j = \sum_{k=1}^{6} W_{j,k}/6$, in which $W_{j,k}$ is subject j's kth food recall, for $k = 1, \ldots, 6$ and $j = 1, \ldots, 271$. Using the six replicate measures for each underlying X_j, we estimate the variance of measurement errors associated with W_j via one sixth of $\sum_{j=1}^{n} \sum_{k=1}^{6} (W_{j,k} - W_j)^2/(5n)$, following equation (4.3) in Carroll et al. (2006). This gives an estimate of the measurement error variance as $\hat{\sigma}^2 = 0.12$, and the corresponding estimated reliability ratio being 0.73.

We carry out the linear mode regression analysis using the naive method, the Monte Carlo corrected score method, the corrected kernel method assuming Laplace and normal measurement error, respectively, and we also implement the local linear mode estimation as the only fully nonparametric method. Table 1.5 presents the estimated regression coefficients from three linear mode regression methods. These results suggest that both proposed methods produce estimates of the covariate effect, β_1, that imply a stronger association between the FFQ intake and the long-term intake than the estimate from the naive method does. In particular, compared to

the naive estimate, the estimated covariate effect from the Monte Carlo corrected score method increases by 29%, and the estimates from the corrected kernel method increase by 38% and 34% when assuming Laplace measurement error and normal measurement error, respectively. This also gives an example where using the Laplace characteristic function and the normal characteristic function in the corrected kernel method yield very similar estimates. Figure 1.6 depicts three of these estimated mode regression lines, omitting the one from the corrected kernel method under the normality assumption, and the estimated mode curve obtained by applying the local linear estimation in Zhou and Huang (2016). Computer codes for implementing the two proposed method for this data set are provided in Appendice D and E.

Table 1.5: Regression coefficient estimates in the linear mode regression model from the naive method, the Monte Carlo corrected score method, and the corrected kernel method (assuming Laplace and normal U, respectively) using the dietary data. Numbers in parentheses are estimated standard deviations of the regression coefficient estimates resulting from 200 bootstrap samples. MCCS, Monte Carlo corrected score method; CK-Laplace, corrected kernel method assuming Laplace U; CK-Normal, corrected kernel method assuming normal U

Method	β_0	β_1
Naive	-0.27 (0.10)	0.36 (0.11)
MCCS	-0.10 (0.05)	0.48 (0.13)
CK-Laplace	-0.07 (0.05)	0.50 (0.12)
CK-Normal	-0.09 (0.06)	0.49 (0.13)

1.7 DISCUSSION

In this chapter, we propose two methods to infer the regression coefficients in a linear mode model for a response given an error-prone covariate. The resultant inference for the covariate effect significantly improve over the naive inference from applying Yao and Li's method without accounting for measurement error. As demonstrated in the real data analysis in Section 1.6, estimating the measurement error variance is

26
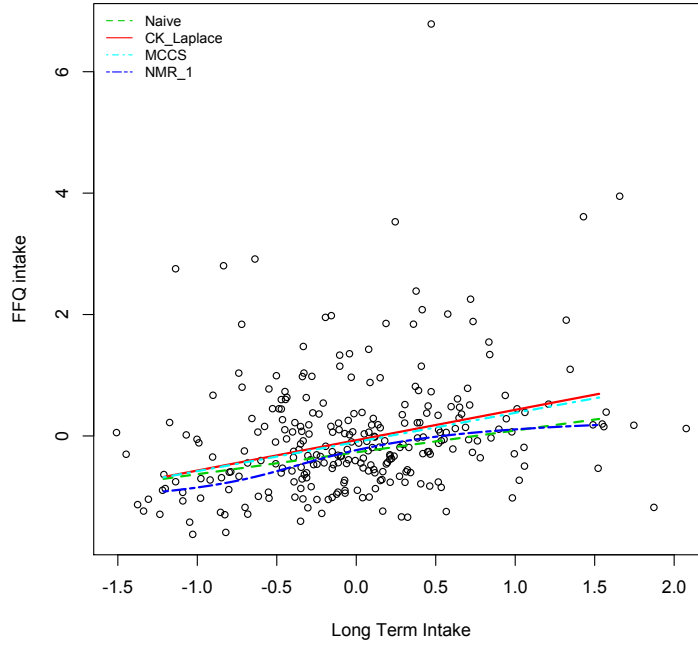
Figure 1.6: Dietary data overlaid with the estimated mode regression line from naively applying Yao and Li's method (green dashed line), the Monte Carlo corrected score method (cyan dot-dashed line), the corrected kernel method assuming Laplace measurement error (red solid line), and a local linear estimate of the mode curve (blue two-dashed line).

trivial when replicate measures of each underlying true covariate value are available. This treatment of unknown σ^2 has been a routine practice in the measurement error literature, where researchers typically observe little impact of estimating σ^2 on the final inference for covariate effects. The measurement error variance is the only piece of information regarding $f_U(u)$ required for implementing the Monte Carlo corrected score method since normal U is assumed for this method. To implement the corrected kernel method, the characteristic function of U, $\phi_U(t)$, is needed, which can also be easily estimated using replicate measures (Delaigle et al., 2008). Moreover, as noted in our simulation study and by several other authors (Meister, 2004; Delaigle, 2008; Delaigle et al., 2009; Zhou and Huang, 2016), simply setting $\phi_U(t)$ as the Laplace characteristic function works well in most scenarios, which frees one from estimating

27

the characteristic function altogether.

Both proposed methods can easily incorporate multiple covariates in the linear mode model. Indeed, Yao and Li's method is developed more generally with multivariate covariates, and the Monte Carlo corrected score method entails evaluating the score function used in Yao and Li's method at simulated contaminated covariate data, hence one only needs to revise MC-1 in the algorithm in Section 1.3 accordingly to implement the Monte Carlo corrected score method with multivariate covariates. To implement the corrected kernel method when there are $p(> 1)$ covariates, some or all of which are prone to nondifferntial measurement error, one uses a multivariate characteristic function of $\boldsymbol{U} = (U_1, \ldots, U_p)^{\mathrm{T}}$ in (1.7) evaluated at $-\boldsymbol{\beta}_1^{\mathrm{T}} s/h$, bearing in mind that $\phi_{U_\ell}(t) = 1$ if the ℓth covariate is error-free, for $\ell \in \{1, 2, \ldots, p\}$.

Although we impose a linear functional form for the mode of Y given $X = x$, the mode residual distribution, $g(\epsilon \mid x)$, is left unspecified, except for that its mode is zero and some mild conditions imposed on it for the study of asymptotics. Hence, the proposed methods are broadly applicable even when one lacks a parametric model for $f_{Y|X}(y \mid x)$. This makes these methods semiparametric in nature. Chapter 2 is to involve semiparametric components in the specification of $y_M(x)$ to relax the linear assumption made in the current study. Yao and Xiang (2016) considered a local polynomial mode estimation that mimics the idea of local polynomial mean estimation (Fan and Gijbels, 1996), and also considered a nonparametric varying coefficient mode regression model. Zhao et al. (2014) proposed a variable selection method based on a partially linear varying coefficient mode regression model. These works, and other existing works on semiparametric mode regression, all assume error-free covariates. We will consider in Chapter 3 partially linear mode regression in the presence of covariate measurement error. To prepare for our exploration on partially linear mode regression with error-prone covariates, we next describe in Chapter 2 methodologies for partially linear mode regression in the absence of covariate measurement error.

CHAPTER 2

PARTIALLY LINEAR MODE REGRESSION

Among the broad range of statistical regression models exploited for studying the association between a response Y and covariates of interest, fully parametric regression models can yield highly efficient inference results, especially when the parametric specifications are parsimonious and correct. But they may lead to misleading inference outcomes when some parametric assumptions are violated. Fully nonparametric regression models are less vulnerable to model misspecification, but they typically suffer from low precision, with precision quickly deteriorating as the number of covariates increases. Partially linear models offer an appealing compromise between parametric models and nonparametric models. They are attractive choices of models when one can envision two separate sets of potentially influential covariates, denoted by T and X, respectively, based on subject-matter knowledge or other scientific grounds, so that the association between Y and X can be well represented by a linear model, whereas the effect of T on Y has an unknown functional form that enters the regression model additively. Among the first to consider such models, Robert F. et al. (1986) employed a partially linear model for investigating effects of weather on electricity demand, where they set T as temperature, and included household income, monthly price of electricity, and other factors in X that are assumed to relate to electricity usage linearly after the possibly nonlinear effect of T on Y is accounted for. The current literature on partially linear models are mainly confined to estimating the conditional mean function of Y (Hyndman et al., 1996; Hardle et al., 2000). There also exists a large body of work on estimating the conditional quantile of the

29

response Y (He and Liang, 2000; Wang et al., 2009) in the partially linear model framework.

As yet another important location measure, partially linear models for the conditional mode of a response is far less investigated, even though linear mode regression models (Yao and Li, 2014) and nonparametric mode regression models (Yao et al., 2012; Chen et al., 2016; Yao and Xiang, 2016) have been studied and adopted in a host of applications, as illustrated in Einbeck and Tutz (2006); Bamford et al. (2008) and Huang and Yao (2012), just to name a few. The only work we are aware of that involves mode regression in a partially linear model is Zhao et al. (2013) and Zhao et al. (2014), where the authors proposed variable selection methods in partially linear varying coefficient models based on mode regression. Instead of conducting variable selection, in this chapter we are more interested in estimating the conditional mode of the response Y by estimating both parametric and nonparametric part. Following the methodology in Zhao et al. (2014), first, we propose a one-stage method to estimate the parametric and nonparametric part simutanously. Based on assumptions in section 2.3, a two-stage method is proposed to estimate the parametric and nonparametric part sequentially.

2.1 DATA AND MODELS

Suppose that the observed data consist of n independent data points, $\{(Y_j, T_j, X_j)\}_{j=1}^n$, where the covariate values, $\{X_j\}_{j=1}^n$ and $\{T_j\}_{j=1}^n$, are scalar covariate values, and $\{Y_j\}_{j=1}^n$ are response variable values. The distribution of Y given (X, T) is specified by

$$Y = g^*(T) + m(X) + \sigma(X, T)\epsilon, \tag{2.1}$$

where $E(\epsilon|X, T) = 0$, and $g^*(T)$ is an unknown smooth function of T, $m(X)$ is a function of X and $\sigma(X, T)$ is a function of X and T. If $m(X)$ is a linear function

30

of X, $\sigma(X, T)$ is a linear function of X and $Mode(\epsilon|X, T)$ is constant free of (T, X), then (2.1) implies a partially linear model for the mode of Y given by,

$$y_M(x, t) = \text{Mode}(Y_j|X_j = x, T_j = t) = g^*(t) + \beta_0 + \beta_1 x, \text{ for } j = 1, \cdots, n, \quad (2.2)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ is the coefficient of the parametric part.

Note that, $g^*(T)$ in (2.2) is an unspecified smooth function. Zhao et al. (2014) proposed an estimation and variable selection procedure based on mode regression, where the nonparametric function is estimated by B-spline basis. He et al. (2002) considered an extension of M-estimators in semiparametirc models, where B-spline is employed to approximate the nonparametric function. By using B-spline, estimating the nonparametric function boils down to estimating the coefficients of basis functions. Following the method in Zhao et al. (2014) and He et al. (2002), we use B-spline to approximate $g^*(T)$. Once the spline basis functions are obtained, both linear parameters and the coefficients of spline basis functions can be estimated by revising the method in Chapter 1. For completeness, we give a brief review of the B-splines methodology in the following section.

2.2   B-spline Methods

As a method to approximate a smooth function, B-spline provides a local support based on basis functions (Boor, 2001). Although the local polynomial estimation method (Yao et al., 2012) gives another way to approximate a smooth function, Zhao et al. (2014) pointed out that the heavy computation involved in local polynomial estimation is a big concern, especially for high dimensional semiparametric partially linear varying coefficient models. In contrast, a small number of knots in B-splines can often provide an excellent approximation. This makes B-spline easier to implement (He et al., 2002) and the method we choose in our study to approximate the smooth function $g^*(T)$ in (2.2). For more details about the construction of basis functions,

31
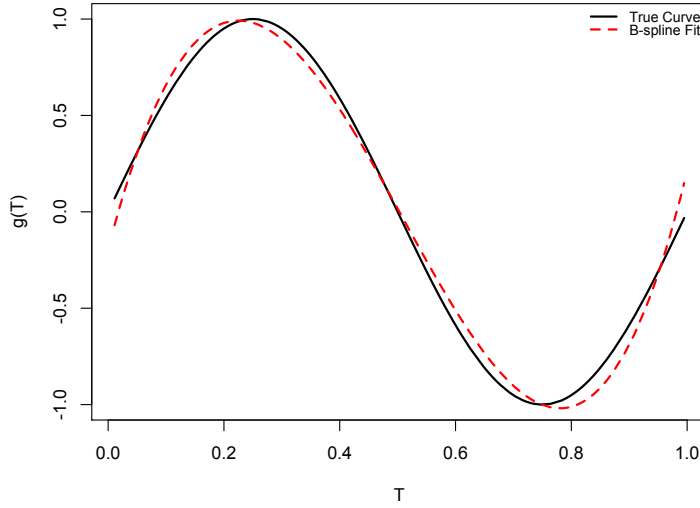
Figure 2.1: An B-spline example

readers are referred to Schumaker (2007, Section 4.3).

Suppose that the smooth function $g^*(T)$ is estimated on the interval $[0, 1]$, where there exists a partition, $0 = s_0 < s_1 < \cdots < s_k < s_{k+1} = 1$. With the order of polynomial spline fixed at ℓ, the spline basis functions are denoted by $B(t) = \{B_1(t), B_2(t), \cdots, B_N(t)\}$, where $N = k + \ell$. Then, $g^*(t)$ can be approximated as

$$g^*(t) \approx B(t)^T \boldsymbol{\gamma}, \tag{2.3}$$

where $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_N)$ are the coefficients associated with spline basis functions. Figure 2.1 presents an example in which $g^*(T) = 2\sin(2\pi T)$, where the black solid line represents the true curve $g^*(T)$, and the red dashed line represents the fitted curve obtained by using the cubic B-spline. Although this approximation is not perfect in Figure 2.1, it can be improved by carefully selecting the tuning parameter as demonstrated in Section 2.4.

2.3 Proposed Estimation Methods

One-Stage Estimation Method

To estimate parameters in the linear part in (2.2), $\boldsymbol{\beta} = (\beta_0, \beta_1)$, and the coefficients $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_N)$, we view the basis functions $B(t)$ as new covariates in the mode regression. Mimicking the objective function based on kernel density estimators of Y given covariates in Chapter 1, we propose to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by maximizing

$$Q_h(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{n} \sum_{j=1}^{n} K_h \left(Y_j - B(T_j)^T \boldsymbol{\gamma} - \beta_0 - \beta_1 X_j \right), \tag{2.4}$$

with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, where $K_h(t) = 1/h \cdot K(t/h)$ is the kernel, and h is the bandwidth. By setting $K(t)$ as the standard normal density function, an EM algorithm proposed by Yao and Li (2014) can be applied to compute the proposed estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. In particular, one may use modal expectation maximization (MEM) algorithm (Li et al., 2007) outlined below to implement the EM algorithm.

- MEM-1: Set $m = 0$, select a starting a point, $\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)} \right)$. For example, one may carry out linear mean regression analysis to regress Y on $B(T)$ and X to obtain such a starting point.

- MEM-2: Compute

$$\pi(j|\boldsymbol{\beta}^{(m)}, \boldsymbol{\gamma}^{(m)}) = \frac{K_h \left\{ Y_j - B(T_j)^T \boldsymbol{\gamma}^{(m)} - \beta_0^{(m)} - \beta_1^{(m)} X_j \right\}}{\sum_{k=1}^{n} K_h \left\{ Y_k - B(T_k)^T \boldsymbol{\gamma}^{(m)} - \beta_0^{(m)} - \beta_1^{(m)} X_k \right\}}, \ j = 1, \cdots, n,$$

- MEM-3: Update $(\boldsymbol{\beta}^{(m+1)}, \boldsymbol{\gamma}^{(m+1)})$ via

$$(\boldsymbol{\beta}^{(m+1)}, \boldsymbol{\gamma}^{(m+1)})$$
$$= \mathrm{argmin}_{(\beta, \gamma)} \sum_{j=1}^{n} \left[\pi(j|\boldsymbol{\beta}^m, \boldsymbol{\gamma}^{(m)}) \log K_h \left(Y_j - B(T_j)^T \boldsymbol{\gamma}^{(m)} - \beta_0^{(m)} - \beta_1^{(m)} X_j \right) \right]$$
$$= \left(\boldsymbol{Z}^T D \boldsymbol{Z} \right)^{-1} \boldsymbol{Z}^T DY, \tag{2.5}$$

where $\boldsymbol{Z} = (\boldsymbol{B(T)}, \boldsymbol{X})$, and D is an $n \times n$ diagonal matrix with diagonal elements $\pi(j|\boldsymbol{\beta}^m, \boldsymbol{\gamma}^{(m)})$, $j = 1, \cdots, n$.

- Set $m = m + 1$. Repeat MEM-2 and MEM-3 till convergence.

It is worth pointing out that

$$\sum_{i=1}^{N} B_i(T) = 1,$$

resulting in perfect multicollinearity in (2.5). This causes the above MEM algorithm to diverge. Therefore, to implement the MEM algorithm, we view model (2.2) as

$$y_M(x, t) = \text{Mode}(Y_j | X_j = x, T_j = t) = g(t) + \beta_1 x, \text{ for } j = 1, \cdots, n. \quad (2.6)$$

where $g(t) = g^*(t) + \beta_0$. By replacing $Q_h(\boldsymbol{\beta}, \boldsymbol{\gamma})$ in (2.4) with

$$Q_h(\beta_1, \boldsymbol{\gamma}) = \frac{1}{n} \sum_{j=1}^{n} K_h \left(Y_j - B(T_j)^T \boldsymbol{\gamma} - \beta_1 X_j \right), \quad (2.7)$$

the MEM algorithm can be used to estimate β_1 and $\boldsymbol{\gamma}$. We refer to this method as the one-stage mode regression method in the absence of measurement error (RO) henceforth, and the estimator is denoted by $(\hat{\beta}_{\text{RO},1}, \hat{\boldsymbol{\gamma}}_{\text{RO}})$.

Two-Stage Estimation Method

Inspired by existing literature on partially linear mean regression, we now formulate a two-stage partially linear mode model induced from a partially linear mean model under certain assumptions that can be satisfied under some practical scenarios. With some abuse of notation, we now state these assumptions using some notations already used in Section 2.1, some of which are also stated following (2.1).

(a1) The mean regression model of Y given (X, T) is

$$Y = \alpha_1 X + g^*(T) + \epsilon^* \quad (2.8)$$

where α_1 is an unknown regression coefficient associated with X. $g^*(\cdot)$ is an unknown function, and ϵ^* is the random error with $E(\epsilon^* | T, X) = 0$;

(a2) $Mode(\epsilon^* | T, X) = \beta_0 + \alpha_2 X$, where β_0 and α_2 are unknown parameters; and

34

(a3) $E(X|T) = 0$.

Assumptions (a1) and (a2) together imply that the mode regression model is also a partially linear model specified by

$$Mode(Y|T, X) = \beta_0 + (\alpha_1 + \alpha_2)X + g^*(T), \qquad (2.9)$$

which is equivalent to (2.6) with $\beta_1 = \alpha_1 + \alpha_2$ and $g(T) = \beta_0 + g^*(T)$. Assumption (a3) coincides with Assumption 3.1 in He and Liang (2000), who pointed out that a sufficient condition for it is the independence of T and X, since $E(X) = 0$ can always be achieved by centering data for X as part of data standardization. Under (a3), (2.8) implicates $E(Y|T) = g^*(T)$.

Besides inducing a partially linear mode model from a partially linear mean model, the above development also reveals that, thanks to (a1) and (a3), one may first estimate $g^*(\cdot)$ by carrying out mean regression analysis of Y on T, which involves no covariate measurement error complication. For this mean regression, one can use any nonparametric mean regression method, such as local polynomial mean regression or spline-based methods designed for error-free data. To be consistent with the approach for approximating the nonparametric part in Section 2.3, we use B-spline mean regression at this stage. After an estimator for $g^*(\cdot)$ is obtained, denoted by $\hat{g}^*(\cdot)$, one may carry out linear mode regression of $Y^* = Y - \hat{g}^*(T)$ on X to estimate β_0 and $\beta_1 = \alpha_1 + \alpha_2$ in (2.9) following the methods proposed by Yao and Li (2014). These two steps accomplish estimating the parametric part, β_1, and the nonparametric part, $g(T) = \beta_0 + g^*(T)$. We refer to the so-obtained estimators the two-stage estimators, and refer to this method the two-stage mode regression method in the absence of measurement error (RT), which we recap in the following algorithm.

(T1) Approximate $g^*(t)$ via cubic splines with k knots, that is, $g^*(t) \approx B'(t)\gamma^*$. Regress Y on $B(T)$ using the least squares method, resulting in an estimator for γ^*, denoted by $\hat{\gamma}^*$. This leads to an estimator for $\hat{g}^*(t) = B'(t)\hat{\gamma}^*$

35

(T2) Define $Y_j^* = Y_j - \hat{g}^*(T_j)$, for $j = 1, \cdots, n$, then implement the EM algorithm on $\{X_j, Y_j^*\}$, for $j = 1, \cdots, n$, in Yao et al. (2012) to estimate $\boldsymbol{\beta}$.

Denote the resultant estimators by $(\hat{\boldsymbol{\beta}}_{\mathrm{RT}}, \hat{\gamma}_{\mathrm{RT}})$.

2.4 TUNING PARAMETER SELECTION

To implement the above two estimation methods, the number of interior knots k and the bandwidth h need to be selected appropriately. In this section, we propose two methods to select the tuning parameters. Similar to the strategy in Zhao et al. (2013) and Zhao et al. (2014), we first consider a two dimensional M-fold cross validation method. This method can be employed to select tuning parameter for both one-stage method and two-stage method. Additionally, in order to reduce the computing time in our one-stage estimation method, we develop a two-layer tuning parameter selection method. This method is designed only for the one-stage method.

TWO-DIMENSIONAL CROSS VALIDATION

As in Zhao et al. (2014), He et al. (2002) and Wang et al. (2009), we use cubic spline basis functions to approximate $g(T)$ with $\ell = 4$. Lower order of spline basis functions can be applied if $g(T)$ is less smooth (He et al., 2002). After fixing the order of spline basis functions, the B-spline method is typically sensitive to the choice of the number of knots k. Besides the interior knots, the performance of all kernel-based methods can be noticeably affected by the choice of bandwidths. To address the choice of k and h, we propose a two dimensional cross validation method that entails minimizing the objective function

$$CV(k, h) = M^{-1} \sum_{m=1}^{M} n_m^{-1} \sum_{i \in \mathcal{I}_m} K_h \left\{ Y_i - \hat{g}^{(-m)}(T_i) - \hat{\beta}_1^{(-m)} X_i \right\}, \qquad (2.10)$$

where M presents the number of partitions of the data set, $\mathcal{I}_m$ is the observation index set associated with the m-th subset of data, n_m is the size of the data set $\mathcal{I}_m$, $m =$

$1, \cdots, M$, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ are estimates from applying the considered estimation method to the observed data after deleting the mth subset. To be more specific, when one chooses (h, k) that go in the one-stage estimation method, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ in (2.10) are $\hat{\beta}_{RO,1}$ and $\hat{g}_{RO}(\cdot)$ computed using data $\{(Y_j, T_j, X_j), j \in \mathcal{I} \backslash \mathcal{I}_m\}$, where $\mathcal{I} = \{1, \cdots, n\}$, and "$\backslash$" is the set subtraction operator, for $m = 1, \cdots, M$. Similarly, when one selects (h, k) that go along with the two-stage estimation method, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ in (2.10) are $\hat{\beta}_{RT,1}$ and $\hat{g}_{RT}(\cdot)$ computed using data $\{(Y_j, T_j, X_j), j \in \mathcal{I} \backslash \mathcal{I}_m\}$, where $\mathcal{I} = \{1, \cdots, n\}$ for $m = 1, \cdots, M$. All these estimators depend on (h, k), the dependence we suppress on the right-hand side of (2.10) for cleaner notations. Following this cross validation procedure, referred to as the two-dimensional CV in the sequel, the chosen number of knots and bandwidth are given by

$$(\hat{k}, \hat{h}) = \max_{k, h} CV(k, h).$$

We follow the strategy used in He et al. (2002) to determine the candidate values for k such that, given a chosen order of B-spline, and thus ℓ is fixed, these values lie in $[\max(0.5n^{1/5} - \ell, 8 + 2n^{1/5} - \ell)]$. The M-fold two dimensional cross validation method can be computationally prohibitive. To ease the computational burden for tuning parameters selection for the one-stage estimation method, we propose another procedure described next for selecting tuning parameters tailored for this estimation method.

TWO LAYER TUNING PARAMETER SELECTION

We observe in extensive simulation studies that the quality of an estimator for β_1 is often noticeably influenced by how well $g(\cdot)$ is estimated, although the other way around is not necessarily true. This motivates our second strategy for selecting (h, k), referred to as the two-layer tuning parameters selection method outlined in the following algorithm.

(L1) For each candidate value of k, k_c, where $c \in \{1, \cdots, C\}$, find an h among its candidate values, $\{h_1, \cdots, h_D\}$, that minimizes the integrated squared error (ISE) of the estimate for $g(\cdot)$,

$$\text{ISE}(h, k_c) = \int_0^1 \{\hat{g}_{RO}(t) - \tilde{g}(t)\}^2 dt, \qquad (2.11)$$

where $\tilde{g}(t)$ is a preliminary estimate for $g(\cdot)$, and $\hat{g}_{RO}(t)$ is the one-stage estimate obtained based on the entired observed data set, whose dependence of on h, after k is fixed at k_c, is suppressed on the right-hand side. Denote by $h^{(c)} = \arg\min_{1 \leq d \leq D} ISE(h_d, h_c)$, for $c = 1, \cdots, C$.

(L2) Compute the M-fold CV criterion in (2.10) evaluated at $(h^{(c)}, k_c)$, for $c = 1, \cdots, C$. The selected values for the tuning parameters used in the one-stage estimation method are given by $(h^{(c^*)}, k_{c^*})$, where $c^* = \arg\min_{1 \leq c \leq C} \text{CV}(h^{(c)}, k_c)$.

This two-layer procedure requires $C(D + M)$ rounds of estimation of β_1 and $g(\cdot)$, in contrast to CDM rounds of such estimation that the two-dimensional CV procedure involves. Hence, besides being well motivated by the empirical evidence that estimating the non-parametric part of the regression model has a greater impact on estimating the parametric part than the influence of the other way around, the two-layer tuning parameters selection method yields a tremendous amount of saving in computing time. The price one pays for such saving is that one needs some pilot estimator for $g(\cdot)$, namely $\tilde{g}(t)$ in (2.11), that can estimate the truth reasonably well. One way to obtain a $\tilde{g}(t)$ is to posit a flexible parametric model for the mode residual, $\epsilon = Y - \beta_1 X - g(T)$, and approximate $g(T)$ via a polynomial function of some order, then estimate the unknowns using the maximum likelihood method. Another option is to use the estimate from the two-stage estimation method, $\hat{g}_{RT}(t)$, as a pilot estimate.

It is worth pointing out that, for the two-stage estimation method, the nonparametric part of the estimation for $\beta_0 + g(t)$ is mostly accomplished in the first stage,

i.e., (T1) in Section 2.3, where $\hat{\beta}_0 + \hat{g}(t)$ is obtained and does not depend on h. Hence, the two-layer tuning parameters selection procedure is not applicable for the two-stage estimation method since one chooses h for estimating the nonparametric part in (L1).

2.5 EMPIRICAL STUDY

SIMULATION STUDY FOR THE ONE-STAGE ESTIMATION METHOD

In the simulation experiment, we consider the following two model configurations:

(E1) $Y = 2\sin(2\pi t) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$, $X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$ and $\text{Corr}(X, T) = 0.83$.

(E2) $Y = \exp\{\sin(\pi t)\} + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$, $X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$ and $\text{Corr}(X, T) = 0.83$.

Given $\{X_j, T_j\}_{j=1}^n$, $\epsilon_M(x, t) \approx 1$ for all (x, t). Hence, in (E1), $y_M(x, t) \approx 2\sin(2\pi t) + 1 + 3x$; in (E2), $y_M(x, t) \approx \exp(\sin(\pi t)) + 1 + 3x$. From the perspective of model (2.6), under configurations (E1) and (E2), ignoring rounding error, the true parametric part coefficient is $\beta_1 = 3$, the true nonparametric function $g(t)$, $t \in [0, 1]$, is equal to $2\sin(2\pi t) + 1$ and $\exp\{\sin(\pi t)\} + 1$, respectively. $g^*(t)$ is equal to $2\sin(2\pi t)$ and $\exp\{\sin(\pi t)\}$, respectively. The simulation settings in both (E1) and (E2) are used to show the performance of the one-stage estimation method. The correlation between X and T is controlled by $\text{Corr}(X, T) = 0.83$. Additionally, to evaluate the performance of $\hat{g}(t)$, we compute the Monte Carlo average of the mean square error of the nonparametric function estimate given by

$$NE^2 = \text{MC average of} \left[n^{-1} \sum_{i=1}^n \{\hat{g}(T_i) - g(T_i)\}^2 \right]. \tag{2.12}$$

Under each model setting, 300 Monte Carlo replicate sets of the form $\{(Y_j, X_j, T_j)\}$ are generated, producing 300 sets of estimates, $\{\hat{\beta}_{RO,1}, NE_{RO}^2\}$ for one-stage method.

39

Table 2.1 presents the Monte Carlo average and the standard error of the estimates obtained from the one-stage method when $n = 200$ and 400, with the tuning parameter is selected by two dimensional cross validation. Compared to the true value of parametric coefficient, the one-stage estimates with a larger sample size are closer to the true value. Furthermore, according to the NE^2 for different sample sizes, the one-stage estimate of the nonparametric part based on $n = 400$ also provides a better approximation.

Table 2.1: Averages of parameter estimates across 300 repetitions from one-stage estimation method with tuning parameters chosen by the two-dimensional cross validation. Numbers in parentheses are $(10 \times \text{standard errors})$ associate with the averages. The truth is $\beta_1 = 3$.

	(E1)				(E2)		
	$n = 200$		$n = 400$		$n = 200$		$n = 400$
β_1	NE^2	β_1	NE^2	β_1	NE^2	β_1	NE^2
2.65	0.38	2.83	0.15	2.66	0.38	2.84	0.15
(0.31)	(0.34)	(0.17)	(0.17)	(0.29)	(0.44)	(0.15)	(0.19)

Table 2.2 repeats the same demonstration as that in Table 2.1 except for that the two-layer tuning parameters selection procedure is usded to choose (h, k), where $\tilde{g}(t)$ in (2.11) is set at the truth for simplicity. Comparing these results with those when the two-dimensional CV procedure is used, we find much improved for both $\hat{\beta}_1$ and $\hat{g}(\cdot)$.

Table 2.2: Averages of parameter estimates across 300 repetitions from one-stage estimation method with tuning parameters chosen by the two-layer tuning parameter selection method. Numbers in parentheses are $(10 \times \text{standard errors})$ associate with the averages. The true $\beta_1 = 3$.

	(E1)				(E2)		
	$n = 200$		$n = 400$		$n = 200$		$n = 400$
β_1	NE^2	β_1	NE^2	β_1	NE^2	β_1	NE^2
2.89	0.07	2.92	0.03	2.88	0.07	2.91	0.03
(0.12)	(0.11)	(0.10)	(0.02)	(0.16)	(0.16)	(0.09)	(0.09)

To test the performance of our two-stage estimation method, we revise the simulation settings (E1) and (E2) slightly to simulate covariates data so that, X and T are independent.

(E3) $Y = 2\sin(2\pi t) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$, $X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$, X and T are independent.

(E4) $Y = \exp\{\sin(\pi t)\} + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$, $X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$, X and T are independent.

Under each model setting, 300 Monte Carlo replicate sets of the form $\{(Y_j, X_j, T_j)\}$ are generated, producing 300 sets of estimates. Table 2.3 represents averages of the estimates $\{\hat{\boldsymbol{\beta}}_{RT}, NE^2_{RO}\}$ across 300 replicates under (E3) and (E4) along with their empirical standard errors when $n = 200$ and 400. Similar as results from the one-stage method, Table 2.3 shows that the estimates for the parametric part with a larger sample size are closer to the true value. The B-spline approximation based on a larger sample size has better performance in terms of NE^2. Comparing results under (E3) and (E4), even with a more complex nonlinear $g(t)$ in (E4) than that in (E3), one can see that the results under (E4) tell the same story as the results under (E3).

2.6 Discussion

Besides the tuning parameter selection strategies presented in Section 2.4, we also experiment on combining $ISE(h)$ and $MSE(h) = (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})^T \Sigma^{-1} (\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta})$ to select the bandwidth. We have found that computing $ISE(h)$ is less time consuming, and that bandwidths chosen via minimizing $ISE(h)$ lead to more accurate estimators than when other criteria are used to select the bandwidth, such as $MSE(h)$, or some

Table 2.3: Averages of parameter estimates from the two-stage estimation method across 300 repetitions. Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truths are $\beta_0 = 1$, $\beta_1 = 3$.

(E3)					
$n = 200$			$n = 400$		
β_0	β_1	NE^2	β_0	β_1	NE^2
0.94	2.79	0.27	0.95	2.83	0.14
(0.15)	(0.17)	(0.12)	(0.11)	(0.14)	(0.05)
(E4)					
$n = 200$			$n = 400$		
β_0	β_1	NE^2	β_0	β_1	NE^2
0.91	2.76	1.27	0.94	2.86	1.15
(0.15)	(0.20)	(0.28)	(0.11)	(0.14)	(0.19)

combination of $ISE(h)$ and $MSE(h)$. We also observe in empirical study that, when it comes to estimating β_1 and $g(t)$ via the one-stage method, the two dimensional cross validation tuning parameter selection method produces similar results as those when the two-layer tuning parameter selection method is used.

When X in model (2.2) is observable, our two proposed method can be implemented to estimate the parametric coefficient and the unspecified smooth function. If, instead of observing X, we observe its error contaminated surrogate W, our two proposed methods need to be revised to account for measurement error in the covariate. In the next chapter, we follow the rationale of the corrected kernel method in Chapter 1 to propose a method for mode regression , assuming partially linear mode models with error-prone X.

CHAPTER 3

PARTIALLY LINEAR MODE REGRESSION WITH ERROR IN COVARIATES

Covariates that cannot be measured precisely or directly are ubiquitous in practice (Carroll et al., 2006). To address the practical issue of error-contaminated covariates, methods accounting for measurement error in partially linear mean regression (Liang et al., 1999, 2007, 2008; Liang and Li, 2009; Koul and Song, 2010) and methods for partially linear quantile regression (He and Liang, 2000; Hardle et al., 2000) in the presence of measurement error have been developed. But, to the best of our knowledge, there exists no published work on partially linear mode regression in the presence of covariate measurement error. The most relevant works so far are that by Zhou and Huang (2016), who proposed methods for nonparametric mode regression with covariate measurement error, and that in Chapter 1, where we developed methods for linear mode regression in the presence of covariate measurement error. We spearheaded in this line of research and present in this chapter methods for inferring the mode of Y conditioning on T and X when X is prone to measurement error.

In this chapter, we propose methods for estimating the conditional mode of a continuous response given covariates of interests, some of which are prone to measurement error and relate to the mode of the response linearly, and some are error-free and relate to the mode via an unknown functional form. We study asymptotic properties of the proposed estimators for the linear part and the nonlinear part of the mode model. Their finite sample properties are investigated via extensive simulation

study, in comparison with their naive counterpart estimators that ignore covariate measurement error, as well as with their error-free counterparts obtained based on data free of covariate measurement error. In the simulation study, two proposed procedures for choosing tuning parameters involved in the estimation methods are also compared empirically. Finally, we apply the proposed methods to data from the Framingham Heart Study.

3.1 DATA AND MODELS

Suppose that the observed data consist of n independent data points, $\{(Y_j, T_j, W_j)\}_{j=1}^n$, where $\{W_j\}_{j=1}^n$ are surrogates of the unobserved covariate values $\{X_j\}_{j=1}^n$, $\{T_j\}_{j=1}^n$ are scalar covariate values, and $\{Y_j\}_{j=1}^n$ are response variable values. In particular, we assume the classical measurement error model (1.2) in Chapter 1,

$$W_j = X_j + U_j.$$

We also assume that there exists a unique largest mode as in Grund and Hall (1995), we aim to infer in this study the mode of the conditional distribution of Y given (T, X), specified by the partially linear model in (2.6). Besides the unspecified function $g(\cdot)$ in (2.6), another nonparametric component of the mode regression model we introduce lies in the mode residual, $\epsilon = Y - Mode(Y|T, X)$, the distribution of which is left unspecified except that its mode is zero conditioning on (T, X). Denote by $f_{\epsilon|(T,X)}(\epsilon|t, x)$ the density of this distribution.

In Chapter 2, where there is no measurement error in the covariates and thus data $\{(Y_j, T_j, X_j)\}_{j=1}^n$ are observed, we consider drawing inference for β_1 and $g(\cdot)$ based on the following kernel density estimator for $f_{\epsilon|(T,X)}(0|t, x)$,

$$Q_h(\beta_1, g) = n^{-1} \sum_{j=1}^n K_h(Y_j - \beta_1 X_j - g(T_j)), \tag{3.1}$$

Sensible estimators for β_1 and $g(\cdot)$ should maximize $Q_h(\beta_1, g)$ for an adequately chosen h since the mode of a distribution is where the corresponding density is maximized.

44

In the presence of measurement error, since $\{X_j\}_{j=1}^n$ are not observed, one cannot use (3.1) to estimate $f_{\epsilon|(T,X)}(\epsilon|t,x)$. Therefore, based on the observed data $\{(Y_j, T_j, W_j)\}_{j=1}^n$, different ways to estimate the distribution of the mode residual ϵ are called for. In this chapter, we still use B-spline to approximate $g(t)$ in (2.6). As for how to create B-spline basis functions, details can be found in Section 2.2. By substituting X_j with W_j in (2.4), a naive objective function that one maximizes with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is given by

$$Q_{h,nv}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = n^{-1} \sum_{j=1}^n K_h(Y_j - B(T_j)^T \boldsymbol{\gamma} - \beta_1 W_j). \tag{3.2}$$

Using this naive objective function in the one-stage estimation method in Section 2.3 yields the one-stage naive estimators (NVO), denoted by $(\hat{\beta}_{\text{NVO},1}, \hat{\boldsymbol{\gamma}}_{\text{NVO}})$. The naive estimators resulting from the two-stage estimation method in Section 2.3 (NVT) are denoted by $(\hat{\beta}_{\text{NVT},1}, \hat{\boldsymbol{\gamma}}_{\text{NVT}})$. To account for measurement error, in the upcoming section, we develop non-naive estimators for β_1 and $g(t)$. This is achieved by modifying the kernel function in (3.2) to obtain an estimator for $f_{\epsilon|(T,X)}(0|t,x)$ acknowledging that W relates to X according to (1.2). This strategy is motivated by the deconvoluting kernel density estimators in the presence of measurement error (Carroll and Hall, 1988; Stefanski and Carroll, 1990) that we elaborate next. To simplify notations, in this chapter, we assume a scalar X henceforth, and defer discussions on generalization to multivariate X at the end of this chapter.

3.2 Proposed Methods

The One-stage Corrected Kernel Estimation Method

Following Carroll and Hall (1988) and Stefanski and Carroll (1990), when U is nondifferential (Carroll et al., 2006, Section 2.5), and the characteristic function of U never vanishes, one has

$$E[K_h^*\{Y - \beta_1 W - g(T)\}] = K\{Y - \beta_1 X - g(T)\}, \tag{3.3}$$

45

where $K_h^*(s) = h^{-1}K^*(s/h)$ is the scaled deconvoluting kernel function defined in (1.7),

$$K^*(s) = \frac{1}{2\pi}\int e^{-ivs}\frac{\phi_K(v)}{\phi_U(-\beta_1 v/h)}ds,$$

in which $\phi_U(v)$ is the characteristic function of U, and $\phi_K(v)$ is the Fourier transform of $K(s)$. All integrals in this chapter are over the real line unless otherwise stated. As in most literature on deconvoluting density estimation, we assume $\phi_U(v)$ known for the majority of the study, and discuss treatments of unknown $\phi_U(v)$ in simulation section. The significance of (3.3) is that it motivates the following estimator for $f_{\epsilon|(T,X)}(0|t,x)$ that accounts for measurement error,

$$Q_h^*(\beta_1, g) = n^{-1}\sum_{j=1}^{n}K_h^*(Y_j - \beta_1 W_j - g(T_j)). \tag{3.4}$$

which has the same bias as that of (3.1).

Using the B-spline approximation for $g(\cdot)$ in (3.4), we obtain non-naive estimators for β_1 and γ by maximizing

$$Q_h^*(\beta_1, \boldsymbol{\gamma}) = n^{-1}\sum_{j=1}^{n}K_h^*(Y_j - B(T_j)^T\boldsymbol{\gamma} - \beta_1 W_j). \tag{3.5}$$

with respect to β_1 and $\boldsymbol{\gamma}$. Because (3.5) involves a deconvoluting kernel that corrects the original kernel in (3.2) for measurement error, and we maximize (3.5) to estimate β_1 and $\boldsymbol{\gamma}$ (and thus $g(t)$) simultaneously, we refer to the so-obtained estimators corrected kernel one-stage estimators, denoted by corrected kernel one-stage estimators, denoted by $(\hat{\beta}_{\mathrm{CKO},1}, \hat{\gamma}_{\mathrm{CKO}})$, respectively, and refer to this method the one-stage corrected kernel estimation method. By comparing the one-stage corrected kernel estimation method proposed here with the one-stage estimation method in Chapter 2, one can easily see that the difference lies in that the kernel function in (2.4) is replaced by the corrected kernel function (1.7) when the observed covariate is W instead of X.

46

Similarly, one may revise the two-stage estimation method in the absence of measurement error in Chapter 2 to estimate the unknowns in a partially linear mode model when X is prone to measurement error. More specifically, after an estimator for $g^*(\cdot)$ is obtained in the first stage, denoted by $\hat{g}^*(\cdot)$, one may carry out linear mode regression of $Y^* = Y - \hat{g}^*(T)$ on X using data $\{(Y_j^*, W_j)\}_{j=1}^n$ to estimate β_0 and $\beta_1 = \alpha_1 + \alpha_2$ in (2.9) following the methods proposed in Chapter 1. These two steps accomplish estimating the parametric part, β_1, and the nonparametric part, $g(T) = \beta_0 + g^*(T)$. We refer to the so-obtained estimators the corrected kernel two-stage (CKT) estimators, and refer to this method the two-stage corrected kernel estimation method. The algorithm for implementing this method is recapped below.

(T*1) Approximate $g^*(t)$ via cubic splines with k knots, that is, $g^*(t) \approx B'(t)\boldsymbol{\gamma}^*$. Regress Y on $B(T)$ using the least squares method, resulting in an estimator for $\boldsymbol{\gamma}^*$, denoted by $\hat{\boldsymbol{\gamma}}^*$. This leads to an estimator for $\hat{g}^*(t) = B'(t)\hat{\boldsymbol{\gamma}}^*$

(T*2) Define $Y_j^* = Y - \hat{g}^*(T_j)$, for $j = 1, \cdots, n$. Maximize the following objective function that involves the corrected kernel with respect to (β_0, β_1),

$$Q_h^*(\beta_0, \beta_1) = n^{-1} \sum_{j=1}^n K_h^*(Y_j^* - \beta_0 - \beta_1 W_j), \qquad (3.6)$$

which is constructed following the same rationale as that of (3.4) and is the same objective function used in Chapter 1 for linear mode regression in the presence of covariate measurement error. Denote the resultant estimators for β_0 and β_1 as $\hat{\beta}_0$ and $\hat{\beta}_{CKT,1}$, respectively.

3.3 ASYMPTOTIC PROPERTIES

In this section, asymptotic properties of the estimator obtained from the one-stage estimation method will be discussed. First, we have the following Theorem 3 which

states the consistency of the proposed one-stage estimator.

Theorem 3. *Under conditions (C^*1)–(C^*5) in Appendix D and conditions in Lemma C, there exists a maximizer of $Q_h^*(\gamma, \beta)$, denoted by $\hat{\boldsymbol{\theta}}_{CK} = (\hat{\gamma}_{CK}, \hat{\boldsymbol{\beta}}_{CK})$, such that, as $n \to \infty$ and $h \to 0$,*

    (i) when U follows an ordinary smooth distribution of order b, if $nh^{7+2b} \to 0$, then

$$\|\hat{\boldsymbol{\theta}}_{CK} - \boldsymbol{\theta}\| = O(h^2) + O_p\left(\sqrt{\frac{1}{nh^{3+2b}}}\right), \qquad (3.7)$$

    (ii) when U follows a super smooth distribution of order b,

        if $\exp(2|\beta_1|^b h^{-b}/d_2))/(nh^{b_6}) \to 0$,

        where $b_6 = \max\{3 - 2\min(b_2, b_3), 5 - 2\min(b_2, b_3, b_4), 7 - 2\min(b_2, b_3, b_4, b_5)\}$,

        in which b_ℓ, for $\ell = 2, 3, 4, 5$, are defined in Lemma C, then

$$\|\hat{\boldsymbol{\theta}}_{CK} - \boldsymbol{\theta}\| = O(h^2) + O_p\left\{\exp\left(\frac{|\beta_1|^b}{d_2 h^b}\right)\sqrt{\frac{1}{nh^{3-2\min(b_2, b_3)}}}\right\}, \qquad (3.8)$$

$\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is the coefficient vector of basis functions $\boldsymbol{B}(t)$ such that $\|g(t) - \boldsymbol{B}(t)\boldsymbol{\gamma}\| = O(k^{-r})$ and $g(t)$ is rth continuously differentiable. The next theorem states that the one-stage estimator for $\boldsymbol{\beta}$ follows a normal distribution asymptotically.

Theorem 4. *Under the same assumptions imposed in Theorem 3,*

    (i) if U follows an ordinary smooth distribution of order b,

$$\sqrt{nh^{3+2b}}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2\mu_2 I^{*-1}Q/4\right) \xrightarrow{d} N(0, I^{*-1}M_L I^{*-1}); \qquad (3.9)$$

    where M_L is constant matrices.

    (ii) if U follows a super smooth distribution of order b,

$$\left\{Var(\hat{\boldsymbol{\beta}}_{CK})\right\}^{-1/2}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2\mu_2 I^{*-1}Q/4\right) \xrightarrow{d} N(0, 1), \qquad (3.10)$$

where

$$Q = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} E\{f^{(3)}(0|X_j, T_j)\tilde{X}_j\},$$

$$I^* = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} E\left\{f^{(2)}(0|X_j, T_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}} + \tilde{X}_j B_j^T \Phi^{-1}\Psi\right\},$$

$Var(\hat{\boldsymbol{\beta}}_{CK}) = O[\exp\{2|\beta_1|^b/(d_2h^b)\}/\{nh^{3-2\min(b_2,b_3)}\}]$, *and* $\boldsymbol{\Sigma}^{-1/2}$ *denotes the inverse of the positive definite square root of a positive definite matrix* $\boldsymbol{\Sigma}$.

The proofs of Theorem 3 and Theorem 4 can be found in Appendix D and Appendix E, respectively.

3.4 Tuning Parameter Selection

As in Chapter 2, cubic spline basis functions are adopted to approximate $g(T)$ with $\ell = 4$ in this chapter. After fixing the order of spline basis functions, similar as the the tuning parameter selection in section 2.4, two-dimensional cross validation method and two-layer method are also employed to select the tuning parameter in this chapter.

Two-dimensional cross validation

Since sensible estimators for β_1 and $g(\cdot)$, using error-prone data maximize (3.4), a natural way to choose h and k is by maximizing the following M-fold cross validation (CV) criterion, which can be computed after one partitions the observed data into M, subsets that are as of equal size as possible,

$$CV(k, h) = M^{-1} \sum_{m=1}^{M} n_m^{-1} \sum_{i \in \mathcal{I}_m} K_h^* \left\{Y_i - \hat{g}^{(-m)}(T_i) - \hat{\beta}_1^{(-m)}X_i\right\}, \qquad (3.11)$$

where M presents the number of partitions of the data set, $\mathcal{I}_m$ is the observation index set associated with the m-th subset of data, n_m is the size of the data set $\mathcal{I}_m$, $m = 1, \cdots, M$, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ are estimators for β_1 and $g(\cdot)$, respectively, based on the raw data of size n excluding the mth subset of size n_m, for

$m = 1, \cdots, M$. To be more specific, when one chose (h, k) that go in the one-stage estimation method, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ in (3.11) are $\hat{\beta}_{CKO,1}$ and $\hat{g}_{CKO}(\cdot)$ computed using data $\{(Y_j, T_j, W_j), j \in \mathcal{I} \backslash \mathcal{I}_m\}$, where $\mathcal{I} = \{1, \cdots, n\}$, and "$\backslash$" is the set subtraction operator, for $m = 1, \cdots, M$. Similarly, when one selects (h, k) that go along with the two-stage estimation method, $\hat{\beta}_1^{(-m)}$ and $\hat{g}^{(-m)}(\cdot)$ in (3.11) are $\hat{\beta}_{CKT,1}$ and $\hat{g}_{CKT}(\cdot)$ computed using data $\{(Y_j, T_j, W_j), j \in \mathcal{I} \backslash \mathcal{I}_m\}$, where $\mathcal{I} = \{1, \cdots, n\}$ for $m = 1, \cdots, M$. All these estimators depend on (h, k), the dependence we suppress on the right-hand side of (3.11) for cleaner notations. Following this cross validation procedure, referred to as the two-dimensional CV in the sequel, the chosen number of knots and bandwidth are followed by

$$(\hat{k}, \hat{h}) = \max_{k,h} CV(k, h).$$

Zhao et al. (2013) and Zhao et al. (2014) employed a similar procedure to choose tuning parameters in partially linear varying coefficient mode regression in the absence of covariate measurement error. In our simulation, FFT method (Bailey and Swarztrauber, 1994) is utilized to calculate the deconvoluting kernel $K^*(\cdot)$ in (3.1). Carrying out the FFT method in conjunction with the two dimensional M-fold cross validation in the one-stage corrected kernel method makes it a computationally expensive procedure. To ease the computational burden for tuning parameters selection for the one-stage corrected kernel estimation method, the two-layer tuning parameters selection method will be utilized.

TWO LAYER TUNING PARAMETER SELECTION

Our proposed methods in this chapter involve calculate the deconvoluting kernel (1.7). Although the FFT method is employed in this calculation, the computing time is much longer than the process in Chapter 2. This motivates us to mimic the second strategy in last chapter and develop the following algorithm.

(L*1) For each candidate value of k, k_c, where $c \in \{1, \cdots, C\}$, find an h among its candidate values, $\{h_1, \cdots, h_D\}$, that minimizes the integrated squared error (ISE) of the estimate for $g(\cdot)$,

$$\text{ISE}(h, k_c) = \int_0^1 \{\hat{g}_{CKO}(t) - \tilde{g}(t)\}^2 dt, \qquad (3.12)$$

where $\tilde{g}(t)$ is a preliminary estimate for $g(\cdot)$, and $\hat{g}_{CKO}(t)$ is the one-stage corrected kernel estimate obtained based on the entired observed data set, whose dependence of on h, after k is fixed at k_c, is suppressed on the right-hand side. Denote by $h^{(c)} = \arg\min_{1 \leq d \leq D} ISE(h_d, k_c)$, for $c = 1, \cdots, C$.

(L*2) Compute the M-fold CV criterion in (3.11) evaluated at $(h^{(c)}, k_c)$, for $c = 1, \cdots, C$. The selected values for the tuning parameters used in the one-stage estimation method are given by $(h^{(c^*)}, k_{c^*})$, where $c^* = \arg\max_{1 \leq c \leq C} \text{CV}(h^{(c)}, k_c)$.

The same as the two-layer tuning parameter method in chapter 2, for the two-stage estimation method, the nonparametric part of the estimation for $g(t) = \beta_0 + g^*(t)$ is mostly accomplished in the first stage, i.e., (T*1) in Section 3.2, where $\hat{g}^*(t)$ is obtained and does not depend on h. Hence, the two-layer tuning parameters selection procedure is not applicable for the two-stage estimation method since one chooses h for estimating the nonparametric part in (L*1).

3.5 EMPIRICAL STUDY

SIMULATION STUDY FOR THE ONE-STAGE ESTIMATION METHOD

To assess finite sample performance of the one-stage estimation method, we compute $\hat{\beta}_{CKO,1}$ and $\hat{g}_{CKO}(t)$ based on data simulated from the following two settings,

(F1) $Y = 2\sin(2\pi t) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$, $X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$, $U \sim M(0, \sigma_u^2)$, Corr(X,T) = 0.83.

(F2) $Y = \exp(\sin(\pi t)) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$,

$X \sim \text{uniform}(-1, 1)$, $T \sim \text{uniform}(0, 1)$, $U \sim M(0, \sigma_u^2)$, $\text{Corr(X,T)} = 0.83$.

In the above two settings, the distribution of M represents normal and Laplace distribution with mean 0 and variance σ_u^2, respectively. To vary $Var(U) = \sigma_u^2$, under each true model configuration, we set the reliability ratio $\lambda = Var(X)/\{Var(X) + \sigma_u^2\}$ equal to 0.75, 0.85, 0.95, respectively. When implementing our proposed methods, the kernel $K(t)$ used for obtaining estimators based on error-free data $\{(Y_j, T_j, X_i)\}_{j=1}^n$ is the standard normal density, as in Zhao et al. (2014); and we use the kernel $K(T) = 48 \cos t(1 - 15/t^2)/(\pi t^4) - 144 \sin t(2 - 5/t^2)/(\pi t^5)$, of which the Fourier transform is $\phi_K(t) = (1 - t^2)^3 I(-1 \leq t \leq 1)$, for the corrected kernel method based on error-prone data $\{(Y_j, T_j, W_j)\}_{j=1}^n$, where $I(\cdot)$ is the indicator function.

Besides the corrected kernel one-stage estimates, $\hat{\beta}_{CKO,1}$ and $\hat{g}_{CKO}(t)$, we also maximize (3.2) to obtain the naive one-stage estimates, denoted by $\hat{\beta}_{NVO,1}$ and $\hat{g}_{NVO}(t)$. Lastly, as benchmark estimates that the proposed estimates and the naive estimates compare with, we maximize (2.7), with $g(\cdot)$ approximated by cubic splines, to obtain the one-stage mode regression estimates, denoted by $\hat{\beta}_{RO,1}$ and $\hat{g}_{RO}(t)$ based on error-free data $\{(Y_j, T_j, X_i)\}_{j=1}^n$. Under each of the simulation setting, 300 Monte Carlo replicate data sets are generated from the true model of $\{Y_j, T_j, W_j\}_{j=1}^n$ with $n = 200$ and 400 respectively, producing 300 sets of estimates from the one-stage estimation method. For the one-stage estimation method, the tuning parameter is selected by using the two dimensional cross validation method and the two-layer method, respectively.

First, the simulation results based on the two dimensional cross validation tuning parameter selection method are presented. When implementing the corrected kernel method for the partially linear model, the fast Fourier transforms (FFT, Bailey and Swarztrauber, 1994) is implemented to approximate relevant integrals, and we use the Laplace characteristic function in (1.7) even when U is actually normally distributed.

Figure 3.1 presents in the left panels boxplots of $\hat{\beta}_{NVO,1}$, $\hat{\beta}_{CKO,1}$, $\hat{\beta}_{RO,1}$ obtained from 300 Monte Carlo (MC) replicate data sets, where the responses are generated according to (F1), $U \sim \text{Laplace}(0, \sigma_u^2)$, and $n = 200$, with tuning parameters chosen via the two-dimensional CV procedure. One can see from these boxplots substantial bias reduction achieved by the proposed estimate $\hat{\beta}_{CKO,1}$ when comparing with its naive counterpart, $\hat{\beta}_{NVO,1}$, which severely underestimates the covariate effect. As one would expect with inference based on error-prone data, $\hat{\beta}_{CKO,1}$ exhibits higher variability than its error-free counterpart, $\hat{\beta}_{RO,1}$, but they become more comparable as error contamination in data lessens, i.e., as λ increases. The right panels in Figure 3.1 provide boxplots of the empirical squared error associated with the estimated nonparametric part, $\text{NE}^2 = \sum_{j=1}^{200} \{\hat{g}(T_j) - g(T_j)\}^2/200$, where $\hat{g}(\cdot)$ denotes generically one of the estimates, $\hat{g}_{NVO}(\cdot)$, $\hat{g}_{CKO}(\cdot)$, and $\hat{g}_{RO}(\cdot)$. These boxplots suggest that the proposed one- stage estimation method also yields improved inference for the nonparametric part of mode regression compared to its naive counterpart.

Figure 3.2 includes boxplots of the same quantities under the same model setting as those in Figure 3.1 except that $U \sim N(0, \sigma_u^2)$. Significant improvement in inference from the proposed one-stage estimation method are again evident when compared with the naive method, although their resemblance to the error-free inference is less impressive than that in the presence of Laplace measurement error when λ is low. This latter phenomenon can be explained by the asymptotic results, which indicate much slower convergence rates of the estimates in the presence of super smooth measurement error than when the measurement error is ordinary smooth.

Figure 3.3 and Figure 3.4 depict boxplots of estimates of β_1 and NE^2 in the presence of Laplace measurement error and normal measurement error, respectively, when $n = 400$. When compared with Figure 3.1 and Figure 3.2, one can see that the variance of the estimates decrease as the sample size increases. Table 3.2 shows averages of parameter estimates across 300 repetitions for Laplace measurement error

53

and normal measurement error under (F1), respectively, when $n = 400$. The same pattern as in Figure 3.3 and Figure 3.4 can be observed in Table 3.2.

Table 3.1: One-stage estimation method with two dimensional cross validation tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 200$ under (F1). Numbers in parentheses are $(10 \times$ standard errors$)$ associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
			$U \sim N(0, \sigma^2)$			
Naive	0.61	2.41	0.94	1.50	1.57	14.06
	(0.32)	(0.67)	(0.40)	(2.34)	(0.43)	(128.99)
CK	2.27	1.05	2.44	0.43	2.73	0.29
	(0.76)	(1.56)	(0.40)	(0.38)	(0.25)	(0.59)
			$U \sim \mathrm{Laplace}(0, \sigma^2)$			
Naive	0.72	1.45	1.01	1.32	1.60	0.94
	(0.35)	(0.52)	(0.41)	(0.82)	(0.47)	(0.73)
CK	2.50	0.13	2.67	0.09	2.51	0.12
	(0.20)	(0.06)	(0.17)	(0.04)	(0.21)	(0.06)
TRUE	2.65	0.38				
	(0.31)	(0.34)				

Table 3.2: One-stage estimation method with two dimensional cross validation tuning parameter selection. Averages of parameter estimates over 300 repetitions when n =400 under (F1). Numbers in parentheses are $(10 \times$ standard errors$)$ associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
			$U \sim N(0, \sigma^2)$			
Naive$_N$	0.47	1.48	0.80	1.24	1.72	0.64
	(0.28)	(0.36)	(0.33)	(0.36)	(0.38)	(0.33)
CK$_N$	2.19	0.56	2.55	0.18	2.79	0.11
	(0.38)	(0.44)	(0.21)	(0.15)	(0.13)	(0.20)
			$U \sim \mathrm{Laplace}(0, \sigma^2)$			
Naive	0.60	1.32	1.03	1.04	1.83	0.56
	(0.30)	(0.34)	(0.37)	(0.37)	(0.39)	(0.31)
CK	2.79	0.35	2.88	0.14	2.83	0.07
	(0.47)	(0.30)	(0.24)	(0.10)	(0.13)	(0.04)
TRUE	2.83	0.15				
	(0.17)	(0.17)				

Figures 3.5 and 3.6 depict counterpart results of those in Figures 3.1 and 3.2 when the responses are simulated according to (F2). Even with a more complex nonlinear $g(t)$ in (F2) than that in (F1), Figure 3.5 and Figure 3.6 tell the same story as before, suggesting that the proposed one-stage estimation method, in conjunction

54
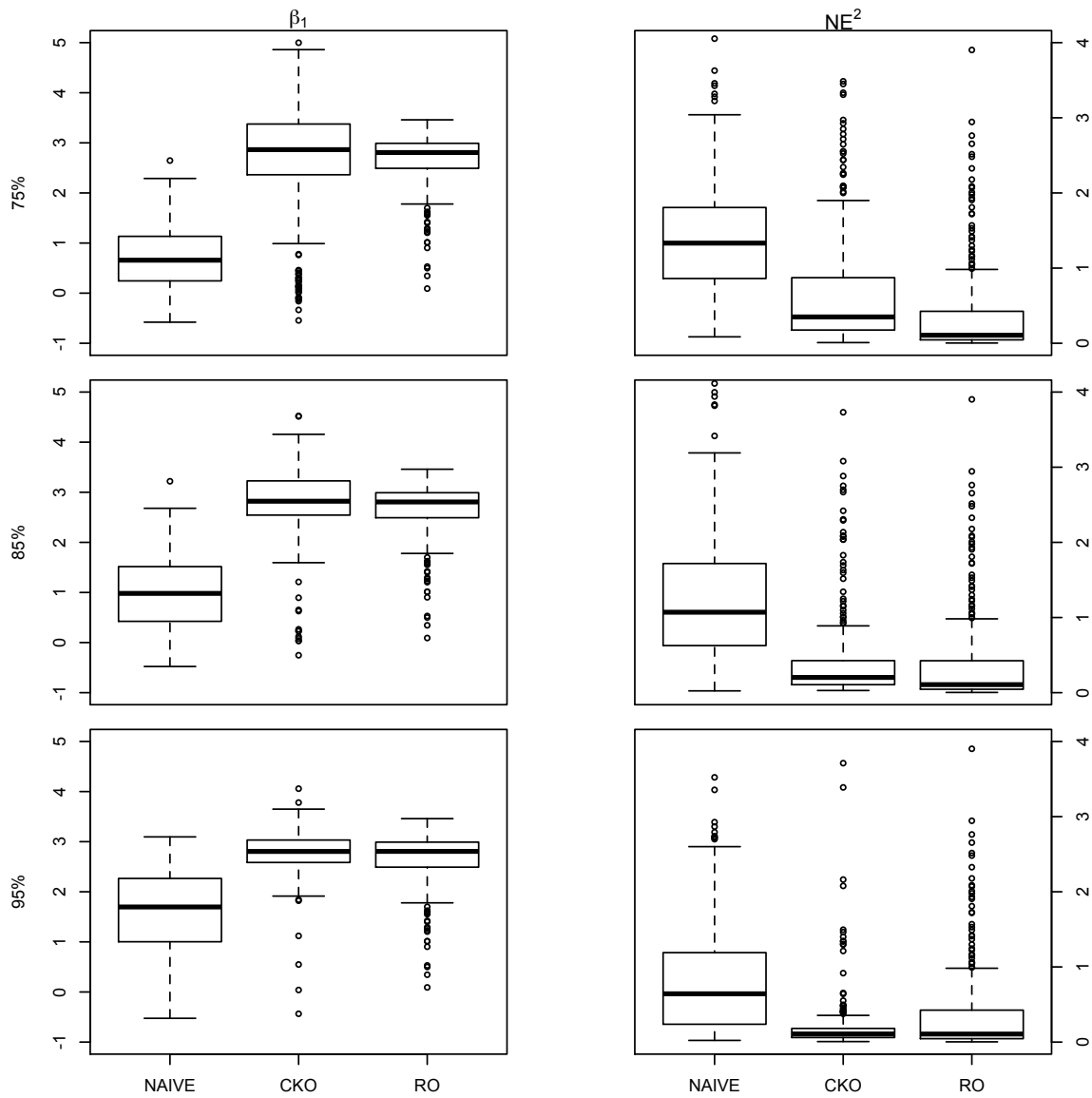
Figure 3.1: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F1) and the sample size $n = 200$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
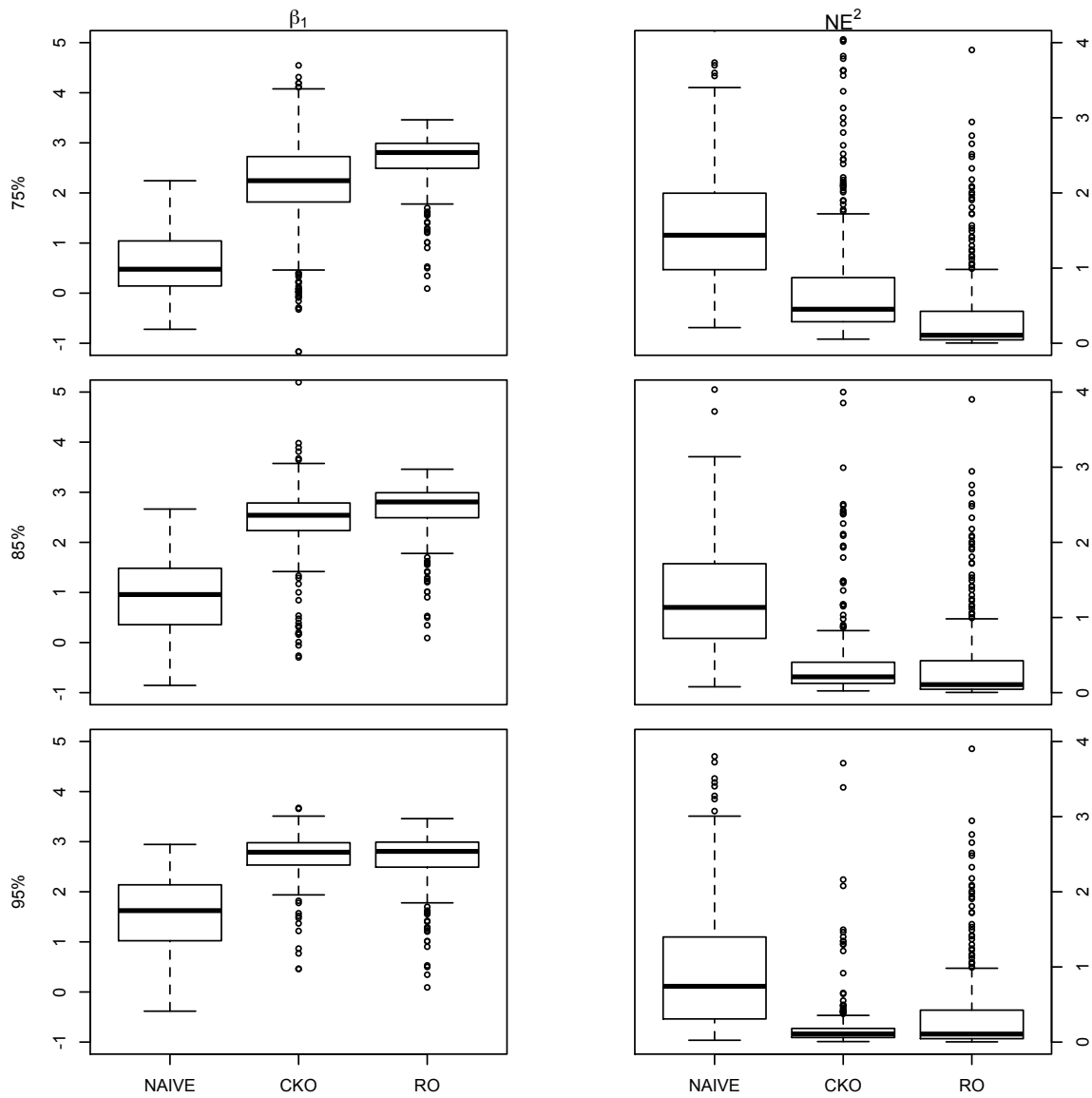
Figure 3.2: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F1) and the sample size $n = 200$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
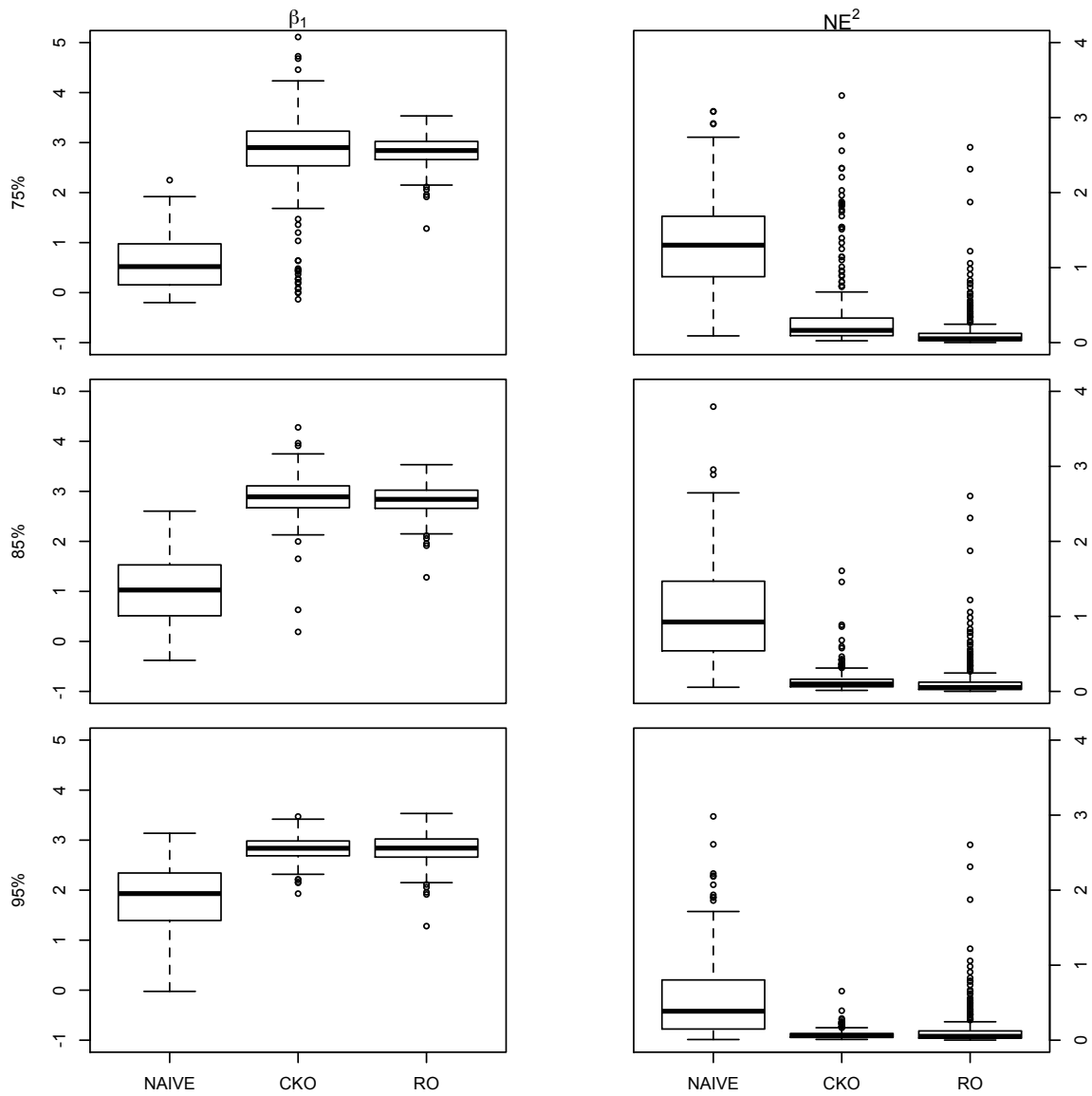
Figure 3.3: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F1) and the sample size $n = 400$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.

57

Figure 3.4: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F1) and the sample size $n = 400$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.

with the two-dimensional CV procedure for tuning parameters selection, outperforms its naive counterpart method, with more impressive improvement in the presence of ordinary smooth measurement error than when the error is super smooth. Numerical summary of these simulation results can be found in Table 3.3. Figure 3.7 and Figure 3.8 describe, after increasing the sample size to $n = 400$, boxplots of estimates of β_1 and estimates of NE^2 in the presence of two types of measurement error. As the sample size increases, less variance of the estimates is also observed. Table 3.4 shows averages of parameter estimates across 300 repetitions for Laplace measurement error and normal measurement error under (F2), respectively, when $n = 400$. The same pattern as in Figure 3.7 and Figure 3.8 can be observed in Table 3.4.

Table 3.3: One-stage estimation method with two dimensional cross validation tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 200$ under (F2). Numbers in parentheses are ($10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
	$U \sim N(0, \sigma^2)$					
Naive	0.63	1.72	0.99	1.29	1.55	0.92
	(0.33)	(2.00)	(0.38)	(0.55)	(0.46)	(0.53)
CK	2.19	0.77	2.46	0.54	2.78	0.23
	(0.55)	(0.45)	(0.36)	(0.74)	(0.18)	(0.32)
	$U \sim \text{Laplace}(0, \sigma^2)$					
Naive	0.72	2.39	1.05	1.39	1.75	1.38
	(0.37)	(0.80)	(0.42)	(1.50)	(0.44)	(5.44)
CK	2.77	0.97	2.74	0.63	2.79	0.23
	(0.71)	(1.03)	(0.42)	(1.57)	(0.23)	(0.30)
TRUE	2.66	0.38				
	(0.29)	(0.44)				

Figures 3.9 and 3.10 repeat the same demonstration as that in Figures 3.1 and 3.2 except for that the two-layer tuning parameters selection procedure is used to choose (h, k), where $\tilde{g}(t)$ in (3.12) is set at the truth for simplicity. Table 3.5 tells the same story as in Figures 3.9 and 3.10. Other figures and tables when $n = 200$ and $n = 400$ under (F2), parallel to Figures 3.4-3.8 and Tables 3.2-3.4 with this tuning parameters selection procedure employed are provided in the Appendix H. Comparing these results with those when the two-dimensional CV procedure is used,

59

Figure 3.5: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F2) and the sample size $n = 200$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.

60
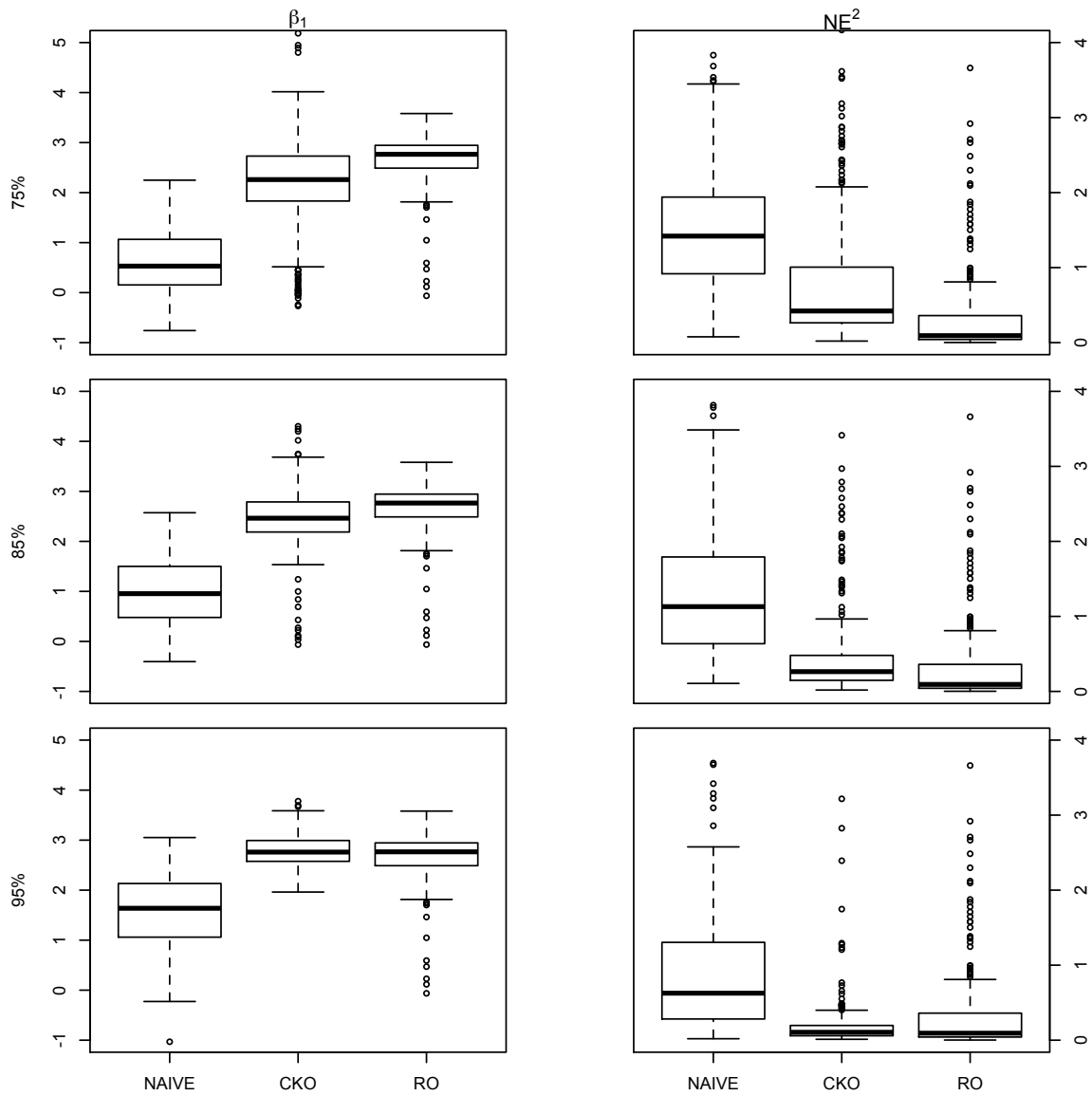
Figure 3.6: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F2) and the sample size $n = 200$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
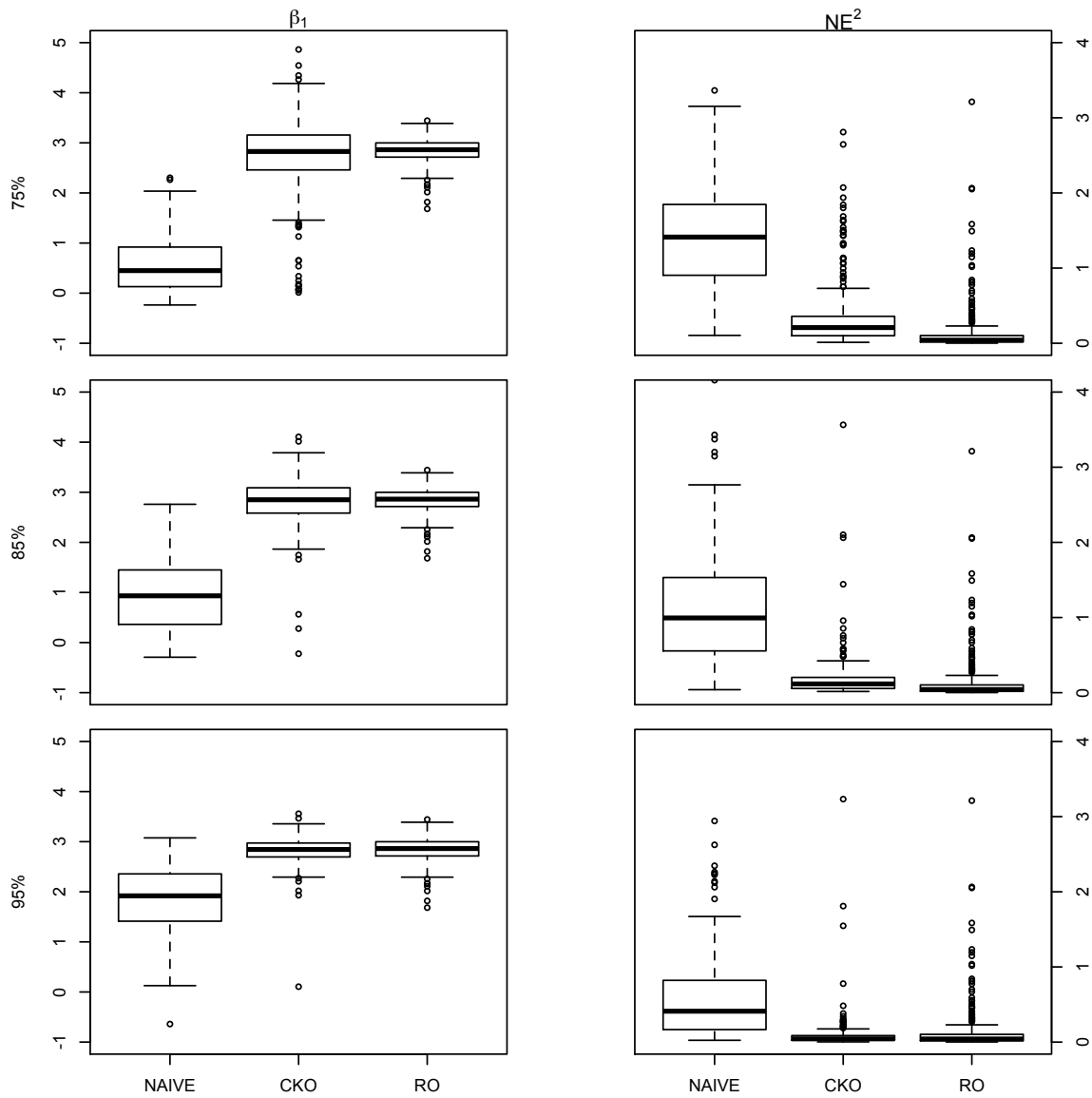
Figure 3.7: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F2) and the sample size $n = 400$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
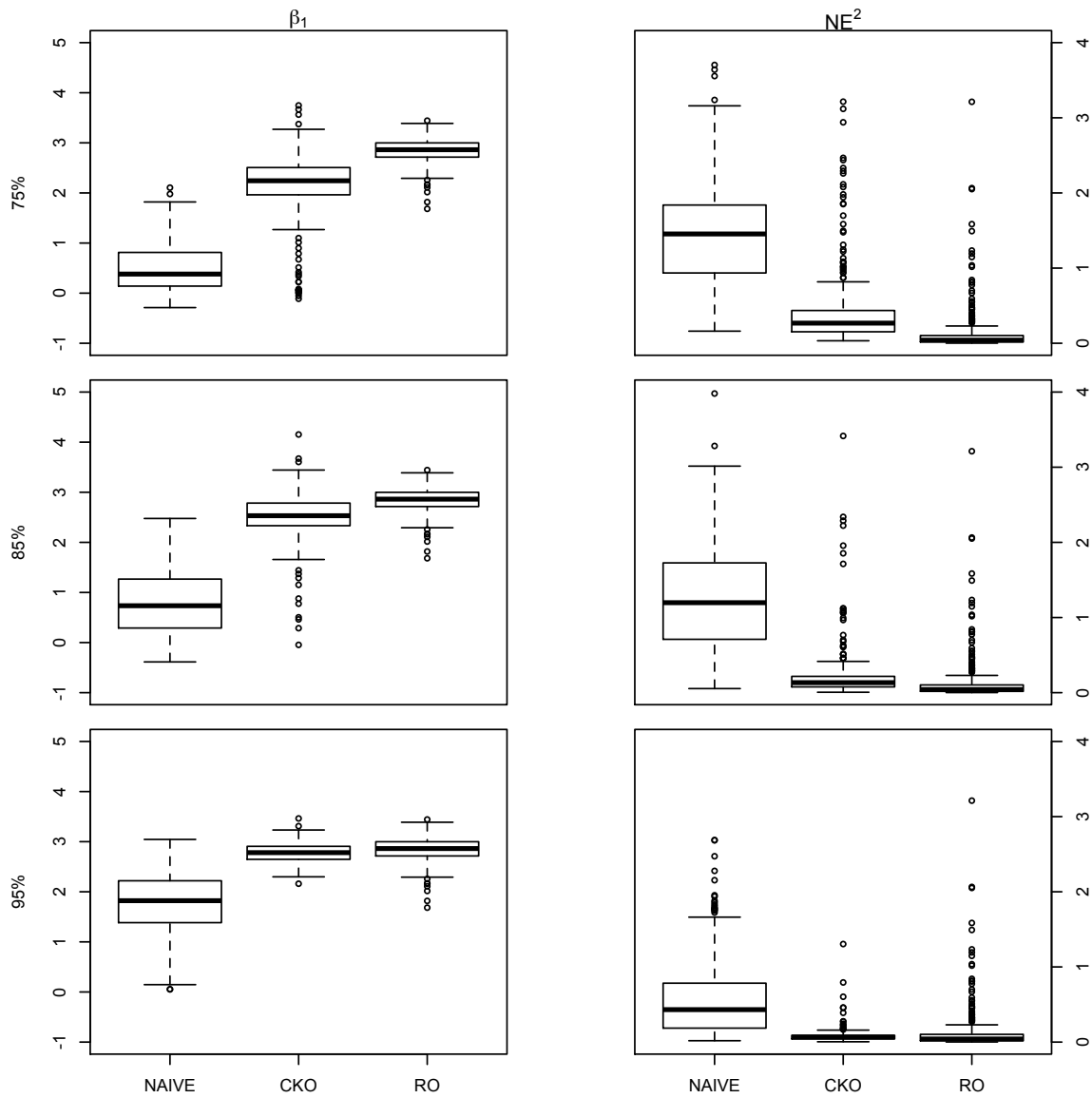
Figure 3.8: Two dimensional cross validation tuning parameter selection method. Under the simulation setting (F2) and the sample size $n = 400$, Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.

Table 3.4: One-stage estimation method with two dimensional cross validation tuning parameter selection. Averages of parameter estimates over 300 repetitions $n = 400$ under (F2). Numbers in parentheses are ($10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
	$U \sim N(0, \sigma^2)$					
Naive$_N$	0.52	1.45	0.81	1.24	1.74	0.57
	(0.28)	(0.38)	(0.35)	(0.38)	(0.36)	(0.30)
CK$_N$	2.16	0.46	2.51	0.23	2.78	0.08
	(0.38)	(0.42)	(0.27)	(0.22)	(0.12)	(0.06)
	$U \sim \text{Laplace}(0, \sigma^2)$					
Naive	0.57	1.42	0.97	1.10	1.82	0.56
	(0.30)	(0.38)	(0.40)	(0.40)	(0.39)	(0.30)
CK	2.75	0.34	2.83	0.18	2.83	0.09
	(0.45)	(0.24)	(0.27)	(0.17)	(0.16)	(0.14)
TRUE	2.84	0.15				
	(0.15)	(0.19)				

we find much improved estimates for $g(\cdot)$, and much less variable but otherwise comparable $\hat{\beta}_{CKO,1}$. This empirical evidence encourages use of the computationally less expensive two-layer procedure when one has a reliable pilot estimate for $g(t)$. As the same size increases from $n = 200$ to $n = 400$, one can see continuing improvement in estimates for the unknowns in a partially linear mode model from the proposed one-stage estimation method, paired with either the two-dimensional CV tuning parameters selection procedure or the two-layer procedure.

Admittedly, under (F1), results shown in Figures 3.9–3.12 where the two-layer procedure is employed to choose tuning parameters, can be overly optimistic, especially in regard to estimation for $g(t)$, because the truth is used as the pilot "estimate" $\tilde{g}(t)$ in the first layer of this tuning parameters selection procedure. In the upcoming subsection, we take a more practical route and use the two-stage estimate $\hat{g}_{CKT}(t)$ as $\tilde{g}(t)$ in this tuning parameters selection procedure. From there, one can see similar patterns in the one-stage estimates observed in Figures 3.9 and 3.10.
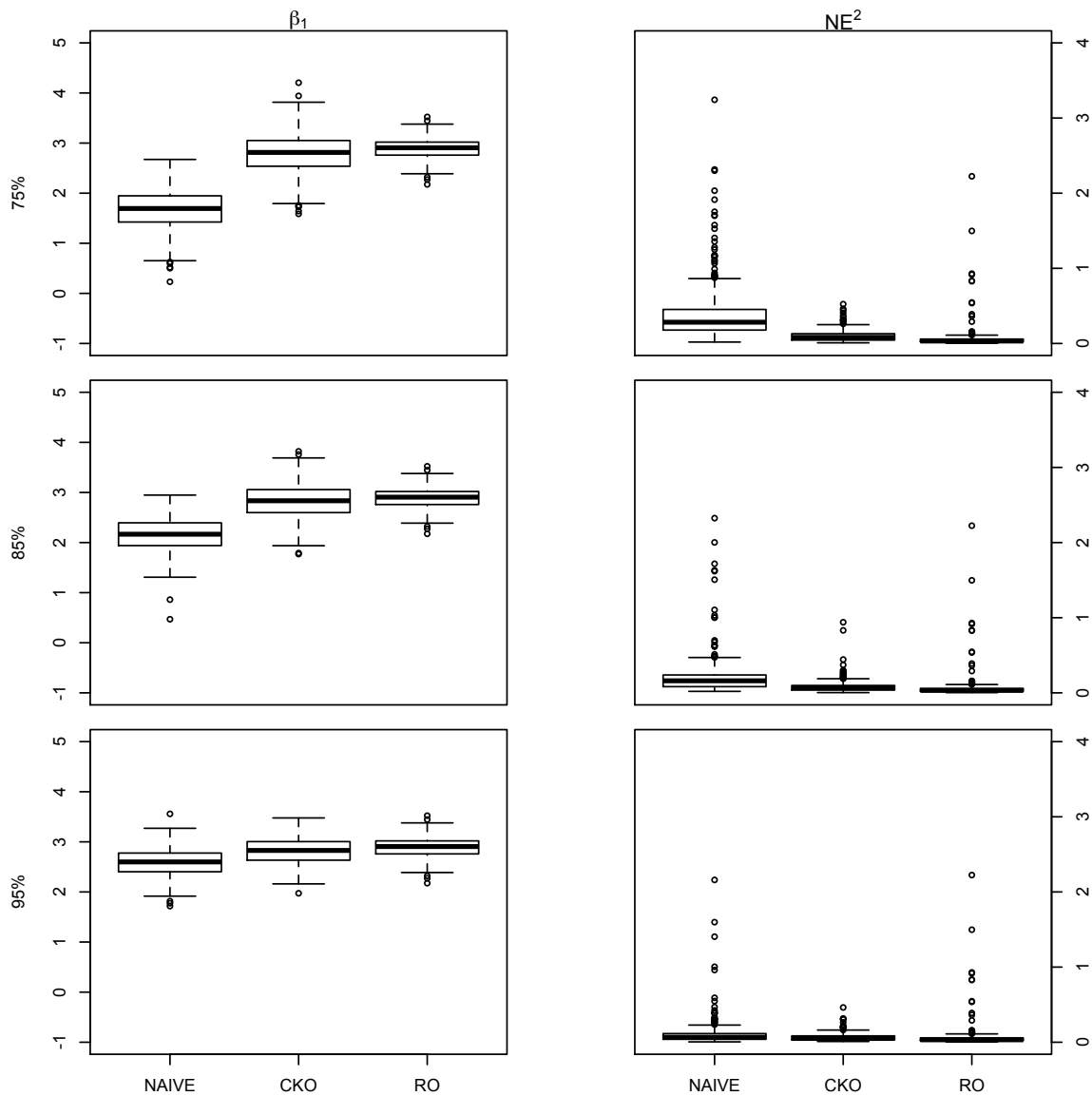
Figure 3.9: Two-layer tuning parameter selection. Under the simulation setting (F1) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.

65
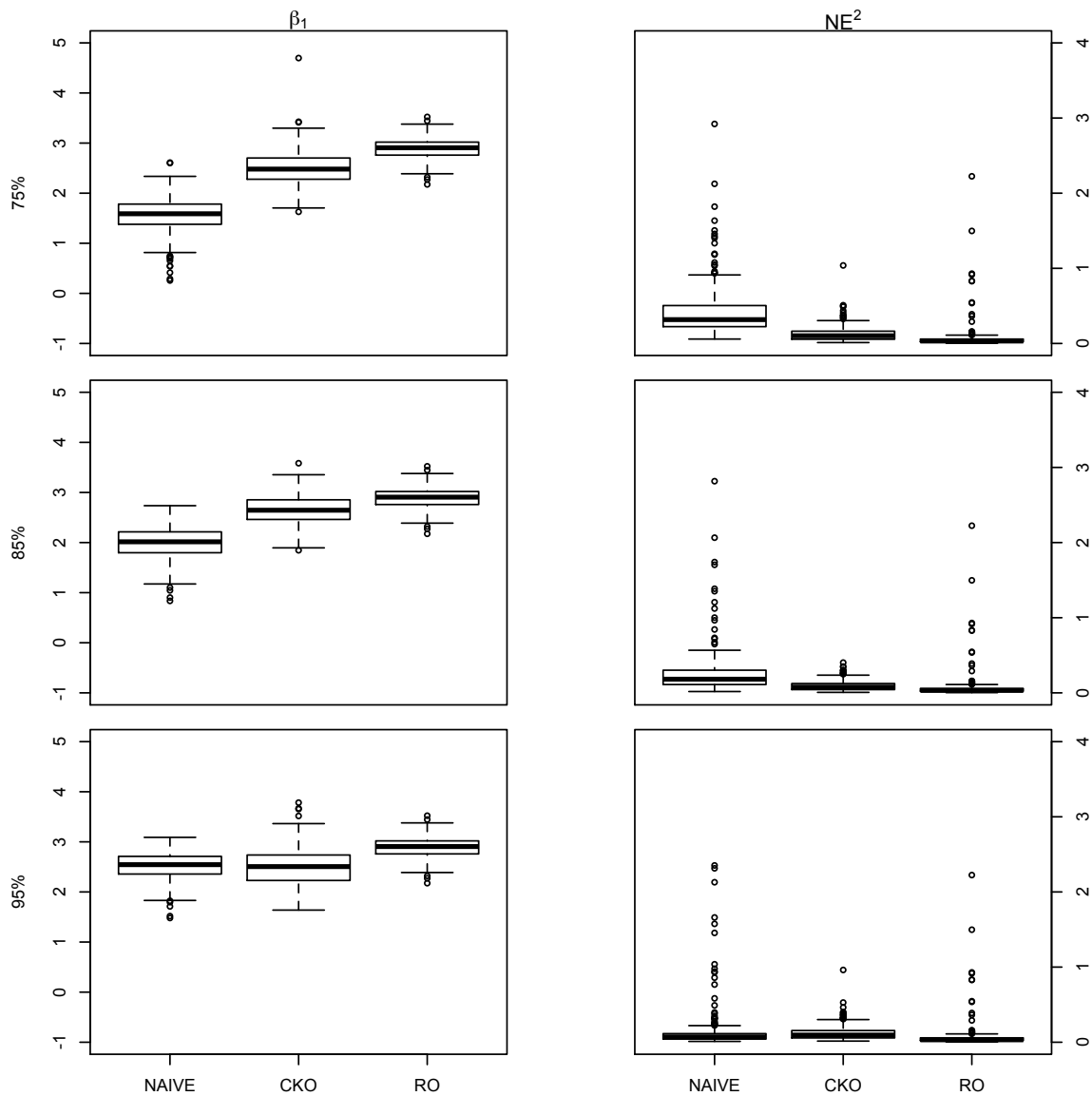
Figure 3.10: Two-layer tuning parameter selection. Under the simulation setting (F1) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage method (RO) in the absence of measurement error, respectively.
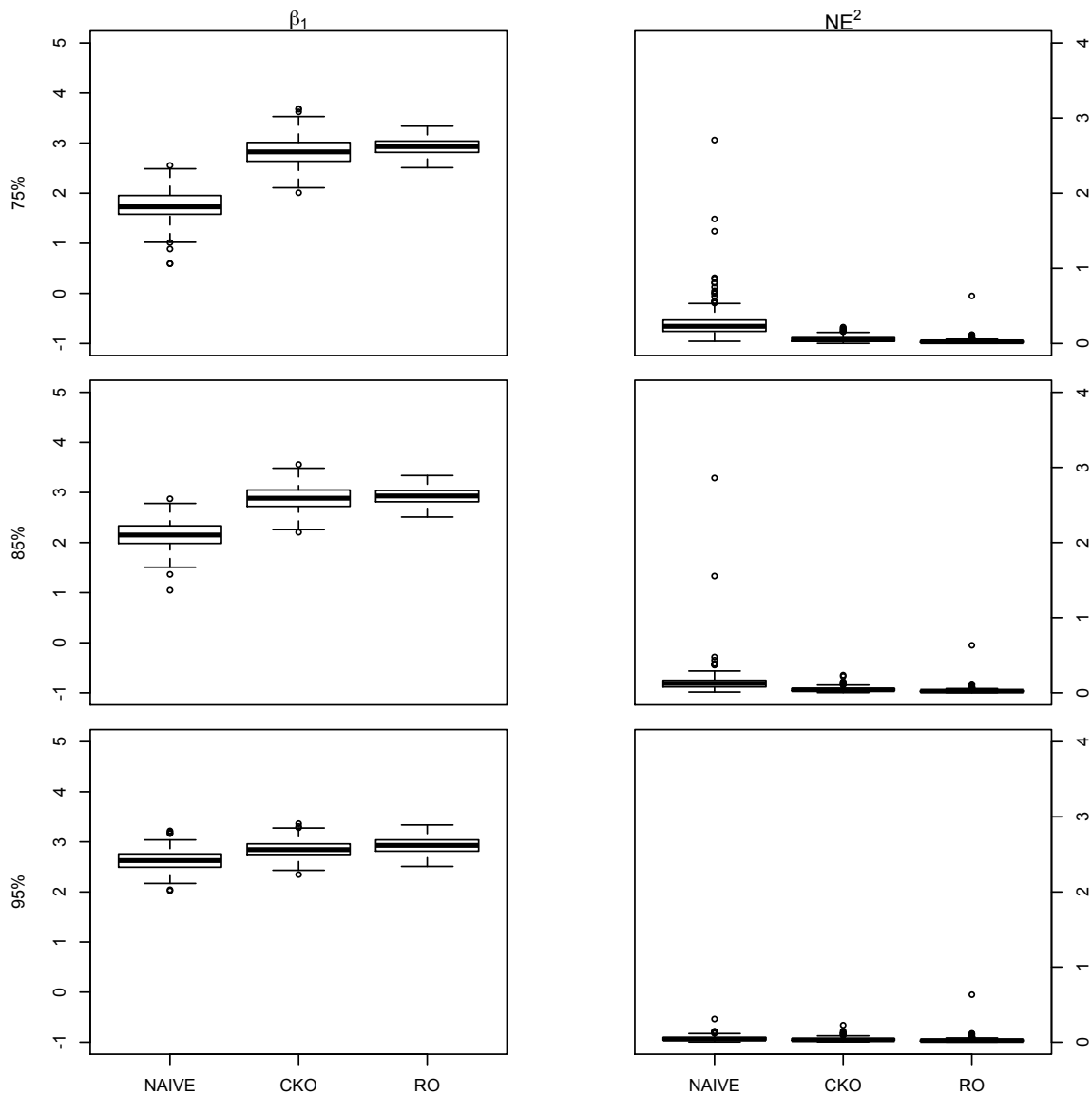
Figure 3.11: Two-layer tuning parameter selection. Under the simulation setting (F1) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
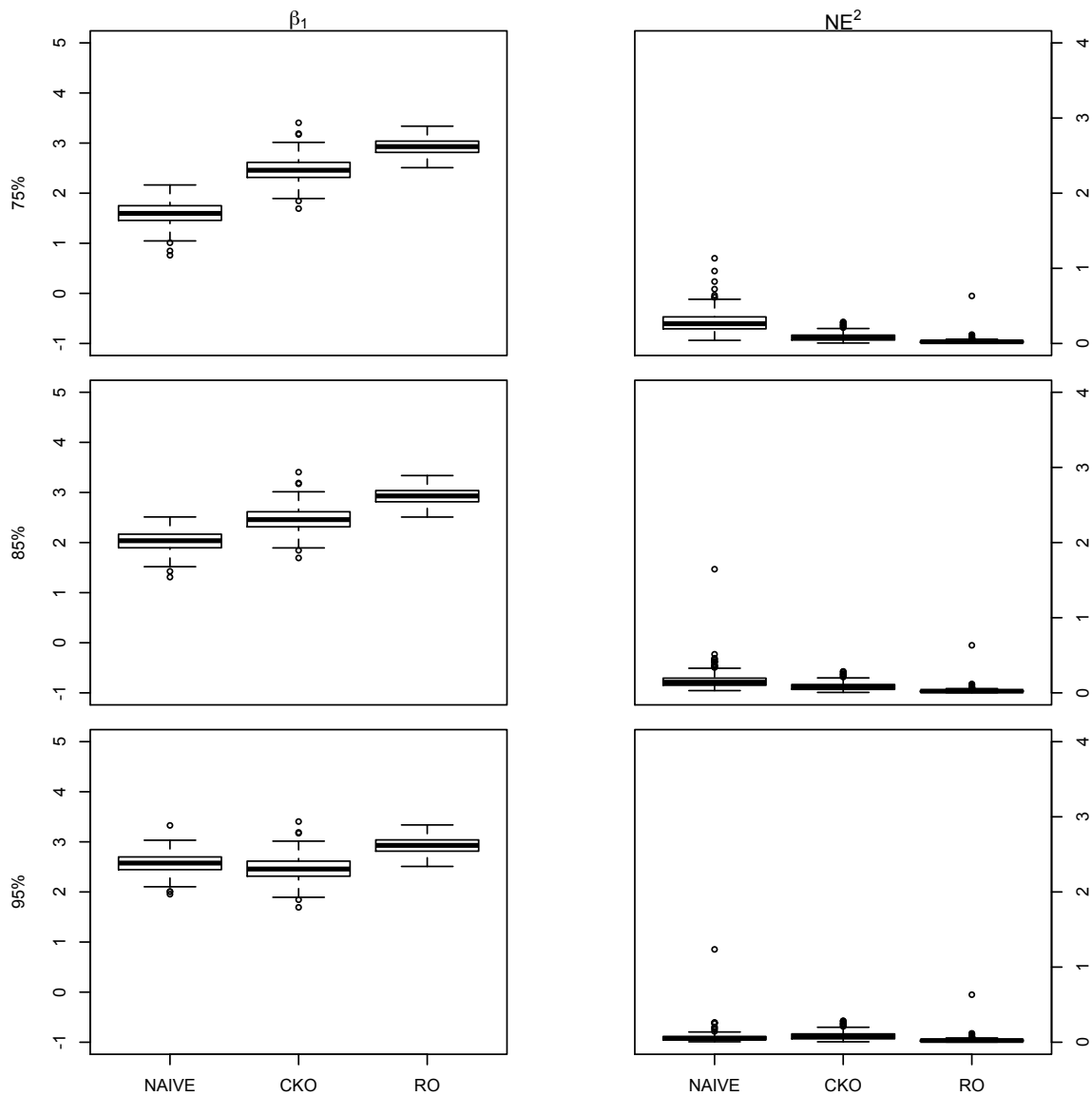
Figure 3.12: Two-layer tuning parameter selection. Under the simulation setting (F1) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage method (RO) in the absence of measurement error, respectively.

68

Table 3.5: One-stage estimation method with two-layer tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 200$ under (F1). Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
			$U \sim N(0, \sigma^2)$			
Naive	1.56	0.41	2.00	0.25	2.53	0.13
	(0.20)	(0.19)	(0.18)	(0.17)	(0.15)	(0.16)
CK	2.51	0.12	2.67	0.10	2.80	0.06
	(0.21)	(0.06)	(0.17)	(0.04)	(0.14)	(0.03)
			$U \sim \text{Laplace}(0, \sigma^2)$			
Naive	1.67	0.40	2.15	0.22	2.59	0.12
	(0.24)	(0.23)	(0.21)	(0.16)	(0.16)	(0.12)
CK	2.80	0.10	2.83	0.08	2.82	0.07
	(0.24)	(0.05)	(0.19)	(0.05)	(0.15)	(0.03)
TRUE	2.89	0.07				
	(0.12)	(0.11)				

Table 3.6: One-stage estimation method with two-layer tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 400$ under (F1). Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
			$U \sim N(0, \sigma^2)$			
Naive$_N$	1.60	0.29	2.03	0.16	2.57	0.06
	(0.13)	(0.08)	(0.12)	(0.07)	(0.11)	(0.05)
CK$_N$	2.47	0.09	2.66	0.05	2.81	0.04
	(0.15)	(0.03)	(0.12)	(0.02)	(0.11)	(0.01)
			$U \sim \text{Laplace}(0, \sigma^2)$			
Naive	1.76	0.27	2.16	0.14	2.63	0.05
	(0.18)	(0.13)	(0.15)	(0.11)	(0.12)	(0.02)
CK	2.83	0.06	2.89	0.05	2.86	0.04
	(0.17)	(0.02)	(0.14)	(0.02)	(0.10)	(0.02)
TRUE	2.92	0.03				
	(0.10)	(0.02)				

SIMULATION STUDY FOR THE TWO-STAGE ESTIMATION METHOD

In experiments where we monitor finite sample performance of the two-stage estimation method, we adopt all simulation settings described in Section 3.5 except that T is independent of X in order to satisfy assumption (a3). For completeness, the two model configurations considered in the two-stage simulation experiment are stated below.

(F3) $Y = 2\sin(2\pi t) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$,

69

$X \sim \text{uniform}(-1,1)$, $T \sim \text{uniform}(0,1)$, $U \sim M(0, \sigma_u^2)$, $X \perp T$.

(F4) $Y = \exp(\sin(\pi t)) + X + (1 + 2X)\epsilon$, where $\epsilon \sim 0.5N(-1, 2, 5^2) + 0.5N(1, 0.5^2)$,

$X \sim \text{uniform}(-1,1)$, $T \sim \text{uniform}(0,1)$, $U \sim M(0, \sigma_u^2)$, $X \perp T$.

In the above two settings, the distribution of M represents normal and Laplace distribution with mean 0 and variance σ_u^2. By using the two-stage corrected kernel estimation method in the absence of measurement error in Section 2.3, it is easy to formulate a counterpart two-stage estimation method when data $\{Y, T, X\}_{j=1}^n$ are available, and also a naive two-stage method based on error-prone data $\{Y, T, W\}_{j=1}^n$. Denote by $(\hat{\beta}_{\text{RT},1}, \hat{g}_{\text{RT}}(\cdot))$ the error-free counterpart of our proposed two-stage estimates, $(\hat{\beta}_{\text{CKT},1}, \hat{g}_{\text{CKT}}(\cdot))$, and by $(\hat{\beta}_{\text{NVT},1}, \hat{g}_{\text{NVT}}(\cdot))$ the naive estimates.

Figures 3.13 and 3.14 depict on the left panels, when $U \sim \text{Laplace}(0, \sigma_u^2)$ and $U \sim N(0, \sigma_u^2)$, respectively, boxplots of $(\hat{\beta}_{\text{NVT},1}, \hat{\beta}_{\text{CKT},1}, \hat{\beta}_{\text{RT},1})$ from 300 MC replicate data sets, with (F3) being the mode regression model for Y. Like what are observed for the one-stage estimation methods, the proposed two-stage estimation method yields more reliable estimator for the covariate effect in the linear part of the mode regression model than that from the naive two-stage method. It is worth pointing out that, because $g(t) = \beta_0 + g^*(t)$ in the context of two-stage estimation, where $g^*(t)$ is estimated in the first stage that is free of measurement error complication, we have the same $\hat{g}^*(t)$ for the proposed method as those from its naive counterpart and error-free counterpart. And thus $\hat{g}_{\text{CKT}}^*(t)$ differs from $\hat{g}_{\text{NVT}}^*(t)$ and $\hat{g}_{\text{RT}}^*(t)$ only due to the differences in estimates for β_0. The right panels in Figures 3.13 and 3.14 are boxplots of the empirical squared errors associated with $(\hat{g}_{\text{NVT}}(\cdot), \hat{g}_{\text{CKT}}(\cdot), \hat{g}_{\text{RT}}(\cdot))$, which indicate improved estimates for β_0 from the proposed two-stage estimation method compared to the naive method. Recall that we only have the two-dimensional CV procedure proposed for the two-stage estimation method to choose (h, k). Table 3.7 shows the numerical result of the two-stage estimation method across 300 repetitions

for Laplace measurement error and normal measurement error under (F3), respectively, when $n = 200$. The same story as in Figures 3.13 and 3.14 can be found in Table 3.7.

**Table 3.7:** Averages of parameter estimates from the two-stage method over 300 repetitions under (F3) when $n = 200$. Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truth is $\beta_0 = 1$, $\beta_1 = 3$.

	75%			85%			95%		
	β_0	β_1	NE^2	β_0	β_1	NE^2	β_0	β_1	NE^2
				$U \sim N(0, \sigma^2)$					
Naive	0.52	1.72	0.30	0.68	2.08	0.30	0.82	2.50	0.28
	(0.19)	(0.27)	(0.18)	(0.11)	(0.25)	(0.12)	(0.17)	(0.23)	(0.12)
CK	0.79	2.39	0.26	0.87	2.59	0.27	0.88	2.76	0.24
	(0.18)	(0.27)	(0.10)	(0.16)	(0.22)	(0.11)	(0.14)	(0.17)	(0.10)
				$U \sim \text{Laplace}(0, \sigma^2)$					
Naive	0.64	1.95	0.28	0.71	2.19	0.29	0.83	2.58	0.25
	(0.18)	(0.28)	(0.11)	(0.18)	(0.25)	(0.12)	(0.16)	(0.22)	(0.10)
CK	0.90	2.77	0.27	0.91	2.81	0.26	0.91	2.78	0.25
	(0.18)	(0.28)	(0.10)	(0.16)	(0.23)	(0.11)	(0.14)	(0.17)	(0.11)
TRUE	0.94	2.79	0.27						
	(0.15)	(0.17)	(0.12)						

As the sample size n increases to 400, Figures 3.15 depicts the boxplots of the estimates $(\hat{\beta}_{\text{CKT},1}, NE^2_{\text{CKT}})$ in the presence of Laplace measurement error. Figure 3.16 shows, when $n = 400$, the estimates $(\hat{\beta}_{\text{CKT},1}, NE^2_{\text{CKT}})$ in the presence of normal measurement error. When comparing with Figures 3.13 and 3.14, the bias and variance in Figures 3.15 and 3.16 are smaller. Table 3.8 shows summary statistics for results from the two-stage estimation method across 300 repetitions for Laplace measurement error and normal measurement error under (F3), respectively, when $n = 400$. It tells the same story as in Figures 3.15 and 3.16. For completeness, we provide in Appendix E figures and tables like Figures 3.13 and 3.14, Tables 3.7 and 3.8 for results when the responses are generated according to (F4).

Finally, we compare our proposed non-naive estimation methods with each other, including the corrected kernel one-stage estimation (CKO) method coupled with the two-dimensional CV procedure for choosing tuning parameters, the same estimation method (CKO) with tuning parameters chosen via the two-layer procedure, and the corrected kernel two-stage estimation (CKT) method. In the two-layer tuning pa-
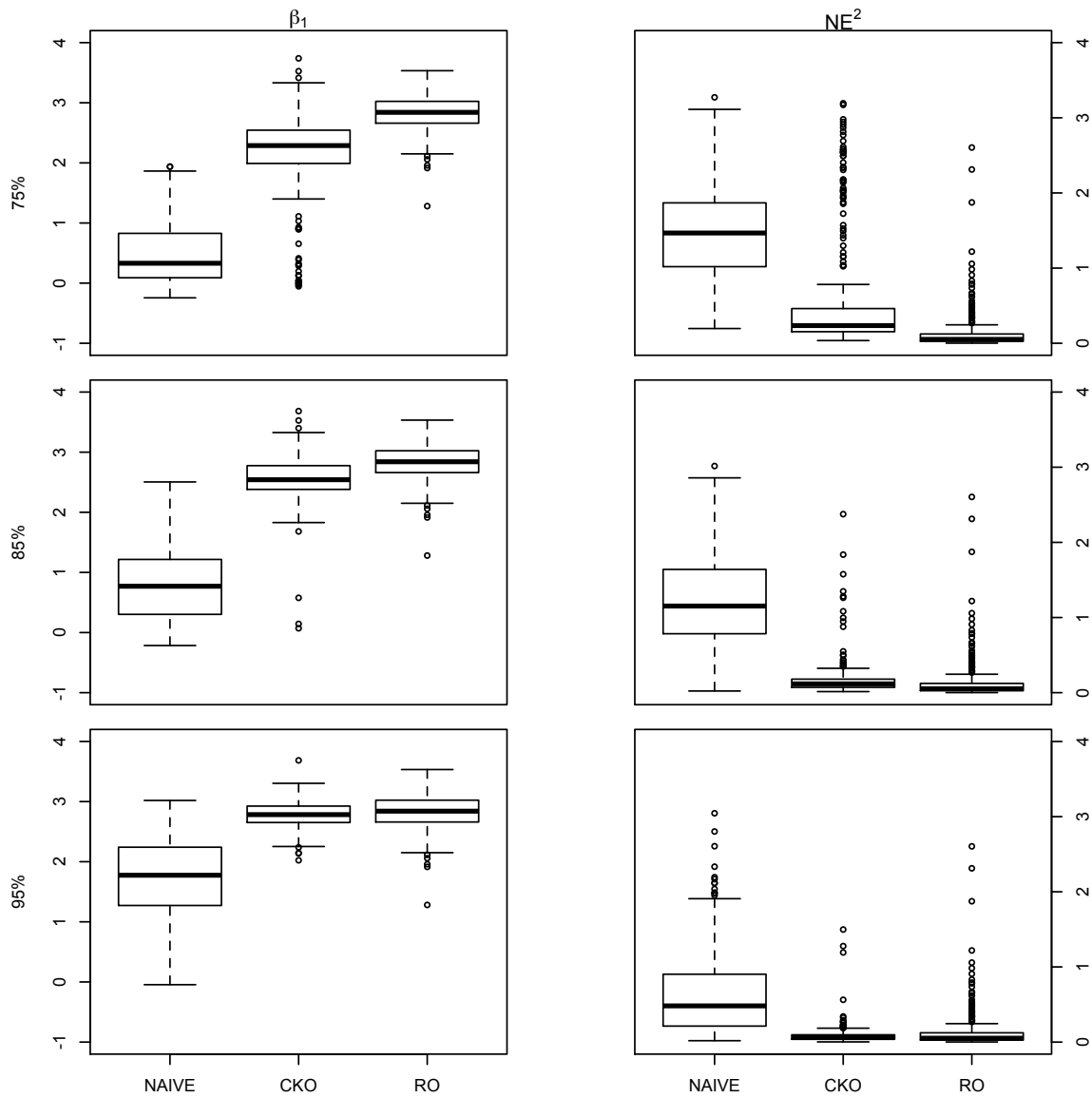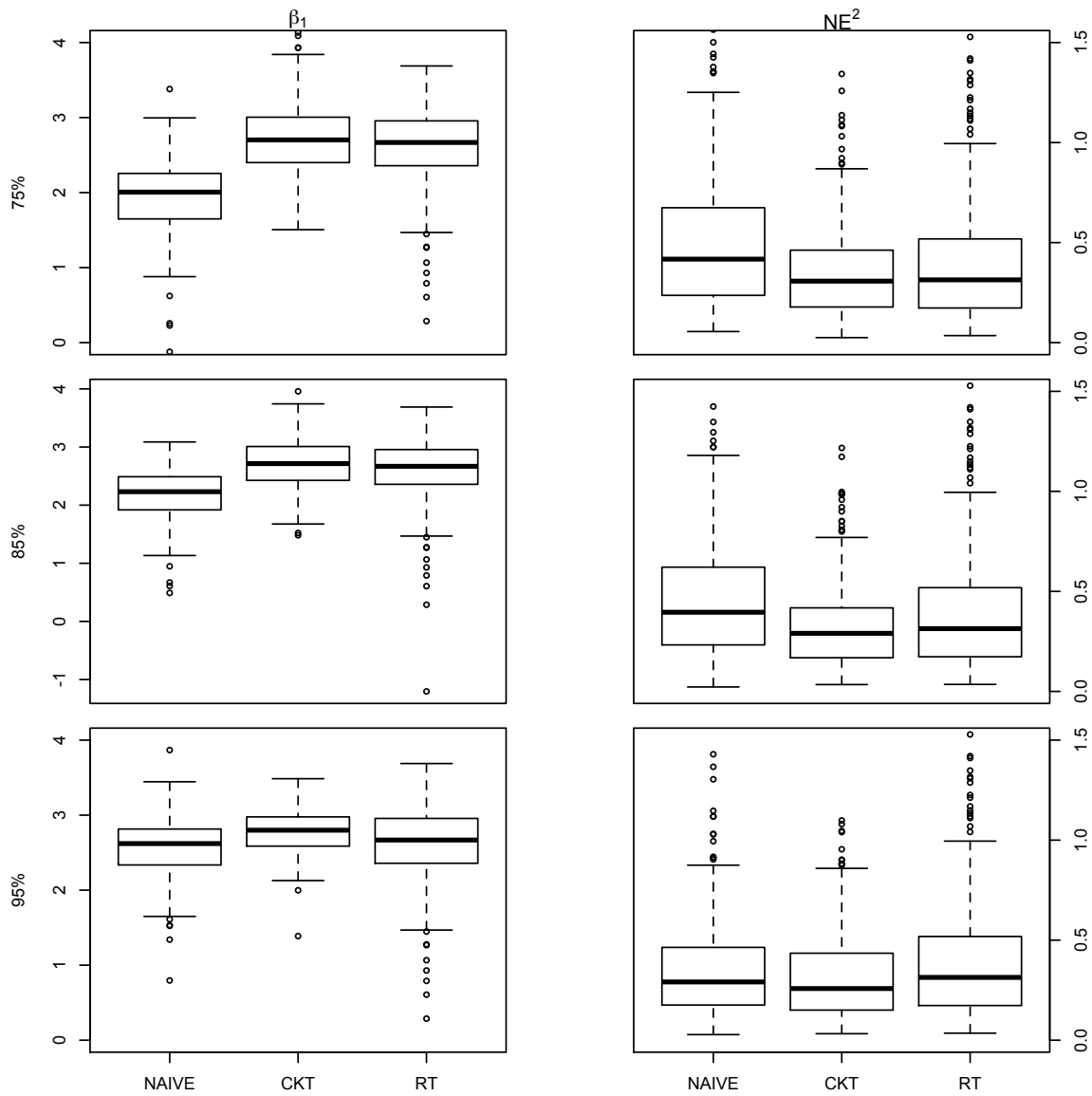
Figure 3.13: Two dimensional cross validation tuning parameter selection. Under the simulation setting (F3) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage (RT) in the absence of measurement error, respectively.
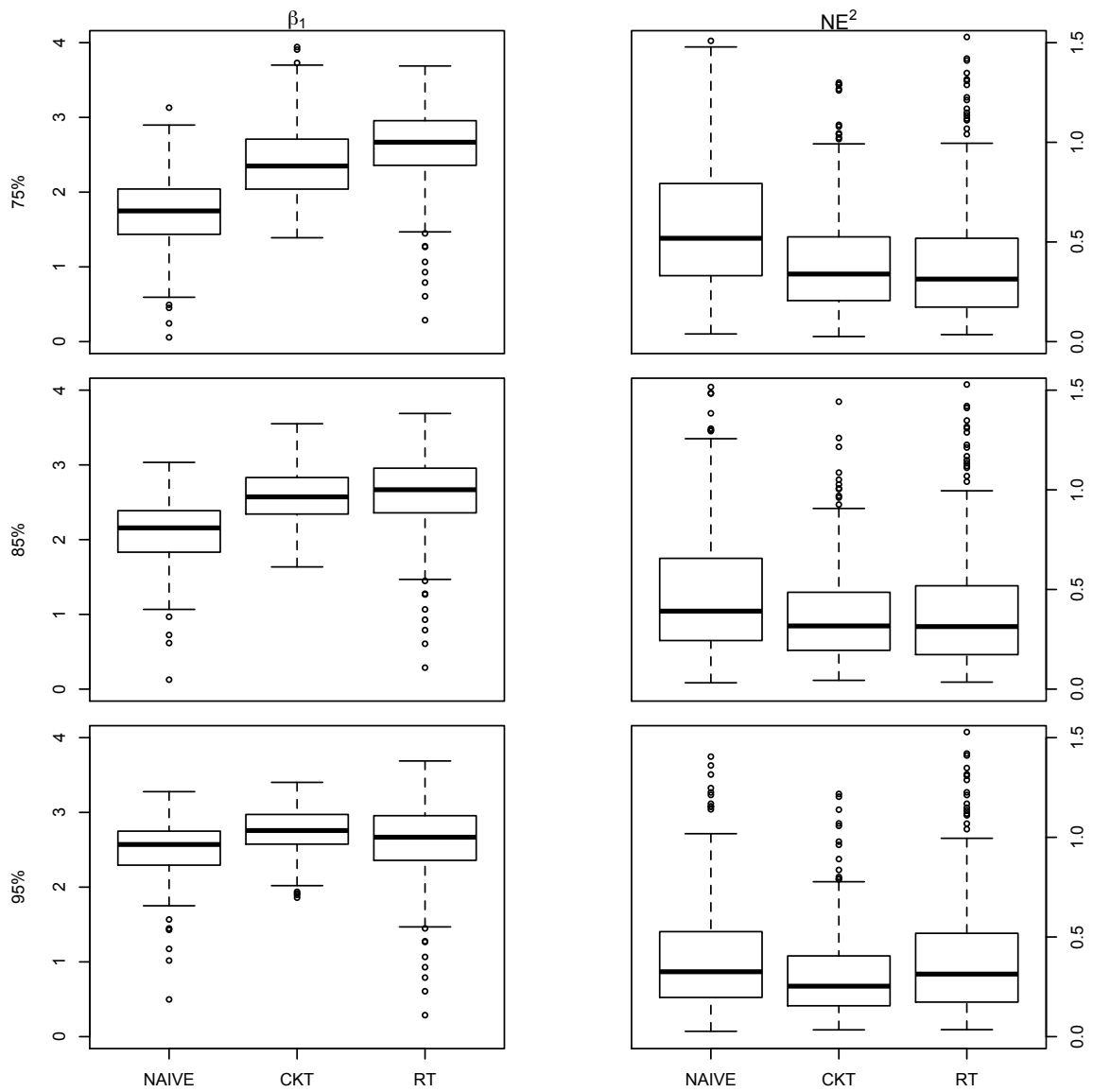
Figure 3.14: Two dimensional cross validation. Under the simulation setting (F3) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage method (RT) in the absence of measurement error, respectively.
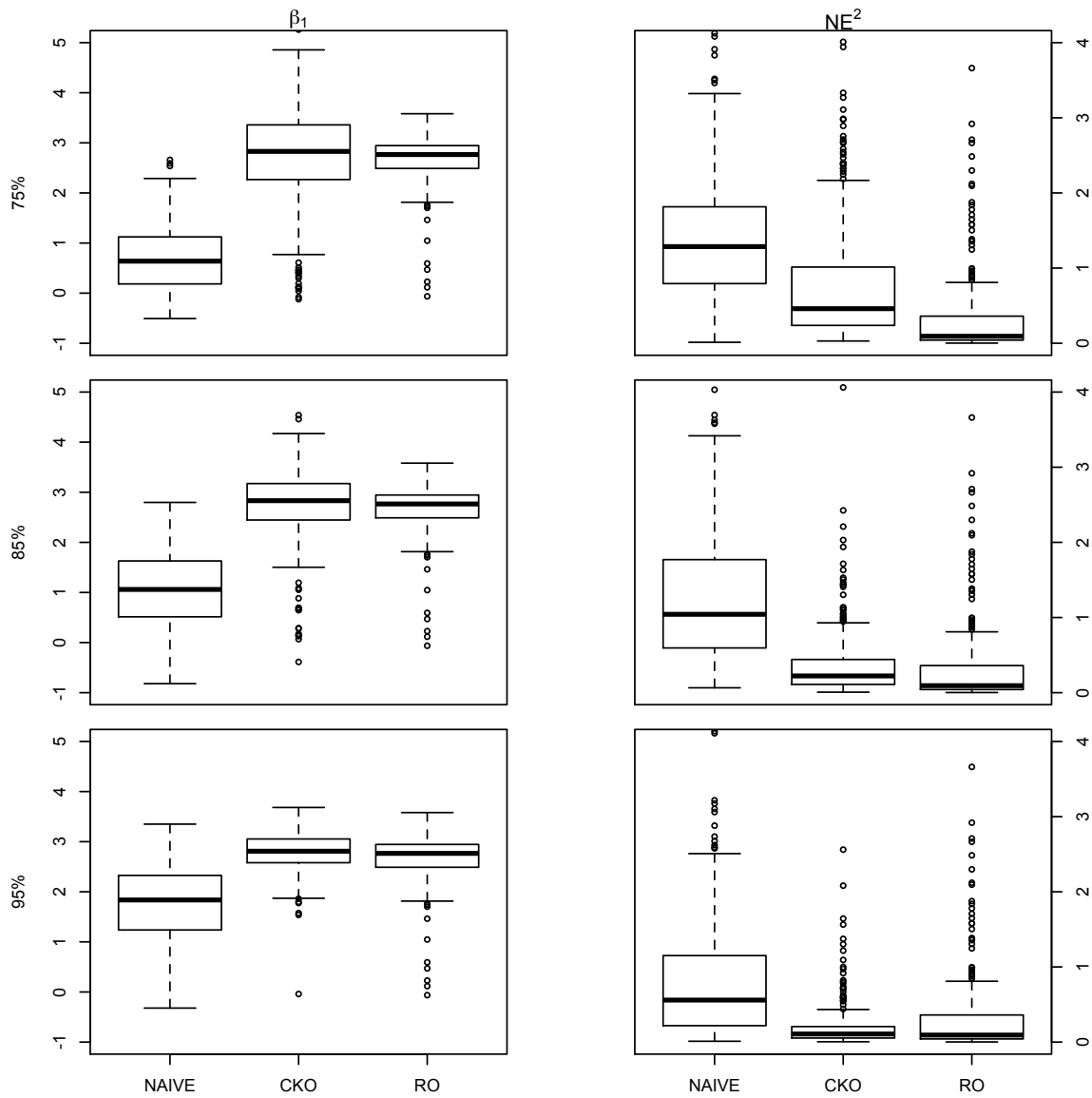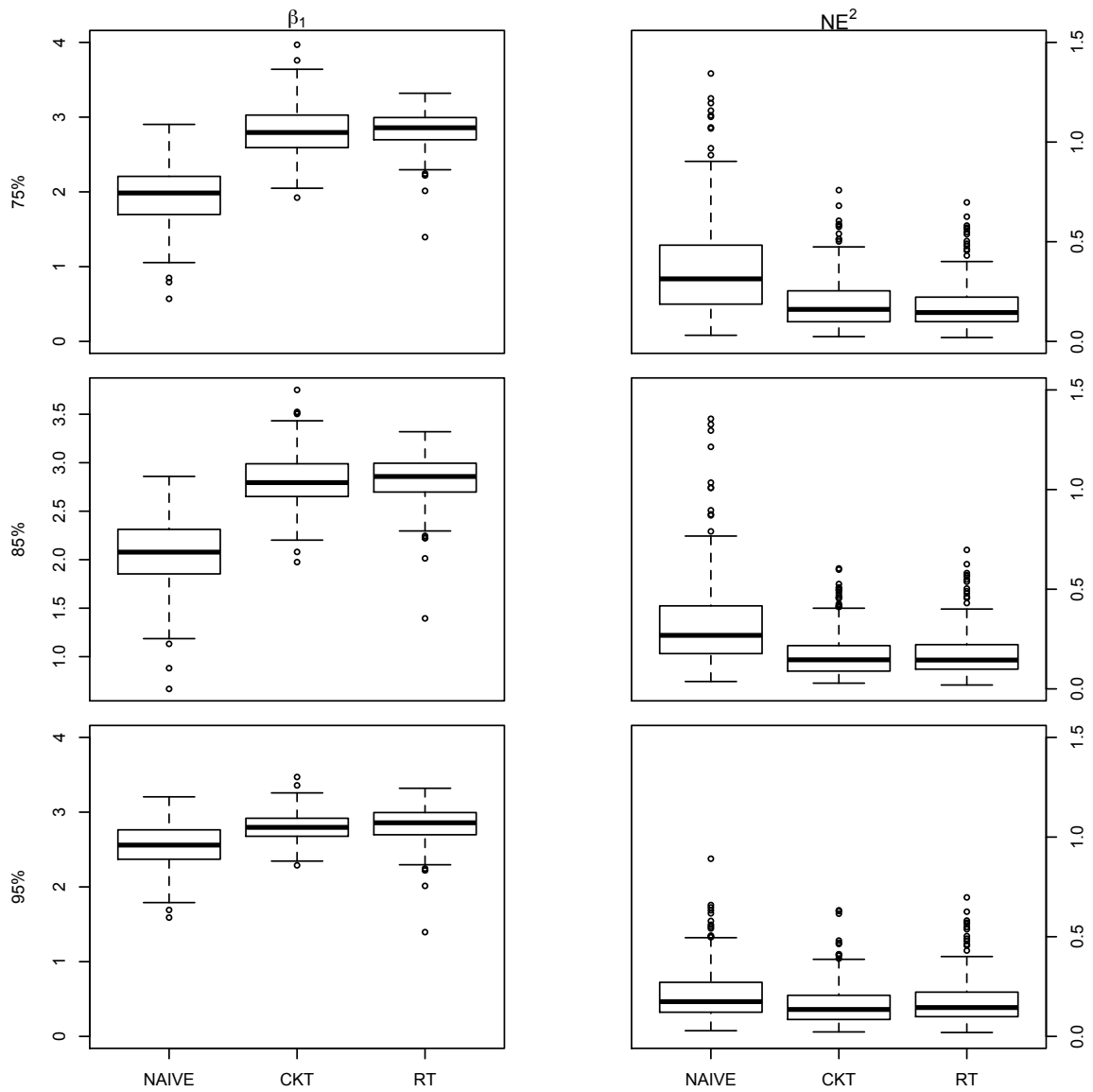
Figure 3.15: Two dimensional cross validation tuning parameter selection. Under the simulation setting (F3) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage (RT) in the absence of measurement error, respectively.
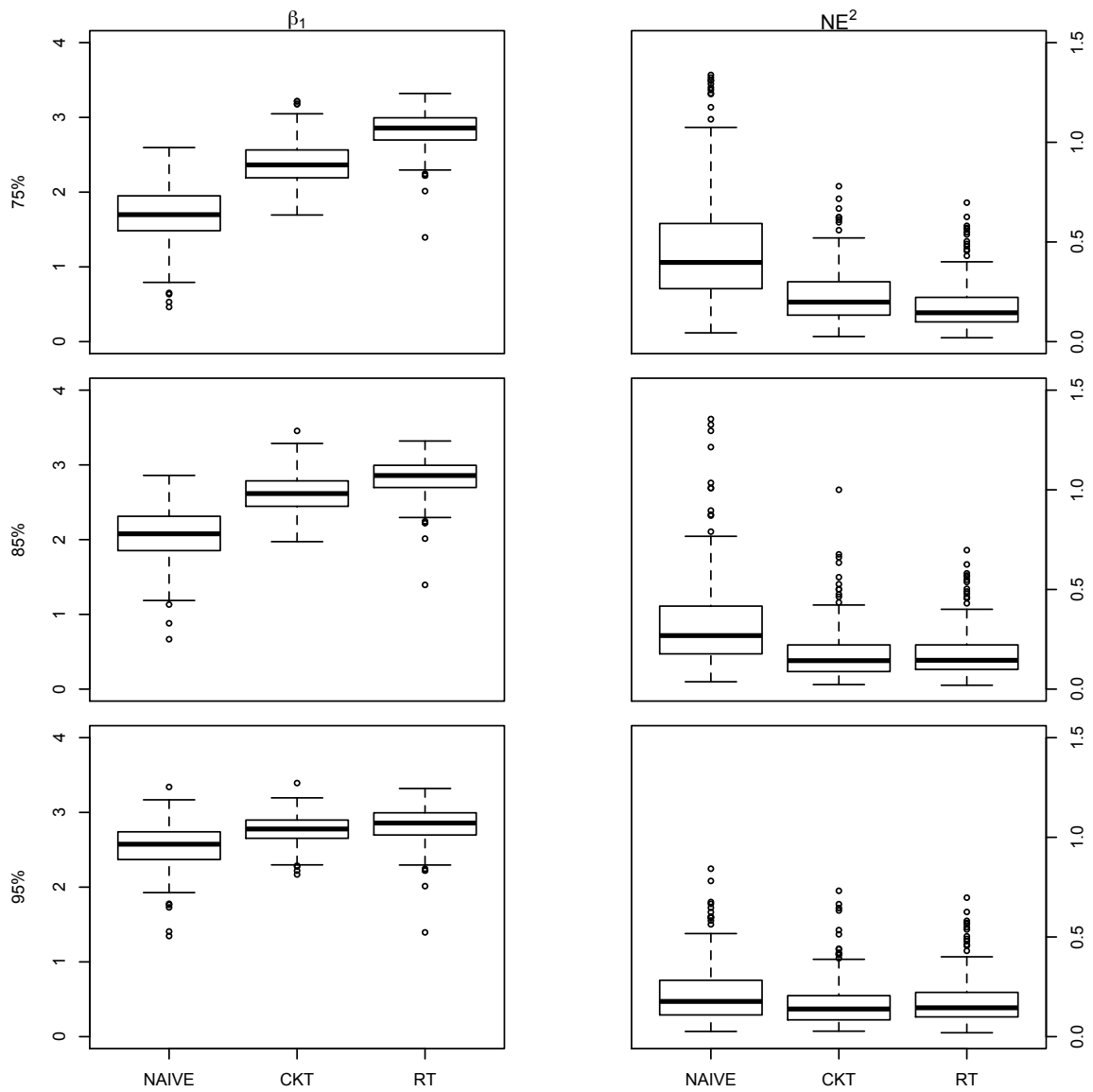
Figure 3.16: Two dimensional cross validation. Under the simulation setting (F3) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage method (RT) in the absence of measurement error, respectively.

75

Table 3.8: Averages of parameter estimates from the two-stage estimation method over 300 repetitions under (F3) when $n = 400$. Numbers in parentheses are ($10 \times$ standard errors) associate with the averages. The truth is $\beta_0 = 1$, $\beta_1 = 3$.

	75%			85%			95%		
	β_0	β_1	NE^2	β_0	β_1	NE^2	β_0	β_1	NE^2
				$U \sim N(0,\sigma^2)$					
Naive	0.52	1.72	0.17	0.66	2.07	0.16	0.83	2.54	0.14
	(0.15)	(0.21)	(0.06)	(0.14)	(0.19)	(0.06)	(0.12)	(0.17)	(0.05)
CK	0.79	2.39	0.14	0.87	2.63	0.12	0.90	2.78	0.12
	(0.12)	(0.17)	(0.05)	(0.11)	(0.15)	(0.05)	(0.10)	(0.10)	(0.05)
				$U \sim \text{Laplace}(0,\sigma^2)$					
Naive	0.59	1.95	0.14	0.66	2.07	0.16	0.82	2.55	0.14
	(0.14)	(0.24)	(0.05)	(0.14)	(0.20)	(0.06)	(0.12)	(0.17)	(0.05)
CK	0.92	2.81	0.13	0.93	2.82	0.14	0.92	2.81	0.12
	(0.13)	(0.20)	(0.04)	(0.11)	(0.15)	(0.05)	(0.10)	(0.10)	(0.05)
TRUE	0.95	2.83	0.14						
	(0.11)	(0.14)	(0.05)						

rameters selection procedure, the estimated function from the two-stage estimation method, $\hat{g}_{\text{CKT}}(t)$, is used as the pilot estimate $\tilde{g}(t)$. Figures 3.17 and 3.18 show such comparison with Laplace measurement error and normally distributed measurement error, respectively, when (F1) is the mean model for the response. When the tuning parameters are chosen via the two-layer procedure, the one-stage estimation method yields similar estimates for β_1 as those from the two-stage estimation method. This may not be surprising because $\hat{g}_{\text{CKT}}(t)$ is treated as a reference for estimating $g(t)$ when choosing the bandwidth h in the first layer of the two-layer tuning parameters selection procedure. This strategy also leads to less variable $\hat{g}_{\text{CKO}}(t)$ compared to when the two-dimensional CV procedure is used to choose (h,k). Similar patterns of the comparison between these three approaches are observed when the response is generated according to (F2).

3.6 REAL DATA ANALYSIS

In this section, the proposed methods is applied to the Framingham data (`https://www.stat.tamu.edu/~carroll/data.php`). The data are from $n = 1615$ males. We focus on studying one's impact of the serum cholesterol (X) and age (T) on his average systolic blood pressure in a fixed two-year period (Y). Since the serum

76
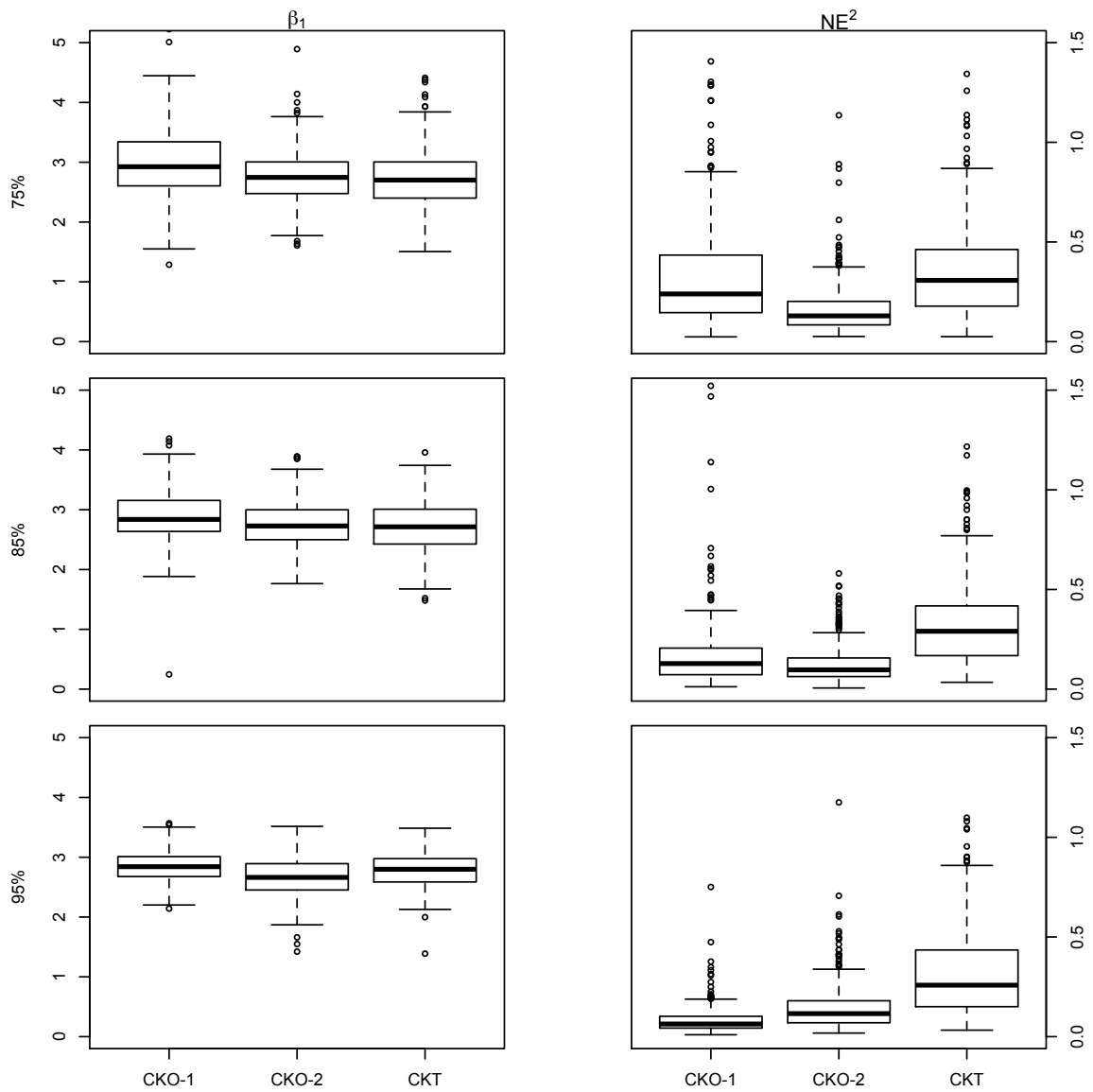
Figure 3.17: Boxplots of estimates for β_1 (on the left panels) and NE^2 (on the right panels) three non-naive approaches: the one-stage estimation method paired with the two-dimensional CV tuning parameters selection (CKO-1), the one-stage estimation method paired with the two-layer tuning parameters selection (CKO-2), and the two-stage estimation method (CKT). Responses are generated according to (F3) and $U \sim \text{Laplace}(0, \sigma_u^2)$, with $\lambda = 0.75$ (top row), 0.85 (middle row), and 0.95 (bottom row).
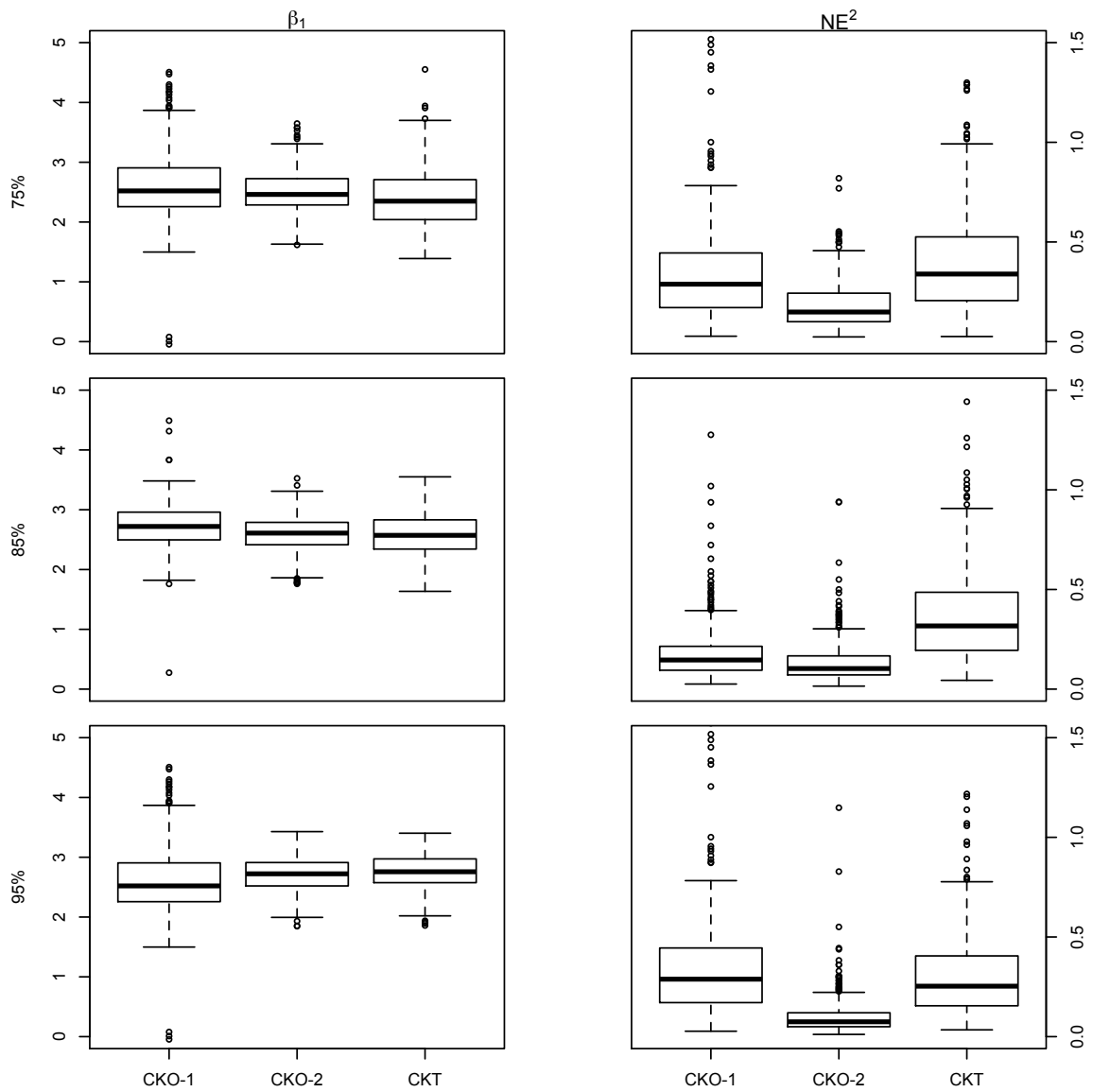
Figure 3.18: Boxplots of estimates for β_1 (on the left panels) and NE^2 (on the right panels) three non-naive approaches: the one-stage estimation method paired with the two-dimensional CV tuning parameters selection (CKO-1), the one-stage estimation method paired with the two-layer tuning parameters selection (CKO-2), and the two-stage estimation method (CKT). Responses are generated according to (F3) and $U \sim N(0, \sigma_u^2)$, with $\lambda = 0.75$ (top row), 0.85 (middle row), and 0.95 (bottom row).

78

cholesterol usually is measured with error, we used the average of the cholesterols as a surrogate (W) of the subject's serum cholesterol. The histogram in Figure 3.19 represents an underlying skewed distribution for the response.
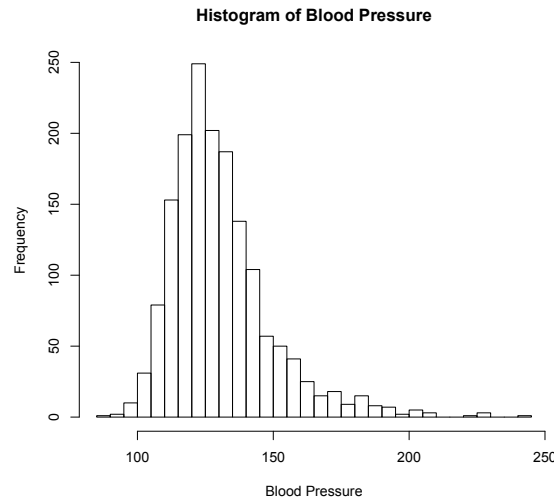
Figure 3.19: The histogram of systolic blood pressure for the Framingham data.

For illustration purpose, we consider a partially linear mode regression model for the mode of Y_j given X_j and T_j, where X_j is not observed but its error-contaminated surrogate W_j is, where $W_j = \sum_{k=1}^{2} W_{j,k}/2$, in which $W_{j,k}$ is subject $j's$ kth serum cholesterol, for $k = 1, 2$ and $j = 1, \cdots, 1615$. Based on the two replicate measures for each underlying X_j, the variance of measurement errors associated with W_j is estimated via one half of $\sum_{j=1}^{n} \sum_{k=1}^{2} (W_{j,k} - W_j)^2/n$. This gives an estimate of the measurement error variance as $\hat{\sigma}^2 = 0.13$, and the corresponding estimated reliability ratio being 0.87.

We carry out the partially linear regression analysis using the naive method of one-stage estimation method and two-stage estimation method, and apply the corrected kernel methods, respectively. When implementing the one-stage estimation method, we choose the tuning parameters using the two-dimensional cross validation method and the two-layer method, respectively. When implementing the two-stage

79

estimation method, we employ the two dimensional cross validation method to select the tuning parameters. Figure 3.20 shows the estimate of $g(T)$ versus T resulting from the two-stage estimation method and the one-stage method with tuning parameters chosen by the two-layer tuning parameter selection, respectively. Figure 3.21 depicts the estimate of $g(T)$ versus T based on the one-stage estimation method with two dimensional cross validation tuning parameter selection and two-layer tuning parameter selection, respectively. In the two-layer tuning parameter selection method, the pilot estimate for $g(T)$ need in the first step is the estimated function from the two-stage estimation method.
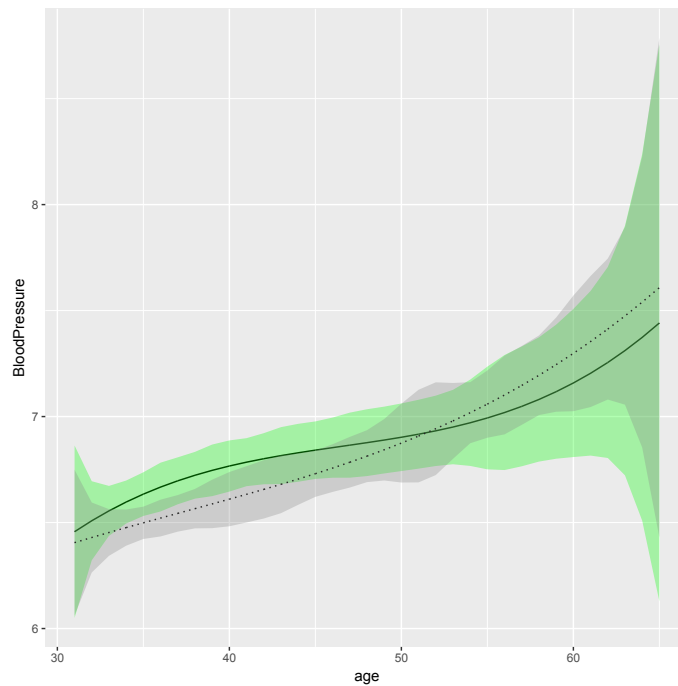
Figure 3.20: Two estimated $g(T)$ from two proposed estimation methods accounting for measurement error. The black solid line represents the estimated curve by using the one-stage corrected kernel method combined with two-layer tuning parameter selection. Under the same estimation method and tuning parameter selection, the green band represents the 95% confidence interval of $\hat{g}(T)$ obtained by bootstrap method. The black dash line represents the estimated curve resulting from the two-stage estimation method. Under the same estimation method and tuning parameter selection, the grey band represents the 95% confidence interval of $\hat{g}(T)$.
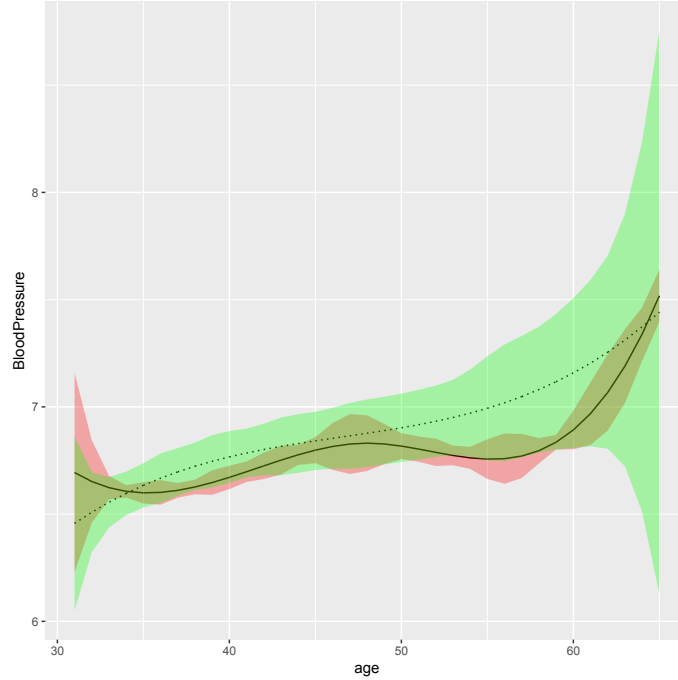
Figure 3.21: Two estimated $g(T)$ from two proposed estimation methods accounting for measurement error. The black dash line represents the estimated curve by using the one-stage corrected kernel method combined with two-layer tuning parameter selection. Under the same estimation method and tuning parameter selection, the green band represents the 95% confidence interval of $\hat{g}(T)$ obtained by a bootstrap method. The black solid line represents the estimated curve result from the one-stage estimation method with two dimensional cross validation method. Under the same estimation method and tuning parameter selection, the red band represents the 95% confidence interval of $\hat{g}(T)$.

Table 3.9 shows the estimates of the covariate effect associated with cholesterols obtained from the naive one-stage estimation method and two-stage estimation method, and estimates from the corrected kernel methods, respectively. These results suggest that both proposed methods produce estimates of the covariate effect that imply a stronger association between the serum cholesterol and the systolic blood pressure than the naive methods. In particular, compared the the one-stage naive method with two dimensional cross validation tuning parameter selection, the estimated covariate effect from the one-stage corrected kernel method with two dimensional cross validation tuning parameter selection increases by 6.5%. Compared

81

with the covariate effect estimate from the one-stage naive method with two-layer tuning parameter selection, the estimate from the one-stage corrected kernel with two-layer tuning parameter selection method increases by 0.86%. Compared with the estimate from the two-stage naive method, the estimate for the covariate effect from the two-stage corrected kernel method increases by 21.0%.

Table 3.9: Regression coefficient estimates resulting from the one-stage naive method and the one-stage corrected kernel method with two dimensional cross validation tuning parameter selection method are under One-CV, the one-stage naive method and the one-stage corrected kernel method with two-layer tuning parameter selection method are under One-ISE-CV, the two-stage naive method and the two-stage corrected kernel method are underTwo-CV. Numbers in parentheses are ($10 \times$ standard errors) associate with the estimates

	One-CV	One-ISE-CV	Two-CV
Naive	0.093(0.044)	0.117(0.014)	0.105(0.020)
CK	0.099(0.006)	0.118(0.018)	0.127(0.022)

3.7 DISCUSSION

We have proposed two methods in this chapter to take care of the case that the parametric part X of the model has measurement error and the nonparametric part T is error-free. An interesting research area is to consider partially linear model models with error-prone T. Furthermore, there is only one nonparametric part T added in the model setting. In practice, a couple of more nonparametric parts can be included in the model. Consequently, another research problem, which focuses on developing the methodology to estimate the covariate effect in the presence of many nonparametric parts, is raised up and remains to be an open problem.

In Chapters 1 and this chapter, mode regression problems involving error-prone covariates are tackled. In addition to unobserved covariates, in a regression model, the response variable sometimes may not be observed either and the estimation based on this type of data is also worth discussing. A practical scenario that gives rise to this complication with unobservable responses is the so-called group testing (Dorfman,

1943). In the next chapter, local polynomial estimation in mean regression models for group testing data with error-free covariates will be discussed.

Chapter 4

Local Polynomial Mean Regression Using

Pooled Responses

Group testing, also known as pooled testing, originally is proposed by Dorfman (1943) to detect sypilis in US soldiers during World War II. It has received increasing attention in disease screening (Gastwirth and Hammick, 1989; Dhand et al., 2010), drug discovery (Remlinger et al., 2006), and genetics (Chi et al., 2009) in the past few decades. In a group testing design, instead of testing the presence of a disease individually, the test is conducted on a pool of individuals. This pooling test has been recognized as a cost-effective strategy to perform large-scale screening for rare infectious diseases. Besides being cost effective, pooling test is also efficient when it comes to measuring assays with limits of detection and can reduce the impact of outliers in the sample (Schisterman et al., 2011). Because of these benefits, lots of research activities are promoted. One interesting research topic is the regression analysis for group testing data. Chen et al. (2009) and McMahan et al. (2013) considered generalized linear regression model using binary pooling responses. Delaigle and Zhou (2015) considered covariates whose values are aggregated and proposed parametric and nonparametric estimators of the conditional prevalence. Other nonparametric regression (Delaigle and Meister, 2011; Delaigle and Hall, 2012) and semi-parametric regression (Delaigle et al., 2014; Wang et al., 2014) methodologies are also developed to analyze group testing data.

However, all research work aforementioned focus on binary responses. Less study

have been done for cases where the response variable is a continuous pooled response (Vexler et al., 2008; Mitchell et al., 2014; Liu et al., 2017). In this chapter, we consider group testing studies where a continuous pooled response is observed for each group, and individual continuous responses in a group are unobserved.

4.1 Models and Data

The goal of our study is to construct nonparametric estimators for the conditional mean function, $m(x) = E(Y|X = x)$, where Y is a continuous response of an experimental unit, e.g., an individual's macrophage inhibitory protein (MIP)-1α level in the Collaborative Perinatal Project (Whitecomb et al., 2007), and X denotes covariates associated with the individual, such as an individual's age, race, and miscarriage status in the CPP. Suppose that there are a total of N individuals in a study, and they are divided into J groups, with group j containing n_j individuals, for $j = 1, \cdots, J$. Define Y_{jk} as the unobserved continuous response of subject k in group j, and assume that X_{jk} follows a distribution specified by the probability density function (pdf) $f_X(x)$, for $j = 1, \cdots, J$, $k = 1, \cdots, c_j$. For ease of exposition, we consider a univariate covariate associated with each subject for the majority of this chapter. Note that, before the test is applied, the specimen (e.g., blood, water, urine) are pooled together. Consequently, Y_{ji} is not observed in studies considered here, and instead the observed result of the test, denoted by Z_j, is

$$Z_j = \frac{1}{c_j} \sum_{k=1}^{c_j} Y_{jk}. \tag{4.1}$$

It is evident from (4.1) that the observed value Z_j obtained from the pooled specimen has less variability than the individual values in the jth pool. In this case, the relationship between Y_{jk} and Z_j can be described by a Berkson measurement error model (Carroll et al., 2006).

Suppose that the regression model of Y_{jk} given X_{jk} is specified as follows,

$$Y_{jk} = m(X_{jk}) + \epsilon_{jk}, \tag{4.2}$$

where $m(X_{jk})$ is the conditional mean function, $E(\epsilon_{jk}|X_{jk}) = 0, Var(\epsilon_{jk}|X_{jk}) = \sigma^2$, X_{jk} and ϵ_{jk}, are independent. The problem of interest in this chapter is to estimate $m(x)$ based on the observed group data $\{(Z_j, \tilde{\boldsymbol{X}}_j)\}_{j=1}^J$, where $\tilde{\boldsymbol{X}}_j = (X_{j,1}, \cdots, X_{j,c_j})$, for $j = 1, \cdots, J$.

4.2 PROPOSED METHODS

When the observations (X_{jk}, Y_{jk}), for $j = 1, \cdots, J$, $k = 1, \cdots, c_j$ are available, Fan and Gijbels (1996) derived a p-th order local polynomial estimator for $m(x)$, which can be succinctly written in the matrix form as follow,

$$\hat{m}(x) = \boldsymbol{e}_1^T \boldsymbol{S}^{-1} \boldsymbol{T}, \tag{4.3}$$

where $\boldsymbol{e}_1$ is a $(p+1) \times 1$ vector with 1 in the first entry and 0 in the remaining p entries,

$$\boldsymbol{S} = \begin{bmatrix} S_{0,0}(x) & \dots & S_{0,p}(x) \\ \vdots & \ddots & \vdots \\ S_{p,0}(x) & \dots & S_{p,p}(x) \end{bmatrix},$$

and $\boldsymbol{T} = (T_{1,0}(x), \dots, T_{1,p}(x))^T$, in which

$$\begin{cases} S_{\ell_1,\ell_2}(x) & = & N^{-1} \sum_{j=1}^J \sum_{k=1}^{c_j} \left(\dfrac{X_{ji} - x}{h} \right)^{\ell_1+\ell_2} K_h(X_{ji} - x), \text{ for } \ell_1, \ell_2 = 0, 1, \dots, p, \\ T_{1,\ell}(x) & = & N^{-1} \sum_{j=1}^J \sum_{k=1}^{c_j} Y_{ij} \left(\dfrac{X_{ji} - x}{h} \right)^{\ell} K_h(X_{ji} - x), \text{ for } \ell = 0, 1, \dots, p, \end{cases} \tag{4.4}$$

and $K_h(x) = h^{-1}K(x/h)$ with $K(\cdot)$ being a symmetric kernel function and h being the bandwidth.

In the practice of group testing, because individual responses Y_{jk}, $j = 1, \cdots, J$, $k = 1, \cdots, c_j$, are not observed, the estimator in (4.3) cannot be calculated. When the

observed data are $\{(Z_j, \tilde{\boldsymbol{X}}_j)\}_{j=1}^{J}$, we consider formulating local polynomial estimators for $m(x)$ from two angles. The first two estimators originate from the mean model $E(Z_j | \tilde{\boldsymbol{X}}_j)$. The third estimator is motivated by the strategy in Lin and Wang (2018), following which a local polynomial estimator of $m(x)$ can be constructed based on the mean model $E(c_j Z_j | X_{jk})$.

The first estimator

By the definition of the pooled response, Z_j, in (4.1), one has

$$E(Z_j | \tilde{\boldsymbol{X}}_j) = \frac{1}{c_j} \sum_{k=1}^{c_j} m(X_{ji}), \ j = 1, \cdots, J, \tag{4.5}$$

where $\tilde{\boldsymbol{X}}_j = (X_{j1}, \cdots, X_{jc_j}), \ j = 1, \cdots, J$. Suppose that the $(p+1)$-th derivative of $m(X_{ji})$ at the point x exists. A p-th order Taylor expansion of $m(X_{jk})$ around x gives, for $k = 1, \ldots, c_j$ and $j = 1, \ldots, J$,

$$m(X_{jk}) \approx m(x) + m^{(1)}(x)(X_{jk} - x) + \cdots + \frac{m^p(x)}{p!}(X_{jk} - x)^p. \tag{4.6}$$

Plugging (4.6) in (4.5), one has

$$E(Z_j | \tilde{\boldsymbol{X}}_j) \approx m(x) + m^{(1)}(x)\frac{1}{c_j} \sum_{k=1}^{c_j} (X_{jk} - x) + \cdots + \frac{m^p(x)}{p!}\frac{1}{c_j} \sum_{k=1}^{c_j} (X_{jk} - x)^p. \tag{4.7}$$

To fit this polynomial, we follow the weighted least squares method in Fan and Gijbels (1996) to estimate $m(x)$ and its derivatives by minimizing

$$Q_1(\boldsymbol{\beta}) = \sum_{j=1}^{J} \left[Z_j - \left\{ \sum_{\ell=0}^{p} \beta_\ell \frac{1}{c_j} \sum_{k=1}^{c_j} (X_{jk} - x)^\ell \right\} \right]^2 \left\{ \frac{1}{c_j} \sum_{k=1}^{c_j} K_h(X_{jk} - x) \right\}, \tag{4.8}$$

where $\beta_\ell = m^{(\ell)}/\ell!$, for $\ell = 0, 1, \cdots, p$. Since there are multiple covariate values within one group, the average weight of group j is defined as $\frac{1}{c_j} \sum_{i=1}^{c_j} K_h(X_{jk} - x), \ j = 1, \cdots, J$. A p-th order local polynomial estimator of $m(x)$ is given by

$$\hat{m}_1(x) = \boldsymbol{e}_1^T \boldsymbol{S}_1^{-1} \boldsymbol{T}_1, \tag{4.9}$$

where

$$\boldsymbol{S}_1(x) = \boldsymbol{D}_1(x)^T \boldsymbol{K}_1(x) \boldsymbol{D}_1(x) = [S_{1,\ell_1,\ell_2}(x)]_{\ell_1,\ell_2=0,\ldots,p},$$

$$\boldsymbol{T}_1(x) = \boldsymbol{D}_1(x)^T \boldsymbol{K}_1(x) \boldsymbol{Z} = (T_{1,0}(x), T_{1,1}(x), \ldots, T_{1,p}(x))^T, \text{ in which}$$

$$\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_J)^T,$$

$$\boldsymbol{D}_1(x) = \begin{bmatrix} 1 & \bar{X}_1 - x & c_1^{-1} \sum_{k=1}^{c_1} (X_{1k} - x)^2 & \ldots & c_1^{-1} \sum_{k=1}^{c_1} (X_{1k} - x)^p \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & \bar{X}_J - x & c_1^{-1} \sum_{k=1}^{c_J} (X_{Jk} - x)^2 & \ldots & c_J^{-1} \sum_{k=1}^{c_J} (X_{Jk} - x)^p \end{bmatrix},$$

$$\boldsymbol{K}_1(x) = \text{diag}\left(c_1^{-1} \sum_{k=1}^{c_1} K_h(X_{jk} - x), \ldots, c_J^{-1} \sum_{k=1}^{c_J} K_h(X_{Jk} - x) \right),$$

in which $\bar{X}_j = c_j^{-1} \sum_{k=1}^{c_j} X_{jk}$, for $j = 1, \ldots, J$. Elaborating the matrix multiplications shows that

$$\begin{cases} S_{1,\ell_1,\ell_2}(x) \\ = \sum_{j=1}^{J} \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell_1} \right\} \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell_2} \right\} \left\{ c_j^{-1} \sum_{k=1}^{c_j} K_h(X_{jk} - x) \right\}, \\ \text{for } \ell_1, \ell_2 = 0, 1, \cdots, p, \\ T_{1,\ell}(x) \\ = \sum_{j=1}^{J} Z_j \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell} \right\} \left\{ c_j^{-1} \sum_{k=1}^{c_j} K_h(X_{jk} - x) \right\}, \text{ for } \ell = 0, 1, \ldots, p. \end{cases}$$

$$(4.10)$$

The second estimator

Instead of using the arithmetic mean $\dfrac{1}{c_j} \sum_{i=1}^{c_j} K_h(X_{jk} - x)$, $j = 1, \cdots, J$, as the weight function in (4.8), we consider to employ the product $\prod_{i=1}^{c_j} K_h(X_{jk} - x)$, $j = 1, \cdots, J$, as the weight to construct a new object function as follows,

$$Q_2(\boldsymbol{\beta}) = \sum_{j=1}^{J} \left[Z_j - \left\{ \sum_{\ell=0}^{p} \beta_\ell \frac{1}{c_j} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell} \right\} \right]^2 \left\{ \frac{1}{c_j} \prod_{k=1}^{c_j} K_h(X_{jk} - x) \right\}.$$

88

By minimizing $Q_2(\boldsymbol{\beta})$, a p-th order local polynomial estimator of $m(x)$ can be given by

$$\hat{m}_2(x) = \boldsymbol{e}_1^T \boldsymbol{S}_2^{-1} \boldsymbol{T}_2, \tag{4.11}$$

where

$$\boldsymbol{S}_2(x) = \boldsymbol{D}_2(x)^T \boldsymbol{K}_2(x) \boldsymbol{D}_2(x) = [S_{2,\ell_1,\ell_2}(x)]_{\ell_1,\ell_2=0,\dots,p},$$

$$\boldsymbol{T}_2(x) = \boldsymbol{D}_2(x)^T \boldsymbol{K}_2(x) \boldsymbol{Z} = (T_{2,0}(x), T_{2,1}(x), \dots, T_{2,p}(x))^T, \text{ in which}$$

$$\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_J)^T,$$

$$\boldsymbol{D}_2(x) = \begin{bmatrix} 1 & \bar{X}_1 - x & c_1^{-1} \sum_{k=1}^{c_1} (X_{1k} - x)^2 & \dots & c_1^{-1} \sum_{k=1}^{c_1} (X_{1k} - x)^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \bar{X}_J - x & c_J^{-1} \sum_{k=1}^{c_J} (X_{Jk} - x)^2 & \dots & c_J^{-1} \sum_{k=1}^{c_J} (X_{Jk} - x)^p \end{bmatrix},$$

$$\boldsymbol{K}_2(x) = \mathrm{diag} \left(\prod_{k=1}^{c_1} K_h(X_{jk} - x), \dots, \prod_{k=1}^{c_J} K_h(X_{Jk} - x) \right).$$

Elaborating the matrix multiplications shows that

$$\begin{cases} S_{2,\ell_1,\ell_2}(x) & = \sum_{j=1}^{J} \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell_1} \right\} \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell_2} \right\} \left\{ \prod_{k=1}^{c_j} K_h(X_{jk} - x) \right\}, \\ & \text{for } \ell_1, \ell_2 = 0, 1, \cdots, p, \\ T_{2,\ell}(x) & = \sum_{j=1}^{J} Z_j \left\{ c_j^{-1} \sum_{k=1}^{c_j} (X_{jk} - x)^{\ell} \right\} \left\{ \prod_{k=1}^{c_j} K_h(X_{jk} - x) \right\}, \text{ for } \ell = 0, 1, \dots, p. \end{cases} \tag{4.12}$$

THE THIRD ESTIMATOR

Instead of considering the conditional mean $E(Z_j | \tilde{\boldsymbol{X}}_j)$ as in the first two proposed methods, one may alternatively consider the conditional mean $E(c_j Z_j | X_{jk})$. By the

89

definition of Z_j in (4.1), one has

$$E(c_j Z_j | X_{jk} = x)$$

$$= \sum_{\ell=1}^{c_j} E(Y_{j\ell} | X_{jk} = x)$$

$$= \sum_{\ell=1, \ell \neq k}^{c_j} E(Y_{j\ell} | X_{jk} = x) + E(Y_{jk} | X_{jk} = x)$$

$$= \sum_{\ell=1, \ell \neq k}^{c_j} E(Y_{j\ell}) + m(x), \text{ by (4.2) and assuming } Y_{j\ell} \perp X_{jk} \text{ for } \ell \neq k,$$

$$= (c_j - 1)\mu + m(x), \text{ assuming } \mu = E(Y_{j\ell}) \text{ for } \ell = 1, \cdots, c_j \text{ and } j = 1, \cdots, J.$$

Hence,

$$E\{c_j Z_j - (c_j - 1)\mu | X_{jk} = x\} = m(x). \tag{4.13}$$

Viewing μ known for now, (4.13) suggests that $m(x)$ is the mean function of a new response, $\tilde{R}_j = c_j Z_j - (c_j - 1)\mu$, given $X_{jk} = x$. One can then follow the construction of the weighted least squares objective function for developing local polynomial estimators in Fan and Gijbels (1996) to formulate an objective function for inferring $m(x)$ using $\{(\tilde{R}_j, \tilde{\boldsymbol{X}}_j)\}_{j=1}^{J}$. With μ actually unknown, a natural surrogate for $\tilde{R}_j$ is $R_j = c_j Z_j - (c_j - 1)\hat{u}$, where $\hat{u} = N^{-1} \sum_{j=1}^{J} c_j Z_j$ and $N = \sum_{j=1}^{J} c_j$. It is worth pointing out that $\tilde{R}_j$ only depends on the j-th pooled response, whereas R_j depends on all J pooled responses. Following the construction of the p-th order local polynomial estimator for $m(x)$, one now has the weighted least squares objective function given by

$$Q_3(\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{k=1}^{c_j} \left\{ R_j - \sum_{\ell=0}^{p} \beta_\ell \left(X_{jk} - x \right)^\ell \right\}^2 K_h(X_{jk} - x). \tag{4.14}$$

Here, we are mainly interested in estimating $m(x) = \beta_0 = \boldsymbol{e}_1^T \boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^T$.

Minimizing $Q_3(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and extracting the first entry of the minimizer yields the following p-th order local polynomial estimator for $m(x)$,

$$\hat{m}_3(x) = \boldsymbol{e}_1^T \boldsymbol{S}_3^{-1}(x) \boldsymbol{T}_3(x), \tag{4.15}$$

where

$$\boldsymbol{S}_3(x) = \boldsymbol{D}_3(x)^T \boldsymbol{K}_3(x) \boldsymbol{D}_3(x) = \left[S_{3,\ell_1,\ell_2}(x) \right]_{\ell_1,\ell_2=0,\dots,p},$$

$$\boldsymbol{T}_3(x) = \boldsymbol{D}_3(x)^T \boldsymbol{K}_3(x) \boldsymbol{R} = (T_{3,0}(x), T_{3,1}(x), \dots, T_{3,p}(x))^T, \text{ in which}$$

$$\boldsymbol{R} = (R_1 \mathbf{1}_{c_1}^T, \dots, R_J \mathbf{1}_{c_J}^T),$$

$$\boldsymbol{D}_3(x) = \begin{bmatrix} 1 & X_{11} - x & (X_{11} - x)^2 & \dots & (X_{11} - x)^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{1,c_1} - x & (X_{1,c_1} - x)^2 & \dots & (X_{1,c_1} - x)^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{J1} - x & (X_{J1} - x)^2 & \dots & (X_{J1} - x)^p \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{J,c_J} - x & (X_{J,c_J} - x)^2 & \dots & (X_{J,c_J} - x)^p \end{bmatrix},$$

$$\boldsymbol{K}_3(x) = \begin{bmatrix} \boldsymbol{K}_{1,1}(x) & \mathbf{0}_{c_1,c_2} & \dots & \mathbf{0}_{c_1,c_J} \\ \mathbf{0}_{c_2,c_1} & \boldsymbol{K}_{1,2}(x) & \dots & \mathbf{0}_{c_2,c_J} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{c_J,c_1} & \mathbf{0}_{c_J,c_2} & \dots & \boldsymbol{K}_{1,J}(x) \end{bmatrix}, \text{ in which}$$

$$\boldsymbol{K}_{1,j} = \mathrm{diag}(K_h(X_{j1} - x), \dots, K_h(X_{j,c_j} - x)), \text{ for } j = 1, \dots, J.$$

In the above, $\mathbf{0}_{s \times t}$ denotes an $s \times t$ matrix of zero's, and $\mathbf{1}_t$ denotes a $t \times 1$ vector of one's. Elaborating the matrix multiplications reveals that

$$\begin{cases} S_{3,\ell_1,\ell_2}(x) &= \sum_{j=1}^{J} \sum_{k=1}^{c_j} (X_{ji} - x)^{\ell_1+\ell_2} K_h (X_{ji} - x), \text{ for } \ell_1, \ell_2 = 0, 1, \dots, p, \\ T_{3,\ell}(x) &= \sum_{j=1}^{J} \sum_{k=1}^{c_j} R_j (X_{ji} - x)^{\ell} K_h (X_{ji} - x), \text{ for } \ell = 0, 1, \dots, p. \end{cases} \tag{4.16}$$

91

4.3 SIMULATION STUDY

BANDWIDTH SELECTION

It has been well acknowledged that it is crucial to select an appropriate bandwidth in kernel-based nonparametric estimation. In this chapter, the leave-one-out cross-validation (CV) method is employed to select the bandwidth h in the local polynomial estimator, with the cross validation criterion one minimizes with respect to h given by

$$\mathrm{CV}(h) = J^{-1} \sum_{j=1}^{J} c_j \left\{ Z_j - c_j^{-1} \sum_{k=1}^{c_j} \hat{m}^{(-j)}(x_{jk}) \right\}^2. \qquad (4.17)$$

Here $\hat{m}^{(-j)}$ denotes the local polynomial estimator, resulting from a considered proposed method, computed based on the observed data excluding the j-th group data from the sample.

SIMULATION RESULTS

In the simulation study, we compare realizations of the three proposed estimators based on data generated according to the following model configuration,

$$[Y|X = x] \sim N(m(x), 0.2^2), \text{ where } m(x) = x^2 + x^3, X \sim \mathrm{uniform}(-0.5, 0.5).$$

To generate pooled observations, we set $N = 1000$ and specify a common group size, $c_j = c$ for all $j = 1, \cdots, J$, where $c = 5$. Different levels of savings can be obtained by choosing different group sizes. For example, when comparing with individual testing in which $c = 1$, the group testing when $c = 5$ indicates an 80% reduction of the testing cost. Under this simulation setting, 500 Monte Carlo (MC) replicates of sample size $N = 1000$ and group size $c = 5$ are generated.

Denote by $\hat{m}_{[\cdot]}$ one of the three estimates. We compare the performance of the three estimators with regard to the quality of the estimation of $m(x)$ at individual x, as well as the quality of the entire regression curve estimation over the domain

Table 4.1: Averages of point-wise absolute error and approximated ISE over 500 repetitions. Numbers in parentheses are $10 \times$ standard deviations associated with the averages. Local constant estimates correspond to $p = 0$. Local linear estimates correspond to $p = 1$.

		x_L	25th	50th	75th	x_U	ISE
	$\hat{m}_1$	0.035	0.029	0.068	0.085	0.245	0.006
		(0.113)	(0.096)	(0.091)	(0.062)	(0.121)	(0.004)
$p = 0$	$\hat{m}_2$	0.060	0.018	0.045	0.022	0.210	0.004
		(0.340)	(0.138)	(0.210)	(0.183)	(0.576)	(0.020)
	$\hat{m}_3$	0.037	0.022	0.021	0.024	0.091	0.001
		(0.270)	(0.172)	(0.161)	(0.180)	(0.478)	(0.007)
	$\hat{m}_1$	0.094	0.015	0.070	0.037	0.187	0.005
		(0.407)	(0.111)	(0.102)	(0.181)	(0.433)	(0.011)
$p = 1$	$\hat{m}_2$	0.072	0.017	0.060	0.043	0.144	0.004
		(0.529)	(0.132)	(0.192)	(0.198)	(0.673)	(0.015)
	$\hat{m}_3$	0.054	0.022	0.024	0.024	0.061	0.001
		(0.432)	(0.190)	(0.194)	(0.178)	(0.486)	(0.009)

of the function. More specifically, first, we calculate the Monte Carlo average of the point-wise mean squared error (MSE) to monitor the quality the estimation of $m(x)$ at the minimum, 25th percentile , 50th percentile, 75th percentile, and the maximum of X. Second, an approximate of the integrated squared error (ISE), ISE $= \int_{x_L}^{x_U} \{\hat{m}_{[\cdot]} - m(x)\}^2 dx$ is employed to describe the quality of the entire regression curve $\hat{m}(x)$. Note that, $[x_L, x_U]$ is the interval of the true covariate $X's$ value of interest. This approximated ISE is given by $\sum_{k=0}^{M} \hat{m}_{[\cdot]}(x_k) - m(x_k)^2 \Delta$, where M is the greatest integer less than $(x_U - x_L)/\Delta$, Δ is the partition resolution, $x_k = x_L + k\Delta$, $k = 1, \cdots, M$.

Table 4.1 presents Monte Carlo averages of the aforementioned point-wise MSE and ISE across 500 replicates along with their standard deviations. By comparing local constant estimator and local linear estimator in terms of ISE, one can see that the local linear estimator provides a better overall fitting. Among all proposed three local linear estimators, based on how their ISEs compare, our third estimator provides the best overall fitting. As for the variability, under the considered simulation setting, our first estimator outperforms the other two proposed method and produces the smallest

93

variability. When inspecting point-wise MSE's at various points associated with the three estimators, one can see the local linear estimators do not necessarily outperform the local constant estimators. For example, the local linear estimator outperform the local constant estimator at $x = x_U$. However, the local constant estimators shows advantage at $x = x_L$ when x is the second quantile.

Figures 4.1 and 4.2 present graphs to describe the overall fitting and point-wise fitting of all three estimators. Similar stories can be found in boxplots in Figures 4.1 and 4.2. Besides these boxplots, one can also see the overall fitting in panel (g), (h) and (f). Each of these three panels presents an estimated curves resulting from a proposed method, where the estimated curve is chosen such that its ISE is the median of all 500 collected ISE's for the corresponding method. In Figure 4.1, our first estimator and second estimator show less advantage at the boundary of the covariate support, the third estimator provides a great fit for the mean of the response. Similar phenomenon can be observed in Figure 4.2.
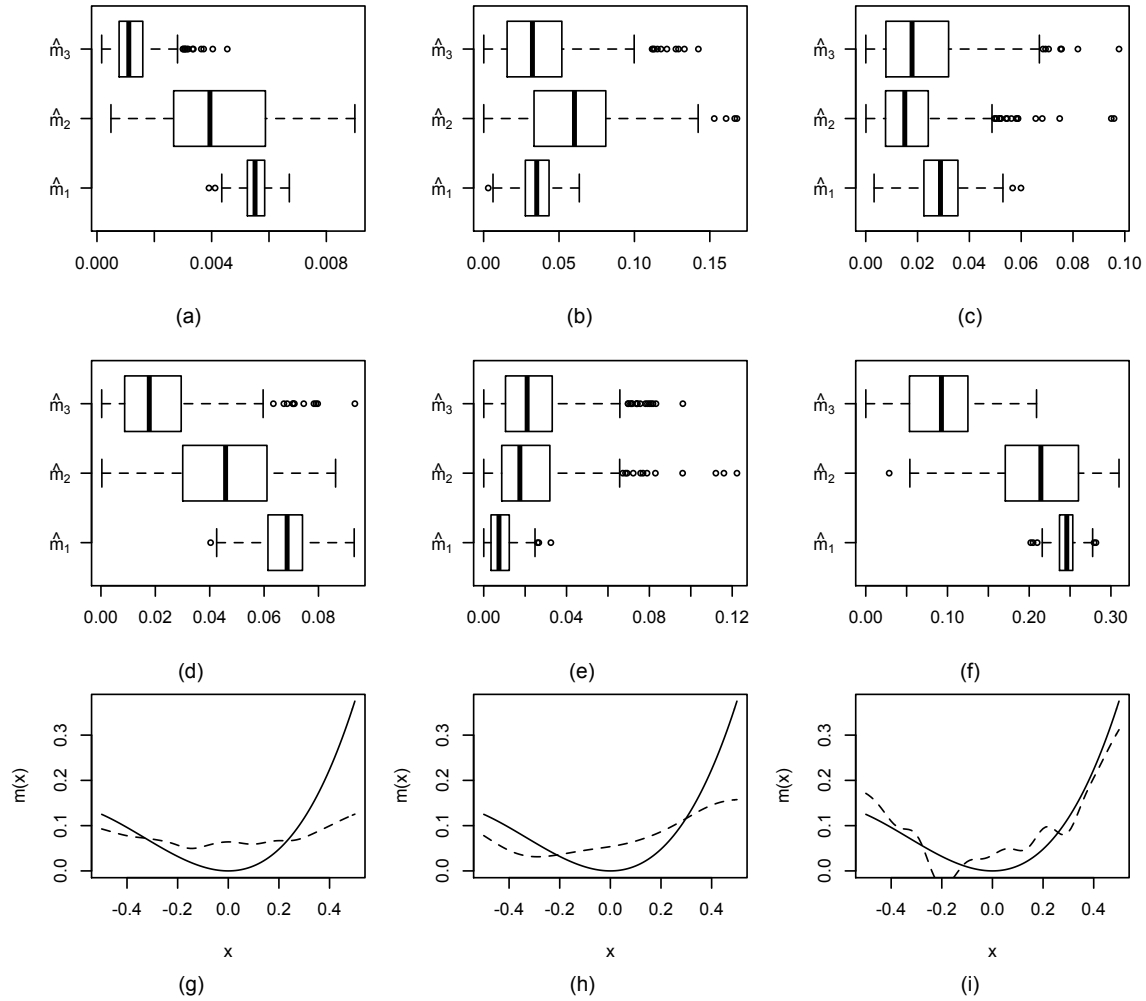
Figure 4.1: Simulation results using local constant regression for three estimators $\hat{m}_1$, $\hat{m}_2$ and $\hat{m}_3$, respectively. Panel (a): boxplots of ISEs. Panels (b) and (f): boxplots of point-wise absolute error at x_L and x_U, respectively. Panels (c), (d) and (e): boxplots of point-wise absolute error at 25th quantile, 50th quantile, 75th quantile, respectively. Panel (g), (h) and (i): Fitted curves whose ISEs equal to Median of ISEs over 500 repetitions of $\hat{m}_1$, $\hat{m}_2$ and $\hat{m}_3$. (dashed lines are the estimated curves for the median ISEs, solid lines for the truth)

95
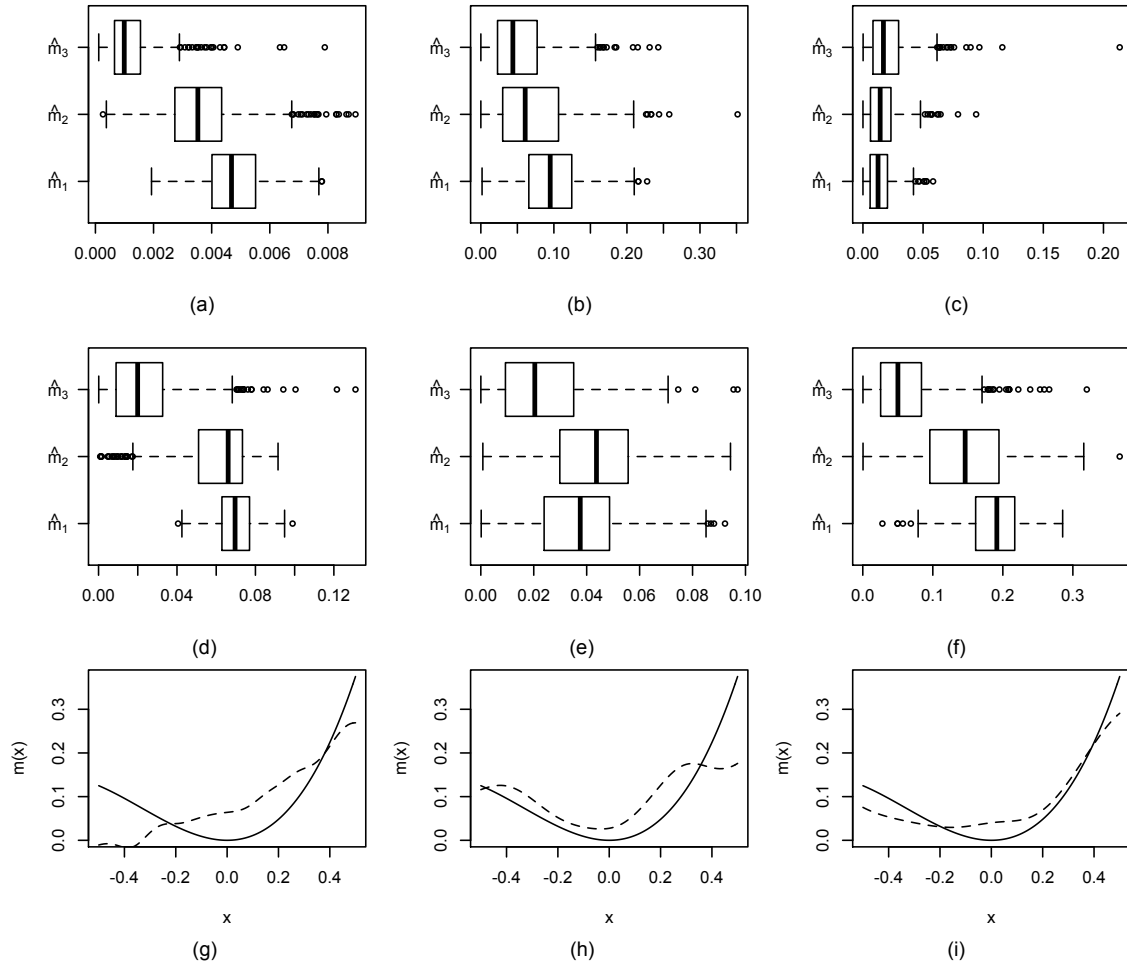
Figure 4.2: Simulation results using local linear regression when $p = 1$ for three estimators $\hat{m}_1$, $\hat{m}_2$ and $\hat{m}_3$, respectively. Panel (a): boxplots of ISEs. Panels (b) and (f): boxplots of point-wise absolute error at x_L and x_U, respectively. Panels (c), (d) and (e): boxplots of point-wise absolute error at 25th quantile, 50th quantile, 75th quantile, respectively. Panel (g), (h) and (i): Fitted curves whose ISEs equal to Median of ISEs over 500 repetitions of $\hat{m}_1$, $\hat{m}_2$ and $\hat{m}_3$. (dashed lines are the estimated curves for the median ISEs, solid lines for the truth)

96

CHAPTER 5

CONCLUSIONS AND FUTURE STUDY

This dissertation considers statistical inference in semiparametric and nonparametric regression in the presence of measurement error. Chapter 1 proposes two methods to infer the regression coefficients in a linear mode model for a response given an error-prone covariate. The resultant inference for the covariate effect significantly improve over the naive inference from applying an existing method without accounting for measurement error. Chapter 2 extends linear mode models to partially linear mode models, and proposes two estimation procedures to estimate parametric coefficients and the nonparametric function in a partially linear mode model. To account for measurement error, following the work in Chapter 1, Chapter 3 apply the idea of the corrected kernel method to obtain consistent estimators based on contaminated data in a partially linear model. Evidence from simulation studies suggest that the two proposed methods can yield estimators for the unknowns in such models that improve over naive estimators substantially. Instead of the classical measurement error considered in the first three chapters, which relates the unobserved covariates with their error-contaminated surrogates, in group testing studies with a continuous response, we consider in Chapter 4 a Berkson measurement error model that relates the unobserved individual response and the aggregated (pooled) response. In particular, Chapter 4 proposes an unobserved individual nonparametric estimation method to estimate the mean of a response variable given individual error-free covariates. Within the scope of this dissertation, as in Chapter 1, properties of the corrected kernel estimators are studied from both empirical and theoretical perspectives. The

97

theoretical properties of both parametric coefficients and nonparametric functions resulting from the one-stage estimation method in Chapter 2 are studied by Zhao et al. (2014) already. Theoretical properties of the proposed estimators accounting for covariate measurement error are investigated in Chapter 3. Moreover, to show the performance of the proposed methods in Chapter 3 under practical scenarios, a real data analysis is presented.

This dissertation focuses on regression problems when the number of covariates is smaller than the sample size. One interesting follow-up research topic, although beyond the scope of this dissertation, is to consider these semi-/non-parametric regression problems that involve high dimensional covariates. In settings with high dimensional covariates, Zhao et al. (2014) propose a variable selection method in partially linear mode regression in the absence of measurement error. But there is no research in the existing literature regarding variable selection in a partially linear mode regression model in the presence of measurement error. Furthermore, in Chapter 4, we only consider cases where covariates are free of measurement error. A similar problem as those considered in Chapters 1 and 3 is to develop consistent estimators for the mean of an unobserved individual response variable given error-prone covariates. This comprises yet another future research branch.

Bibliography

D. H. Bailey and P. N. Swarztrauber. A fast method for the numerical evaluation of continuous fourier and laplace transforms. *SIAM Journal on Scientific Computing*, 15(5):1105–1110, 1994.

S. P. Bamford, A. L. Rojas, R. C. Nichol, C. J. Miller, L. Wasserman, C. R. Genovese, and P. E. Freeman. Revealing components of the galaxy population through non-parametric techniques. *Monthly Notices of the Royal Astronomical Society*, 391 (2):607–616, 2008.

P. Billingsley. *Probability and measure.* 2008.

R. P. Boas. *Entire functions*, volume 5. 2011.

C. d. Boor. *A practical guide to splines*, volume 27. Springer, 2001.

J. P. Buonaccorsi. Prediction in the presence of measurement error: general discussion and an example predicting defoliation. *Biometrics*, 51(4):1562–1569, 1995.

R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu. *Measurement error in nonlinear models: a modern perspective*, volume 105. Chapman & Hall/CRC, 2006.

R. J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988.

R. J. Carroll, L. Freedman, and D. Pee. Design aspects of calibration studies in nutrition, with analysis of missing data in linear measurement error models. *Biometrics*, pages 1440–1457, 1997.

P. Chen, J. M. Tebbs, and C. R. Bilder. Group testing regression models with fixed and random effects. *Biometrics*, 65:1270–1278, 2009.

Y.-C. Chen, C. R. Genovese, R. J. Tibshirani, L. Wasserman, et al. Nonparametric modal regression. *The Annals of Statistics*, 44(2):489–514, 2016.

X. Chi, X. Lou, M. C. Yang, and Q. Shu. An optimal dna pooling strategy for pregressive fine mapping. *Genetica*, 135:267–281, 2009.

J. R. Cook and L. A. Stefanski. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association*, 89 (428):1314–1328, 1994.

H. Cramer and H. Wold. Some theorems on distribution functions. *Journal of the London Mathematical Society*, pages s1–11:290–294, 1936.

A. Delaigle. An alternative view of the deconvolution problem. *Statistica Sinica*, pages 1025–1045, 2008.

A. Delaigle and I. Gijbels. Frequent problems in calculating integrals and optimizing objective functions: a case study in density deconvolution. *Statistics and Computing*, 17(4):349–355, 2007.

A. Delaigle and P. Hall. Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *Journal of the American Statistical Association*, 103(481): 280–287, 2008.

A. Delaigle and P. Hall. Nonparametric regression with homogeneous group testing data. *The Annals of Statistics*, 40:131–158, 2012.

A. Delaigle and A. Meister. Nonparametric regression analysis for group testing data. *Journal of the American Statistical Association*, 106:640–650, 2011.

A. Delaigle and W. Zhou. Nonparametric and parametric estimators of prevalence from group testing data with aggregated covariates. *Journal of the American Statistical Association*, 110:1785–1796, 2015.

A. Delaigle, P. Hall, and A. Meister. On deconvolution with repeated measurements. *The Annals of Statistics*, pages 665–685, 2008.

100

A. Delaigle, J. Fan, and R. J. Carroll. A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104(485):348–359, 2009.

A. Delaigle, P. Hall, and J. R. Wishart. New approaches to non- and semi-parametric regression for univariate and multivariate group testing data. *Biometrika*, 101: 567–585, 2014.

N. K. Dhand, W. O. Johnson, and J.-A. L. M. L. Toribio. A bayesian approach to estimate ojd prevalence from pooled fecal samples of variable pool size. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:452–473, 2010.

R. Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, (14):436–440, 1943.

J. Einbeck and G. Tutz. Modelling beyond regression functions: an application of multimodal regression to speed–flow data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(4):461–475, 2006.

J. Fan. Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 97–110, 1991a.

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272, 1991b.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

J. Fan and E. Masry. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *Journal of multivariate analysis*, 43 (2):237–271, 1992.

J. Fan and Y. K. Truong. Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925, 1993.

W. A. Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

J. L. Gastwirth and P. A. Hammick. Estimation of the prevalence of a rare disease, preserving the anonymity of the subjects by group testing: application to estimating the prevalence of aids antibodies in blood donors. *Journal of Statistical Planning and Inference*, 22:15–27, 1989.

B. Grund and P. Hall. On the minimisation of L^p error in mode estimation. *The Annals of Statistics*, 23(6):2264–2284, 1995.

W. Hardle, H. Liang, and G. Jiti. *Partially Linear Models*. Physica-Verlag Heidelberg, 2000.

X. He and H. Liang. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica*, 10:129–140, 2000.

X. He, Z. Zhu, and W. K. Fung. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89:579–590, 2002.

S. B. Hedges and P. Shah. Comparison of mode estimation methods and application in molecular clock analysis. *BMC bioinformatics*, 4(1):31, 2003.

M. Huang and W. Yao. Mixture of regression models with varying mixing proportions: a semiparametric approach. *Journal of the American Statistical Association*, 107 (498):711–724, 2012.

X. Huang and H. Zhou. An alternative local polynomial estimator for the error-in-variables problem. *Journal of Nonparametric Statistics*, 29:301–325, 2017.

R. J. Hyndman, D. M. Bashtannyk, and G. K. Grunwald. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5(4): 315–336, 1996.

G. C. Kemp and J. S. Silva. Regression towards the mode. *Journal of Econometrics*, 170(1):92–101, 2012.

R. Koenker. *Quantile regression*. Number 38. Cambridge university press, 2005.

H. L. Koul and W. Song. Model checking in partial linear regression models with berkson measurement errors. *Statistica Sinica*, pages 1551–1579, 2010.

102

M.-j. Lee. Mode regression. *Journal of Econometrics*, 42(3):337 – 349, 1989.

M.-j. Lee. Quadratic mode regression. *Journal of Econometrics*, 57(1-3):1–19, 1993.

J. Li, S. Ray, and B. G. Lindsay. A nonparametric statistical approach to clustering via mode identification. *Journal of Machine Learning Research*, 8(8):1687–1723, 2007.

H. Liang and R. Li. Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association*, 104(485):234–248, 2009.

H. Liang, W. Hardle, and R. J. Carroll. Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics*, 27(5):1519–1535, 1999.

H. Liang, S. Wang, and R. J. Carroll. Partially linear models with missing response variables and error-prone covariates. *Biometrika*, 94(1):185–198, 2007.

H. Liang, S. W. Thurston, D. Ruppert, T. Apanasovich, and R. Hauser. Additive partial linear models with measurement errors. *Biometrika*, 95(3):667–678, 2008.

J. Lin and D. Wang. Single-index regression for pooled biomarker data. *To appear in Journal of Nonparametric Statistics*, 2018.

Y. Liu, C. S. McMahan, and C. Gallagher. A general framework for the regression analysis of pooled bimarker assessments. *Statistics in Medicine*, 14:2363?2377, 2017.

E. Masry. Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of multivariate analysis*, 44(1):47–68, 1993.

C. S. McMahan, J. M. Tebbs, and C. R. Bilder. Regression models for group testing data with pool dilution effects. *Biostatistics*, 14:284?298, 2013.

A. Meister. On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32(4):439–449, 2004.

E. Mitchell, R. Lyles, A. Manatunga, M. Danaher, N. Perkins, and E. Schisterman. Regression for skewed biomarker outcomes subject to pooling. *Biometrics*, 70: 202–211, 2014.

T. Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990.

S. J. Novick and L. A. Stefanski. Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97 (458):472–481, 2002.

E. Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

K. S. Remlinger, S. S. Young, J. M. Hughes-Oliver, and R. L. Lam. Statistical design of pools using optimal coverage and minimal collision. *Technometrics*, 48:133–143, 2006.

E. Robert F., G. C.W.J., R. John, and W. Andrew. Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320, 1986.

E. F. Schisterman, A. Vexler, A. Ye, and N. J. Perkins. A combined efficient design for biomarker data subject to a limit of detection due to measuring instrument sensitivity. *The Annals of Applied Statistics*, 5:2651–2667, 2011.

L. L. Schumaker. *Spline Functions: Basic Theory*. Cambridge University Press, 2007.

B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

X. Song and Y. Huang. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, 61(3):702–714, 2005.

L. A. Stefanski and R. J. Carroll. Deconvolving kernel density estimators. *Statistics: A Journal of Theoretical and Applied Statistics*, 21(2):169–184, 1990.

104

L. A. Stefanski and J. R. Cook. Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, 90(432):1247–1256, 1995.

B. Van Es and H.-W. Uh. Asymptotic normality of kernel-type deconvolution estimators. *Scandinavian Journal of Statistics*, 32(3):467–483, 2005.

A. Vexler, E. Schisterman, and A. Liu. Estimation of roc curves based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Statistics in Medicine*, 27:280–296, 2008.

C. Wang. Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, pages 235–253, 2006.

D. Wang, C. McMahan, C. Gallagher, and K. Kulasekera. Semiparametric group testing regression models. *Biometrika*, 101:587–598, 2014.

H. J. Wang, Z. Zhu, and J. Zhou. Quantile regression in partially linear varying coefficient models. *The Annals of Statistics*, 37(6B):3841–3866, 2009.

H. J. Wang, L. A. Stefanski, and Z. Zhu. Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*, 99(2):405, 2012.

Y. Wei and R. J. Carroll. Quantile regression with measurement error. *Journal of the American Statistical Association*, 104(487):1129–1143, 2009.

B. W. Whitecomb, E. F. Schisterman, M. A. Klebanoff, M. Baumgarten, A. Rhoton-Vlasak, X. Luo, and N. Chegini. Circulating chemokine levels and miscarriage. *American journal of epidemiology*, 166(3):323–331, 2007.

W. Yao and L. Li. A new regression model: modal linear regression. *Scandinavian Journal of Statistics*, 41(3):656–671, 2014.

W. Yao and S. Xiang. Nonparametric and varying coefficient modal regression. *arXiv preprint arXiv:1602.06609*, 2016.

W. Yao, B. G. Lindsay, and R. Li. Local modal regression. *Journal of Nonparametric Statistics*, 24(3):647–663, 2012.

C.-H. Zhang. Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, pages 806–831, 1990.

W. Zhao, R. Zhang, J. Liu, and Y. Lv. Robust estimation and variable selection for semi- parametric partially linear varying coefficient model based on modal regression. *Journal of Nonparametric Statistics*, 25(2):523–544, 2013.

W. Zhao, R. Zhang, J. Liu, and Y. Lv. Robust and efficient variable selection for semi-parametric partially linear varying coefficient model based on modal regression. *Annals of the Institute of Statistical Mathematics*, 66(1):165–191, 2014.

H. Zhou and X. Huang. Nonparametric modal regression in the presence of measurement error. *Electronic Journal of Statistics*, 10(2):3579–3620, 2016.

D. M. Zucker and D. Spiegelman. Corrected score estimation in the proportional hazards model with misclassified discrete covariates. *Statistics in medicine*, 27 (11):1911–1933, 2008.

Appendix A

Proof of Theorem 1

For completeness, we repeat the conditions regarding $g(\epsilon|x)$ and the covariate stated in Section 1.3 in the Chapter 2 below. For a scalar s, denote by $\tilde{s}$ the vector $(1, s)^{\mathrm{T}}$.

(C1) The ℓ-th partial derivative of $g(\epsilon|x)$ with respect to (w.r.t.) ϵ, $g^{(\ell)}(\epsilon|x)$, is continuously differentiable around $\epsilon = 0$, for $\ell = 0, 1, 2, 3$, and $g^{(1)}(0|x) = 0$, for all x.

(C2) As $n \to \infty$, $n^{-1} \sum_{j=1}^{n} g(0|X_j) \tilde{X}_j \tilde{X}_j^{\mathrm{T}}$ and $n^{-1} \sum_{j=1}^{n} g^{(3)}(0|X_j) \tilde{X}_j$ converge in probability, and $n^{-1} \sum_{j=1}^{n} g^{(2)}(0|X_j) \tilde{X}_j \tilde{X}_j^{\mathrm{T}}$ converges in probability to a negative definite matrix.

(C3) As $n \to \infty$, $n^{-1} \sum_{j=1}^{n} \|\tilde{X}_j\|^4 = O_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm.

Addition conditions on the characteristic function of U and conditions on $K(t)$ are stated next. The first set of conditions given next are needed for proving Theorems 1 and 2 when U is ordinary smooth.

Conditions O:

(O1) As $|t| \to \infty$, $cb|t|^{-b-1}/2 \leq |\phi_U^{(1)}(t)| \leq 2cb|t|^{-b-1}$.

(O2) As $|t| \to \infty$, $cb(b+1)|t|^{-b-2}/2 \leq |\phi_U^{(2)}(t)| \leq 2cb(b+1)|t|^{-b-2}$.

(O3) As $|t| \to \infty$, $cb(b+1)(b+2)|t|^{-b-3}/2 \leq |\phi_U^{(3)}(t)| \leq 2cb(b+1)(b+2)|t|^{-b-3}$.

(O4) $\int t^{2(2+b)}\phi_K^2(t)dt < \infty$, and $\int t^6\phi_K^2(t)dt < \infty$.

The second set of conditions stated next are needed for proving Theorems 1 and 2 when U is super smooth.

Conditions S:

(S1) As $|t| \to \infty$, $d_0^{(1)}|t|^{b_0^{(1)}}\exp(-|t|^b/d_2) \leq |\phi_U^{(1)}(t)| \leq d_1^{(1)}|t|^{b_1^{(1)}}\exp(-|t|^b/d_2)$, for some positive constants $b_0^{(1)}$, $b_1^{(1)}$, $d_0^{(1)}$, and $d_1^{(1)}$, where $b_0^{(1)} \geq b_0$ and $b_1^{(1)} \geq b_1$.

(S2) As $|t| \to \infty$, $d_0^{(2)}|t|^{b_0^{(2)}}\exp(-|t|^b/d_2) \leq |\phi_U^{(2)}(t)| \leq d_1^{(2)}|t|^{b_1^{(2)}}\exp(-|t|^b/d_2)$, for some positive constants $b_0^{(2)}$, $b_1^{(2)}$, $d_0^{(2)}$, and $d_1^{(2)}$, where $b_0^{(2)} \geq b_0^{(1)}$ and $b_1^{(2)} \geq b_1^{(1)}$.

(S3) As $|t| \to \infty$, $d_0^{(3)}|t|^{b_0^{(3)}}\exp(-|t|^b/d_2) \leq |\phi_U^{(3)}(t)| \leq d_1^{(3)}|t|^{b_1^{(3)}}\exp(-|t|^b/d_2)$, for some positive constants $b_0^{(3)}$, $b_1^{(3)}$, $d_0^{(3)}$, and $d_1^{(3)}$, where $b_0^{(3)} \geq b_0^{(2)}$ and $b_1^{(3)} \geq b_1^{(2)}$.

(S4) The support of $\phi_K(t)$ is $[-1, 1]$.

The third set of conditions listed next are additional conditions needed to establish the asymptotic normality claimed in Theorem 2.

Conditions N:

(N1) The expectation $E\{U^\ell g(\beta_1 U|x)\}$, for $\ell = 0, 1, 2$, exists for all x.

(N2) The expectation $E[\{g^{(3)}(0|X)X\}^2]$ exists.

(N3) As $|t| \to \infty$, $\phi_U(t)$ is either purely real or pure imaginary.

(N4) There exist positive constants δ, e_1, e_2, and q such that $|\phi_K(t)| \leq e_1(1-t)^q$ and $\phi_K(t) \geq e_2(1-t)^q$ for $t \in (1-\delta, 1)$.

Among the above four additional conditions, (N1) and (N2) for needed to establish asymptotic normality of $\hat{\beta}_{CK}$ when U is ordinary smooth; (N3) and (N4) are needed

108

when U is super smooth, which are equivalent to Condition 3.1 iii)–v) in Fan and Masry (1992).

As in the Chapter 2, all integrations are over the entire real line $\mathbb{R}$ unless otherwise specified.

Theorem 1. *Under conditions (C1)–(C3) and conditions in Lemma C, there exists a maximizer of $Q_h^*(\boldsymbol{\beta})$, denoted by $\hat{\boldsymbol{\beta}}_{CK}$, such that, as $n \to \infty$ and $h \to 0$,*

(i) when U follows an ordinary smooth distribution of order b, if $nh^{7+2b} \to 0$, then

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left(\sqrt{\frac{1}{nh^{3+2b}}}\right); \tag{A.1}$$

(ii) when U follows a super smooth distribution of order b,
   if $\exp(2|\beta_1|^b h^{-b}/d_2))/(nh^{b_6}) \to 0$,
   where $b_6 = \max\{3 - 2\min(b_2, b_3), 5 - 2\min(b_2, b_3, b_4), 7 - 2\min(b_2, b_3, b_4, b_5)\}$,
   in which b_ℓ, for $\ell = 2, 3, 4, 5$, are defined in Lemma C, then

$$\|\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta}\| = O(h^2) + O_p\left\{\exp\left(\frac{|\beta_1|^b}{d_2 h^b}\right)\sqrt{\frac{1}{nh^{3-2\min(b_2, b_3)}}}\right\}. \tag{A.2}$$

Proof. Define a series of integrals that involve in the integrand the cosine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s) = -1/\phi_U(s)$ evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IC}_j^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \cos\left(\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j}{h} t\right) \phi_K(t) \tau^{(\ell_2)}(\beta_1 t/h) dt, \tag{A.3}$$

for nonnegative integers ℓ_1 and ℓ_2. For example, one can show using Euler's formula that, under the assumption stated in the Chapter 2 that $\phi_K(t)$ and $\phi_U(t)$ are even functions, the deconvoluting kernel evaluated at $(Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j)/h$,

$$K^*\left(\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j}{h}\right) = \frac{1}{2\pi} \int \exp\left(-i\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j}{h} t\right) \frac{\phi_K(t)}{\phi_U(-\beta_1 t/h)} dt,$$

is equal to $-\mathrm{IC}_j^{(0,0)}$. Similarly define another series of integrals that involve in the integrand the sine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s)$

evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IS}_j^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \sin\left(\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j}{h} t\right) \phi_K(t)\tau^{(\ell_2)}(\beta_1 t/h) dt. \tag{A.4}$$

To derive the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$, the following derivatives of $\mathrm{IC}_j^{(\ell_1,\ell_2)}$ and $\mathrm{IS}_j^{(\ell_1,\ell_2)}$ with respect to $\boldsymbol{\beta}$ are needed,

$$
\begin{aligned}
\frac{\partial \mathrm{IC}_j^{(\ell_1,\ell_2)}}{\partial \boldsymbol{\beta}} &= h^{-1}\left(\mathrm{IS}_j^{(\ell_1+1,\ell_2)}\begin{bmatrix}1\\W_j\end{bmatrix} + \mathrm{IC}_j^{(\ell_1+1,\ell_2+1)}\begin{bmatrix}0\\1\end{bmatrix}\right), \\
\frac{\partial \mathrm{IS}_j^{(\ell_1,\ell_2)}}{\partial \boldsymbol{\beta}} &= h^{-1}\left(-\mathrm{IC}_j^{(\ell_1+1,\ell_2)}\begin{bmatrix}1\\W_j\end{bmatrix} + \mathrm{IS}_j^{(\ell_1+1,\ell_2+1)}\begin{bmatrix}0\\1\end{bmatrix}\right).
\end{aligned}
\tag{A.5}
$$

To reveal the convergence rate of $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$, denoted by r_n, we aim to establish a sufficient condition for $\|\hat{\boldsymbol{\beta}}_{\mathrm{CK}} - \boldsymbol{\beta}\| = O_p(r_n)$, which states that, for any given $\delta > 0$, there exists a constant c such that

$$P\left\{\sup_{\|\boldsymbol{d}\|=c} Q_h^*(\boldsymbol{\beta} + r_n\boldsymbol{d}) < Q_h^*(\boldsymbol{\beta})\right\} \geq 1 - \delta. \tag{A.6}$$

This sufficient condition motivates us to consider the difference $\Delta(r_n) = Q_h^*(\boldsymbol{\beta} + r_n\boldsymbol{d}) - Q_h^*(\boldsymbol{\beta})$. In particular, a third-order Taylor expansion of $\Delta(r_n)$ around zero gives

$$\Delta(r_n) = \Delta(0) + r_n\Delta^{(1)}(0) + 0.5r_n^2\Delta^{(2)}(0) + 6^{-1}r_n^3\Delta^{(3)}(r^*),$$

where r^* lies between zero and r_n, $\Delta(0) = 0$,

$$
\begin{aligned}
\Delta^{(1)}(0) &= \left.\frac{\partial Q_h^*(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}^{\mathrm{T}}}\right|_{\tilde{\beta}=\beta}\boldsymbol{d} = K_n^{\mathrm{T}}\boldsymbol{d}, \\
\Delta^{(2)}(0) &= \boldsymbol{d}^{\mathrm{T}}\left.\frac{\partial^2 Q_h^*(\tilde{\boldsymbol{\beta}})}{\partial \tilde{\boldsymbol{\beta}}\partial \tilde{\boldsymbol{\beta}}^{\mathrm{T}}}\right|_{\tilde{\beta}=\beta}\boldsymbol{d} = \boldsymbol{d}^{\mathrm{T}}J_n\boldsymbol{d}, \\
\Delta^{(3)}(r^*) &= \boldsymbol{d}^{\mathrm{T}}L_n\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d},
\end{aligned}
$$

in which, by (A.5),

$$
K_n = \frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{nh} \sum_{j=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} K^* \left(\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}} \tilde{W}_j}{h} \right)
$$

$$
= -\frac{1}{nh^2} \sum_{j=1}^{n} \left(\mathrm{IS}_j^{(1,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_j^{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \tag{A.7}
$$

$$
J_n = \frac{\partial^2 Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\mathrm{T}}} \tag{A.8}
$$

$$
= -\frac{1}{nh^3} \sum_{j=1}^{n} \left(-\mathrm{IC}_j^{(2,0)} \begin{bmatrix} 1 & W_j \\ W_j & W_j^2 \end{bmatrix} + \mathrm{IS}_j^{(2,1)} \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix} + \mathrm{IC}_j^{(2,2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right), \tag{A.9}
$$

$$
L_n = -\frac{1}{nh^3} \sum_{i=1}^{n} \left(-\begin{bmatrix} 1 & W_j \\ W_j & W_j^2 \end{bmatrix} \frac{\partial \mathrm{IC}_j^{(2,0)}}{\partial \boldsymbol{\beta}} + \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix} \frac{\partial \mathrm{IS}_j^{(2,1)}}{\partial \boldsymbol{\beta}} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \frac{\partial \mathrm{IC}_j^{(2,2)}}{\partial \boldsymbol{\beta}} \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}
$$

$$
= \frac{1}{nh^4} \sum_{j=1}^{n} \left(\mathrm{IS}_j^{(3,0)} \begin{bmatrix} 1 + W_j^2 \\ W_j(1+W_j^2) \end{bmatrix} + \mathrm{IC}_j^{(3,1)} \begin{bmatrix} 2W_j \\ 1 + 3W_j^2 \end{bmatrix} \right.
$$

$$
\left. - \mathrm{IS}_j^{(3,2)} \begin{bmatrix} 1 \\ 3W_j \end{bmatrix} - \mathrm{IC}_j^{(3,3)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}
$$

$$
= \frac{1}{nh^4} \sum_{j=1}^{n} \begin{bmatrix} (1+W_j^2)\mathrm{IS}_j^{(3,0)} + 2W_j \mathrm{IC}_j^{(3,1)} - \mathrm{IS}_j^{(3,2)} \\ W_j(1+W_j^2)\mathrm{IS}_j^{(3,0)} + (1+3W_j^2)\mathrm{IC}_j^{(3,1)} - 3W_j \mathrm{IS}_j^{(3,2)} - \mathrm{IC}_j^{(3,3)} \end{bmatrix} \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*},
$$

$$
\tag{A.10}
$$

with $\boldsymbol{\beta}^*$ lying between $\boldsymbol{\beta}$ and $\boldsymbol{\beta} + r_n \boldsymbol{d}$, corresponding to r^* lying between zero and r_n.

In summary, we have

$$
Q_h^*(\boldsymbol{\beta} + r_n \boldsymbol{d}) - Q_h^*(\boldsymbol{\beta}) = r_n K_n^{\mathrm{T}} \boldsymbol{d} + 0.5 r_n^2 \boldsymbol{d}^{\mathrm{T}} J_n \boldsymbol{d} + 6^{-1} r_n^3 \boldsymbol{d}^{\mathrm{T}} L_n \boldsymbol{d}^{\mathrm{T}} \boldsymbol{d}. \tag{A.11}
$$

In order to reveal r_n that satisfies (A.6), we study the orders of J_n, K_n, and L_n based on the mean-variance decomposition given by $V = E(V) + O_p\{\sqrt{\mathrm{Var}(V)}\}$, for a random variable V under regularity conditions. The means and variances of J_n, K_n and L_n are derived in Appendix C, with results summarized in Lemma C.

111

Define $\mu_2 = \int t^2 K(t)dt$. If the measurement error distribution is ordinary of order b, by Lemma C, we have

$$
\begin{cases}
K_n &= \dfrac{\mu_2 h^2}{2n} \displaystyle\sum_{j=1}^{n} g^{(3)}(0|X_j)\tilde{X}_j\{1 + o_p(1)\} + O_p\left(1/\sqrt{nh^{3+2b}}\right), \\[2mm]
J_n &= \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} g^{(2)}(0|X_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}\{1 + o_p(1)\} + O_p\left(1/\sqrt{nh^{5+2b}}\right), \\[2mm]
L_n &= O_p(1) + O_p\left(1/\sqrt{nh^{7+2b}}\right).
\end{cases}
\tag{A.12}
$$

Based on these rates, by setting $r_n = h^2 + 1/\sqrt{nh^{3+2b}}$, one can show that (A.11) is dominated by the second term for a large enough c, which is negative definite in probability by condition (C2). More specifically, the first term in the right-hand side of (A.11) is of order $O(r_n^2)$, the second terms is $0.5 r_n^2 \boldsymbol{d}^{\mathrm{T}} J^* \boldsymbol{d}\{1 + o_p(1)\}$, where $J^* = \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} g^{(2)}(0|_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}$, and the third term is of order $o_p(r_n^2)$ provided that $nh^{7+2b} \to 0$. Hence, for a large enough c, (A.6) holds for $r_n = h^2 + 1/\sqrt{nh^{3+2b}}$. In fact, the rate r_n is determined by the rate of K_n, which has h^2 as the order of the bias of $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$ and $1/\sqrt{nh^{3+2b}}$ as the order of its standard error. This leads to (A.1).

If the measurement error distribution is super smooth of order b, by Lemma C, we have

$$
\begin{cases}
K_n &= \dfrac{\mu_2 h^2}{2} \displaystyle\sum_{j=1}^{n} g^{(3)}(0|X_j)\tilde{X}_j\{1 + o_p(1)\} \\[2mm]
& \quad + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}\right\}, \\[2mm]
J_n &= \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} g^{(2)}(0|X_j)\tilde{X}_j\tilde{X}_j^{\mathrm{T}}\{1 + o_p(1)\} \\[2mm]
& \quad + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{5-2\min(b_2,b_3,b_4)})}\right\}, \\[2mm]
L_n &= O_p(1) + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{7-2\min(b_2,b_3,b_4,b_5)})}\right\},
\end{cases}
\tag{A.13}
$$

where b_ℓ, for $\ell = 2,3,4,5$, are defined in Lemma C. Based on these rates, by setting $r_n = h^2 + \sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}$, one can show that (A.11) is again dominated by the second term for a large enough c. Thus, for a large enough c, (A.6) holds for $r_n = h^2 + \sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}$. This proves (A.2).

$\square$

112

Appendix B

Proof of Theorem 2

Theorem 2. *Under the same assumptions imposed in Theorem 1,*

(i) if U follows an ordinary smooth distribution of order b,

$$\sqrt{nh^{3+2b}}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2\mu_2 J^{*-1}Q/4\right) \xrightarrow{d} N(0,\, J^{*-1}K_L J^{*-1}); \qquad \text{(B.1)}$$

(ii) if U follows a super smooth distribution of order b,

$$\left\{Var(\hat{\boldsymbol{\beta}}_{CK})\right\}^{-1/2}\left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2\mu_2 J^{*-1}Q/4\right) \xrightarrow{d} N(0,\, 1), \qquad \text{(B.2)}$$

where $Var(\hat{\boldsymbol{\beta}}_{CK}) = O[\exp\{2|\beta_1|^b/(d_2 h^b)\}/\{nh^{3-2\min(b_2,b_3)}\}]$, and $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of the positive definite square root of a positive definite matrix $\boldsymbol{\Sigma}$.

Proof. Because $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$ maximizes $Q_h^*(\boldsymbol{\beta}) = (nh)^{-1}\sum_{j=1}^n K^*\{(Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j)/h\}$, one has

$$
\begin{aligned}
\frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \mathbf{0} \\
&= \frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \left.\frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\beta = \hat{\beta}_{\mathrm{CK}}} \\
&= \left.\frac{\partial^2 Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}}\right|_{\beta = \beta^*}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\mathrm{CK}}),
\end{aligned}
$$

where the last equality is by the mean value theorem, with $\boldsymbol{\beta}^*$ lying between $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{CK}}$. Thus,

$$\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\mathrm{CK}} = \left\{\left.\frac{\partial^2 Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathbf{T}}}\right|_{\beta=\beta^*}\right\}^{-1}\frac{\partial Q_h^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (J_n + L_n^*)^{-1}K_n, \qquad \text{(B.3)}$$

113

where K_n is given by (A.7), J_n is given by (A.9), and the last expression results from a first-order Taylor approximation to be elaborated next, from which the expression of L_n^* becomes clear.

In (B.3), L_n^* is the residual term when one uses J_n to approximate $(\partial^2/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}})Q_h^*(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}^*$ via a first-order Taylor expansion. To see how this residual is obtained, it is helpful to define $\tilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta} + t(\boldsymbol{\beta}^* - \boldsymbol{\beta})$, where $t \in [0, 1]$, and to introduce a function mapping $t \in [0, 1]$ to a 2×2 matrix defined by

$$
G(t) = \left. \frac{\partial^2 Q_h^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathbf{T}}} \right|_{\beta=\tilde{\beta}^*}, \text{ next recall (A.9),}
$$

$$
= -\frac{1}{nh^3} \sum_{j=1}^n \left\{ -\mathrm{IC}_j^{(2,0)} \begin{bmatrix} 1 & W_j \\ W_j & W_j^2 \end{bmatrix} + \mathrm{IS}_j^{(2,1)} \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix} + \mathrm{IC}_j^{(2,2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \Bigg|_{\beta=\tilde{\beta}^*}.
$$

Since $\tilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta} + t(\boldsymbol{\beta}^* - \boldsymbol{\beta})$, $G(0)$ is equal to J_n given in (A.9), and $G(1)$ is equal to $(\partial^2/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}})Q_h^*(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}^*$, that is, the quantity inside the curly brackets in (B.3). It follows that a first-order Taylor approximation of $G(1)$ around $t = 0$ leads to

$$
G(1) = G(0) + G'(t^*)(1 - 0) = J_n + G'(t^*),
$$

where $G'(t^*)$ is the derivative of $G(t)$ evaluated at $t = t^*$, for some t^* lying between 0 and 1, and $G'(t^*)$ is the residual L_n^* in (B.3). More specifically, one can apply the chain rule to obtain $G'(t)$ as follows. Firstly, one differentiates $G(t)$ with respect to $\tilde{\boldsymbol{\beta}}^*$, which amounts to differentiating $\mathrm{IC}_j^{(2,0)}$, $\mathrm{IS}_j^{(2,1)}$, and $\mathrm{IC}_j^{(2,2)}$ with respect to $\boldsymbol{\beta}$, then one evaluates the resultant expressions at $\tilde{\boldsymbol{\beta}}^*$. Secondly, one differentiates $\tilde{\boldsymbol{\beta}}^*$ with respect to t, producing the factor $\boldsymbol{\beta}^* - \boldsymbol{\beta}$. These two steps lead to the following expression of L_n^*,

$$
L_n^* = -\frac{1}{nh^3} \sum_{j=1}^n \left\{ -\begin{bmatrix} 1 & W_j \\ W_j & W_j^2 \end{bmatrix} \frac{\partial\mathrm{IC}_j^{(2,0)}}{\partial\boldsymbol{\beta}} + \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix} \frac{\partial\mathrm{IS}_j^{(2,1)}}{\partial\boldsymbol{\beta}} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \frac{\partial\mathrm{IC}_j^{(2,2)}}{\partial\boldsymbol{\beta}} \right\} \Bigg|_{\beta=\tilde{\beta}^*}
$$

$$
(\boldsymbol{\beta}^* - \boldsymbol{\beta})^{\mathrm{T}}.
$$

114

Since $\|\tilde{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\| \leq \|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| \leq \|\hat{\boldsymbol{\beta}}_{\text{CK}} - \boldsymbol{\beta}\| = o_p(1)$, using the results in Lemma C regarding L_n given by (A.10), one has $L_n^* = o_p(1)$.

Besides $L_n^* = o_p(1)$, Lemma C indicates that J_n converges to a finite constant matrix in probability. Hence, by (B.3), to establish the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\text{CK}} - \boldsymbol{\beta}$, it suffices to show the asymptotic normality for K_n. This is proved in the following two parts, the first part for establishing (B.1), and the second part for showing (B.2).

Part (I): Show (B.1).

When U is ordinary smooth, define $K_n^* = \sqrt{nh^{3+2b}}K_n$. We next show that,

$$\{\boldsymbol{p}^{\text{T}}\text{Cov}(K_n^*)\boldsymbol{p}\}^{1/2}\boldsymbol{p}^{\text{T}}\{K_n^* - E(K_n^*)\} \xrightarrow{d} N(0,1),$$

$$\text{for any unit vector } \boldsymbol{p} = (p_1, p_2)^{\text{T}} \in \mathbb{R}^2, \tag{B.4}$$

where "$\xrightarrow{d}$" refers to convergence in distribution. Once (B.4) is proved, we conclude the asymptotic normality of K_n^* by the Cramér-Wold Theorem (Cramer and Wold, 1936), and thus the asymptotic normality of K_n.

By (A.7), $\boldsymbol{p}^{\text{T}}K_n^* = \sum_{j=1}^n \xi_j$, where

$$\xi_j = -\frac{h^b}{\sqrt{nh}}\boldsymbol{p}^{\text{T}}\left(\text{IS}_j^{(1,0)}\begin{bmatrix}1\\W_j\end{bmatrix} + \text{IC}_j^{(1,1)}\begin{bmatrix}0\\1\end{bmatrix}\right).$$

Define $m_j = E(\xi_j|X_j)$ and $S_n^2 = \sum_{j=1}^n \text{Var}(\xi_j|X_j)$. We next use the Lyapunov Central Limit Theorem to show the asymptotic normality of $\boldsymbol{p}^{\text{T}}K_n^*$. This requires proving the Lyapunov's conditions (Billingsley, 2008), which states that

$$\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{j=1}^n E|\xi_j - m_j|^{2+\delta} = 0, \text{ for some } \delta > 0. \tag{B.5}$$

In particular, we show that (B.5) is satisfied for $\delta = 1$.

First, because $\xi_1, \ldots, \xi_n$ are independent,

$$S_n^2 = \mathrm{Var}\left(\sum_{j=1}^n \xi_j \middle| \boldsymbol{X}\right)$$

$$= \boldsymbol{p}^{\mathrm{T}}\mathrm{Var}(K_n^*|\boldsymbol{X})\boldsymbol{p}, \text{ next use (A.7)},$$

$$= h^{2b-1}\boldsymbol{p}^{\mathrm{T}}\mathrm{Var}\left(\mathrm{IS}_j^{(1,0)}\begin{bmatrix}1\\W_j\end{bmatrix} + \mathrm{IC}_j^{(1,1)}\begin{bmatrix}0\\1\end{bmatrix}\middle| X_j\right)\boldsymbol{p}, \quad (\text{B.6})$$

The variance in (B.6) involve the following expectations,

$$E\left\{\left(\mathrm{IS}_j^{(1,0)}\right)^2\middle| X_j\right\}, \quad E\left\{\left(\mathrm{IS}_j^{(1,0)}\right)^2 W_j\middle| X_j\right\}, \quad E\left\{\left(\mathrm{IS}_j^{(1,0)}\right)^2 W_j^2\middle| X_j\right\},$$
$$E\left(\mathrm{IS}_j^{(1,0)}\mathrm{IC}_j^{(1,1)}\middle| X_j\right), \quad E\left(\mathrm{IS}_j^{(1,0)}\mathrm{IC}_j^{(1,1)}W_j\middle| X_j\right), \quad E\left\{\left(\mathrm{IC}_j^{(1,1)}\right)^2\middle| X_j\right\}. \quad (\text{B.7})$$

Derivations of the limits of h^{2b-1} times these expectations as $n \to \infty$ and $h \to 0$ make use of intermediate results revealed in the proof for Lemma C in Appendix C, and Lemma 2.1 in Fan (1991a). In what follows, we elaborate the derivation of $\lim_{n\to\infty} h^{2b-1}E\{(\mathrm{IS}_j^{(1,0)})^2 U^\ell | X_j\}$, for $\ell = 0, 1, 2$, which relates to the first three expectations in (B.7).

By the arguments following (F.20) in Appendix C,
$h^b\{1/(2\pi)\}t\sin(vt)\phi_K(t)/\phi_U(\beta_1 t/h)$ is integrable, and thus, by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty} h^b \frac{1}{2\pi}\int t\sin(vt)\frac{\phi_K(t)}{\phi_U(\beta_1 t/h)}dt$$
$$= \frac{1}{2\pi}\int \lim_{n\to\infty} h^b t\sin(vt)\phi_K(t)\frac{|\beta_1 t/h|^b}{|\beta_1 t/h|^b\phi_U(\beta_1 t/h)}I(|\beta_1 t| > Mh)dt,$$
$$= \frac{|\beta_1|^b}{2\pi c_0}\int t\sin(vt)|t|^b\phi_K(t)dt, \quad (\text{B.8})$$

where M is a positive constant $c_0 = \lim_{|t|\to\infty} |t|^b \phi_U(t)$. It follows that

$$\lim_{n\to\infty} h^{2b-1} E\left\{ \left(\text{IS}_j^{(1,0)}\right)^2 U^\ell \middle| X_j \right\}$$

$$= \lim_{n\to\infty} h^{2b} \int f_U(u) u^\ell \int \left\{ \frac{1}{2\pi} \int t\sin(vt) \frac{\phi_K(t)}{\phi_U(\beta_1 t/h)} dt \right\}^2 g(\beta_1 u - hv|X_j) dv du$$

$$= \int f_U(u) u^\ell \lim_{n\to\infty} \int \left\{ h^b \frac{1}{2\pi} \int t\sin(vt) \frac{\phi_K(t)}{\phi_U(\beta_1 t/h)} dt \right\}^2 g(\beta_1 u - hv|X_j) dv du$$

$$= \int f_U(u) u^\ell g(\beta_1 u|X_j) \frac{\beta_1^{2b}}{c_0^2} \int \left\{ \frac{1}{2\pi} \int t\sin(vt)|t|^b \phi_K(t) dt \right\}^2 dv du \tag{B.9}$$

$$= \frac{\beta_1^{2b}}{c_0^2} E\{U^\ell g(\beta_1 U|X_j)\} \frac{1}{2\pi} \int t^{2+2b} \phi_K^2(t) dt, \tag{B.10}$$

where (B.9) is obtained by using Lemma 2.1 in Fan (1991a), along with the result in (B.8), and the expectation in (B.10) is with respect to the distribution of U. Following similar lines of arguments leading to (B.8) and (B.9), one can show that, relating to the last three expectations in (B.7),

$$\lim_{n\to\infty} h^{2b-2} E\left(\text{IS}_j^{(1,0)} \text{IC}_j^{(1,1)} U_j^\ell \middle| X_j \right) =$$
$$\frac{|\beta_1|^b c_1}{c_0^3} E\{U^\ell g(\beta_1 U|X_j)\} \frac{1}{2\pi} \int t^{1+2b} \phi_K^2(t) dt, \text{ for } \ell = 0, 1, \tag{B.11}$$
$$\lim_{n\to\infty} h^{2b-3} E\left\{ \left(\text{IC}_j^{(1,0)}\right)^2 \middle| X_j \right\} = \frac{\beta_1^{2b-2} c_1^2}{c_0^4} E\{g(\beta_1 U|X_j)\} \frac{1}{2\pi} \int t^{2b} \phi_K^2(t) dt,$$

where $c_1 = \lim_{|t|\to\infty} |t|^{b+1} |\phi_U^{(1)}(t)|$.

Using (B.10) and (B.11), along with relevant expectations derived in Appendix

117

C, one has

$$\lim_{n\to\infty} \mathrm{Var}(K_n^*|\boldsymbol{X})$$

$$= \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} h^{2b-1} \tag{B.12}$$

$$\begin{bmatrix} \mathrm{Var}\left(\mathrm{IS}_j^{(1,0)} \,\middle|\, X_j\right) & \mathrm{Cov}\left(\mathrm{IS}_j^{(1,0)}, \, \mathrm{IS}_j^{(1,0)} W_j + \mathrm{IC}_j^{(1,1)} \,\middle|\, X_j\right) \\ \mathrm{Cov}\left(\mathrm{IS}_j^{(1,0)}, \, \mathrm{IS}_j^{(1,0)} W_j + \mathrm{IC}_j^{(1,1)} \,\middle|\, X_j\right) & \mathrm{Var}\left(\mathrm{IS}_j^{(1,0)} W_j + \mathrm{IC}_j^{(1,1)} \,\middle|\, X_j\right) \end{bmatrix} \tag{B.13}$$

$$= \lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \frac{\beta_1^{2b}}{c_0^2} \frac{1}{2\pi} \int t^{2+2b} \phi_K^2(t) dt \times$$

$$\begin{bmatrix} E\{g(\beta_1 U|X_j)\} & X_j E\{g(\beta_1 U|X_j)\} + E\{U g(\beta_1 U|X_j)\} \\ X_j E\{g(\beta_1 U|X_j)\} + E\{U g(\beta_1 U|X_j)\} & T_4 \end{bmatrix}, \tag{B.14}$$

where $T_4 = X_j^2 E\{g(\beta_1 U|X_j)\} + 2X_j E\{U g(\beta_1 U|X_j)\} + E\{U^2 g(\beta_1 U|X_j)\}$, which is well defined under conditions (O4) and (N1). Hence, $\lim_{n\to\infty} S_n^2 = \boldsymbol{p}^{\mathrm{T}} K_{\mathrm{L}} \boldsymbol{p}$, where K_{L} is the expectation (with respect to $\boldsymbol{X}$) of (B.14).

Second, following similar derivations of expectations elaborated in Appendix C, one can show that multiplying the following three expectations by $n^{-1/2} h^{3(b-1/2)}$ all lead to quantities that tend to zero as $n \to \infty$,

$$E\left(\left|\mathrm{IS}_j^{(1,0)}\right|^3 \,\middle|\, X_j\right), \quad E\left(\left|\mathrm{IS}_j^{(1,0)} W_j\right|^3 \,\middle|\, X_j\right), \quad E\left(\left|\mathrm{IC}_j^{(1,1)}\right|^3 \,\middle|\, X_j\right). \tag{B.15}$$

These imply that, as $n \to \infty$,

$$n E(|\xi_j|^3|X_j) = n^{-1/2} h^{3(b-1/2)} E\left(\left|p_1 \mathrm{IS}_j^{(1,0)} + p_2 \mathrm{IS}_j^{(1,0)} W_j + p_2 \mathrm{IC}_j^{(1,1)}\right|^3 \,\middle|\, X_j\right) \to 0.$$

It follows that $\lim_{n\to\infty} \sum_{j=1}^{n} |\xi_j - m_j|^3 = 0$. This and the previous conclusion regarding S_n^2 together lead to the Lyapunov's conditions in (B.5), which is a sufficient condition for, taking into account the asymptotic mean of K_n according to Lemma C,

$$\sqrt{n h^{3+2b}} \left(K_n - \mu_2 Q h^2/2\right) \xrightarrow{d} N(0, K_{\mathrm{L}}).$$

118

Finally, by the Slutsky's Theorem, we have (B.1).

Part (II): Show (B.2).

When U is super smooth, define $K_n^* = \exp\{-2|\beta_1|^b/(d_2 h^b)\}\sqrt{nh^{3-2\min(b_2,b_3)}}K_n$. Now we have $\boldsymbol{p}^{\mathrm{T}}K_n^* = \sum_{j=1}^n \xi_j$ with

$$\xi_j = -\exp\left(-\frac{2|\beta_1|^b}{d_2 h^b}\right)\frac{1}{\sqrt{nh^{1+\min(b_2,b_3)}}}\boldsymbol{p}^{\mathrm{T}}\left(\mathrm{IS}_j^{(1,0)}\begin{bmatrix}1\\W_j\end{bmatrix}+\mathrm{IC}_j^{(1,1)}\begin{bmatrix}0\\1\end{bmatrix}\right).$$

The remaining task is to show (B.5) for the so-defined ξ_j. This leads one to look into the expectations in (B.7) and those in (B.15). In particular, relating to (B.15), one can show that $\sum_{j=1}^n |\xi_j - m_j|^3$ converges to a quantity of order $n^{-1/2}\exp\{-3|\beta_1|^b/(d_2 h^b)\}$. As for the expectations in (B.7), the key lies in establishing lower bounds for $|\mathrm{IS}_j^{(1,0)}|$, $|\mathrm{IC}_j^{(1,1)}|$, and $|\mathrm{IS}_j^{(1,0)}\mathrm{IC}_j^{(1,1)}|$. We adopt the strategies used in the proof for Lemma 2.3 in Fan (1991a) and the proof for Lemma 3.1 in Fan and Masry (1992), including splitting the region of an integration, shrinking or magnifying parts of an integrand, to find lower bounds of these quantities. Take $|\mathrm{IS}_j^{(1,0)}|$ as an example, it suffices to zoom in on the following integral, as $h \to 0$,

$$\begin{aligned}
&\frac{1}{2\pi}\int_{|t|\in(1-h^b,1)} t\sin(vt)\frac{\phi_K(t)}{\phi_U(\beta_1 t/h)}dt\\
=&\frac{1}{\pi}\sin(v)\{1+o(1)\}\int_{1-h^b}^1 t\frac{\phi_K(t)}{\phi_U(\beta_1 t/h)}dt\\
\geq&C\sin(v)h^{b_1}\int_{1-h^b}^1 t^{1-b_1}(1-t)^q\exp\left(\frac{|\beta_1|^b t^b}{d_2 h^b}\right)dt,\\
\geq&C\sin(v)h^{b_1}(1-h^b)^{1-b_1}\exp\left\{\frac{|\beta_1|^b(1-h^b)^b}{d_2 h^b}\right\}\int_{1-h^b}^1(1-t)^q dt &\text{(B.16)}\\
\geq&C\sin(v)h^{b_1+(q+1)b}\exp\left\{\frac{|\beta_1|^b}{d_2 h^b}\right\}, &\text{(B.17)}
\end{aligned}$$

where C denotes a generic constant in the above derivations; (B.16) is obtained because, if $b_1 \leq 1$, $t^{1-b_1}\exp\{|\beta_1|^b t^b/(d_2 h^b)\}$ is increasing for all $t > 0$, and, if $b_1 > 1$, it is increasing in $t \in (1-h^b,1)$ for an h close to zero; (B.17) is obtained due to the fact that $(1-h^b)^b \geq 1 - bh^b/2$ for a small enough h^b. Under (N3) and

119

www.manaraa.com

(N4), $(2\pi)^{-1}\int_{|t|\leq 1-h^b} t\sin(vt)\phi_K(t)/\phi_U(\beta_1 t/h)dt$ is dominated by the above integral as $h\to 0$. This suggests that $|\mathrm{IS}_j^{(1,0)}|^2$ is bounded from below by a quantity of order $h^{2b_1+2(q+1)b}\exp\{2|\beta_1|^b/(d_2 h^b)\}$. Similarly, assuming $0\leq 2b_1-b_0^{(1)}\leq 1$, under (N3), (N4), (S1), and the definition of super smoothness, one can show that $|\mathrm{IC}_j^{(1,1)}|^2$ is bounded from below by a quantity of order

$h^{4b_1-2b_0^{(1)}+2(q+1)b}\exp\{2|\beta_1|^b/(d_2 h^b)\}$, and $|\mathrm{IS}_j^{(1,0)}\mathrm{IC}_j^{(1,1)}|$ is bounded from below by a quantity of order

$h^{b_1+2b_1-b_0^{(1)}+2(q+1)b}\exp\{2|\beta_1|^b/(d_2 h^b)\}$. It follows that S_n^3 is bounded from below by a quantity of order $h^{3\{(q+1)b-0.5-\min(b_2,b_3)+\min(b_1,2b_1-b_0^{(1)})\}}\exp\{-3|\beta_1|^b/(d_2 h^b)\}$ as $n\to\infty$. This, along with the aforementioned result regarding $\sum_{j=1}^n|\xi_j-m_j|^3$, implies (B.5).

Unlike in Part (I) above, here we are unable to find the limits in (B.13). We state under Theorem 2-(ii) the asymptotic normality result for the case with super smooth measurement error in the form of (B.2), followed by the rate of $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{CK}})$. Deriving limits similar to those in (B.13) in the presence of super smooth measurement error has been a long-standing hurdle in establishing the central limit theorem of the classical form, such as the form in (B.1), for estimators like ours that involved deconvoluting kernels. Existing works where authors faced similar hurdles include Zhang (1990), Fan (1991a), Fan and Masry (1992), Masry (1993), Delaigle et al. (2009) and Delaigle and Zhou (2015), among many others, in which asymptotic normality results are presented in ways without involving explicit dominating terms of asymptotic variance of relevant estimators. Van Es and Uh (2005) derived the dominating terms in the asymptotic variance of deconvoluting kernel density estimators when the order of super smoothness b is larger than one. One may be able to follow a similar strategy to obtain the dominating term of $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{CK}})$ in the presence of super smooth measurement error of order $b>1$, which we do not pursue in the current work.

$\square$

Appendix C

Lemmas referenced in Appendices A and B

Lemma C. *Assume $E(U^4) < \infty$, and assume conditions stated in Lemmas C.1, C.2, and C.3 hold. Define $\mu_2 = \int t^2 K(t) dt$ and $\boldsymbol{X} = (X_1, \ldots, X_n)$. For K_n, J_n and L_n given by (A.7), (A.9), and (A.10), respectively, we have*

$$E(K_n | \boldsymbol{X}) = \frac{\mu_2 h^2}{2n} \sum_{j=1}^{n} g^{(3)}(0|X_j) \tilde{X}_j \{1 + o_p(1)\}, \tag{C.1}$$

$$E(J_n | \boldsymbol{X}) = \frac{1}{n} \sum_{j=1}^{n} g^{(2)}(0|X_j) \tilde{X}_j \tilde{X}_j^{\mathrm{T}} \{1 + o_p(1)\}, \tag{C.2}$$

$$E(L_n | \boldsymbol{X}) = O_p(1). \tag{C.3}$$

And when U is ordinary smooth of order b, under Conditions O, we have

$$Var(K_n | \boldsymbol{X}) = O_p\{1/(nh^{3+2b})\}, \tag{C.4}$$

$$Var(J_n | \boldsymbol{X}) = O_p\{1/(nh^{5+2b})\}, \tag{C.5}$$

$$Var(L_n | \boldsymbol{X}) = O_p\{1/(nh^{7+2b})\}; \tag{C.6}$$

when U is super smooth of order b, under Conditions S, we have

$$Var(K_n | \boldsymbol{X}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})\}, \tag{C.7}$$

$$Var(J_n | \boldsymbol{X}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{5-2\min(b_2,b_3,b_4)})\}, \tag{C.8}$$

$$Var(L_n | \boldsymbol{X}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{7-2\min(b_2,b_3,b_4,b_5)})\}, \tag{C.9}$$

where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)}) I(2b_0 - b_1^{(1)} < 0.5)$, $b_4 = \min\{(2b_0 - b_1^{(2)}) I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)}) I(3b_0 - 2b_1^{(1)} < 0.5)\}$, and $b_5 = \min\{(2b_0 - b_1^{(3)}) I(2b_0 - b_1^{(3)} < 0.5), (3b_0 - b_1^{(1)} - b_1^{(2)}) I(3b_0 - b_1^{(1)} - b_1^{(2)} < 0.5), (4b_0 - b_1^{(1)} - 2b_1^{(2)}) I(4b_0 - b_1^{(1)} - 2b_1^{(2)} < 0.5)\}$.

121

Proof. The proof consists of six parts to establish (C.1)–(C.9). We first define a series of integrals as follows, with integrands involving t^{ℓ_1} and the ℓ_2-th derivative of $\tau(s) = -1/\phi_U(s)$ evaluated it $s = \beta_1 t/h$,

$$\mathcal{F}_{\ell_1,\ell_2}(s) = \frac{1}{2\pi}\int e^{ist}t^{\ell_1}\phi_K(t)\tau^{(\ell_2)}(\beta_1 t/h)dt, \qquad (\text{C.10})$$

for nonnegative integers ℓ_1 and ℓ_2.

Part (I): Show (C.1).

By (A.7),

$$E(K_n|\boldsymbol{X}) = -\frac{1}{nh^2}\sum_{j=1}^n \begin{bmatrix} E\left(\mathrm{IS}_j^{(1,0)}\Big| X_j\right) \\ E\left(\mathrm{IS}_j^{(1,0)}W_j + \mathrm{IC}_j^{(1,1)}\Big| X_j\right) \end{bmatrix}.$$

Recall that, given $X = x$, the mode residual, $\epsilon = Y - \boldsymbol{\beta}^{\mathrm{T}}\tilde{X}$, follows a distribution specified by the pdf $g(\epsilon|x)$, and $f_{Y|X}(y|x) = g(y - \boldsymbol{\beta}^{\mathrm{T}}\tilde{x}|x)$. It follows that, for the first element in the 2×1 summand above, we have

$$\begin{aligned}
&E\left(\mathrm{IS}_j^{(1,0)}\Big| X_j\right)\\
=&E\left\{\frac{1}{2\pi}\int t\sin\left(\frac{Y_j - \boldsymbol{\beta}^{\mathrm{T}}\tilde{W}_j}{h}t\right)\phi_K(t)\tau(\beta_1 t/h)dt\,\bigg|\,X_j\right\}\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau(\beta_1 t/h)\int \sin\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau(\beta_1 t/h)\phi_U(-\beta_1 t/h)\sin(\epsilon t/h)dt\,d\epsilon, \text{ by Lemma C.1,}\\
=&-\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\sin(\epsilon t/h)dt\,d\epsilon,\\
=&-h\cdot\frac{1}{2\pi}\int g(sh|X_j)\int t\phi_K(t)\sin(st)dt\,ds\\
=&-h\cdot\frac{1}{2\pi}\int\left\{g(0|X_j) + 0.5g^{(2)}(0|X_j)s^2h^2\right.\\
&\left.+6^{-1}g^{(3)}(0|X_j)s^3h^3 + 24^{-1}g^{(4)}(s^*|X_j)s^4h^4\right\}\cdot\\
&\int t\phi_K(t)\sin(st)dt\,ds, \text{ for some } s^* \text{ lying between 0 and } sh,\\
=&-0.5\mu_2 h^4 g^{(3)}(0|X_j) + O_p(h^5), \text{ by Lemma C.2.}
\end{aligned}$$

The second element in the 2×1 summand of $E(K_n|\boldsymbol{X})$ above is

$$E\left(\mathrm{IS}_j^{(1,0)}W_j + \mathrm{IC}_j^{(1,1)}\Big| X_j\right) = X_j E\left(\mathrm{IS}_j^{(1,0)}\Big| X_j\right) + E\left(\mathrm{IS}_j^{(1,0)}U_j\Big| X_j\right) + E\left(\mathrm{IC}_j^{(1,1)}\Big| X_j\right),$$

122

where the first expectation is derived above, and the latter two expectations sum to zero, as we show next. Following a similar elaboration of $E(\mathrm{IS}_j^{(1,0)}|X_j)$ above, one has

$$
\begin{aligned}
&E\left(\mathrm{IS}_j^{(1,0)}U_j\middle|X_j\right) + E\left(\mathrm{IC}_j^{(1,1)}\middle|X_j\right)\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau(\beta_1 t/h)\int u\sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon+\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int\cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau(\beta_1 t/h)\cdot(-1)\cdot\phi_U^{(1)}(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon+\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon,\ \text{by Lemma C.1,}\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t\phi_K(t)\left\{\tau(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h)+\tau^{(1)}(\beta_1 t/h)\right.\\
&\left.\phi_U(\beta_1 t/h)\right\}\cos(\epsilon t/h)dt\,d\epsilon,
\end{aligned}
$$

where, by Lemma C.3, the expression within the curly brackets is equal to zero. Hence, the summand of $E(K_n|\boldsymbol{X})$ is equal to $-0.5\mu_2 h^4 g^{(3)}(0|X_j)\tilde{X}_j + O_p(h^5)$. This proves (C.1).

Part (II): Show (C.4) and (C.7).

By (A.7), the order of $\mathrm{Var}(K_n|\boldsymbol{X})$ is determined by

$$
\frac{1}{nh^4}E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2\begin{bmatrix}1 & W_j\\ W_j & W_j^2\end{bmatrix}+\mathrm{IS}_j^{(1,0)}\mathrm{IC}_j^{(1,1)}\begin{bmatrix}0 & 1\\ 1 & 2W_j\end{bmatrix}+\left(\mathrm{IC}_j^{(1,1)}\right)^2\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right|X_j\right\}.
\tag{C.11}
$$

We next derive the six expectations involved in (C.11), also listed in (B.7) for proving Theorem 2. Among these six expectations,

$$
\begin{aligned}
E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2 W_j\right|X_j\right\} &= E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2 U_j\right|X_j\right\}+E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2\right|X_j\right\}X_j,\\
E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2 W_j^2\right|X_j\right\} &= E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2 U_j^2\right|X_j\right\}+2E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2 U_j\right|X_j\right\}X_j\\
&\quad+E\left\{\left.\left(\mathrm{IS}_j^{(1,0)}\right)^2\right|X_j\right\}X_j^2.
\end{aligned}
$$

It can be shown that, provided that the first two moments of U exist, $E\{(\mathrm{IS}_j^{(1,0)})^2 U_j|X_j\}$ and $E\{(\mathrm{IS}_j^{(1,0)})^2 U_j^2|X_j\}$ have the same rate as $E\{(\mathrm{IS}_j^{(1,0)})^2|X_j\}$. Hence,

123

$E\{(\text{IS}_j^{(1,0)})^2 W_j | X_j\}$ and $E\{(\text{IS}_j^{(1,0)})^2 W_j^2 | X_j\}$ have the same rate as $E\{(\text{IS}_j^{(1,0)})^2 | X_j\}$. Similarly, $E(\text{IS}_j^{(1,0)} \text{IC}_j^{(1,1)} W_j | X_j)$ is of the same order as $E(\text{IS}_j^{(1,0)} \text{IC}_j^{(1,1)} | X_j)$, which, according to the Cauchy-Schwarz inequality, is bounded from above by $\sqrt{E\{(\text{IS}_j^{(1,0)})^2 | X_j\} E\{(\text{IC}_j^{(1,0)})^2 | X_j\}}$. Hence, we only need to focus on $E\{(\text{IS}_j^{(1,0)})^2 | X_j\}$ and $E\{(\text{IC}_j^{(1,0)})^2 | X_j\}$ next.

First, by (A.4),

$$
\begin{aligned}
E\left\{\left(\text{IS}_j^{(1,0)}\right)^2 \middle| X_j = x\right\} \\
= E\left[\left\{\frac{1}{2\pi} \int t \sin\left(\frac{Y_j - \boldsymbol{\beta}_0^{\mathrm{T}} \tilde{W}_j}{h} t\right) \frac{\phi_K(t)}{\phi_U(\beta_1 t/h)} dt\right\}^2 \middle| X_j = x\right] \\
= \int f_U(u) \int \left\{\frac{1}{2\pi} \int t \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) \frac{\phi_K(t)}{\phi_U(\beta_1 t/h)} dt\right\}^2 g(\epsilon | x) d\epsilon du \\
\leq B_g h \int f_U(u) \int |-\mathcal{F}_{1,0}(s)|^2 ds du \\
= B_g h \int |\mathcal{F}_{1,0}(s)|^2 ds, \quad\quad\quad (\text{C.12})
\end{aligned}
$$

where $B_g = \sup_x \sup_\epsilon g(\epsilon | x)$. When U is ordinary smooth of order b, by Definition 1 in the Chapter 2, there exists a constant $M > 0$ such that

$$
\begin{aligned}
\int |\mathcal{F}_{1,0}(s)|^2 ds = \frac{1}{2\pi} \int t^2 \frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)} dt, \text{ by the Parseval's Theorem,} \quad (\text{C.13}) \\
= \frac{1}{2\pi} \int_{|\beta_1 t| \leq Mh} t^2 \frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)} dt + \frac{1}{2\pi} \int_{|\beta_1 t| > Mh} t^2 \frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)} dt \\
\leq \frac{1}{2\pi} \left\{\inf_{|\beta_1 t| \leq Mh} |\phi_U(\beta_1 t/h)|\right\}^{-2} \int_{|\beta_1 t| \leq Mh} t^2 \phi_K^2(t) dt + \quad (\text{C.14}) \\
\frac{1}{2\pi} \int_{|\beta_1 t| > Mh} t^2 \frac{\phi_K^2(t)}{(c^2/4)|\beta_1 t/h|^{-2b}} dt \\
= O(h^{-2b}), \text{ under Condition (O4).}
\end{aligned}
$$

Hence, for ordinary smooth U, (C.12) suggests $E\{(\text{IS}_j^{(1,0)})^2 | X_j\} = O(h^{1-2b})$. When U is super smooth of order b, by Definition 2 in the Chapter 2, following (F.20), one

124

has,

$$\int |\mathcal{F}_{1,0}(s)|^2 ds$$

$$\leq \frac{1}{2\pi} \left\{ \inf_{|\beta_1 t| \leq Mh} |\phi_U(\beta_1 t/h)| \right\}^{-2} \int_{|\beta_1 t| \leq Mh} t^2 \phi_K^2(t) dt$$

$$+ \frac{1}{2\pi} \int_{Mh < |\beta_1 t| \leq |\beta_1|} t^2 \frac{\phi_K^2(t)}{d_0^2 |\beta_1 t/h|^{2b_0} \exp(-2|\beta_1 t/h|^b/d_2)} dt$$

$$\leq C_1 \int_{|\beta_1 t| \leq Mh} t^2 \phi_K^2(t) dt + C_2 h^{2b_0} \exp(2|\beta_1|^b h^{-b}/d_2) \int_{Mh < |\beta_1 t| \leq |\beta_1|} t^2 \frac{\phi_K^2(t)}{|t|^{2b_0}} dt$$

$$= O\{h^{2b_2} \exp(2|\beta_1|^b h^{-b}/d_2)\},$$

where $b_2 = b_0 I(b_0 < 0.5)$. Hence, for super smooth U,

$$E\{(\mathrm{IS}_j^{(1,0)})^2 | X_j\} = O\{h^{1+2b_2} \exp(2|\beta_1|^b h^{-b}/d_2)\}.$$

Similarly,

$$E\left\{ \left(\mathrm{IC}_j^{(1,1)}\right)^2 \middle| X_j = x \right\} \leq B_g h \int f_U(u) \int |\mathcal{F}_{1,1}(s)|^2 ds du = B_g h \int |\mathcal{F}_{1,1}(s)|^2 ds. \tag{C.15}$$

When U is ordinary smooth of order b, using the definition of $\tau^{(1)}(s)$,

$$\int |\mathcal{F}_{1,1}(s)|^2 ds$$

$$= \frac{1}{2\pi} \int t^2 \phi_K^2(t) \left\{ \frac{\phi_U^{(1)}(\beta_1 t/h)}{\phi_U^2(\beta_1 t/h)} \right\}^2 dt, \text{ next use Definition 1 and Condition (O1),} \tag{C.16}$$

$$\leq \frac{1}{2\pi} \sup_{|\beta_1 t| \leq Mh} |\tau^{(1)}(\beta_1 t/h)|^2 \int_{|\beta_1 t| \leq Mh} t^2 \phi_K^2(t) dt$$

$$+ \frac{1}{2\pi} \int_{|\beta_1 t| > Mh} t^2 \phi_K^2(t) \left(\frac{2cb|\beta_1 t/h|^{-b-1}}{c^2 |\beta_1 t/h|^{-2b}/4} \right)^2 dt$$

$$= O(h^{2-2b}), \text{ under Condition (O4).}$$

Hence, for ordinary smooth U, by (C.15), $E\{(\mathrm{IC}_j^{(1,1)})^2 | X_j = x\} = O(h^{3-2b})$. When U

is super smooth of order b, following (C.16) and using Defition 2 and Condition (S1),

$$
\int |\mathcal{F}_{1,1}(s)|^2 ds \leq \frac{1}{2\pi} \sup_{|\beta_1 t| \leq Mh} |\tau^{(1)}(\beta_1 t/h)|^2 \int_{|\beta_1 t| \leq Mh} t^2 \phi_K^2(t) dt
$$
$$
+ \frac{1}{2\pi} \int_{Mh < |\beta_1 t| \leq |\beta_1|} t^2 \phi_K^2(t) \left\{ \frac{d_1^{(1)} |\beta_1 t/h|^{b_1^{(1)}} \exp(-|\beta_1 t/h|^b/d_2)}{d_0^2 |\beta_1 t/h|^{2b_0} \exp(-2|\beta_1 t/h|^b/d_2)} \right\}^2 dt
$$
$$
= O \left\{ h^{2b_3} \exp(2|\beta_1|^b h^{-b}/d_2) \right\},
$$

where $b_3 = (2b_0 - b_1^{(1)})I(2b_0 - b_1^{(1)} < 0.5)$. Hence, for super smooth U, (C.15) indicates $E\{(\mathrm{IC}_j^{(1,1)})^2 | X_j = x\} = O\{h^{2b_3+1} \exp(2|\beta_1|^b h^{-b}/d_2)\}$.

Having the rates of the expectations in (B.7) derived above under the two types of measurement error distributions, we are now ready to return to (C.11) and conclude the rate of $\mathrm{Var}(K_n|\boldsymbol{X})$. In particular, (C.11) implies (C.4) if U is ordinary smooth, and it implies (C.7) if U is super smooth.

Up to this point, we have established the rates of $E(K_n|\boldsymbol{X})$ and $\mathrm{Var}(K_n|\boldsymbol{X})$. One will see later that the theme used above to establish these rates is repeatedly used to derive the rates of $E(J_n|\boldsymbol{X})$, $\mathrm{Var}(J_n|\boldsymbol{X})$, $E(L_n|\boldsymbol{X})$, and $\mathrm{Var}(L_n|\boldsymbol{X})$. Before moving forward to proving the next result, we shall summarize two patterns learnt from Parts (I) and (II) that can be helpful for later parts of the proof. The first pattern pertains to deriving the rate of the mean of K_n, or J_n, or L_n. As seen in Part (I), the order of such mean mostly depends on $E(\mathrm{IC}_j^{(\ell_1,\ell_2)} W_j^{\ell_3} | X_j)$ and $E(\mathrm{IS}_j^{(\ell_1,\ell_2)} W_j^{\ell_3} | X_j)$, for some nonnegative integers ℓ_1, ℓ_2 and ℓ_3. These expectations are derived in the same way for ordinary smooth U and for super smooth U. The second pattern relates to deriving the rate of the variance of K_n, or J_n, or L_n. Such rate mainly depends on $E\{(\mathrm{IC}_j^{(\ell_1,\ell_2)})^2 W_j^{\ell_3} | X_j)$ and $E\{(\mathrm{IS}_j^{(\ell_1,\ell_2)})^2 W_j^{\ell_3} | X_j)$, for some nonnegative integers ℓ_1, ℓ_2 and ℓ_3. Moreover, raising the power ℓ_3 from zero does not affect the rate, hence one may focus on $E\{(\mathrm{IC}_j^{(\ell_1,\ell_2)})^2 | X_j)$ and $E\{(\mathrm{IS}_j^{(\ell_1,\ell_2)})^2 | X_j)$. As seen in Part (II), each of these expectations is bounded from above by $B_g h \int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$. Discussions of the rate of $\int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$ need to be carried out separately for ordinary smooth U and super smooth U. This rate mostly relies on ℓ_2, and the rate is derived under

the assumption that $\int t^{2\ell_1}\phi_K^2(t)dt < \infty$, besides other assumptions. If U is ordinary smooth, for $\ell_2 < \ell_2'$, $\int |\mathcal{F}_{\ell_1',\ell_2'}(s)|^2 ds / \int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds = o(1)$, where ℓ_1 and ℓ_1' can be the same or different. If U is super smooth, the comparison between the rate of $\int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$ and that of $\int |\mathcal{F}_{\ell_1',\ell_2'}(s)|^2 ds$ is less clear-cut, and requires more careful inspection.

Part (III): Show (C.2).

By (A.9), one has

$$
\begin{aligned}
&E(J_n|\boldsymbol{X}) \\
&= -\frac{1}{nh^3}\sum_{j=1}^n \left\{ E\left(-\mathrm{IC}_j^{(2,0)} \begin{bmatrix} 1 & W_j \\ W_j & W_j^2 \end{bmatrix} \middle| X_j \right) + E\left(\mathrm{IS}_j^{(2,1)} \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix} \middle| X_j \right) \right. \quad\quad (C.17) \\
&\quad \left. + E\left(\mathrm{IC}_j^{(2,2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \middle| X_j \right) \right\} \\
&= -\frac{1}{nh^3}\sum_{j=1}^n \begin{bmatrix} \eta_{j,1} & \eta_{j,2} \\ \eta_{j,2} & \eta_{j,3} \end{bmatrix}, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (C.18)
\end{aligned}
$$

where

$$
\begin{cases}
\eta_{j,1} &= -E\left(\mathrm{IC}_j^{(2,0)} \middle| X_j \right), \\
\eta_{j,2} &= -E\left(\mathrm{IC}_j^{(2,0)} \middle| X_j \right) X_j - E\left(\mathrm{IC}_j^{(2,0)} U_j \middle| X_j \right) + E\left(\mathrm{IS}_j^{(2,1)} \middle| X_j \right), \quad (C.19) \\
\eta_{j,3} &= -E\left(\mathrm{IC}_j^{(2,0)} W_j^2 \middle| X_j \right) + 2E\left(\mathrm{IS}_j^{(2,1)} W_j \middle| X_j \right) + E\left(\mathrm{IC}_j^{(2,2)} \middle| X_j \right).
\end{cases}
$$

To derive (C.18), in what follows, we first show that $\eta_{j,2} = \eta_{j,1}X_j$ and $\eta_{j,3} = \eta_{j,1}X_j^2$. Then we focus on deriving $\eta_{j,1}$.

By (C.19), $\eta_{j,2}$ contains two extra expectations besides the one defined as $\eta_{j,1}$.

These two expectations together give

$$
\begin{aligned}
&- E\left(\mathrm{IC}_j^{(2,0)} U_j \middle| X_j\right) + E\left(\mathrm{IS}_j^{(2,1)} \middle| X_j\right) \\
&= -\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \int u \cos\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon + \\
&\quad \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \int \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon \\
&= -\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \phi_U^{(1)}(-\beta_1 t/h) \sin(\epsilon t/h) dt\, d\epsilon + \\
&\quad \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \phi_U(-\beta_1 t/h) \sin(\epsilon t/h) dt\, d\epsilon, \text{ by Lemma C.1,} \\
&= \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \left\{ \tau(\beta_1 t/h) \phi_U^{(1)}(\beta_1 t/h) \right. \\
&\quad \left. + \tau^{(1)}(\beta_1 t/h) \phi_U(\beta_1 t/h) \right\} \sin(\epsilon t/h) dt\, d\epsilon,
\end{aligned}
$$

where, by Lemma C.3, the part of the integrand insides the curly brackets is equal to zero. Therefore, $\eta_{j,2} = \eta_{j,1} X_j$.

Elaborating $\eta_{j,3}$ in (C.19) yields

$$
\begin{aligned}
\eta_{j,3} &= -E\left\{ \mathrm{IC}_j^{(2,0)} U_j^2 \middle| X_j \right\} + 2E\left\{ \mathrm{IS}_j^{(2,1)} U_j \middle| X_j \right\} + E\left\{ \mathrm{IC}_j^{(2,2)} \middle| X_j \right\} + \\
&\quad 2\left\{ -E\left(\mathrm{IC}_j^{(2,0)} U_j \middle| X_j \right) + E\left(\mathrm{IS}_j^{(2,1)} \middle| X_j \right) \right\} + \eta_{j,1} X_j^2 \\
&= -E\left\{ \mathrm{IC}_j^{(2,0)} U_j^2 \middle| X_j \right\} + 2E\left\{ \mathrm{IS}_j^{(2,1)} U_j \middle| X_j \right\} + E\left\{ \mathrm{IC}_j^{(2,2)} \middle| X_j \right\} + \eta_{j,1} X_j^2,
\end{aligned}
$$

where, to reach the last step, we drop the part in the first step that reduces to zero according to the preceding derived results in regard to $\eta_{j,2}$. As for the remaining

128

three expectations in front of $\eta_{j,1} X_j^2$ above, we have

$$
- E\left\{\mathrm{IC}_j^{(2,0)} U_j^2 \Big| X_j\right\} + 2 E\left\{\mathrm{IS}_j^{(2,1)} U_j \Big| X_j\right\} + E\left\{\mathrm{IC}_j^{(2,2)} \Big| X_j\right\}
$$

$$
= -\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \int u^2 \cos\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon +
$$

$$
2 \cdot \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \int u \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon +
$$

$$
\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(2)}(\beta_1 t/h) \int \cos\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon
$$

$$
= \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \phi_U^{(2)}(-\beta_1 t/h) \cos(\epsilon t/h) dt\, d\epsilon +
$$

$$
2 \cdot \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \cdot (-1) \cdot \phi_U^{(1)}(-\beta_1 t/h) \cos(\epsilon t/h) dt\, d\epsilon +
$$

$$
\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau^{(2)}(\beta_1 t/h) \phi_U(-\beta_1 t/h) \cos(\epsilon t/h) dt\, d\epsilon, \text{ by Lemma C.1,}
$$

$$
= \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \left\{\tau(\beta_1 t/h) \phi_U^{(2)}(\beta_1 t/h) + 2\tau^{(1)}(\beta_1 t/h) \phi_U^{(1)}(\beta_1 t/h) + \right.
$$

$$
\left. \tau^{(2)}(\beta_1 t/h) \phi_U(\beta_1 t/h)\right\} \cos(\epsilon t/h) dt\, d\epsilon,
$$

in the last step of which we use the assumption that $\phi_U(\cdot)$ is even, thus $\phi_U^{(1)}(\cdot)$ is odd and $\phi_U^{(2)}(\cdot)$ is even. By Lemma C.3, the part inside the curly brackets in the last step reduces to zero. Hence, $\eta_{j,3} = \eta_{j,1} X_j^2$.

Now that (C.19) only hinges on $\eta_{j,1}$, we shall derive this quantity next,

$$
\eta_{j,1} = -E\left(\mathrm{IC}_j^{(2,0)} \Big| X_j\right)
$$

$$
= -\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \int \cos\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u) du\, dt\, d\epsilon
$$

$$
= -\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \tau(\beta_1 t/h) \phi_U(-\beta_1 t/h) \cos(\epsilon t/h) dt\, d\epsilon, \text{ by Lemma C.1,}
$$

$$
= \frac{1}{2\pi} \int g(\epsilon|X_j) \int t^2 \phi_K(t) \cos(\epsilon t/h) dt\, d\epsilon, \text{ since } \tau(s) = -1/\phi_U(s),
$$

$$
= h \cdot \frac{1}{2\pi} \int g(hs|X_j) \int t^2 \phi_K(t) \cos(st) dt\, ds
$$

$$
= h \cdot \frac{1}{2\pi} \int \left\{g(0|X_j) + g^{(2)}(0|X_j) h^2 s^2/2 + O_p(h^3)\right\} \int t^2 \phi_K(t) \cos(st) dt\, ds
$$

$$
= h \cdot g(0|X_j) \cdot \frac{1}{2\pi} \int \int t^2 \phi_K(t) \cos(st) dt\, ds +
$$

$$
\frac{h^3}{2} g^{(2)}(0|X_j) \cdot \frac{1}{2\pi} \int \int t^2 \phi_K(t) \cos(st) dt\, s^2\, ds + O_p(h^4)
$$

$$
= -h^3 g^{(2)}(0|X_j) + O_p(h^4), \text{ by Lemma C.2.}
$$

129

To this end, one can conclude that

$$
\begin{cases}
\eta_{j,1} &= -h^3 g^{(2)}(0|X_j)\{1 + o_p(1)\}, \\
\eta_{j,2} &= -h^3 g^{(2)}(0|X_j)X_j\{1 + o_p(1)\}, \\
\eta_{j,3} &= -h^3 g^{(2)}(0|X_j)X_j^2\{1 + o_p(1)\}.
\end{cases}
\tag{C.20}
$$

Using (C.20) in (C.18) leads to (C.2).

Part (IV): Show (C.5) and (C.8).

By (A.9), the order of $\mathrm{Var}(J_n)$ is determined by the rates of $E\{(\mathrm{IC}_j^{(2,0)})^2 W_j^k|X_j\}$ for $k = 0,1,2,3,4$, the rates of $E\{(\mathrm{IS}_j^{(2,1)})^2 W_j^k|X_j\}$ for $k = 0,1,2$, and that of $E\{(\mathrm{IC}_j^{(2,2)})^2|X_j\}$. It can be shown that, if $E(U^4) < \infty$, $E\{(\mathrm{IC}_j^{(2,0)})^2 W_j^k|X_j\}$, for $k = 1,2,3,4$, have the same rate as that of $E\{(\mathrm{IC}_j^{(2,0)})^2|X_j\}$, and $E\{(\mathrm{IS}_j^{(2,1)})^2 W_j^k|X_j\}$, for $k = 1,2$, have the same rate as that of $E\{(\mathrm{IS}_j^{(2,1)})^2|X_j\}$. In what follows, we focus on deriving the rates of $E\{(\mathrm{IC}_j^{(2,0)})^2|X_j\}$, $E\{(\mathrm{IS}_j^{(2,1)})^2|X_j\}$, and $E\{(\mathrm{IC}_j^{(2,2)})^2|X_j\}$.

We first find an upper bound for each of these three expectations as follows,

$$
\begin{aligned}
& E\left\{ \left(\mathrm{IC}_j^{(2,0)}\right)^2 \middle| X_j \right\} \\
&= \int f_U(u) \int \left\{ \frac{1}{2\pi} \int t^2 \cos\left(\frac{\epsilon - \beta_1 u}{h}t\right) \phi_K(t)\tau(\beta_1 h/t)dt \right\}^2 g(\epsilon|X_j)d\epsilon\, du \\
&= h \int f_U(u) \int |-\mathcal{F}_{2,0}(s)|^2 g(\beta_1 u + sh|X_j)ds\, du \\
&\leq B_g h \int |\mathcal{F}_{2,0}(s)|^2 ds.
\end{aligned}
$$

Similarly,

$$
E\left\{ \left(\mathrm{IS}_j^{(2,1)}\right)^2 \middle| X_j \right\} \leq B_g h \int |\mathcal{F}_{2,1}(s)|^2 ds,
$$
$$
E\left\{ \left(\mathrm{IC}_j^{(2,2)}\right)^2 \middle| X_j \right\} \leq B_g h \int |\mathcal{F}_{2,2}(s)|^2 ds.
$$

When U is ordinary smooth, recalling the pattern stated at the end of Part (II), $\int |\mathcal{F}_{2,0}(s)|^2 ds$ dominates $\int |\mathcal{F}_{2,1}(s)|^2 ds$ and $\int |\mathcal{F}_{2,2}(s)|^2 ds$, and $\int |\mathcal{F}_{2,0}(s)|^2 ds$ is of the same order as $\int |\mathcal{F}_{1,0}(s)|^2 ds$, which we looked into in Part (II). Using the

130

results regarding $\int |\mathcal{F}_{1,0}(s)|^2 ds$, we have $\int |\mathcal{F}_{2,0}(s)|^2 ds = O(h^{-2b})$ if U is ordinary smooth, and thus $E\{(\mathrm{IC}_j^{(2,0)})^2|X_j\} = O(h^{1-2b})$, which dominates $E\{(\mathrm{IS}_j^{(2,1)})^2|X_j\}$, and $E\{(\mathrm{IC}_j^{(2,2)})^2|X_j\}$. These rates lead to (C.5).

If U is super smooth, under Conditions S, similar to the proof in Part (II), one can show that $E\{(\mathrm{IC}_j^{(2,0)})^2|X_j\} = O\{h^{1+2b_2} \exp(2|\beta_1|^b h^{-b}/d_2)\}$, $E\{(\mathrm{IS}_j^{(2,1)})^2|X_j\} = O\{h^{1+2b_3} \exp(2|\beta_1|^b h^{-b}/d_2)\}$, and $E\{(\mathrm{IC}_j^{(2,2)})^2|X_j\} = O\{h^{1+2b_4} \exp(2|\beta_1|^b h^{-b}/d_2)\}$, where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)}) I(2b_0 - b_1^{(1)} < 0.5)$, and $b_4 = \min\{(2b_0 - b_1^{(2)}) I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)}) I(3b_0 - 2b_1^{(1)} < 0.5)\}$. Using these rates one can establish (C.8).

Part (V): Show (C.3).

Because $\|\boldsymbol{\beta}^* - \boldsymbol{\beta}\| \le c r_n$, under conditions that guarantee boundedness of $\mathrm{IS}_j^{(4,\ell_2)}$ and $\mathrm{IC}_j^{(4,\ell_2)}$, for $\ell_2 = 0, 1, 2, 3, 4$, L_n evaluated at $\boldsymbol{\beta}^*$ and L_n evaluated at $\boldsymbol{\beta}$ are of the same order. Hence, we focus on studying L_n evaluated at $\boldsymbol{\beta}$ in the sequel.

Given X_j, the first entry in the 2×1 vector in (A.10) has mean given by

$$(1 + X_j^2) E\left(\mathrm{IS}_j^{(3,0)}\Big| X_j\right) + E\left(U_j^2 \mathrm{IS}_j^{(3,0)}\Big| X_j\right) + 2X_j E\left(U_j \mathrm{IS}_j^{(3,0)}\Big| X_j\right) +$$

$$2X_j E\left(\mathrm{IC}_j^{(3,1)}\Big| X_j\right) + 2E\left(U_j \mathrm{IC}_j^{(3,1)}\Big| X_j\right) - E\left(\mathrm{IS}_j^{(3,2)}\Big| X_j\right)$$

$$= (1 + X_j^2) E\left(\mathrm{IS}_j^{(3,0)}\Big| X_j\right) + 2X_j \left\{ E\left(U_j \mathrm{IS}_j^{(3,0)}\Big| X_j\right) + E\left(\mathrm{IC}_j^{(3,1)}\Big| X_j\right) \right\} + \quad (\text{C.21})$$

$$E\left(U_j^2 \mathrm{IS}_j^{(3,0)}\Big| X_j\right) + 2E\left(U_j \mathrm{IC}_j^{(3,1)}\Big| X_j\right) - E\left(\mathrm{IS}_j^{(3,2)}\Big| X_j\right). \quad (\text{C.22})$$

We next show that the two terms together inside the curly brackets in (C.21) is zero, so is the three terms together in (C.22).

By the definition of $\mathrm{IS}_j^{(3,0)}$ and $\mathrm{IC}_j^{(3,1)}$, the two terms in (C.21) can be elaborated

as follows,

$$
\begin{aligned}
&E\left(U_j\mathrm{IS}_j^{(3,0)}\Big|\,X_j\right) + E\left(\mathrm{IC}_j^{(3,1)}\Big|\,X_j\right)\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\int u\sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon+\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int \cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)(-1)\phi_U^{(1)}(-\beta_1 t/h)\cos(\epsilon t/h)\,dt\,d\epsilon+\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)\,dt\,d\epsilon,\ \text{by Lemma C.1,}\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\left\{\tau(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h)\right.\\
&\left.+\ \tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\right\}\cos(\epsilon t/h)\,dt\,d\epsilon\\
=&0,
\end{aligned}
$$

where the last step is reached because the terms inside the curly brackets in the second-to-the-last step is equal to zero by Lemma C.3. For (C.22), we have the following elaboration by the definitions of $\mathrm{IS}_j^{(3,0)}$, $\mathrm{IC}_j^{(3,1)}$, and $\mathrm{IS}_j^{(3,2)}$,

$$
\begin{aligned}
&E\left(U_j^2\mathrm{IS}_j^{(3,0)}\Big|\,X_j\right) + 2E\left(U_j\mathrm{IC}_j^{(3,1)}\Big|\,X_j\right) - E\left(\mathrm{IS}_j^{(3,2)}\Big|\,X_j\right)\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\int u^2\sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon+\\
&2\cdot\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int u\cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon-\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(2)}(\beta_1 t/h)\int \sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon\\
=&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\cdot(-1)\cdot\phi_U^{(2)}(-\beta_1 t/h)\sin(\epsilon t/h)\,dt\,d\epsilon+\\
&2\cdot\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U^{(1)}(-\beta_1 t/h)\sin(\epsilon t/h)\,dt\,d\epsilon-\\
&\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\tau^{(2)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\sin(\epsilon t/h)\,dt\,d\epsilon,\ \text{by Lemma C.1,}\\
=&-\frac{1}{2\pi}\int g(\epsilon|X_j)\int t^3\phi_K(t)\sin(\epsilon t/h)\cdot\\
&\left\{\tau(\beta_1 t/h)\phi_U^{(2)}(\beta_1 t/h) + 2\tau^{(1)}(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h) + \tau^{(2)}(\beta_1 t/h)\phi_U(\beta_1 t/h)\right\}\,dt\,d\epsilon\\
=&0,\ \text{by Lemma C.3.}
\end{aligned}
$$

In conclusion, the mean of the first entry of the 2×1 vector in (A.10) given X_j boils down to $(1 + X_j^2)$ times the following expectation,

$$
\begin{aligned}
&E\left(\mathrm{IS}_j^{(3,0)}\Big|\, X_j\right) \\
={}&\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^3 \phi_K(t) \tau(\beta_1 t/h) \int \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u)\, du\, dt\, d\epsilon \\
={}&\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^3 \phi_K(t) \tau(\beta_1 t/h) \phi_U(-\beta_1 t/h) \sin(\epsilon t/h)\, dt\, d\epsilon, \text{ by Lemma C.1,} \\
={}&-\frac{1}{2\pi} \int g(\epsilon|X_j) \int t^3 \phi_K(t) \sin(\epsilon t/h)\, dt\, d\epsilon, \text{ since } \tau(s) = -1/\phi_U(s), \\
={}&-h \cdot \frac{1}{2\pi} \int g(hs|X_j) \int t^3 \phi_K(t) \sin(st)\, dt\, ds \\
={}&-h \cdot \frac{1}{2\pi} \int \left\{ g(0|X_j) + g^{(2)}(0|X_j) h^2 s^2/2 + g^{(3)}(s^*|X_j) h^3 s^3/6 \right\} \\
&\int t^3 \phi_K(t) \sin(st)\, dt\, ds, \\
={}&-\frac{h^4}{6} \cdot \frac{1}{2\pi} \int g^{(3)}(s^*|X_j) s^3 \int t^3 \phi_K(t) \sin(st)\, dt\, ds,
\end{aligned}
$$

where we use the third-order Taylor expansion of $g(sh|X_j)$ around zero in the second to the last step, with s^* lying between 0 and sh; then we use identities in Lemma C.2 to simplify the integrals and keep the only part that is not necessarily equal to zero. Assuming $g^{(3)}(\cdot|X_j)$ bounded and using Lemma C.2, the last expression above suggests that $E(\mathrm{IS}_j^{(3,0)}|X_j)$ is of order $O(h^4)$. Hence, the first component of the summand in (A.10) is of order $O(h^4)$.

Given X_j, the mean of the second component of the summand in (A.10) can be derived similarly as above, using Lemmas C.1, C.2, and C.3. And one can show that this expectation is equal to $X_j(1 + X_j^2)E(\mathrm{IS}_j^{(3,0)}|X_j)$, and thus is of the same order of $E(\mathrm{IS}_j^{(3,0)}|X_j)$ derived above. Therefore, the summand in (A.10) has expectation of order $O(h^4)$. This proves (C.3), provided Conditions (C2) and (C3) hold.

Part (VI): Show (C.6) and (C.9).

By (A.10), the order of $\mathrm{Var}(L_n)$ is determined by the orders of $E\{(\mathrm{IS}_j^{(3,0)})^2|X_j\}$, $E\{(\mathrm{IC}_j^{(3,1)})^2|X_j\}$, $E\{(\mathrm{IS}_j^{(3,2)})^2|X_j\}$, and $E\{(\mathrm{IC}_j^{(3,3)})^2|X_j\}$. Following similar arguments

as those in Parts (II) and (IV), one can show that these expectations are bounded from above by $B_g h \int |\mathcal{F}_{3,\ell_2}|^2 ds$, for $\ell_2 = 0, 1, 2, 3$, respectively. If U is ordinary smooth, $\int |\mathcal{F}_{3,0}|^2 ds$ dominates the other three according to the patterns pointed out at the end of Part (II), which of the same order as $\int |\mathcal{F}_{1,0}|^2 ds$ derived there. Therefore, using results from Part (II), we have the order of $\mathrm{Var}(L_n)$ being $O_p\{1/(nh^{7+2b})\}$ if U is ordinary smooth. This proves (C.6). If U is super smooth, one can show that

$$E\{(\mathrm{IS}_j^{(3,0)})^2 | X_j\} = O\{h^{1+2b_2} \exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\mathrm{IC}_j^{(3,1)})^2 | X_j\} = O\{h^{1+2b_3} \exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\mathrm{IS}_j^{(3,2)})^2 | X_j\} = O\{h^{1+2b_4} \exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\mathrm{IC}_j^{(3,3)})^2 | X_j\} = O\{h^{1+2b_5} \exp(2|\beta_1|^b h^{-b}/d_2)\},$$

where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)}) I(2b_0 - b_1^{(1)} < 0.5)$, $b_4 = \min\{(2b_0 - b_1^{(2)}) I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)}) I(3b_0 - 2b_1^{(1)} < 0.5)\}$, and $b_5 = \min\{(2b_0 - b_1^{(3)}) I(2b_0 - b_1^{(3)} < 0.5), (3b_0 - b_1^{(1)} - b_1^{(2)}) I(3b_0 - b_1^{(1)} - b_1^{(2)} < 0.5), (4b_0 - b_1^{(1)} - 2b_1^{(2)}) I(4b_0 - b_1^{(1)} - 2b_1^{(2)} < 0.5)\}$. Putting these rates together gives (C.9).

$\square$

Lemma C.1. *If the density of U, $f_U(u)$, is an even function, then the following identities hold, where a and b are two constants,*

$$E\{\sin(a + bU)\} = \phi_U(b) \sin(a), \quad E\{\cos(a + bU)\} = \phi_U(b) \cos(a),$$

$$E\{U \sin(a + bU)\} = -\phi_U^{(1)}(b) \cos(a), \quad E\{U \cos(a + bU)\} = \phi_U^{(1)}(b) \sin(a),$$

$$E\{U^2 \sin(a + bU)\} = -\phi_U^{(2)}(b) \sin(a), \quad E\{U^2 \cos(a + bU)\} = -\phi_U^{(2)}(b) \cos(a),$$

$$E\{U^3 \sin(a + bU)\} = \phi_U^{(3)}(b) \cos(a), \quad E\{U^3 \cos(a + bU)\} = -\phi_U^{(3)}(b) \sin(a),$$

in which $\phi_U^{(\ell)}(b)$ is the ℓ-th derivatives of $\phi_U(b)$, for $\ell = 1, 2, 3$.

Proof. If $f_U(u)$ is an even function, its characteristic function reduces to

$$\phi_U(t) = \int e^{itu} f_U(u) du = \int \cos(tu) f_U(u) du. \tag{C.23}$$

134

in which Euler's formula is used to reach the last expression. It follows that

$$E\{\sin(a + bU)\} = \int \{\sin(a)\cos(bu) + \cos(a)\sin(bu)\} f_U(u)du$$

$$= \sin(a) \cdot \int \cos(bu) f_U(u)du, \text{ since } \sin(bu) f_U(u) \text{ is an odd function,}$$

$$= \phi_U(b)\sin(a), \text{ by (C.23).}$$

The second identity in Lemma C.1 can be shown by differentiating both sides of the first identity with respect to a. Each of the latter six identities can be shown by differentiating an earlier identity on both sides with respect to b. □

Lemma C.2. *Let $K(t)$ be a symmetric kernel and denote by $\phi_K(t)$ the Fourier transform of it. If $t^2 K(t)$ and $t^3 K^{(3)}(t)$ are integrable, $K^{(1)}(t) = o(1/t^4)$ and $K^{(2)}(t) = o(1/t^3)$, as $|t| \to \infty$, then*

$$\frac{1}{2\pi} \int \int t\sin(st)\phi_K(t)dts^3 ds = 3\mu_2, \text{ where } \mu_2 = \int t^2 K(t)dt, \tag{C.24}$$

$$\frac{1}{2\pi} \int \int t^3 \sin(st)\phi_K(t)dts^2 ds = 0, \tag{C.25}$$

$$\frac{1}{2\pi} \int \int t^3 \sin(st)\phi_K(t)dts^3 ds = -6, \tag{C.26}$$

$$\frac{1}{2\pi} \int \int t^2 \cos(st)\phi_K(t)dtds = 0, \tag{C.27}$$

$$\frac{1}{2\pi} \int \int t^2 \cos(st)\phi_K(t)dts^2 ds = -2. \tag{C.28}$$

Proof. To show (C.24), one first observes that

$$\frac{1}{2\pi} \int \int t\sin(st)\phi_K(t)dts^3 ds$$

$$= \frac{1}{i} \cdot \frac{1}{2\pi} \int \int te^{ist}\phi_K(t)dts^3 ds$$

$$= \frac{1}{i^2} \int s^3 \frac{\partial}{\partial s}\left\{\frac{1}{2\pi} \int e^{ist}\phi_K(t)dt\right\} ds$$

$$= -\int s^3 \frac{\partial}{\partial s}K(s)ds, \text{ by the Fourier inverse theorem,}$$

$$= -\int s^3 K^{(1)}(s)ds$$

$$= 3\int s^2 K(s)ds, \text{ since } K^{(1)}(s) = o(1/s^4) \text{ as } |s| \to \infty,$$

$$= 3\mu_2.$$

To show (C.25) and (C.26), one may first derive the common inner integral as
follows,

$$
\begin{aligned}
& \frac{1}{2\pi} \int t^3 \sin(st)\phi_K(t)dt \\
= & \frac{1}{i} \cdot \frac{1}{2\pi} \int t^3 e^{ist}\phi_K(t)dt \\
= & \frac{1}{i^4} \cdot \frac{1}{2\pi} \int \frac{\partial^3}{\partial s^3} e^{ist}\phi_K(t)dt \\
= & \frac{\partial^3}{\partial s^3} \left\{ \frac{1}{2\pi} \int e^{ist}\phi_K(t)dt \right\} \\
= & \frac{\partial^3}{\partial s^3} K(s), \text{ by the Fourier inversion theorem,} \\
= & K^{(3)}(s).
\end{aligned} \tag{C.29}
$$

Because $K(s)$ is an even function, $K^{(3)}(s)$ is an odd function. Hence, the integral
in (C.25) reduces to $\int s^2 K^{(3)}(s)ds = 0$. This proves (C.25). As for (C.26), one uses
(C.29) again to simplify the integral in (C.26) to be

$$
\begin{aligned}
& \int s^3 K^{(3)}(s)ds \\
= & -3\int s^2 K^{(2)}(s)ds, \text{ since } s^3 K^{(3)}(s) \text{ is integrable and } K^{(2)}(s) = o(1/s^3), \\
= & 6\int s K^{(1)}(s)ds, \text{ since } K^{(2)}(s) = o(1/s^3) \text{ and } K^{(1)}(s) = o(1/s^4), \\
= & -6\int K(s)ds, \\
= & -6.
\end{aligned}
$$

This proves (C.26).

Lastly, the common inner integral of (C.27) and (C.28) is

$$
\begin{aligned}
& \frac{1}{2\pi} \int t^2 \cos(st)\phi_K(t)dt \\
= & \frac{1}{2\pi} \int t^2 e^{ist}\phi_K(t)dt \\
= & \frac{1}{i^2} \cdot \frac{1}{2\pi} \int \frac{\partial^2}{\partial s^2} e^{ist}\phi_K(t)dt \\
= & -\frac{\partial^2}{\partial s^2} \left\{ \frac{1}{2\pi} \int e^{ist}\phi_K(t)dt \right\} \\
= & -\frac{\partial^2}{\partial s^2} K(s), \text{ by the Fourier inversion theorem,} \\
= & -K^{(2)}(s).
\end{aligned}
$$

136

Hence, the integral in (C.27) is equal to $-\int K^{(2)}(s)ds = 0$ since $K^{(1)}(s) = o(1/s^4)$ as $|s| \to \infty$. And the integral in (C.28) is equal to

$$-\int s^2 K^{(2)}(s)ds = 2\int sK^{(1)}(s)ds = -2\int K(s)ds = -2.$$

□

Lemma C.3. *Define* $\tau(t) = -1/\phi_U(t)$, *where* $\phi_U(t)$ *is a non-vanishing even charac-teristic function of* U *that is third-order continuously differentiable. Then*

$$\tau(t)\phi_U^{(1)}(t) + \tau^{(1)}(t)\phi_U(t) = 0,$$

$$\tau(t)\phi_U^{(2)}(t) + 2\tau^{(1)}(t)\phi_U^{(1)}(t) + \tau^{(2)}(t)\phi_U(t) = 0,$$

$$\tau(t)\phi_U^{(3)}(t) + 3\tau^{(1)}(t)\phi_U^{(2)}(t) + 3\tau^{(2)}(t)\phi_U^{(1)}(t) + \tau^{(3)}(t)\phi_U(t) = 0,$$

where $\tau^{(\ell)}(t)$ *and* $\phi_U^{(\ell)}(t)$ *are the* ℓ-*th derivatives of* $\tau(t)$ *and* $\phi_U(t)$, *respectively, for* $\ell = 0, 1, 2, 3$.

Proof. Starting from $\tau(t) = -1/\phi_U(t)$, one has

$$\tau^{(1)}(t) = \frac{\phi_U^{(1)}(t)}{\phi_U^2(t)},$$

$$\tau^{(2)}(t) = \frac{\phi_U^{(2)}(t)}{\phi_U^2(t)} - \frac{2\left\{\phi_U^{(1)}(t)\right\}^2}{\phi_U^3(t)},$$

$$\tau^{(3)}(t) = \frac{\phi_U^{(3)}(t)}{\phi_U^2(t)} - \frac{6\phi_U^{(1)}(t)\phi_U^{(2)}(t)}{\phi_U^3(t)} + \frac{6\left\{\phi_U^{(1)}(t)\right\}^3}{\phi_U^4(t)}.$$

Using these three derivatives, one can show the three identities in Lemma C.3 via straightforward algebra. □

Appendix D

Proof of Theorem 3

For completeness, we repeat the conditions regarding $f(\epsilon|x,t)$ and the covariate stated in chapter 5 below. For a scalar s, denote by $\tilde{s}$ the vector $(1,s)^{\mathrm{T}}$.

(C*1) The ℓ-th partial derivative of $f(\epsilon|x,t)$ with respect to (w.r.t.) ϵ, $f^{(\ell)}(\epsilon|x,t)$, is continuously differentiable around $\epsilon = 0$, for $\ell = 0,1,2,3$, and $f^{(1)}(0|x,t) = 0$, for all x and t.

(C*2) As $n \to \infty$, $n^{-1}\sum_{j=1}^{n} f(0|X_j,T_j)\tilde{Z}_j\tilde{Z}_j^{\mathrm{T}}$ and $n^{-1}\sum_{j=1}^{n} f^{(3)}(0|X_j,T_j)\tilde{Z}_j$ converge in probability, and $n^{-1}\sum_{j=1}^{n} f^{(2)}(0|X_j,T_j)\tilde{Z}_j\tilde{Z}_j^{\mathrm{T}}$ converges in probability to a negative definite matrix, where $\tilde{Z}_j^T = (\tilde{X}_j^T, \tilde{B}_j^T)$.

(C*3) As $n \to \infty$, $n^{-1}\sum_{j=1}^{n} \|\tilde{X}_j\|^4 = O_p(1)$, where $\|\cdot\|$ denotes the Euclidean norm.

(C*4) The unspecified smooth function $g(t)$ is rth continuously differentiable on $[0,1]$, where $r > 2$,

(C*5) $\{g(T) - \tilde{B}\boldsymbol{\gamma}\}h^{-2} = o_p(1)$, as $n \to \infty$.

Addition conditions on the characteristic function of U and conditions on $K(t)$ are stated next. The first set of conditions given next are needed for proving Theorems 1 and 2 when U is ordinary smooth.

Conditions O:

(O1) As $|t| \to \infty$, $cb|t|^{-b-1}/2 \le |\phi_U^{(1)}(t)| \le 2cb|t|^{-b-1}$.

138

(O2) As $|t| \to \infty$, $cb(b+1)|t|^{-b-2}/2 \le |\phi_U^{(2)}(t)| \le 2cb(b+1)|t|^{-b-2}$.

(O3) As $|t| \to \infty$, $cb(b+1)(b+2)|t|^{-b-3}/2 \le |\phi_U^{(3)}(t)| \le 2cb(b+1)(b+2)|t|^{-b-3}$.

(O4) $\int t^{2(2+b)} \phi_K^2(t) dt < \infty$, and $\int t^6 \phi_K^2(t) dt < \infty$.

The second set of conditions stated next are needed for proving Theorems 3 and 4 when U is super smooth.

Conditions S:

(S1) As $|t| \to \infty$, $d_0^{(1)}|t|^{b_0^{(1)}} \exp(-|t|^b/d_2) \le |\phi_U^{(1)}(t)| \le d_1^{(1)}|t|^{b_1^{(1)}} \exp(-|t|^b/d_2)$, for some positive constants $b_0^{(1)}$, $b_1^{(1)}$, $d_0^{(1)}$, and $d_1^{(1)}$, where $b_0^{(1)} \ge b_0$ and $b_1^{(1)} \ge b_1$.

(S2) As $|t| \to \infty$, $d_0^{(2)}|t|^{b_0^{(2)}} \exp(-|t|^b/d_2) \le |\phi_U^{(2)}(t)| \le d_1^{(2)}|t|^{b_1^{(2)}} \exp(-|t|^b/d_2)$, for some positive constants $b_0^{(2)}$, $b_1^{(2)}$, $d_0^{(2)}$, and $d_1^{(2)}$, where $b_0^{(2)} \ge b_0^{(1)}$ and $b_1^{(2)} \ge b_1^{(1)}$.

(S3) As $|t| \to \infty$, $d_0^{(3)}|t|^{b_0^{(3)}} \exp(-|t|^b/d_2) \le |\phi_U^{(3)}(t)| \le d_1^{(3)}|t|^{b_1^{(3)}} \exp(-|t|^b/d_2)$, for some positive constants $b_0^{(3)}$, $b_1^{(3)}$, $d_0^{(3)}$, and $d_1^{(3)}$, where $b_0^{(3)} \ge b_0^{(2)}$ and $b_1^{(3)} \ge b_1^{(2)}$.

(S4) The support of $\phi_K(t)$ is $[-1, 1]$,

(S5) $0 \le b^{(1)} - b_0 \le 1$.

Conditions N*:

(N*1) The expectation $E\{U^\ell f(\beta_1 U)\}$, for $\ell = 0, 1, 2$, exists.

(N*2) The expectation $E\left[\left\{f^{(3)}(0|X)X\right\}^2\right]$.

(N*3) As $|t| \to \infty$, $\phi_U(t)$ is either purely real or pure imaginary.

(N*4) There exist positive constants $\delta, e_1, e_2,$ and q such that $\phi_K(t) \le e_1(1-t)^q$ and $\phi_K(t) \ge e_2(1-t)^q$.

As in chapter 5, all integrations are over the entire real line $\mathbb{R}$ unless otherwise specified.

Theorem 3. *Under conditions (C^*1)–(C^*5) and conditions in Lemma C, there exists a maximizer of $Q_h^*(\boldsymbol{\gamma}, \boldsymbol{\beta})$, denoted by $\boldsymbol{\theta}_{CK} = (\hat{\boldsymbol{\gamma}}_{CK}, \hat{\boldsymbol{\beta}}_{CK})$, such that, as $n \to \infty$ and $h \to 0$,*

(i) when U follows an ordinary smooth distribution of order b, if $nh^{7+2b} \to 0$, then

$$\|\hat{\boldsymbol{\theta}}_{CK} - \boldsymbol{\theta}\| = O(h^2) + O_p\left(\sqrt{\frac{1}{nh^{3+2b}}}\right), \tag{D.1}$$

(ii) when U follows a super smooth distribution of order b,

if $\exp(2|\beta_1|^b h^{-b}/d_2))/(nh^{b_6}) \to 0$, where

$b_6 = \max\{3 - 2\min(b_2, b_3), 5 - 2\min(b_2, b_3, b_4), 7 - 2\min(b_2, b_3, b_4, b_5)\}$, in which b_ℓ, for $\ell = 2, 3, 4, 5$, are defined in Lemma C, then

$$\|\hat{\boldsymbol{\theta}}_{CK} - \boldsymbol{\theta}\| = O(h^2) + O_p\left\{\exp\left(\frac{|\beta_1|^b}{d_2 h^b}\right)\sqrt{\frac{1}{nh^{3-2\min(b_2,b_3)}}}\right\}, \tag{D.2}$$

Proof. Define a series of integrals that involve in the integrand the cosine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s) = -1/\phi_U(s)$ evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IC}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)} = \frac{1}{2\pi}\int t^{\ell_1}\cos\left(\frac{Y_j - \tilde{B}_j^T\boldsymbol{\gamma} - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta}}{h}t\right)\phi_K(t)\tau^{(\ell_2)}(\beta_1 t/h)dt, \tag{D.3}$$

for nonnegative integers ℓ_1 and ℓ_2. For example, one can show using Euler's formula that, under the assumption stated in the main article that $\phi_K(t)$ and $\phi_U(t)$ are even functions, the deconvoluting kernel evaluated at $(Y_j - \tilde{B}_j^T\boldsymbol{\gamma} - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta})/h$,

$$K^*\left(\frac{Y_j - \tilde{B}_j^T\boldsymbol{\gamma} - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta}}{h}\right) = \frac{1}{2\pi}\int\exp\left(-i\frac{Y_j - \tilde{B}_j^T\boldsymbol{\gamma} - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta}}{h}t\right)\frac{\phi_K(t)}{\phi_U(-\beta_1 t/h)}dt,$$

is equal to $-\mathrm{IC}_{j(\gamma,\beta)}^{(0,0)}$. Similarly define another series of integrals that involve in the integrand the sine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s)$

evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IS}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \sin\left(\frac{Y_j - \tilde{B}_j^T \boldsymbol{\gamma} - \tilde{W}_j^{\mathrm{T}} \boldsymbol{\beta}}{h} t\right) \phi_K(t) \tau^{(\ell_2)}(\beta_1 t/h) dt. \tag{D.4}$$

To derive the convergence rate of $(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}}, \hat{\boldsymbol{\beta}}_{\mathrm{CK}})$, the following derivatives of $\mathrm{IC}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)}$ and $\mathrm{IS}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)}$ with respect to $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})^T$ are needed,

$$
\begin{aligned}
\frac{\partial \mathrm{IC}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)}}{\partial \boldsymbol{\theta}} &= h^{-1} \left(\mathrm{IS}_{j(\gamma,\beta)}^{(\ell_1+1,\ell_2)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j(\gamma,\beta)}^{(\ell_1+1,\ell_2+1)} \begin{bmatrix} \tilde{0}_{q\times 1} \\ \tilde{e}_1 \end{bmatrix} \right), \\
\frac{\partial \mathrm{IS}_{j(\gamma,\beta)}^{(\ell_1,\ell_2)}}{\partial \boldsymbol{\theta}} &= h^{-1} \left(-\mathrm{IC}_{j(\gamma,\beta)}^{(\ell_1+1,\ell_2)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} + \mathrm{IS}_{j(\gamma,\beta)}^{(\ell_1+1,\ell_2+1)} \begin{bmatrix} \tilde{0}_{q\times 1} \\ \tilde{e}_1 \end{bmatrix} \right),
\end{aligned}
\tag{D.5}
$$

where $\tilde{e}_1 = (0,1)^T$. To reveal the convergence rate of $\hat{\boldsymbol{\theta}}_{\mathrm{CK}}$, denoted by δ_n, we aim to establish a sufficient condition for $\|\hat{\boldsymbol{\theta}}_{\mathrm{CK}} - \boldsymbol{\theta}\| = O_p(\delta_n)$, which states that, for any given $\rho > 0$, there exists a constant c such that

$$P\left\{ \sup_{\|\boldsymbol{d}\|=c} Q_h^*(\boldsymbol{\theta} + \delta_n \boldsymbol{d}) < Q_h^*(\boldsymbol{\theta}) \right\} \geq 1 - \rho. \tag{D.6}$$

This sufficient condition motivates us to consider the difference $\Delta(\delta_n) = Q_h^*(\boldsymbol{\theta} + \delta_n \boldsymbol{d}) - Q_h^*(\boldsymbol{\theta})$. In particular, a third-order Taylor expansion of $\Delta(\delta_n)$ around zero gives

$$\Delta(\delta_n) = \Delta(0) + \delta_n \Delta^{(1)}(0) + 0.5 \delta_n^2 \Delta^{(2)}(0) + 6^{-1} r_n^3 \Delta^{(3)}(\delta^*),$$

where δ^* lies between zero and δ_n, $\Delta(0) = 0$,

$$
\begin{aligned}
\Delta^{(1)}(0) &= \left. \frac{\partial Q_h^*(\tilde{\boldsymbol{\theta}})}{\partial \tilde{\boldsymbol{\theta}}^{\mathrm{T}}} \right|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \boldsymbol{d} = K_n^{\mathrm{T}} \boldsymbol{d}, \\
\Delta^{(2)}(0) &= \boldsymbol{d}^{\mathrm{T}} \left. \frac{\partial^2 Q_h^*(\tilde{\boldsymbol{\theta}})}{\partial \tilde{\boldsymbol{\theta}} \partial \tilde{\boldsymbol{\theta}}^{\mathrm{T}}} \right|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}} \boldsymbol{d} = \boldsymbol{d}^{\mathrm{T}} J_n \boldsymbol{d}, \\
\Delta^{(3)}(r^*) &= \boldsymbol{d}^{\mathrm{T}} L_n \boldsymbol{d}^{\mathrm{T}} \boldsymbol{d},
\end{aligned}
$$

in which, by (D.5),

$$K_n = \frac{\partial Q_h^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{nh} \sum_{j=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} K^* \left(\frac{Y_j - (\tilde{B}_j^T, \tilde{W}_j^T)\boldsymbol{\theta}}{h} \right)$$

$$= -\frac{1}{nh^2} \sum_{j=1}^{n} \left(\mathrm{IS}_{j(\gamma,\beta)}^{(1,0)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j(\gamma,\beta)}^{(1,1)} \begin{bmatrix} 0_{q \times 1} \\ \tilde{e}_1 \end{bmatrix} \right), \tag{D.7}$$

$$J_n = \frac{\partial^2 Q_h^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} = -\frac{1}{nh^3} \sum_{j=1}^{n} \left(-\mathrm{IC}_{j(\gamma,\beta)}^{(2,0)} \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + \mathrm{IS}_{j(\gamma,\beta)}^{(2,1)} \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} \right.$$

$$\left. +\mathrm{IC}_{j(\gamma,\beta)}^{(2,2)} \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right), \tag{D.8}$$

$$L_n = -\frac{1}{nh^3} \sum_{i=1}^{n} \left(-\begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} \frac{\partial \mathrm{IC}_{j(\gamma,\beta)}^{(2,0)}}{\partial \boldsymbol{\theta}} + \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} \frac{\partial \mathrm{IS}_{j(\gamma,\beta)}^{(2,1)}}{\partial \boldsymbol{\theta}} \right.$$

$$\left. + \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \frac{\partial \mathrm{IC}_{j(\gamma,\beta)}^{(2,2)}}{\partial \boldsymbol{\theta}} \right) \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$$

$$= \frac{1}{nh^4} \sum_{j=1}^{n} \left(\mathrm{IS}_{j(\gamma,\beta)}^{(3,0)} \begin{bmatrix} A_1\tilde{B}_j + C_1\tilde{W}_j \\ C_1^T\tilde{B}_j + D_1\tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j(\gamma,\beta)}^{(3,1)} \begin{bmatrix} C_1\tilde{e}_1 + A_2\tilde{B}_j + C_2\tilde{W}_j \\ D_1\tilde{e}_1 + C_2^T\tilde{B}_j + D_2\tilde{W}_j \end{bmatrix} - \right.$$

$$\left. \mathrm{IS}_{j(\gamma,\beta)}^{(3,2)} \begin{bmatrix} C_2\tilde{e}_1 + A_3\tilde{B}_j + C_3\tilde{W}_j \\ D_2\tilde{e}_1 + C_3^T\tilde{B}_j + D_3\tilde{W}_j \end{bmatrix} - \mathrm{IC}_{j(\gamma,\beta)}^{(3,3)} \begin{bmatrix} C_3\tilde{e}_1 \\ D_3\tilde{e}_1 \end{bmatrix} \right) \Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}, \tag{D.9}$$

where

$$\begin{cases} A_1 = \tilde{B}_j\tilde{B}_j^T, \ C_1 = \tilde{B}_j\tilde{W}_j^T, \ D_1 = \tilde{W}_j\tilde{W}_j^T, \\[2mm] A_2 = \tilde{0}_{q \times q}, \ C_2 = \begin{bmatrix} 0_{q \times 1}, \ \tilde{B}_j \end{bmatrix}, D_2 = \begin{bmatrix} 0 & 1 \\ 1 & 2W_j \end{bmatrix}, \\[4mm] A_3 = \tilde{0}_{q \times q}, \ C_3 = 0_{q \times 2}, D_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \end{cases} \tag{D.10}$$

$\boldsymbol{\theta}^*$ lying between $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \delta_n\boldsymbol{d}$, corresponding to δ^* lying between zero and δ_n.

In summary, we have

$$Q_h^*(\boldsymbol{\theta} + \delta_n\boldsymbol{d}) - Q_h^*(\boldsymbol{\theta}) = \delta_n K_n^{\mathrm{T}}\boldsymbol{d} + 0.5\delta_n^2\boldsymbol{d}^{\mathrm{T}}J_n\boldsymbol{d} + 6^{-1}\delta_n^3\boldsymbol{d}^{\mathrm{T}}L_n\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}. \tag{D.11}$$

142

In order to reveal δ_n that satisfies (D.6), we study the orders of J_n, K_n, and L_n based on the mean-variance decomposition given by $V = E(V) + O_p\{\sqrt{\mathrm{Var}(V)}\}$, for a random variable V under regularity conditions. The means and variances of J_n, K_n and L_n are derived in Appendix F, with results summarized in Lemma C.

Define $\mu_2 = \int t^2 K(t)dt$. If the measurement error distribution is ordinary smooth of order b, by Lemma C, we have

$$
\left\{
\begin{array}{rcl}
K_n & = & \dfrac{\mu_2 h^2}{2n} \sum_{j=1}^{n} f^{(3)}(0|X_j,\ T_j)\tilde{Z}_j\{1 + o_p(1)\} + O_p\left(1/\sqrt{nh^{3+2b}}\right), \\[2mm]
J_n & = & \dfrac{1}{n} \sum_{j=1}^{n} f^{(2)}(0|X_j,\ T_j)\tilde{Z}_j\tilde{Z}_j^{\mathrm{T}}\{1 + o_p(1)\} + O_p\left(1/\sqrt{nh^{5+2b}}\right), \\[2mm]
L_n & = & O_p(1) + O_p\left(1/\sqrt{nh^{7+2b}}\right),
\end{array}
\right. \qquad \text{(D.12)}
$$

where $\tilde{Z}_j^T = \left(\tilde{B}_j^T, \tilde{X}_j^T\right)$. Based on these rates, by setting $\delta_n = h^2 + 1/\sqrt{nh^{3+2b}}$, one can show that (D.11) is dominated by the second term for a large enough c, which is negative definite in probability by condition (C^*2). More specifically, the first term in the right-hand side of (D.11) is of order $O(\delta_n^2)$, the second terms is $0.5\delta_n^2 \boldsymbol{d}^{\mathrm{T}} J^* \boldsymbol{d}\{1 + o_p(1)\}$, where $J^* = \lim_{n\to\infty} n^{-1}\sum_{j=1}^{n} f^{(2)}(0|_j)\tilde{Z}_j\tilde{Z}_j^{\mathrm{T}}$, and the third term is of order $o_p(\delta_n^2)$ provided that $nh^{7+2b} \to 0$. Hence, for a large enough c, (D.6) holds for $\delta_n = h^2 + 1/\sqrt{nh^{3+2b}}$. In fact, the rate δ_n is determined by the rate of K_n, which has h^2 as the order of the bias of $\hat{\boldsymbol{\theta}}_{\mathrm{CK}}$ and $1/\sqrt{nh^{3+2b}}$ as the order of its standard error. This leads to (D.1).

If the measurement error distribution is super smooth of order b, by Lemma C, we have

$$
\left\{
\begin{array}{rcl}
K_n & = & \dfrac{\mu_2 h^2}{2} \sum_{j=1}^{n} f^{(3)}(0|X_j,\ T_j)\tilde{Z}_j\{1 + o_p(1)\} \\[2mm]
& & + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}\right\}, \\[2mm]
J_n & = & \dfrac{1}{n} \sum_{j=1}^{n} f^{(2)}(0|X_j,\ T_j)\tilde{Z}_j\tilde{Z}_j^{\mathrm{T}}\{1 + o_p(1)\} \\[2mm]
& & + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{5-2\min(b_2,b_3,b_4)})}\right\}, \\[2mm]
L_n & = & O_p(1) + O_p\left\{\sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{7-2\min(b_2,b_3,b_4,b_5)})}\right\},
\end{array}
\right.
$$

143

where b_ℓ, for $\ell = 2, 3, 4, 5$, are defined in Lemma C. Based on these rates, by setting $\delta_n = h^2 + \sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}$, one can show that (D.11) is again dominated by the second term for a large enough c. Thus, for a large enough c, (D.6) holds for $\delta_n = h^2 + \sqrt{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})}$. This proves (D.2).

$\square$

Appendix E

Proof of Theorem 4

Theorem 4. *Under the same assumptions imposed in Theorem 3,*

(i) if U follows an ordinary smooth distribution of order b,

$$\sqrt{nh^{3+2b}} \left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2 \mu_2 I^{*-1} Q/4 \right) \xrightarrow{d} N(0,\, I^{*-1} M_L I^{*-1}); \qquad \text{(E.1)}$$

where M_L is constant matrices.

(ii) if U follows a super smooth distribution of order b,

$$\left\{ Var(\hat{\boldsymbol{\beta}}_{CK}) \right\}^{-1/2} \left(\hat{\boldsymbol{\beta}}_{CK} - \boldsymbol{\beta} - h^2 \mu_2 I^{*-1} Q/4 \right) \xrightarrow{d} N(0,\, 1), \qquad \text{(E.2)}$$

where

$$Q = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} E\{f^{(3)}(0|X_j, T_j)\tilde{X}_j\},$$

$$I^* = \lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} E\left\{ f^{(2)}(0|X_j, T_j)\tilde{X}_j \tilde{X}_j^{\mathrm{T}} + \tilde{X}_j B_j^T \Phi^{-1}\Psi \right\},$$

$Var(\hat{\boldsymbol{\beta}}_{CK}) = O[\exp\{2|\beta_1|^b/(d_2 h^b)\}/\{nh^{3-2\min(b_2,b_3)}\}]$, *and $\boldsymbol{\Sigma}^{-1/2}$ denotes the inverse of the positive definite square root of a positive definite matrix $\boldsymbol{\Sigma}$.*

Proof. Define a series of integrals that involve in the integrand the cosine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s) = -1/\phi_U(s)$ evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \cos\left(\frac{Y_j - g(T_j) - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta}}{h} t \right) \phi_K(t)\tau^{(\ell_2)}(\beta_1 t/h)dt, \qquad \text{(E.3)}$$

$$\mathrm{IC}_{j\eta_j^*}^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \cos\left(\eta_j^* t \right) \phi_K(t)\tau^{(\ell_2)}(\beta_1 t/h)dt, \qquad \text{(E.4)}$$

145

where η_j^* is between $\dfrac{Y - g(T_j) - \tilde{W}_j^T \boldsymbol{\beta}}{h}$ and $\dfrac{Y - \tilde{B}_j^T \hat{\boldsymbol{\gamma}}_{\mathrm{CK}} - \tilde{W}_j^T \hat{\boldsymbol{\beta}}_{\mathrm{CK}}}{h}$. Similarly define another series of integrals that involve in the integrand the sine function, a power function t^{ℓ_1}, and the ℓ_2-th derivative of $\tau(s)$ evaluated at $\beta_1 t/h$ as follows,

$$\mathrm{IS}_{j(g,\beta)}^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \sin\left(\frac{Y_j - g(T_j) - \tilde{W}_j^{\mathrm{T}}\boldsymbol{\beta}}{h} t\right) \phi_K(t) \tau^{(\ell_2)}(\beta_1 t/h) dt. \tag{E.5}$$

$$\mathrm{IS}_{j\eta_j^*}^{(\ell_1,\ell_2)} = \frac{1}{2\pi} \int t^{\ell_1} \sin\left(\eta_j^* t\right) \phi_K(t) \tau^{(\ell_2)}(\beta_1 t/h) dt, \tag{E.6}$$

Because $\hat{\boldsymbol{\theta}}_{\mathrm{CK}} = (\hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\gamma}}_{\mathrm{CK}})$ maximizes $Q_h^*(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (nh)^{-1} \sum_{j=1}^{n} K^*\{(Y_j - \tilde{B}_j \boldsymbol{\gamma} - \tilde{W}_j \boldsymbol{\beta})/h\}$, one has

$$\begin{aligned}
&\mathcal{L}_1(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\gamma}}_{\mathrm{CK}}) \\
&= -\frac{1}{nh^2} \sum_{j=1}^{n} \left(\mathrm{IS}_{j(\gamma,\beta)}^{(1,0)} \Big|_{\gamma=\hat{\gamma}_{\mathrm{CK}},\beta=\hat{\beta}_{\mathrm{CK}}} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j(\gamma,\beta)}^{(1,1)} \Big|_{\gamma=\hat{\gamma}_{\mathrm{CK}},\beta=\hat{\beta}_{\mathrm{CK}}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\
&= \tilde{0}_{2\times 1},
\end{aligned}$$

$$\begin{aligned}
&\mathcal{L}_2(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\gamma}}_{\mathrm{CK}}) \\
&= -\frac{1}{nh^2} \sum_{j=1}^{n} \mathrm{IS}_{j(\gamma,\beta)}^{(1,0)} \Big|_{\gamma=\hat{\gamma}_{\mathrm{CK}},\beta=\hat{\beta}_{\mathrm{CK}}} \tilde{B}_j = \tilde{0}_{q\times 1},
\end{aligned} \tag{E.7}$$

where $\mathcal{L}_1 = \partial Q_h^*(\boldsymbol{\beta}, \boldsymbol{\gamma})/\partial \boldsymbol{\beta}$ and $\mathcal{L}_2 = \partial Q_h^*(\boldsymbol{\beta}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}$. Let $R_j = g(T_j) - \tilde{B}_j^T \boldsymbol{\gamma}$, by doing taylor expansion around $\dfrac{Y_j - g(T_j) - \tilde{W}_j^T \boldsymbol{\beta}}{h} t$, some simple calculation yields,

$$\begin{aligned}
&\mathcal{L}_1(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\gamma}}_{\mathrm{CK}}) = \\
&-\frac{1}{nh^2} \sum_{j=1}^{n} \left[\left(\mathrm{IS}_{j(g,\beta)}^{(1,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j(g,\beta)}^{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) - \left(\mathrm{IS}_{j(g,\beta)}^{(2,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} - \mathrm{IC}_{j(g,\beta)}^{(2,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right. \\
&\left\{ R_j h^{-1} - \tilde{W}_j^T \left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}} - \boldsymbol{\beta} \right) h^{-1} - \tilde{B}_j^T \left(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}} - \boldsymbol{\gamma} \right) h^{-1} \right\} \\
&-\frac{1}{2} \left(\mathrm{IS}_{j\eta_j^*}^{(3,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j\eta_j^*}^{(3,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\
&\left. \left\{ R_j h^{-1} - \tilde{W}_j^T \left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}} - \boldsymbol{\beta} \right) h^{-1} - \tilde{B}_j^T \left(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}} - \boldsymbol{\gamma} \right) h^{-1} \right\}^2 \right] \\
&+ o_p \left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}} - \boldsymbol{\beta} \right) = \tilde{0},
\end{aligned} \tag{E.8}$$

Similarly,

$$
\begin{aligned}
&\mathcal{L}_2(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}, \hat{\boldsymbol{\gamma}}_{\mathrm{CK}}) \\
={}& -\frac{1}{nh^2}\sum_{j=1}^{n}\Big[\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\tilde{B}_j \\
&\quad +\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)}\tilde{B}_j\left\{R_jh^{-1}-\tilde{W}_j^T\left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}-\boldsymbol{\beta}\right)h^{-1}-\tilde{B}_j^T\left(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}}-\boldsymbol{\gamma}\right)h^{-1}\right\} \\
&\quad -\frac{1}{2}\mathrm{IS}_{j\eta_j^*}^{(3,0)}\tilde{B}_j\left\{R_jh^{-1}-\tilde{W}_j^T\left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}-\boldsymbol{\beta}\right)h^{-1}-\tilde{B}_j^T\left(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}}-\boldsymbol{\gamma}\right)h^{-1}\right\}^2\Big] \\
&\quad +o_p\left(\hat{\boldsymbol{\gamma}}_{\mathrm{CK}}-\boldsymbol{\gamma}\right)=\tilde{0}, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (\mathrm{E.9})
\end{aligned}
$$

Let $\Phi = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f^{(2)}(0|X_j,T_j)\tilde{B}_j\tilde{B}_j^T$ and $\Psi = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}f^{(2)}(0|X_j,T_j)\tilde{B}_j\tilde{X}_j^T$, then, by the result of (F.13) and (F.16), after some calculations based on (E.9), it follows that

$$
\hat{\boldsymbol{\gamma}}_{\mathrm{CK}}-\boldsymbol{\gamma} = (\Phi + o_p(1))^{-1}\left\{-\Psi\left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}-\boldsymbol{\beta}\right)+o_p(1)\right\} \quad\quad\quad (\mathrm{E.10})
$$

Substituting (E.10) into (E.8), we have

$$
M_n + o_p(1) = I_n\left(\hat{\boldsymbol{\beta}}_{\mathrm{CK}}-\boldsymbol{\beta}\right), \quad\quad\quad\quad\quad\quad\quad (\mathrm{E.11})
$$

where

$$
M_n = \frac{1}{nh^2}\sum_{j=1}^{n}\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\tilde{W}_j + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\tilde{e}_1\right), \quad\quad\quad\quad (\mathrm{E.12})
$$

$$
I_n = \frac{1}{nh^3}\sum_{j=1}^{n}\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)}\tilde{W}_j - \mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(2,1)}\tilde{e}_1\right)\left(\tilde{W}_j^T + B_j^T\Phi^{-1}\Psi\right). \quad\quad (\mathrm{E.13})
$$

Part (I): Show (E.1). When U is ordinary smooth, define $M_n^* = \sqrt{nh^{3+2b}}M_n$. We next show that,

$$
\left\{\boldsymbol{p}^{\mathrm{T}}\mathrm{Cov}(M_n^*)\boldsymbol{p}\right\}^{-1/2}\boldsymbol{p}^{\mathrm{T}}\{M_n^* - E(M_n^*)\} \xrightarrow{d} N(0,1),
$$
$$
\text{for any unit vector } \boldsymbol{p} = (p_1,p_2)^{\mathrm{T}}\in\mathbb{R}^2, \quad\quad\quad (\mathrm{E.14})
$$

where "$\xrightarrow{d}$" refers to convergence in distribution. Once (E.14) is proved, we conclude the asymptotic normality of M_n^* by the Cramér-Wold Theorem (Cramer and Wold,

1936), and thus the asymptotic normality of M_n. By (D.7), $\boldsymbol{p}^{\mathrm{T}} M_n^* = \sum_{j=1}^{n} \xi_j$, where

$$\xi_j = -\frac{h^b}{\sqrt{nh}} \boldsymbol{p}^{\mathrm{T}} \left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

Define $m_j = E(\xi_j | X_j)$ and $S_n^2 = \sum_{j=1}^{n} \mathrm{Var}(\xi_j | X_j)$. We next use the Lyapunov Central Limit Theorem to show the asymptotic normality of $\boldsymbol{p}^{\mathrm{T}} M_n^*$. This requires proving the Lyapunov's conditions (Billingsley, 2008), which states that

$$\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{j=1}^{n} E|\xi_j - m_j|^{2+\delta} = 0, \text{ for some } \delta > 0. \tag{E.15}$$

In particular, we show that (E.15) is satisfied for $\delta = 1$.

First, because $\xi_1, \ldots, \xi_n$ are independent,

$$\begin{aligned} S_n^2 &= \mathrm{Var}\left(\sum_{j=1}^{n} \xi_j \right) \\ &= E\left\{ \mathrm{Var}\left(\sum_{j=1}^{n} \xi_j \,\middle|\, \boldsymbol{X} \right) \right\} + \mathrm{Var}\left\{ E\left(\sum_{j=1}^{n} \xi_j \,\middle|\, \boldsymbol{X} \right) \right\} \\ &= \boldsymbol{p}^{\mathrm{T}} E\left\{ \mathrm{Var}\left(M_n^* | \boldsymbol{X} \right) \right\} \boldsymbol{p} + \boldsymbol{p}^{\mathrm{T}} \mathrm{Var}\left\{ E\left(M_n^* | \boldsymbol{X} \right) \right\} \boldsymbol{p} \end{aligned} \tag{E.16}$$

From Part (I) of the proof for Lemma C in Appendix C, one can have ,under condition (N2), the second term in (E.16) converges to zero as $n \to \infty$. The first term is equal to the expectation of, by (E.13)

$$h^{2b-1} \boldsymbol{p}^{\mathrm{T}} \mathrm{Var}\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \,\middle|\, X_j, T_j \right) \boldsymbol{p} \tag{E.17}$$

which involve h^{2b-1} multiplying the following expectations,

$$\begin{aligned} &E\left\{ \left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \right)^2 \,\middle|\, X_j \right\}, \quad E\left\{ \left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \right)^2 W_j \,\middle|\, X_j \right\}, \quad E\left\{ \left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \right)^2 W_j^2 \,\middle|\, X_j \right\}, \\ &E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \,\middle|\, X_j \right), \quad E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} W_j \,\middle|\, X_j \right), \quad E\left\{ \left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \right)^2 \,\middle|\, X_j \right\}. \end{aligned} \tag{E.18}$$

These expectations are studied in details in the proof for Lemma C in Appendix C. For ordinary smooth measurement error, based on the results derived there, one has

148

$S_n^2 = \boldsymbol{p}^{\mathrm{T}} M_{\mathrm{L}} \boldsymbol{p} + o(1)$, where M_{L} is a constant matrix as the limit of $\mathrm{Var}(M_n^* | \boldsymbol{X})$ as $n \to \infty$.

Second, following similar derivations for the aforementioned expectations, one can show that multiplying the following three expectations by $n^{-1/2} h^{3(b-1/2)}$ all lead to quantities that tend to zero as $n \to \infty$,

$$E\left(\left|\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right|^3 \middle| X_j, T_j\right), \quad E\left(\left|\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} W_j\right|^3 \middle| X_j, T_j\right), \quad E\left(\left|\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\right|^3 \middle| X_j, T_j\right). \quad \text{(E.19)}$$

These imply that, as $n \to \infty$,

$$nE(|\xi_j|^3 | X_j, T_j) = n^{-1/2} h^{3(b-1/2)} E\left(\left|p_1 \mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} + p_2 \mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} W_j + p_2 \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\right|^3 \middle| X_j, T_j\right)$$

$$\to 0.$$

It follows that $\lim_{n \to \infty} \sum_{j=1}^n |\xi_j - m_j|^3 = 0$. This and the previous conclusion regarding S_n^2 together lead to the Lyapunov's conditions in (E.15), which is a sufficient condition for, taking into account the asymptotic mean of M_n according to Lemma C,

$$\sqrt{nh^{3+2b}} \left(M_n - \mu_2 Q h^2 / 2\right) \xrightarrow{d} N(0, M_{\mathrm{L}}).$$

Finally, by the Slutsky's Theorem, we have (E.1).

Part (II): Show (E.2).

When U is super smooth, define $M_n^* = \exp\{-2|\beta_1|^b/(d_2 h^b)\} \sqrt{nh^{3-2\min(b_2,b_3)}} M_n$. Now we have $\boldsymbol{p}^{\mathrm{T}} M_n^* = \sum_{j=1}^n \xi_j$ with

$$\xi_j = -\exp\left(-\frac{2|\beta_1|^b}{d_2 h^b}\right) \frac{1}{\sqrt{nh^{1+\min(b_2,b_3)}}} \boldsymbol{p}^{\mathrm{T}} \left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \begin{bmatrix} 1 \\ W_j \end{bmatrix} + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

The remaining task is to show (E.15) for the so-defined ξ_j. This leads one to look into the expectations in (E.18) and those in (E.19), following the strategies used in the proof for Lemma C, under conditions (N3), (N4) and (S1), one can have $|IS_{j(g,\boldsymbol{\beta})}^{(0,1)}|^2$ is bounded from below by a quantity of order $h^{2b_1 + 2(q+1)b} \exp\left\{2|\beta_1|^b/(d_2 h^b)\right\}$. Similarly,

149

$|IC_{j(g,\beta)}^{(1,1)}|^2$ is bounded from below by a quantity of order

$h^{4b_1 - 2b_0^{(1)} + 2(q+1)b} \exp\left\{2|\beta_1|^b/(d_2 h^b)\right\}$, and $|IS_{j(g,\beta)}^{(1,0)} IC_{j(g,\beta)}^{(1,1)}|$ is bounded from below by

a quantity of order

$h^{b_1 + 2b_1 - b_0^{(1)} + 2(q+1)b} \exp\left\{2|\beta_1|^b/(d_2 h^b)\right\}$. It follows that S_n^3 is bounded from below by a

quantity of order

$h^{3\{(q+1)b - 0.5 - \min(b_2, b_3) + \min(b_1, 2b_1 - b_0^{(1)})\}} \exp\left\{-3|\beta_1|^b/(d_2 h^b)\right\}$ as $n \to \infty$. This, along with

the aforementioned result regarding $\lim_{n \to \infty} \sum_{j=1}^{n} |\xi_j^* - m_j|^3$, implies (E.15).

Furtherm assuming that $\lim_{n \to \infty} \text{Var}(M_n^* | \boldsymbol{X})$ exists, and denoting its expectation by

M_N, which leads to

$$\exp\left(-\frac{2|\beta_1|^b}{d_2 h^b}\right) \sqrt{nh^{3 - 2\min(b_2, b_3)}} \left(M_n - \mu_2 Q h^2/2\right) \xrightarrow{d} N(0,\ M_{\text{N}}).$$

Lastly, one uses the Slutsky's Theorem to conclude (E.2).

$\square$

Appendix F

Lemmas referenced in Appendices D and E

Lemma F. *Assume $E(U^4) < \infty$, and assume conditions stated in Lemmas C.1, C.2, and C.3 hold. Define $\mu_2 = \int t^2 K(t) dt$, $\boldsymbol{X} = (X_1, \ldots, X_n)$, $\boldsymbol{T} = (T_1, \ldots, T_n)$ and $\tilde{Z}_j^T = (\tilde{X}_j^T, \tilde{B}_j^T)$. For K_n, J_n and L_n given by (D.7), (D.8), and (D.9), respectively, we have*

$$E(K_n|\boldsymbol{X}, \boldsymbol{T}) = \frac{\mu_2 h^2}{2n} \sum_{j=1}^n g^{(3)}(0|X_j, T_j)\tilde{Z}_j\{1 + o_p(1)\}, \tag{F.1}$$

$$E(J_n|\boldsymbol{X}, \boldsymbol{T}) = \frac{1}{n} \sum_{j=1}^n g^{(2)}(0|X_j, T_j)\tilde{Z}_j \tilde{Z}_j^{\mathrm{T}}\{1 + o_p(1)\}, \tag{F.2}$$

$$E(L_n|\boldsymbol{X}, \boldsymbol{T}) = O_p(1). \tag{F.3}$$

And when U is ordinary smooth of order b, under Conditions O, we have

$$Var(K_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{1/(nh^{3+2b})\}, \tag{F.4}$$

$$Var(J_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{1/(nh^{5+2b})\}, \tag{F.5}$$

$$Var(L_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{1/(nh^{7+2b})\}; \tag{F.6}$$

when U is super smooth of order b, under Conditions S, we have

$$Var(K_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{3-2\min(b_2,b_3)})\}, \tag{F.7}$$

$$Var(J_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{5-2\min(b_2,b_3,b_4)})\}, \tag{F.8}$$

$$Var(L_n|\boldsymbol{X}, \boldsymbol{T}) = O_p\{\exp(2|\beta_1|^b h^{-b}/d_2)/(nh^{7-2\min(b_2,b_3,b_4,b_5)})\}, \tag{F.9}$$

where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)})I(2b_0 - b_1^{(1)} < 0.5)$, $b_4 = \min\{(2b_0 - b_1^{(2)})I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)})I(3b_0 - 2b_1^{(1)} < 0.5)\}$, and $b_5 = \min\{(2b_0 - $

$b_1^{(3)})I(2b_0 - b_1^{(3)} < 0.5),\ (3b_0 - b_1^{(1)} - b_1^{(2)})I(3b_0 - b_1^{(1)} - b_1^{(2)} < 0.5),\ (4b_0 - b_1^{(1)} - 2b_1^{(2)})I(4b_0 - b_1^{(1)} - 2b_1^{(2)} < 0.5)\}.$

Proof. The proof consists of six parts to establish (F.1)–(F.9). We first define a series of integrals as follows, with integrands involving t^{ℓ_1} and the ℓ_2-th derivative of $\tau(s) = -1/\phi_U(s)$ evaluated it $s = \beta_1 t/h$,

$$\mathcal{F}_{\ell_1,\ell_2}(s) = \frac{1}{2\pi} \int e^{ist} t^{\ell_1} \phi_K(t) \tau^{(\ell_2)}(\beta_1 t/h)\, dt, \tag{F.10}$$

for nonnegative integers ℓ_1 and ℓ_2.

Part (I): Show (F.1). Note that $R_j = g(T_j) - \tilde{B}_j^T \boldsymbol{\gamma}$, by Corollary 6.21 in Schumaker (2007), one has

$$\|R_j\| = O(n^{-r/(2r+1)}) \tag{F.11}$$

By (D.7), Taylor expansion of K_n around $\dfrac{Y_j - g(T_j) - \tilde{W}_j^T \boldsymbol{\beta}}{h} t$ yields

$$K_n = -\frac{1}{nh^2} \sum_{j=1}^n \left(\mathrm{IS}_{j(g,\beta)}^{(1,0)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j(g,\beta)}^{(1,1)} \begin{bmatrix} 0_{q\times1} \\ \tilde{e}_1 \end{bmatrix} \right)$$

$$\quad - \frac{1}{nh^2} \sum_{j=1}^n \left(\mathrm{IC}_{j(g,\beta)}^{(2,0)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} - \mathrm{IS}_{j(g,\beta)}^{(2,1)} \begin{bmatrix} 0_{q\times1} \\ \tilde{e}_1 \end{bmatrix} \right) h^{-1} R_j$$

$$\quad + \frac{1}{2nh^2} \sum_{j=1}^n \left(\mathrm{IS}_{j\eta_j^*}^{(3,0)} \begin{bmatrix} \tilde{B}_j \\ \tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j\eta_j^*}^{(3,1)} \begin{bmatrix} 0_{q\times1} \\ \tilde{e}_1 \end{bmatrix} \right) h^{-2} R_j^2 + o_p(1) \tag{F.12}$$

where η_j^* lying between $\dfrac{Y - g(T_j) - \tilde{W}_j^T \boldsymbol{\beta}}{h}$ and $\dfrac{Y - \tilde{B}_j^T \hat{\boldsymbol{\gamma}}_{\mathrm{CK}} - \tilde{W}_j^T \hat{\boldsymbol{\beta}}_{\mathrm{CK}}}{h}$. By (F.12)

and assumptions $C4$ and $C5$,

$$
\begin{aligned}
E(K_n|\boldsymbol{X},\boldsymbol{T}) = & -\frac{1}{nh^2}\sum_{j=1}^{n}
\begin{bmatrix}
E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\tilde{B}_j\,\middle|\,X_j,T_j\right) \\
E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\tilde{W}_j + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\tilde{e}_1\,\middle|\,X_j,T_j\right)
\end{bmatrix} \\
& -\frac{1}{nh^2}\sum_{j=1}^{n}
\begin{bmatrix}
E\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)}\tilde{B}_j\,\middle|\,X_j,T_j\right) \\
E\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,0)}\tilde{W}_j - \mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,1)}\tilde{e}_1\,\middle|\,X_j,T_j\right)
\end{bmatrix}h^{-1}R_j \\
& +\frac{1}{2nh^2}\sum_{j=1}^{n}
\begin{bmatrix}
E\left(\mathrm{IS}_{j\eta_j^*}^{(3,0)}\tilde{B}_j\,\middle|\,X_j,T_j\right) \\
E\left(\mathrm{IS}_{j\eta_j^*}^{(3,0)}\tilde{W}_j + \mathrm{IC}_{j\eta_j^*}^{(1,1)}\tilde{e}_1\,\middle|\,X_j,T_j\right)
\end{bmatrix}h^{-2}R_j^2.
\end{aligned}
$$

Recall that, given $X = x$ and $T = t$, the mode residual, $\epsilon = Y - g(T) - \tilde{X}^T\boldsymbol{\beta}$, follows a distribution specified by the pdf $f(\epsilon|x)$, and $f_{Y|X}(y|x) = f\left\{y - g(T) - \tilde{x}^T\boldsymbol{\beta}|x,t\right\}$. It follows that, for the first element in the 2×1 summand of the first term of $E\left(K_n|X_j,T_j\right)$ above, we have

$$
\begin{aligned}
& E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\,\middle|\,X_j,T_j\right) \\
=& E\left\{\frac{1}{2\pi}\int t\sin\left(\frac{Y_j - g(T) - \tilde{W}^T\boldsymbol{\beta}}{h}t\right)\phi_K(t)\tau(\beta_1 t/h)dt\,\middle|\,X_j,T_j\right\} \\
=& \frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t\phi_K(t)\tau(\beta_1 t/h)\int \sin\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon \\
=& \frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t\phi_K(t)\tau(\beta_1 t/h)\phi_U(-\beta_1 t/h)\sin(\epsilon t/h)dt\,d\epsilon,\ \text{by Lemma C.1,} \\
=& -\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t\phi_K(t)\sin(\epsilon t/h)dt\,d\epsilon, \\
=& -h\cdot\frac{1}{2\pi}\int f(sh|X_j,T_j)\int t\phi_K(t)\sin(st)dt\,ds \\
=& -h\cdot\frac{1}{2\pi}\int\left\{f(0|X_j,T_j) + 0.5f^{(2)}(0|X_j,T_j)s^2h^2 + 6^{-1}f^{(3)}(0|X_j,T_j)s^3h^3\right. \\
& \left. + 24^{-1}f^{(4)}(s^*|X_j,T_j)s^4h^4\right\}\cdot \\
& \int t\phi_K(t)\sin(st)dt\,ds,\ \text{for some } s^* \text{ lying between 0 and } sh, \\
=& -0.5\mu_2 h^4 f^{(3)}(0|X_j,T_j) + O_p(h^5),\ \text{by Lemma C.2.} \hspace{2cm} \text{(F.13)}
\end{aligned}
$$

The second element in the 2×1 summand of the first term of $E(K_n|\boldsymbol{X},\boldsymbol{T})$ above

involves

$$E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}W_j + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\Big|\, X_j, T_j\right) = X_j E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\Big|\, X_j, T_j\right) + E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}U_j\Big|\, X_j, T_j\right)$$
$$+ E\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\Big|\, X_j, T_j\right),$$

where the first expectation is derived above, and the latter two expectations sum to zero, as we show next. Following a similar elaboration of $E(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}|X_j, T_j)$ above, one has

$$E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}U_j\Big|\, X_j, T_j\right) + E\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\Big|\, X_j, T_j\right)$$
$$= \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t\phi_K(t)\tau(\beta_1 t/h)\int u\sin\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon +$$
$$\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int \cos\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon$$
$$= \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t\phi_K(t)\tau(\beta_1 t/h)\cdot(-1)\cdot\phi_U^{(1)}(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon +$$
$$\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon, \text{ by Lemma C.1,}$$
$$= \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t\phi_K(t)\left\{\tau(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h)\right.$$
$$\left.+\tau^{(1)}(\beta_1 t/h)\phi_U(\beta_1 t/h)\right\}\cos(\epsilon t/h)dt\,d\epsilon,$$

where, by Lemma C.3, the expression within the curly brackets is equal to zero. Hence, the summand of the first term of $E(K_n|\boldsymbol{X}, \boldsymbol{T})$ is equal to $-0.5\mu_2 h^4 f^{(3)}(0|X_j, T_j)\tilde{Z}_j + O_p(h^5)$. Because the assumption (C*5), under conditions that guarantee boundedness of $\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(3,0)}$ and $\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}$, similar as the strategy in the first term of EK_n, one can show,

$$E\left(\mathrm{IS}_{j\eta_j^*}^{(3,0)}\tilde{B}_j\Big|\, X_j, T_j\right) = o_p(h^4),$$
$$E\left(\mathrm{IS}_{j\eta_j^*}^{(3,0)}\tilde{W}_j + \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\tilde{e}_1\Big|\, X_j, T_j\right) = o_p(h^4). \tag{F.14}$$

Based on (F.14) and assumption (C*5), one can have the third term of $E(K_n|\boldsymbol{X}, \boldsymbol{T})$ is $o_p(1)$. Next, we show the second terms of $E(K_n|\boldsymbol{X}, \boldsymbol{T})$ is $o_p(1)$. Note that, the first

element in the 2×1 summand is

$$
- E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\Big| X_j, T_j\right) \tag{F.15}
$$

$$
= -\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\int \cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)du\, dt\, d\epsilon
$$

$$
= -\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)dt\, d\epsilon,\ \text{by Lemma C.1,}
$$

$$
= \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\cos(\epsilon t/h)dt\, d\epsilon,\ \text{since } \tau(s) = -1/\phi_U(s),
$$

$$
= h\cdot\frac{1}{2\pi}\int f(hs|X_j, T_j)\int t^2\phi_K(t)\cos(st)dt\, ds
$$

$$
= h\cdot\frac{1}{2\pi}\int \left\{f(0|X_j, T_j) + f^{(2)}(0|X_j, T_j)h^2 s^2/2 + O_p(h^3)\right\}\int t^2\phi_K(t)\cos(st)dt\, ds
$$

$$
= h\cdot f(0|X_j, T_j)\cdot\frac{1}{2\pi}\int\int t^2\phi_K(t)\cos(st)dt\, ds +
$$

$$
\frac{h^3}{2}f^{(2)}(0|X_j, T_j)\cdot\frac{1}{2\pi}\int\int t^2\phi_K(t)\cos(st)dt\, s^2\, ds + O_p(h^4)
$$

$$
= -h^3 f^{(2)}(0|X_j) + O_p(h^4),\ \text{by Lemma C.2.} \tag{F.16}
$$

In addition, the second element in the 2×1 summand involves

$$
- E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j\Big| X_j, T_j\right) + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}\Big| X_j, T_j\right)
$$

$$
= -\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\int u\cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)du\, dt\, d\epsilon +
$$

$$
\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int \sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)du\, dt\, d\epsilon
$$

$$
= -\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\phi_U^{(1)}(-\beta_1 t/h)\sin(\epsilon t/h)dt\, d\epsilon +
$$

$$
\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\sin(\epsilon t/h)dt\, d\epsilon,\ \text{by Lemma C.1,}
$$

$$
= \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\left\{\tau(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h)\right.
$$

$$
\left. +\tau^{(1)}(\beta_1 t/h)\phi_U(\beta_1 t/h)\right\}\sin(\epsilon t/h)dt\, d\epsilon, \tag{F.17}
$$

where, by Lemma C.3, the part of the integrand insides the curly brackets is equal to zero. Based on (F.16), (F.17) and assumption (C^*5), one can have the second term of EK_n is $o_p(1)$. After combining three summands in EK_n, one can have (F.1).

Part (II): Show (F.4) and (F.7).

155

By (F.12) and assumption (C*5), the order of $\text{Var}(K_n|\boldsymbol{X})$ is determined by

$$
\frac{1}{nh^4} E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 \begin{bmatrix} \tilde{B}_j \tilde{B}_j^T & \tilde{B}_j \tilde{W}_j^T \\ \tilde{W}_j \tilde{B}_j^T & \tilde{W}_j \tilde{W}_j^T \end{bmatrix} + \text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \text{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} \begin{bmatrix} \tilde{0}_{q\times q} & \tilde{B}_j \tilde{e}_1^T \\ \tilde{0}_{2\times q} & \tilde{W}_j \tilde{e}_1^T \end{bmatrix} \right.
$$

$$
\left. + \left(\text{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\right)^2 \begin{bmatrix} \tilde{0}_{q\times q} & \tilde{0}_{q\times 2} \\ \tilde{0}_{2\times q} & \tilde{e}_1 \tilde{e}_1^T \end{bmatrix} \,\middle|\, X_j, T_j \right\}. \tag{F.18}
$$

We next derive the six expectations involved in (F.18), also listed in (E.18) for proving Theorem 2. Among these six expectations,

$$
E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 W_j \,\middle|\, X_j, T_j \right\}
$$

$$
= E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 U_j \,\middle|\, X_j, T_j \right\} + E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 \,\middle|\, X_j, T_j \right\} X_j,
$$

$$
E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 W_j^2 \,\middle|\, X_j, T_j \right\}
$$

$$
= E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 U_j^2 \,\middle|\, X_j, T_j \right\} + 2E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 U_j \,\middle|\, X_j, T_j \right\} X_j
$$

$$
+ E\left\{ \left(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2 \,\middle|\, X_j, T_j \right\} X_j^2.
$$

It can be shown that, provided that the first two moments of U exist, $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 U_j | X_j, T_j\}$ and $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 U_j^2 | X_j, T_j\}$ have the same rate as $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 | X_j, T_j\}$. Hence, $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 W_j \, | X_j, T_j\}$ and $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 W_j^2 | X_j, T_j\}$ have the same rate as $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 | X_j, T_j\}$. Similarly, $E(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \text{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} W_j | X_j, T_j)$ is of the same order as $E(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)} \text{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)} | X_j, T_j)$, which, according to the Cauchy-Schwarz inequality, is bounded from above by $\sqrt{E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 | X_j, T_j\}} \cdot \sqrt{E\{(\text{IC}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 | X_j, T_j\}}$. Hence, we only need to focus on $E\{(\text{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 | X_j, T_j\}$ and $E\{(\text{IC}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2 \, | X_j, T_j\}$ next.

First, by (D.4),

$$E\left\{\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)}\right)^2\bigg| X_j = x, T_j = z\right\}$$

$$= E\left[\left\{\frac{1}{2\pi}\int t\sin\left(\frac{Y_j - g(T_j) - \tilde{W}_j^T\boldsymbol{\beta}}{h}t\right)\frac{\phi_K(t)}{\phi_U(\beta_1 t/h)}dt\right\}^2\bigg| X_j = x, T_j = z\right]$$

$$= \int f_U(u)\int\left\{\frac{1}{2\pi}\int t\sin\left(\frac{\epsilon - \beta_1 u}{h}t\right)\frac{\phi_K(t)}{\phi_U(\beta_1 t/h)}dt\right\}^2 f(\epsilon|x)d\epsilon du$$

$$\leq B_f h\int f_U(u)\int|-\mathcal{F}_{1,0}(s)|^2 ds du$$

$$= B_f h\int|\mathcal{F}_{1,0}(s)|^2 ds, \tag{F.19}$$

where $B_f = \sup_x\sup_\epsilon f(\epsilon|x)$. When U is ordinary smooth of order b, by Definition 1 in the main article, there exists a constant $M > 0$ such that

$$\int|\mathcal{F}_{1,0}(s)|^2 ds = \frac{1}{2\pi}\int t^2\frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)}dt,\ \text{by the Parseval's Theorem,} \tag{F.20}$$

$$= \frac{1}{2\pi}\int_{|\beta_1 t|\leq Mh} t^2\frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)}dt + \frac{1}{2\pi}\int_{|\beta_1 t|>Mh} t^2\frac{\phi_K^2(t)}{\phi_U^2(\beta_1 t/h)}dt$$

$$\leq \frac{1}{2\pi}\left\{\inf_{|\beta_1 t|\leq Mh}|\phi_U(\beta_1 t/h)|\right\}^{-2}\int_{|\beta_1 t|\leq Mh} t^2\phi_K^2(t)dt$$

$$+ \frac{1}{2\pi}\int_{|\beta_1 t|>Mh} t^2\frac{\phi_K^2(t)}{(c^2/4)|\beta_1 t/h|^{-2b}}dt$$

$$= O(h^{-2b}),\ \text{under Condition (O4)}.$$

Hence, for ordinary smooth U, (F.19) suggests $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2|X_j, T_j\} = O(h^{1-2b})$. When U is super smooth of order b, by Definition 2 in the main article, following (F.20), one has,

$$\int|\mathcal{F}_{1,0}(s)|^2 ds$$

$$\leq \frac{1}{2\pi}\left\{\inf_{|\beta_1 t|\leq Mh}|\phi_U(\beta_1 t/h)|\right\}^{-2}\int_{|\beta_1 t|\leq Mh} t^2\phi_K^2(t)dt$$

$$+ \frac{1}{2\pi}\int_{Mh<|\beta_1 t|\leq|\beta_1|} t^2\frac{\phi_K^2(t)}{d_0^2|\beta_1 t/h|^{2b_0}\exp(-2|\beta_1 t/h|^b/d_2)}dt$$

$$\leq C_1\int_{|\beta_1 t|\leq Mh} t^2\phi_K^2(t)dt + C_2 h^{2b_0}\exp(2|\beta_1|^b h^{-b}/d_2)\int_{Mh<|\beta_1 t|\leq|\beta_1|} t^2\frac{\phi_K^2(t)}{|t|^{2b_0}}dt$$

$$= O\{h^{2b_2}\exp(2|\beta_1|^b h^{-b}/d_2)\},$$

where $b_2 = b_0 I(b_0 < 0.5)$. Hence, for super smooth U, $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(1,0)})^2|X_j, T_j\} = O\{h^{1+2b_2}\exp(2|\beta_1|^b h^{-b}/d_2)\}$.

Similarly,

$$E\left\{\left(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)}\right)^2\middle|X_j = x, T_j = z\right\}$$
$$\leq B_f h \int f_U(u)\int |\mathcal{F}_{1,1}(s)|^2 ds du = B_f h \int |\mathcal{F}_{1,1}(s)|^2 ds. \qquad \text{(F.21)}$$

When U is ordinary smooth of order b, using the definition of $\tau^{(1)}(s)$,

$$\int |\mathcal{F}_{1,1}(s)|^2 ds$$
$$= \frac{1}{2\pi}\int t^2\phi_K^2(t)\left\{\frac{\phi_U^{(1)}(\beta_1 t/h)}{\phi_U^2(\beta_1 t/h)}\right\}^2 dt, \text{ next use Definition 1 and Condition (O1)},$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(F.22)}$$
$$\leq \frac{1}{2\pi}\sup_{|\beta_1 t|\leq Mh}|\tau^{(1)}(\beta_1 t/h)|^2\int_{|\beta_1 t|\leq Mh}t^2\phi_K^2(t)dt$$
$$+ \frac{1}{2\pi}\int_{|\beta_1 t|>Mh}t^2\phi_K^2(t)\left(\frac{2cb|\beta_1 t/h|^{-b-1}}{c^2|\beta_1 t/h|^{-2b}/4}\right)^2 dt$$
$$= O(h^{2-2b}), \text{ under Condition (O4)}.$$

Hence, for ordinary smooth U, by (F.21), $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(1,1)})^2|X_j, T_j\} = O(h^{3-2b})$. When U is super smooth of order b, following (F.22) and using Defition 2 and Condition (S1),

$$\int |\mathcal{F}_{1,1}(s)|^2 ds \leq \frac{1}{2\pi}\sup_{|\beta_1 t|\leq Mh}|\tau^{(1)}(\beta_1 t/h)|^2\int_{|\beta_1 t|\leq Mh}t^2\phi_K^2(t)dt$$
$$+ \frac{1}{2\pi}\int_{Mh<|\beta_1 t|\leq|\beta_1|}t^2\phi_K^2(t)\left\{\frac{d_1^{(1)}|\beta_1 t/h|^{b_1^{(1)}}\exp(-|\beta_1 t/h|^b/d_2)}{d_0^2|\beta_1 t/h|^{2b_0}\exp(-2|\beta_1 t/h|^b/d_2)}\right\}^2 dt$$
$$= O\left\{h^{2b_3}\exp(2|\beta_1|^b h^{-b}/d_2)\right\},$$

where $b_3 = (2b_0 - b_1^{(1)})I(2b_0 - b_1^{(1)} < 0.5)$. Hence, for super smooth U, (F.21) indicates $E\{(\mathrm{IC}_j^{(1,1)})^2|X_j = x\} = O\{h^{2b_3+1}\exp(2|\beta_1|^b h^{-b}/d_2)\}$.

Having the rates of the expectations in (E.18) derived above under the two types of measurement error distributions, we are now ready to return to (F.18) and conclude the rate of $\mathrm{Var}(K_n|\boldsymbol{X})$. In particular, (F.18) implies (F.4) if U is ordinary smooth, and it implies (F.7) if U is super smooth.

158

Up to this point, we have established the rates of $E(K_n|\boldsymbol{X},\boldsymbol{T})$ and $\mathrm{Var}(K_n|\boldsymbol{X},\boldsymbol{T})$. One will see later that the theme used above to establish these rates is repeatedly used to derive the rates of $E(J_n|\boldsymbol{X},\boldsymbol{T})$, $\mathrm{Var}(J_n|\boldsymbol{X},\boldsymbol{T})$, $E(L_n|\boldsymbol{X},\boldsymbol{T})$, and $\mathrm{Var}(L_n|\boldsymbol{X},\boldsymbol{T})$. Before moving forward to proving the next result, we shall summarize two patterns learnt from Parts (I) and (II) that can be helpful for later parts of the proof. The first pattern pertains to deriving the rate of the mean of K_n, or J_n, or L_n. As seen in Part (I), the order of such mean mostly depends on $E(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)} W_j^{\ell_3}|X_j,T_j)$ and $E(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)} W_j^{\ell_3}|X_j,T_j)$, for some nonnegative integers ℓ_1,ℓ_2 and ℓ_3. These expectations are derived in the same way for ordinary smooth U and for super smooth U. The second pattern relates to deriving the rate of the variance of K_n, or J_n, or L_n. Such rate mainly depends on $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)})^2 W_j^{\ell_3}|X_j,T_j)$ and $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)})^2 W_j^{\ell_3}|X_j,T_j)$, for some nonnegative integers ℓ_1,ℓ_2 and ℓ_3. Moreover, raising the power ℓ_3 from zero does not affect the rate, hence one may focus on $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)})^2|X_j,T_j)$ and $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(\ell_1,\ell_2)})^2|X_j,T_j)$. As seen in Part (II), each of these expectations is bounded from above by $B_f h \int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$. Discussions of the rate of $\int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$ need to be carried out separately for ordinary smooth U and super smooth U. This rate mostly relies on ℓ_2, and the rate is derived under the assumption that $\int t^{2\ell_1}\phi_K^2(t)dt < \infty$, besides other assumptions. If U is ordinary smooth, for
$\ell_2 < \ell_2'$, $\int |\mathcal{F}_{\ell_1',\ell_2'}(s)|^2 ds / \int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds = o(1)$, where ℓ_1 and ℓ_1' can be the same or different. If U is super smooth, the comparison between the rate of $\int |\mathcal{F}_{\ell_1,\ell_2}(s)|^2 ds$ and that of $\int |\mathcal{F}_{\ell_1',\ell_2'}(s)|^2 ds$ is less clear-cut, and requires more careful inspection.

Part (III): Show (F.2).

159

By (D.8), taylor expansion of Jn yields

J_n

$$
= -\frac{1}{nh^3} \sum_{j=1}^{n} \left(-\mathrm{IC}_{j(g,\beta)}^{(2,0)} \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + \mathrm{IS}_{j(g,\beta)}^{(2,1)} \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} + \mathrm{IC}_{j(g,\beta)}^{(2,2)} \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right)
$$

$$
- \frac{1}{nh^3} \sum_{j=1}^{n} \left(\mathrm{IS}_{j(g,\beta)}^{(3,0)} \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + \mathrm{IC}_{j(g,\beta)}^{(3,1)} \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} - \mathrm{IS}_{j(g,\beta)}^{(3,2)} \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right) h^{-1} R_j
$$

$$
- \frac{1}{nh^3} \sum_{j=1}^{n} \left(\mathrm{IC}_{\eta_j^*}^{(4,0)} \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} - \mathrm{IS}_{\eta_j^*}^{(4,1)} \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} - \mathrm{IC}_{\eta_j^*}^{(4,2)} \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right) h^{-2} R_j^2
$$

$$
+ o_p(1), \tag{F.23}
$$

where η_j^* lying between $\dfrac{Y - g(T_j) - \tilde{W}_j^T \boldsymbol{\beta}}{h}$ and $\dfrac{Y - \tilde{B}_j^T \hat{\boldsymbol{\gamma}}_{\mathrm{CK}} - \tilde{W}_j^T \hat{\boldsymbol{\beta}}_{\mathrm{CK}}}{h}$. By (F.23),

$E\left(J_n | \boldsymbol{X}, \boldsymbol{T}\right)$

$$
= -\frac{1}{nh^3} \sum_{j=1}^{n} \left\{ -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)} | X_j, T_j\right) \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)} | X_j, T_j\right) \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} \right.
$$

$$
\left. + E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)} | X_j, T_j\right) \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right\}
$$

$$
- \frac{1}{nh^3} \sum_{j=1}^{n} \left\{ E\left(\mathrm{IS}_{j(g,\beta)}^{(3,0)} | X_j, T_j\right) \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + E\left(\mathrm{IC}_{j(g,\beta)}^{(3,1)} | X_j, T_j\right) \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} \right.
$$

$$
\left. - E\left(\mathrm{IS}_{j(g,\beta)}^{(3,2)} | X_j, T_j\right) \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right\} h^{-1} R_j
$$

$$
- \frac{1}{nh^3} \sum_{j=1}^{n} \left\{ E\left(\mathrm{IC}_{\eta_j^*}^{(4,0)} | X_j, T_j\right) \begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} - E\left(\mathrm{IS}_{\eta_j^*}^{(4,1)} | X_j, T_j\right) \begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix} \right.
$$

$$
\left. - E\left(\mathrm{IC}_{\eta_j^*}^{(4,2)} | X_j, T_j\right) \begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix} \right\} h^{-2} R_j^2, \tag{F.24}
$$

where $A_i, \; C_i, \; D_i, \; i = 1, 2, 3$ are defined in (D.10).

160

By (F.24), the first summand of $E\left(J_n|\boldsymbol{X},\boldsymbol{T}\right)$ yields

$$-\frac{1}{nh^3}\sum_{j=1}^{n}\left\{E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}|X_j,T_j\right)\begin{bmatrix}A_1 & C_1 \\ C_1^T & D_1\end{bmatrix} + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}|X_j,T_j\right)\begin{bmatrix}A_2 & C_2 \\ C_2^T & D_2\end{bmatrix}\right.$$

$$\left.-E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)}|X_j,T_j\right)\begin{bmatrix}A_3 & C_3 \\ C_3^T & D_3\end{bmatrix}\right\} = -\frac{1}{nh^3}\sum_{j=1}^{n}\begin{bmatrix}\kappa_{1,j,1} & \kappa_{1,j,2} \\ \kappa_{1,j,2}^T & \kappa_{1,j,3}^1\end{bmatrix},$$

where

$$\begin{cases}\kappa_{1,j,1} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\middle|X_j,T_j\right)A_1, \\[2mm]\kappa_{1,j,2} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\middle|X_j,T_j\right)C_1 + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}\middle|X_j,T_j\right)C_2 \\[2mm] &\quad+ E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)}\middle|X_j,T_j\right)C_3, \\[2mm]\kappa_{1,j,3} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\middle|X_j,T_j\right)D_1 + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}\middle|X_j,T_j\right)D_2 \\[2mm] &\quad+ E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)}\middle|X_j,T_j\right)D_3.\end{cases} \tag{F.25}$$

By (F.16) and (F.17), one can show $\kappa_{1,j,1} = -h^3 f^{(2)}(0|X_j,T_j)\tilde{B}_j\tilde{B}_j\{1+o_p(1)\}$ and $\kappa_{1,j,2} = -h^3 f^{(2)}(0|X_j,T_j)\tilde{B}_j\tilde{X}_j^T\{1+o_p(1)\}$. Then we focus on derive the rate of $\kappa_{1,j,3}$. Note that

$$\kappa_{1,j,3} = E\left(-\mathrm{IC}_{j(g,\beta)}^{(2,0)}\begin{bmatrix}1 & W_j \\ W_j & W_j^2\end{bmatrix}\middle|X_j,T_j\right) + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}\begin{bmatrix}0 & 1 \\ 1 & 2W_j\end{bmatrix}\middle|X_j,T_j\right)$$

$$+E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)}\begin{bmatrix}0 & 0 \\ 0 & 1\end{bmatrix}\middle|X_j,T_j\right)$$

$$=\begin{bmatrix}\eta_{j,1} & \eta_{j,2} \\ \eta_{j,2} & \eta_{j,3}\end{bmatrix}, \tag{F.26}$$

where

$$\begin{cases}\eta_{j,1} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\middle|X_j,T_j\right), \\[2mm]\eta_{j,2} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}\middle|X_j,T_j\right)X_j - E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j\middle|X_j,T_j\right) \\[2mm] &\quad+ E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}\middle|X_j,T_j\right), \\[2mm]\eta_{j,3} &= -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}W_j^2\middle|X_j,T_j\right) + 2E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)}W_j\middle|X_j,T_j\right) \\[2mm] &\quad+ E\left(\mathrm{IC}_{j(g,\beta)}^{(2,2)}\middle|X_j,T_j\right).\end{cases} \tag{F.27}$$

161

To derive (F.26), in what follows, we first show that $\eta_{j,2} = \eta_{j,1}X_j$. Then we focus on deriving $\eta_{j,3} = \eta_{j,1}X_j^2$.

By (F.27), $\eta_{j,2}$ contains two extra expectations besides the one defined as $\eta_{j,1}$. By (F.17), these two expectations is equal to zero. Therefore, $\eta_{j,2} = \eta_{j,1}X_j$.

Elaborating $\eta_{j,3}$ in (F.27) yields

$$
\begin{aligned}
\eta_{j,3} = {} & -E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j^2 \middle| X_j, T_j \right\} + 2E\left\{ \mathrm{IS}_{j(g,\beta)}^{(2,1)}U_j \middle| X_j, T_j \right\} + E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,2)} \middle| X_j \right\} + \\
& 2\left\{ -E\left(\mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j \middle| X_j, T_j \right) + E\left(\mathrm{IS}_{j(g,\beta)}^{(2,1)} \middle| X_j, T_j \right) \right\} + \eta_{j,1}X_j^2 \\
= {} & -E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j^2 \middle| X_j, T_j \right\} + 2E\left\{ \mathrm{IS}_{j(g,\beta)}^{(2,1)}U_j \middle| X_j, T_j \right\} \\
& + E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,2)} \middle| X_j, T_j \right\} + \eta_{j,1}X_j^2,
\end{aligned}
$$

where, to reach the last step, we drop the part in the first step that reduces to zero according to the preceding derived results in regard to $\eta_{j,2}$. As for the remaining three expectations in front of $\eta_{j,1}X_j^2$ above, we have

$$
\begin{aligned}
& -E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,0)}U_j^2 \middle| X_j, T_j \right\} + 2E\left\{ \mathrm{IS}_{j(g,\beta)}^{(2,1)}U_j \middle| X_j, T_j \right\} + E\left\{ \mathrm{IC}_{j(g,\beta)}^{(2,2)} \middle| X_j, T_j \right\} \quad \text{(F.28)} \\
= {} & -\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\int u^2\cos\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon + \\
& 2\cdot\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int u\sin\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon + \\
& \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(2)}(\beta_1 t/h)\int \cos\left(\frac{\epsilon - \beta_1 u}{h}t\right)f_U(u)du\,dt\,d\epsilon \\
= {} & \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau(\beta_1 t/h)\phi_U^{(2)}(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon + \\
& 2\cdot\frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(1)}(\beta_1 t/h)\cdot(-1)\cdot\phi_U^{(1)}(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon + \\
& \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\tau^{(2)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)dt\,d\epsilon, \\
= {} & \frac{1}{2\pi}\int f(\epsilon|X_j, T_j)\int t^2\phi_K(t)\left\{ \tau(\beta_1 t/h)\phi_U^{(2)}(\beta_1 t/h) + 2\tau^{(1)}(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h) + \right. \\
& \left. \tau^{(2)}(\beta_1 t/h)\phi_U(\beta_1 t/h) \right\}\cos(\epsilon t/h)dt\,d\epsilon,
\end{aligned}
$$

in the last step of which we use the assumption that $\phi_U(\cdot)$ is even, thus $\phi_U^{(1)}(\cdot)$ is odd and $\phi_U^{(2)}(\cdot)$ is even. By Lemma C.3, the part inside the curly brackets in the last step

162

reduces to zero. Hence, $\eta_{j,3} = \eta_{j,1}X_j^2$. To this end, by (F.16), one can conclude that

$$\begin{cases} \eta_{j,1} & = & -h^3 f^{(2)}(0|X_j)\{1 + o_p(1)\}, \\[2mm] \eta_{j,2} & = & -h^3 f^{(2)}(0|X_j)X_j\{1 + o_p(1)\}, \\[2mm] \eta_{j,3} & = & -h^3 f^{(2)}(0|X_j)X_j^2\{1 + o_p(1)\}. \end{cases} \qquad (\text{F.29})$$

Using (F.29) in (F.26) leads to $\kappa_{1,j,3} = -h^3 f^{(2)}(0|X_j,T_j)\tilde{X}_j\tilde{X}_j\{1 + o_p(1)\}$. By $\kappa_{1,j,i}$, $i = 1,2,3$, the first summand of $E\left(J_n|\boldsymbol{X},\boldsymbol{T}\right)$ is equal to $-h^3 f^{(2)}(0|X_j,T_j)\tilde{Z}_j\tilde{Z}_j\{1 + o_p(1)\}$, where $Z_j = (B_j^T, X_j^T)$.

Similar as the above proof, by the assumption (C^*5), one can show the rate of third summand of $E\left(J_n|\boldsymbol{X},\boldsymbol{T}\right)$ is equal to the rate of first summand of $E\left(J_n|\boldsymbol{X},\boldsymbol{T}\right)$. Then, we only focus on deriving the rate of the second summand of $E\{J_n|\boldsymbol{X},\boldsymbol{T}\}$.

By (F.24), the second summand of $E\{J_n|\boldsymbol{X},\boldsymbol{T}\}$ yields

$$-\frac{1}{nh^3}\sum_{j=1}^{n}\left\{E\left(\text{IS}_{j(g,\beta)}^{(3,0)}|X_j,T_j\right)\begin{bmatrix} A_1 & C_1 \\ C_1^T & D_1 \end{bmatrix} + E\left(\text{IC}_{j(g,\beta)}^{(3,1)}|X_j,T_j\right)\begin{bmatrix} A_2 & C_2 \\ C_2^T & D_2 \end{bmatrix}\right.$$
$$\left. -E\left(\text{IS}_{j(g,\beta)}^{(3,2)}|X_j,T_j\right)\begin{bmatrix} A_3 & C_3 \\ C_3^T & D_3 \end{bmatrix}\right\} = -\frac{1}{nh^3}\sum_{j=1}^{n}\begin{bmatrix} \kappa_{2,j,1} & \kappa_{2,j,2} \\ \kappa_{2,j,2}^T & \kappa_{2,j,3} \end{bmatrix},$$

where

$$\begin{cases} \kappa_{2,j,1} & = & E\left(\text{IS}_{j(g,\beta)}^{(3,0)}\middle| X_j,T_j\right)A_1, \\[2mm] \kappa_{2,j,2} & = & E\left(\text{IS}_{j(g,\beta)}^{(3,0)}\middle| X_j,T_j\right)C_1 + E\left(\text{IC}_{j(g,\beta)}^{(3,1)}\middle| X_j,T_j\right)C_2 \\[2mm] & - & E\left(\text{IS}_{j(g,\beta)}^{(3,2)}\middle| X_j,T_j\right)C_3, \\[2mm] \kappa_{2,j,3} & = & E\left(\text{IS}_{j(g,\beta)}^{(3,0)}\middle| X_j,T_j\right)D_1 + E\left(\text{IC}_{j(g,\beta)}^{(3,1)}\middle| X_j,T_j\right)D_2 \\[2mm] & - & E\left(\text{IS}_{j(g,\beta)}^{(3,2)}\middle| X_j,T_j\right)D_3. \end{cases} \qquad (\text{F.30})$$

163

First, in $\kappa_{2,j,1}$,

$$E\left(\mathrm{IS}_{j(g,\beta)}^{(3,0)}\Big|X_j,T_j\right) \tag{F.31}$$

$$=\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\int \sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon$$

$$=\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\phi_U(-\beta_1 t/h)\sin(\epsilon t/h)\,dt\,d\epsilon,\ \text{by Lemma C.1,}$$

$$=-\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\sin(\epsilon t/h)\,dt\,d\epsilon,\ \text{since } \tau(s)=-1/\phi_U(s),$$

$$=-h\cdot\frac{1}{2\pi}\int f(hs|X_j,T_j)\int t^3\phi_K(t)\sin(st)\,dt\,ds$$

$$=-h\cdot\frac{1}{2\pi}\int\left\{f(0|X_j,T_j)+f^{(2)}(0|X_j)h^2s^2/2\right.$$

$$\left.+f^{(3)}(s^*|X_j,T_j)h^3s^3/6\right\}\int t^3\phi_K(t)\sin(st)\,dt\,ds,$$

$$=-\frac{h^4}{6}\cdot\frac{1}{2\pi}\int f^{(3)}(s^*|X_j,T_j)s^3\int t^3\phi_K(t)\sin(st)\,dt\,ds,$$

where we use the third-order Taylor expansion of $f(sh|X_j)$ around zero in the second to the last step, with s^* lying between 0 and sh; then we use identities in Lemma C.2 to simplify the integrals and keep the only part that is not necessarily equal to zero. Assuming $f^{(3)}(\cdot|X_j)$ bounded and using Lemma C.2, the last expression above suggests that $E(\mathrm{IS}_j^{(3,0)}|X_j)$ is of order $O(h^4)$. Hence, $\kappa_{2,j,1}$ is of order $O(h^4)$.

By the definition of $\mathrm{IS}_{j(g,\beta)}^{(3,0)}$ and $\mathrm{IC}_{j(g,\beta)}^{(3,1)}$, the two terms are involved in $\kappa_{2,j,2}$ by,

$$E\left(U_j\mathrm{IS}_{j(g,\beta)}^{(3,0)}\Big|X_j,T_j\right)+E\left(\mathrm{IC}_{j(g,\beta)}^{(3,1)}\Big|X_j,T_j\right) \tag{F.32}$$

$$=\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)\int u\sin\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon+$$

$$\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\int \cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)f_U(u)\,du\,dt\,d\epsilon$$

$$=\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau(\beta_1 t/h)(-1)\phi_U^{(1)}(-\beta_1 t/h)\cos(\epsilon t/h)\,dt\,d\epsilon+$$

$$\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\cos(\epsilon t/h)\,dt\,d\epsilon,$$

$$=\frac{1}{2\pi}\int f(\epsilon|X_j,T_j)\int t^3\phi_K(t)\left\{\tau(\beta_1 t/h)\phi_U^{(1)}(\beta_1 t/h)\right.$$

$$\left.+\tau^{(1)}(\beta_1 t/h)\phi_U(-\beta_1 t/h)\right\}\cos(\epsilon t/h)\,dt\,d\epsilon$$

$$=0,$$

where the last step is reached because the terms inside the curly brackets in the second-to-the-last step is equal to zero by Lemma C.3. For $\kappa_{2,j,3}$, we have the following elaboration by the definitions of $\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(3,0)}$, $\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(3,1)}$, and $\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(3,2)}$,

$$
\begin{aligned}
& E\left(U_j^2 \mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(3,0)} \,\big|\, X_j, T_j\right) + 2E\left(U_j \mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(3,1)} \,\big|\, X_j, T_j\right) - E\left(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(3,2)} \,\big|\, X_j, T_j\right) \qquad \text{(F.33)} \\
=& \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau(\beta_1 t/h) \int u^2 \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u)\, du\, dt\, d\epsilon + \\
& 2 \cdot \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \int u \cos\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u)\, du\, dt\, d\epsilon - \\
& \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau^{(2)}(\beta_1 t/h) \int \sin\left(\frac{\epsilon - \beta_1 u}{h} t\right) f_U(u)\, du\, dt\, d\epsilon \\
=& \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau(\beta_1 t/h) \cdot (-1) \cdot \phi_U^{(2)}(-\beta_1 t/h) \sin(\epsilon t/h)\, dt\, d\epsilon + \\
& 2 \cdot \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau^{(1)}(\beta_1 t/h) \phi_U^{(1)}(-\beta_1 t/h) \sin(\epsilon t/h)\, dt\, d\epsilon - \\
& \frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \tau^{(2)}(\beta_1 t/h) \phi_U(-\beta_1 t/h) \sin(\epsilon t/h)\, dt\, d\epsilon, \\
=& -\frac{1}{2\pi} \int f(\epsilon|X_j, T_j) \int t^3 \phi_K(t) \sin(\epsilon t/h) \cdot \\
& \left\{ \tau(\beta_1 t/h) \phi_U^{(2)}(\beta_1 t/h) + 2\tau^{(1)}(\beta_1 t/h) \phi_U^{(1)}(\beta_1 t/h) + \tau^{(2)}(\beta_1 t/h) \phi_U(\beta_1 t/h) \right\}\, dt\, d\epsilon
\end{aligned}
$$

$=0$, by Lemma C.3.

Therefore, combine $\kappa_{2,j,i}$, $i = 1, 2, 3$, the second summand of J_n in (D.8) has expectation of order $O(h^4)$.

By assumption (C*5) and the rate of the second summand of $E(J_n|\boldsymbol{X}, \boldsymbol{T})$, one can have the rate of the second term in (F.24) is $o_p(1)$. Similarly, the rate of the third summand of $E(J_n|\boldsymbol{X}, \boldsymbol{T})$ is also equal to $o_p(1)$. In other words, the first term dominates the other two term in (F.24), which leads to (F.2).

Part (IV): Show (F.5) and (F.8).

By (D.8), the order of $\mathrm{Var}(J_n)$ is determined by the first term and the order of the first term is dominated by the rates of $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)})^2 W_j^k | X_j, T_j\}$ for $k = 0, 1, 2, 3, 4$, the rates of $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(2,1)})^2 W_j^k | X_j, T_j\}$ for $k = 0, 1, 2$, and that of $E\{(\mathrm{IC}_j^{(2,2)})^2 | X_j, T_j\}$. It can be shown that, if $E(U^4) < \infty$, $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)})^2 W_j^k | X_j, T_j\}$, for $k = 1, 2, 3, 4$, have the same rate as that of $E\{(\mathrm{IC}_{j(g,\boldsymbol{\beta})}^{(2,0)})^2 | X_j, T_j\}$, and $E\{(\mathrm{IS}_{j(g,\boldsymbol{\beta})}^{(2,1)})^2 W_j^k | X_j, T_j\}$, for $k = 1, 2$,

have the same rate as that of $E\{(\mathrm{IS}^{(2,1)}_{j(g,\beta)})^2|X_j,T_j\}$. In what follows, we focus on deriving the rates of $E\{(\mathrm{IC}^{(2,0)}_{j(g,\beta)})^2|X_j,T_j\}$, $E\{(\mathrm{IS}^{(2,1)}_{j(g,\beta)})^2|X_j,T_j\}$, and $E\{(\mathrm{IC}^{(2,2)}_j)^2|X_j,T_j\}$.

We first find an upper bound for each of these three expectations as follows,

$$
E\left\{\left(\mathrm{IC}^{(2,0)}_{j(g,\beta)}\right)^2\Big|X_j,T_j\right\} = \int f_U(u)\int\left\{\frac{1}{2\pi}\int t^2\cos\left(\frac{\epsilon-\beta_1 u}{h}t\right)\phi_K(t)\tau(\beta_1 h/t)dt\right\}^2
$$
$$
f(\epsilon|X_j,T_j)d\epsilon\,du
$$
$$
= h\int f_U(u)\int |-\mathcal{F}_{2,0}(s)|^2 f(\beta_1 u+sh|X_j,T_j)ds\,du
$$
$$
\leq B_f h\int |\mathcal{F}_{2,0}(s)|^2 ds.
$$

Similarly,

$$
E\left\{\left(\mathrm{IS}^{(2,1)}_{j(g,\beta)}\right)^2\Big|X_j,T_j\right\} \leq B_f h\int|\mathcal{F}_{2,1}(s)|^2 ds,
$$
$$
E\left\{\left(\mathrm{IC}^{(2,2)}_{j(g,\beta)}\right)^2\Big|X_j,T_j\right\} \leq B_f h\int|\mathcal{F}_{2,2}(s)|^2 ds.
$$

When U is ordinary smooth, recalling the pattern stated at the end of Part (II), $\int|\mathcal{F}_{2,0}(s)|^2 ds$ dominates $\int|\mathcal{F}_{2,1}(s)|^2 ds$ and $\int|\mathcal{F}_{2,2}(s)|^2 ds$, and $\int|\mathcal{F}_{2,0}(s)|^2 ds$ is of the same order as $\int|\mathcal{F}_{1,0}(s)|^2 ds$, which we looked into in Part (II). Using the results regarding $\int|\mathcal{F}_{1,0}(s)|^2 ds$, we have $\int|\mathcal{F}_{2,0}(s)|^2 ds = O(h^{-2b})$ if U is ordinary smooth, and thus $E\{(\mathrm{IC}^{(2,0)}_{j(g,\beta)})^2|X_j,T_j\} = O(h^{1-2b})$, which dominates $E\{(\mathrm{IS}^{(2,1)}_j)^2|X_j,T_j\}$, and $E\{(\mathrm{IC}^{(2,2)}_{j(g,\beta)})^2|X_j,T_j\}$. These rates lead to (F.5).

If U is super smooth, under Conditions S, similar to the proof in Part (II), one can show that

$$
E\{(\mathrm{IC}^{(2,0)}_{j(g,\beta)})^2|X_j,T_j\} = O\{h^{1+2b_2}\exp(2|\beta_1|^b h^{-b}/d_2)\},
$$
$$
E\{(\mathrm{IS}^{(2,1)}_{j(g,\beta)})^2|X_j,T_j\} = O\{h^{1+2b_3}\exp(2|\beta_1|^b h^{-b}/d_2)\},
$$
$$
E\{(\mathrm{IC}^{(2,2)}_{j(g,\beta)})^2|X_j,T_j\} = O\{h^{1+2b_4}\exp(2|\beta_1|^b h^{-b}/d_2)\},
$$

where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)})I(2b_0 - b_1^{(1)} < 0.5)$, and $b_4 = \min\{(2b_0 - b_1^{(2)})I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)})I(3b_0 - 2b_1^{(1)} < 0.5)\}$. Using these rates one can establish (F.8).

Part (V): Show (F.3).

Because $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq c\delta_n$, under conditions that guarantee boundedness of $\mathrm{IS}_{j(\gamma,\beta)}^{(4,\ell_2)}$ and $\mathrm{IC}_{j(\gamma,\beta)}^{(4,\ell_2)}$, for $\ell_2 = 0,1,2,3,4$, L_n evaluated at $\boldsymbol{\theta}^*$ and L_n evaluated at $\boldsymbol{\theta}$ are of the same order. Hence, we focus on studying L_n evaluated at $\boldsymbol{\theta}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T)$ in the sequel.

By taylor expansion, (D.9) yields

$$
\begin{aligned}
L_n =& \frac{1}{nh^4}\sum_{j=1}^{n}\left(\mathrm{IS}_{j(g,\beta)}^{(3,0)}\begin{bmatrix} A_1\tilde{B}_j + C_1\tilde{W}_j \\ C_1^T\tilde{B}_j + D_1\tilde{W}_j \end{bmatrix} + \mathrm{IC}_{j(g,\beta)}^{(3,1)}\begin{bmatrix} C_1\tilde{e}_1 + A_2\tilde{B}_j + C_2\tilde{W}_j \\ D_1\tilde{e}_1 + C_2^T\tilde{B}_j + D_2\tilde{W}_j \end{bmatrix} \right. \\
& \left. - \mathrm{IS}_{j(g,\beta)}^{(3,2)}\begin{bmatrix} C_2\tilde{e}_1 + A_3\tilde{B}_j + C_3\tilde{W}_j \\ D_2\tilde{e}_1 + C_3^T\tilde{B}_j + D_3\tilde{W}_j \end{bmatrix} - \mathrm{IC}_{j(g,\beta)}^{(3,3)}\begin{bmatrix} C_3\tilde{e}_1 \\ D_3\tilde{e}_1 \end{bmatrix} \right) \\
& + \frac{1}{nh^4}\sum_{j=1}^{n}\left(\mathrm{IC}_{j(g,\beta)}^{(4,0)}\begin{bmatrix} A_1\tilde{B}_j + C_1\tilde{W}_j \\ C_1^T\tilde{B}_j + D_1\tilde{W}_j \end{bmatrix} - \mathrm{IS}_{j(g,\beta)}^{(4,1)}\begin{bmatrix} C_1\tilde{e}_1 + A_2\tilde{B}_j + C_2\tilde{W}_j \\ D_1\tilde{e}_1 + C_2^T\tilde{B}_j + D_2\tilde{W}_j \end{bmatrix} \right. \\
& \left. - \mathrm{IC}_{j(g,\beta)}^{(4,2)}\begin{bmatrix} C_2\tilde{e}_1 + A_3\tilde{B}_j + C_3\tilde{W}_j \\ D_2\tilde{e}_1 + C_3^T\tilde{B}_j + D_3\tilde{W}_j \end{bmatrix} + \mathrm{IS}_{j(g,\beta)}^{(4,3)}\begin{bmatrix} C_3\tilde{e}_1 \\ D_3\tilde{e}_1 \end{bmatrix} \right)h^{-1}R_j \\
& - \frac{1}{2nh^4}\sum_{j=1}^{n}\left(\mathrm{IS}_{\eta_j^*}^{(5,0)}\begin{bmatrix} A_1\tilde{B}_j + C_1\tilde{W}_j \\ C_1^T\tilde{B}_j + D_1\tilde{W}_j \end{bmatrix} + \mathrm{IC}_{\eta_j^*}^{(5,1)}\begin{bmatrix} C_1\tilde{e}_1 + A_2\tilde{B}_j + C_2\tilde{W}_j \\ D_1\tilde{e}_1 + C_2^T\tilde{B}_j + D_2\tilde{W}_j \end{bmatrix} \right. \\
& \left. - \mathrm{IS}_{\eta_j^*}^{(5,2)}\begin{bmatrix} C_2\tilde{e}_1 + A_3\tilde{B}_j + C_3\tilde{W}_j \\ D_2\tilde{e}_1 + C_3^T\tilde{B}_j + D_3\tilde{W}_j \end{bmatrix} - \mathrm{IC}_{\eta_j^*}^{(5,3)}\begin{bmatrix} C_3\tilde{e}_1 \\ D_3\tilde{e}_1 \end{bmatrix} \right)h^{-2}R_j^2
\end{aligned}
$$

By (F.31), (F.32) and (F.33), similar as proof in Part(I) and Part(III), the order of the first summand of expectation of L_n and the order the third summand of expectation of L_n is equal to $O_p(h^4)$. By (F.16) and (F.28), the order of the second summand of expectation of L_n is equal to $O_p(h^3)$. Based on the assumption (C*5), the order of expectation of L_n is $O_p(1)$, which leads to (F.3).

Part (VI): Show (F.6) and (F.9).

By (D.9), the order of $\mathrm{Var}(L_n)$ is determined by the orders of $E\{(\mathrm{IS}_{j(g,\beta)}^{(3,0)})^2|X_j, T_j\}$, $E\{(\mathrm{IC}_{j(g,\beta)}^{(3,1)})^2$

167

$|X_j, T_j\}$, $E\{(\text{IS}_{j(g,\beta)}^{(3,2)})^2|X_j, T_j\}$, and $E\{(\text{IC}_j^{(3,3)})^2|X_j, T_j\}$. Following similar arguments as those in Parts (II) and (IV), one can show that these expectations are bounded from above by $B_f h \int |\mathcal{F}_{3,\ell_2}|^2 ds$, for $\ell_2 = 0, 1, 2, 3$, respectively. If U is ordinary smooth, $\int |\mathcal{F}_{3,0}|^2 ds$ dominates the other three according to the patterns pointed out at the end of Part (II), which of the same order as $\int |\mathcal{F}_{1,0}|^2 ds$ derived there. Therefore, using results from Part (II), we have the order of $\text{Var}(L_n)$ besing $O_p\{1/(nh^{7+2b})\}$ if U is ordinary smooth. This proves (F.6). If U is super smooth, one can show that

$$E\{(\text{IS}_{j(g,\beta)}^{(3,0)})^2|X_j, T_j\} = O\{h^{1+2b_2}\exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\text{IC}_{j(g,\beta)}^{(3,1)})^2|X_j, T_j\} = O\{h^{1+2b_3}\exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\text{IS}_{j(g,\beta)}^{(3,2)})^2|X_j, T_j\} = O\{h^{1+2b_4}\exp(2|\beta_1|^b h^{-b}/d_2)\},$$

$$E\{(\text{IC}_{j(g,\beta)}^{(3,3)})^2|X_j, T_j\} = O\{h^{1+2b_5}\exp(2|\beta_1|^b h^{-b}/d_2)\},$$

where $b_2 = b_0 I(b_0 < 0.5)$, $b_3 = (2b_0 - b_1^{(1)})I(2b_0 - b_1^{(1)} < 0.5)$, $b_4 = \min\{(2b_0 - b_1^{(2)})I(2b_0 - b_1^{(2)} < 0.5), (3b_0 - 2b_1^{(1)})I(3b_0 - 2b_1^{(1)} < 0.5)\}$, and $b_5 = \min\{(2b_0 - b_1^{(3)})I(2b_0 - b_1^{(3)} < 0.5), (3b_0 - b_1^{(1)} - b_1^{(2)})I(3b_0 - b_1^{(1)} - b_1^{(2)} < 0.5), (4b_0 - b_1^{(1)} - 2b_1^{(2)})I(4b_0 - b_1^{(1)} - 2b_1^{(2)} < 0.5)\}$. Putting these rates together gives (F.9).

$\square$

168

Appendix G

Computer codes for analyzing the dietary data using the Monte Carlo corrected score method

The main MATLAB code for carrying out linear mode regression analysis of the dietary data using the Monte Carlo corrected score method is given first. The dietary data is saved in wishers.csv. The main code calls three functions, named CV_1, CV_2, and MCCS. The MCCS function calls another function named phi_mcb. These four functions are given after the main code next.

(I) Main code for applying the MCCS method:

```
%*****************************************************************%
%                                                               %
%  MATLAB code to analyze dietary data by using Monte Carlo     %
%  Corrected Score method (MCCS).                               %
%                                                               %
%*****************************************************************%
clear;clc;
%*****************************************************************%
% Set path to read the dietary data
% Specify your path to read the data
%*****************************************************************%
addpath('Your Path to Load Data')
```

```matlab
%*********************************************************************%
% Part I: Data cleaning
%*********************************************************************%
% Data claim:
% y_up: FFQ intake measured as the percent calories from fat,
%       Response Variable Y.
% w_n1: average of six recalls from each subject as a surrogate
%       of this subject's long-term intake, covariate W.
% fri_n,i=1,...,6: six 24-hour food recalls on randomly selected days.

% Read data from 'wishreg.csv'
data = readtable('wishreg.csv');
s_ize = size(data);
% Scale and center all variables
fr1_n = (data.fr1 - mean(data.fr1)) / std(data.fr1);
fr2_n = (data.fr2 - mean(data.fr2)) / std(data.fr2);
fr3_n = (data.fr3 - mean(data.fr3)) / std(data.fr3);
fr4_n = (data.fr4 - mean(data.fr4)) / std(data.fr4);
fr5_n = (data.fr5 - mean(data.fr5)) / std(data.fr5);
fr6_n = (data.fr6 - mean(data.fr6)) / std(data.fr6);
y=data.ffq;
y_up = (y-mean(y)) / std(y);
% Obtain contaminated covariate W
w_n1 = ( fr1_n + fr2_n + fr3_n + fr4_n + fr5_n + fr6_n ) / 6 ;

%*********************************************************************%
```

170

```matlab
% Part II: Estimate the variance of measurement errros
%********************************************************************%
fr_t = cat(2,fr1_n,fr2_n,fr3_n,fr4_n,fr5_n,fr6_n);
gama_u2 = sum( sum((fr_t - repmat(w_n1,[1,6]) ).^2,2)) / 5 / s_ize(1);
sigma_u2 = gama_u2 / 6;


%********************************************************************%
% Part III: Implement MCCS method
%********************************************************************%


x_total = w_n1';
y_total = y_up';
% B and options are parameters in CV_1 and CV_2 function.
B=10;
options = optimset('Display','off');


x_con = x_total;
y = y_total;
% Bandwidth Selection using SIMEX method,
% (-0.2672,0.3627) is naive estimate as the starting point
parms = [-0.2672;0.3627];
f1 = @(Lambda)CV_1(parms, x_con, y,Lambda, B, s_ize(1) ,options );
f2 = @(Lambda)CV_2(parms, x_con, y,Lambda, B, s_ize(1) ,options );
h_1 = fminbnd(f1,0.6,1.6);
h_2 = fminbnd(f2,0.6,1.7);
h = h_1^2/h_2;
```

171

```
% Obtain final estimates from MCCS method
if isnan(h)

    beta = NaN(1,2);

else

    u_b = normrnd(0,0.3390,1000,s_ize(1));

    f = @(beta)MCCS(beta,h,x_con,y,u_b,s_ize(1));

    beta = fsolve(f,[-0.2672;0.3627],options);

end


table(beta(1),beta(2),'VariableNames',{'beta_0','beta_1'})
```

(II) Function CV_1:

```
********************************************************************%
% CV_1 function computes the value of h_{1} in SM_2 step
%*******************************************************************%
% Arguments:
% parms = starting points of intercept and slope paramters
% Lambda = Bandwidth
% x_cont = Contaminated covariate W
% y_o = Response variable
% B = the number of further contaminated covariate data in Section 3.3
% n_o = sample size
% options
%= control whether or not to show results from "fslove" function
% Outputs:
% out: MSE
function [ out ] = CV_1( parms, x_cont, y_o,Lambda, B,n_o,options)
```

```
beta_1 = zeros(B,2);

beta_2 = zeros(B,2);


for i=1:B
        u_b =  normrnd(0,0.3390,1000,n_o);
        u_b1 =  normrnd(0,0.3390,1000,n_o);
        w_st = x_cont +  normrnd(0,0.3390,1,n_o);
        f_1 = @(beta)MCCS(beta,Lambda,x_cont,y_o,u_b,n_o);
        beta_1(i,:) = fsolve(f_1,parms,options);
        f_2 = @(beta)MCCS(beta,Lambda,w_st,y_o,u_b1,n_o);
        beta_2(i,:) = fsolve(f_2,parms,options);
end
  beta_f = beta_2-beta_1;
  beta_f( (beta_1(:,1)==2 & beta_1(:,2)==5)
              |(beta_2(:,1)==2 & beta_2(:,2)==5),:)=[];
  dim_beta = size(beta_f);
  if dim_beta(1) <= 1
      d_st_1 = NaN(1);
  else
      S_star = cov(beta_f);
      d_st_1 = diag((beta_f) / S_star * (beta_f)');
  end
  out = mean(d_st_1);
end

(III) Function CV_2:
```

```matlab
%*******************************************************************%
% CV_2 function computes the value of h_{2} in SM_4 step
%*******************************************************************%
% Arguments:
% parms = starting points of intercept and slope paramters
% Lambda = Bandwidth
% x_cont = Contaminated covariate W
% y_o = Response variable
% B = the number of further contaminated covariate data in Section 3.3
% n_o = sample size
% options
% = control whether or not to show results from "fslove" function
% Outputs:
% out: MSE
function [ out ] = CV_2( parms, x_cont, y_o,Lambda, B, n_o, options  )

 beta_2 = zeros(B,2);
 beta_3 = zeros(B,2);

 for i=1:B
        u_b =  normrnd(0,0.3390,1000,n_o);
        u_b1 =  normrnd(0,0.3390,1000,n_o);
        w_st = x_cont +  normrnd(0,0.3390,1,n_o);
        w_st2 = w_st +  normrnd(0,0.3390,1,n_o);
        f_1 = @(beta)MCCS(beta,Lambda,w_st,y_o,u_b,n_o);
        beta_2(i,:) = fsolve(f_1,parms,options);
        f_2 = @(beta)MCCS(beta,Lambda,w_st2,y_o,u_b1,n_o);
```

```matlab
        beta_3(i,:) = fsolve(f_2,parms,options);
    end


    beta_f = beta_3-beta_2;
    beta_f( (beta_2(:,1)==2 & beta_2(:,2)==5)
             | (beta_3(:,1)==2 & beta_3(:,2)==5) , :)=[];
    dim_beta = size(beta_f);
    if dim_beta(1) <= 1
        d_st_2 = NaN(1);
    else
        S_star2 = cov(beta_f);
        d_st_2 = diag((beta_f) / S_star2 * (beta_f)');
    end
    out = mean(d_st_2);
end
```

(IV) MCCS function:

```matlab
%********************************************************************%
% MCCS function computes the sum of \Psi_{MC,B} in MC_4 step
%********************************************************************%
% Arguments:
% beta = (\beta_{0},\beta_{1}) intercept and slope paramters
% Lambda = Bandwidth
% x_cont = Contaminated covariate W
% y_o = Response variable
% u_b = Indepedent random errors generated from Normal distribution
% n_o = sample size
```

```matlab
% Outputs:
% out: sum of \Psi_{MC,B} in MC_4 step


function [ out ] = MCCS( beta,Lambda, x_cont, y_o,u_b,n_o)
f = zeros(n_o,2);
for i=1:n_o
% phi_mcb function is used to compute \Psi_{MC,B} in MC_3 Step
  f(i,:) =  phi_mcb(beta,Lambda,x_cont(i),y_o(i),u_b(:,i));
end
out = sum(f,1);
end
```

(V) phi_mcb function:

```matlab
%*******************************************************************%
% phi_mcb function computes the value of \Psi_{MC,B} in MC_3 step
%*******************************************************************%
% Arguments:
% beta = (\beta_{0},\beta_{1}) intercept and slope paramters
% Lambda = Bandwidth
% x_cont = Contaminated covariate W
% y_o = Response variable
% u_b = Indepedent random errors generated from Normal distribution
% Outputs:
% out: the value of \Psi_{MC,B} in MC_3 step


function [out]=phi_mcb(beta,Lambda,x_cont,y_o,u_b)
w_con = complex(x_cont,u_b);
```

```
I_one = ones(1000,1);

x_new = [I_one,w_con];

const = 1/sqrt(2*pi)*exp( - ( y_o - x_new*beta).^2/2/Lambda^2 ).*

(( y_o - x_new*beta)/Lambda^2 );

const_1 = [const,const];

re = real(x_new.*const_1);

out = mean(re,1);

end
```

Appendix E: Computer codes for analyzing the dietary data using the corrected kernel method

The main R code for carrying out linear mode regression analysis of the dietary data using the corrected kernel method assuming Laplace measurement error is given first. The main code calls FFT.cpp that is given following the main code.

(I) The main R code for implementing the corrected kernel method:

```
############################################################
#   R Code to analyze dietary data using the corrected kernel method
#   assuming Laplace measurement error.
############################################################



##############################################################
# Set the path to read the dietary data
# Specify your path to read the data and 'FFT.cpp' file
##############################################################
setwd("Your Path to Load the Data")
```

177

```
###################################################################
# required packages
###################################################################
library(Rcpp)

library(RcppArmadillo)

library(rmutil)

library(nloptr)



##################################################################
# Rcpp is used to compute the value of objective function in mode
# regression for a given beta
###############################################################
sourceCpp('FFT.cpp')


###############################################################
# Arguments in FFT_AP function:
# input  - m input values of K^{*}(t), m = 2^16 in our simulation.
# mconst -
# positive and negative sign corresponding to m input values in FFT.
# beta   - the interval of t for the m input values of K^{*}(t).
# m      - the number of input values of K^{*}(t).
###########################################################
m = 2^16

beta = sqrt((2*pi)/m)

input  = seq(-m*beta/2, m*beta/2-beta, by=beta)
```

```
mconst = (-1)^(0:(m-1))


##############################################################
# Part I: Bandwidth selection using SIMEX method.
##############################################################
# CV_1 function
#computes the value of h_{1} in SM_2 step in the manuscript.
# Arguments:
# start = staring point of estimates
# x_con = contaminated covariate W
# y = response variable
# B = the number of estimators
# n_o = sample size
# Outputs:
# M1/M2: MSE
cV1 <- function(start, lambda, x_con, y, B,n_o)
{
h <- lambda
beta_1 <- beta_2 <- matrix(0,B,2)
d_st_1 <- array(0,B)
        beta_11 <- bobyqa(start,fn=FFT_AP,x_cont=x_con,y_o=y,
                                   input=input, Lambda=h,
          beta=beta, mconst=mconst)$par
for (i in 1:B)
{
w_st <- x_con + rlaplace(n_o,0,0.3390/sqrt(2))
    beta_1[i,] <- beta_11
```

```r
        beta_2[i,] <- bobyqa(start,fn=FFT_AP,x_cont=w_st,y_o=y,
                                        input=input,Lambda=h,
        beta=beta, mconst=mconst)$par
        }


        S_star <- cov(beta_2-beta_1)
        d_st_1 <- diag( (beta_2-beta_1)\%*\%solve(S_star)
                        \%*\%t(beta_2-beta_1) )


        M1 <- mean(d_st_1)
        return(M1)


}


# CV_2 function computes the value of h_{2} in SM_4 step
# Arguments:
# start = staring point of estimtors
# x_con = contaminated covariate W
# y = response varialbe
# B = the number of estimators
# n_o = sample size
# Outputs:
# M1/M2: MSE
cV2 <- function(start, lambda, x_con, y, B, n_o)
{
h <- lambda
        beta_2 <- beta_3 <- matrix(0,B,2)
```

```
d_st_2 <- array(0,B)

for (i in 1:B)

{

    w_st <- x_con + rlaplace(n_o,0,0.3390/sqrt(2))

    w_st2 <- w_st + rlaplace(n_o,0,0.3390/sqrt(2))

    beta_2[i,] <- bobyqa(start,fn=FFT_AP,x_cont=w_st,y_o=y,

                                        input=input, Lambda=h,

  beta=beta, mconst=mconst)$par

    beta_3[i,] <- bobyqa(start,fn=FFT_AP,x_cont=w_st2,y_o=y,

                                        input=input, Lambda=h,

            beta=beta, mconst=mconst)$par

}


  S_star2 <- cov(beta_3-beta_2)

d_st_2 <- diag( (beta_3-beta_2)\%*\%solve(S_star2)

                \%*\%t(beta_3-beta_2) )

  M2 <-  mean(d_st_2)


  return(M2)

}


###########################################################

# Part II: Data cleaning

###########################################################

# Data claim:

# y_total: FFQ intake measured as the percent calories from fat,

# Response Variable Y.
```

```r
# x_total: average of six recalls from each subject as
# a surrogate of this subject's  long-term intake, covariate W.
# (fr1_n, fr2_n, fr3_n, fr4_n, fr5_n, fr6_n):
#  six 24-hour food recalls on randomly selected days.


# read data from 'wishreg.csv'
data = read.csv('wishreg.csv', header=TRUE,sep=",")
n = dim(data)[1]
# scale and center all variables
y_total = (data$ffq-mean(data$ffq)) / sqrt(var(data$ffq))
fr1_n = (data$fr1 - mean(data$fr1)) / sqrt(var(data$fr1));
fr2_n = (data$fr2 - mean(data$fr2)) / sqrt(var(data$fr2));
fr3_n = (data$fr3 - mean(data$fr3)) / sqrt(var(data$fr3));
fr4_n = (data$fr4 - mean(data$fr4)) / sqrt(var(data$fr4));
fr5_n = (data$fr5 - mean(data$fr5)) / sqrt(var(data$fr5));
fr6_n = (data$fr6 - mean(data$fr6)) / sqrt(var(data$fr6));
# obtain contaminated covariate W
x_total = ( fr1_n + fr2_n + fr3_n + fr4_n + fr5_n + fr6_n ) / 6 ;


##################################################
# Part III: Implement the corrected kernel method assuming
# Laplace measurement error
##################################################


# Bandwidth selection using SIMEX method,
# (-0.2672, 0.3627) is the naive estimator,
# used as starting point
```

```r
y <- y_total

x_con <- x_total

h_1 <- optimize(cV1,c(0.5,1.5),start=c(-0.2672,0.3627),

                      x_con = x_con, y = y,

B=10, n)$minimum

h_2 <- optimize(cV2,c(0.5,1.7),start=c(-0.2672,0.3627),

                      x_con = x_con, y = y,

        B=10, n)$minimum

h_select_n <- h_1^2/h_2

# obtain final estimates

beta_e <- bobyqa(c(-0.2672,0.3627),fn=FFT_AP,x_cont=x_con,y_o=y,

                      input=input, Lambda=h_select_n,

                      beta=beta, mconst=mconst)$par


# Compute estimates \beta_{0} and \beta_{1}

beta_e <- matrix(beta_e,1,2)

colnames(beta_e) = c("beta_0","beta_1")

rownames(beta_e) = c("")

round(beta_e,4)
```

(II) FFT.cpp:

```
//********************************************************//
//  This document includes two functions:
//  1) double func(double x, double beta1, double Lambda)
//   is used to calculate \phi_{K} / \phi_{U}
// in the corrected kernel K^{*}.
//  2) double FFT_AP  is used to compute
```

183

```
//  the objective function in the corrected kernel method
//   using fast Fourier transformation.
//*****************************************************//


//*****************************************************//
//required library
//*****************************************************//
#include <RcppArmadillo.h>
#include <Rcpp.h>


// [[Rcpp::depends(RcppArmadillo)]]
using namespace arma;
using namespace Rcpp;
using namespace std;


//*****************************************************//
// func is used to compute \phi_{K}/\phi_{U}
// in the corrected kenrel K^{*}
// Arguments:
// beta1  - slope parameter in mode regression
// Lambda - bandwidth in the corrected kernel method
//*****************************************************//
// [[Rcpp::export]]
double func(double x, double beta1, double Lambda)
{
    if (  x <=1 && x >=-1 ) {
        return pow(1-pow(x,2),3) / (2*pow(Lambda,2) /
```

184

```cpp
        ( 2*pow(Lambda,2) + pow(0.3390,2)*pow(beta1,2)*pow(x,2) ));

    }else{

        return 0;

    }

}


//***********************************************//

// FFT_AP computes the objective function in the corrected

//  kernel method using fast Fourier transformation (FFT).

// Arguments:

// parms  - intercept and slope parameters in mode regression.

// x_cont - contaminated covariate W.

// y_o    - response variable Y.

// Lambda - bandwidth in the corrected kernel method.

// input  - m input values of K^{*}(t), m = 2^16 in the simulation.

// mconst - positive and negative sign corresponding to m

input values in FFT.

// beta   - the interval of t for the m input values of K^{*}(t).

//***********************************************//

// [[Rcpp::export]]

double FFT_AP(NumericVector& parms, NumericVector& x_cont,

NumericVector& y_o, const NumericVector& input, double Lambda,

double beta, const NumericVector& mconst)

{

    int m = input.size();

    NumericVector FKoutput(m);

    NumericVector re = (y_o-parms[0]-parms[1]*x_cont)/Lambda;
```

185

```
int m_mid = m/2;

FKoutput[m_mid] = func(input[m_mid],parms[1],Lambda);

int i_d = 1;

double indicator1 = 1;

double indicator2 = 1;

while ( ( abs(indicator1) > 1e-30

         || abs(indicator2) > 1e-30 ) && (i_d<m/2)){

    indicator1 = func(input[m_mid-i_d],parms[1],Lambda);

    indicator2 = func(input[m_mid+i_d],parms[1],Lambda);

    FKoutput[m_mid-i_d] = indicator1;

    FKoutput[m_mid+i_d] = indicator2;

    i_d=i_d+1;

}

arma::vec FK_op(FKoutput.begin(), m, false);

arma::vec FK_i_f(m,fill::zeros);

arma::cx_vec FK_f(FK_op,FK_i_f);

arma::vec mcon(const_cast<NumericVector

                      &>(mconst).begin(), m, false);


arma::cx_vec fXF = mcon*beta%arma::ifft(mcon%FK_f)/2/PI*m;


arma::vec fhat = arma::real(fXF);


int n_re = re.size();

NumericVector f_estimate(n_re);

for (int i=0; i<n_re; i++) {
```

```
        int ind = (int)(round(re[i]/beta+m_mid));

        f_estimate[i] = fhat[ind];

    }

    double f_e = Rcpp::sum(f_estimate)/n_re/-Lambda ;


    return f_e;

}
```

Supplement Materials for Graphs and Tables

Table H.1: Two-layer tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 200$ under (F2). Numbers in parentheses are ($10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
	$U \sim N(0, \sigma^2)$					
Naive	1.56	0.43	2.01	0.29	2.57	0.12
	(0.21)	(0.24)	(0.19)	(0.25)	(0.14)	(0.16)
CK	2.49	0.13	2.65	0.09	2.79	0.08
	(0.18)	(0.05)	(0.18)	(0.05)	(0.15)	(0.12)
	$U \sim \text{Laplace}(0, \sigma^2)$					
Naive	1.70	0.42	2.14	0.24	2.61	0.10
	(0.26)	(0.26)	(0.21)	(0.22)	(0.16)	(0.13)
CK	2.78	0.11	2.83	0.09	2.82	0.06
	(0.24)	(0.05)	(0.20)	(0.05)	(0.14)	(0.03)
TRUE	2.88	0.07				
	(0.16)	(0.16)				

Table H.2: Two-layer tuning parameter selection. Averages of parameter estimates over 300 repetitions when $n = 400$ under (F2). Numbers in parentheses are ($10 \times$ standard errors) associate with the averages. The truth is $\beta_1 = 3$.

	75%		85%		95%	
	β_1	NE^2	β_1	NE^2	β_1	NE^2
	$U \sim N(0, \sigma^2)$					
Naive$_N$	1.58	0.33	2.05	0.15	2.57	0.06
	(0.14)	(0.10)	(0.11)	(0.06)	(0.10)	(0.05)
CK$_N$	2.47	0.08	2.68	0.05	2.81	0.04
	(0.15)	(0.03)	(0.13)	(0.02)	(0.10)	(0.02)
	$U \sim \text{Laplace}(0, \sigma^2)$					
Naive	1.77	0.25	2.18	0.13	2.63	0.06
	(0.17)	(0.10)	(0.14)	(0.05)	(0.12)	(0.07)
CK	2.84	0.06	2.85	0.04	2.88	0.03
	(0.16)	(0.03)	(0.13)	(0.02)	(0.10)	(0.01)
TRUE	2.91	0.03				
	(0.09)	(0.09)				
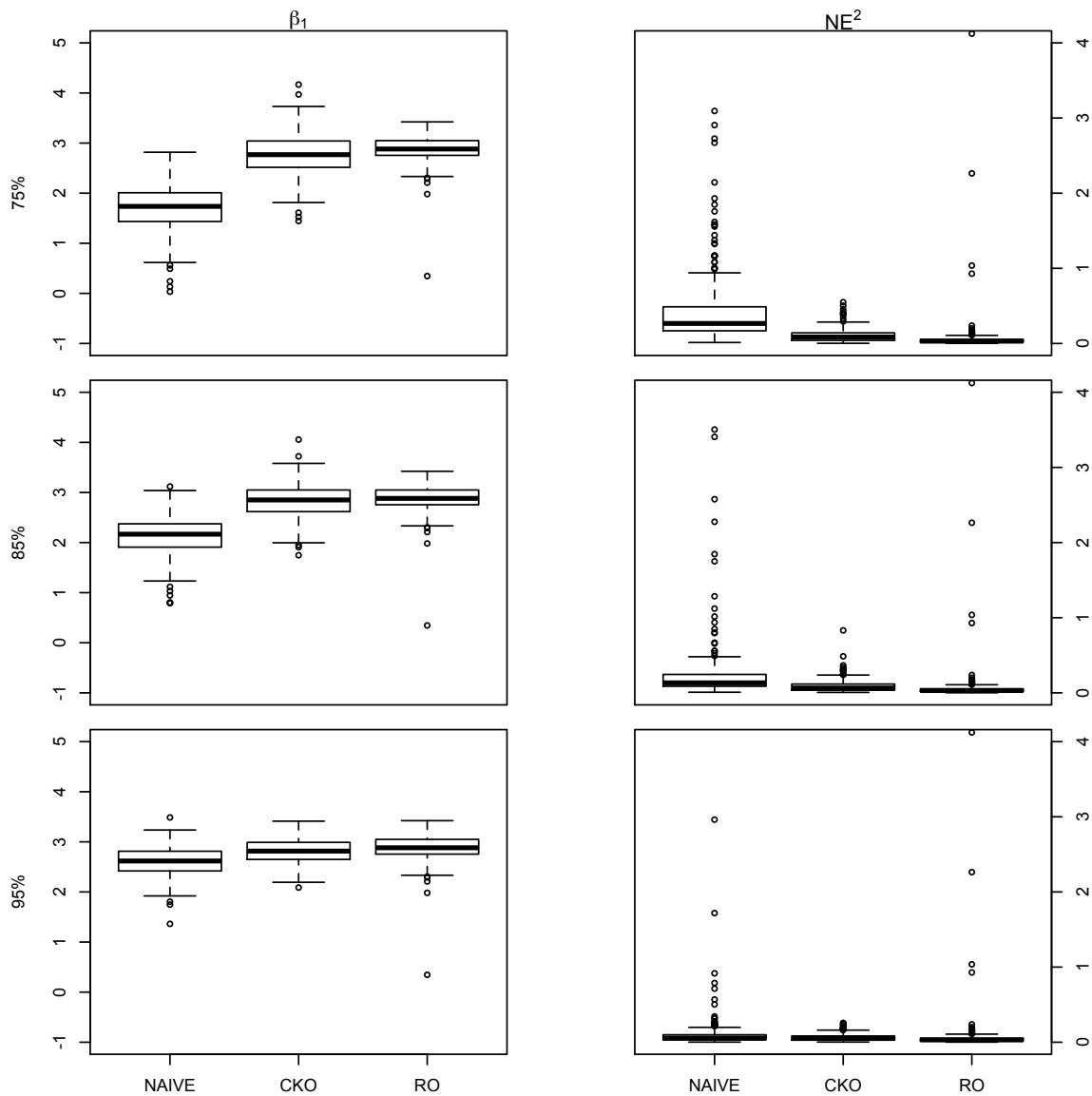
Figure H.1: Two-layer tuning parameter selection. Under the simulation setting (F2) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
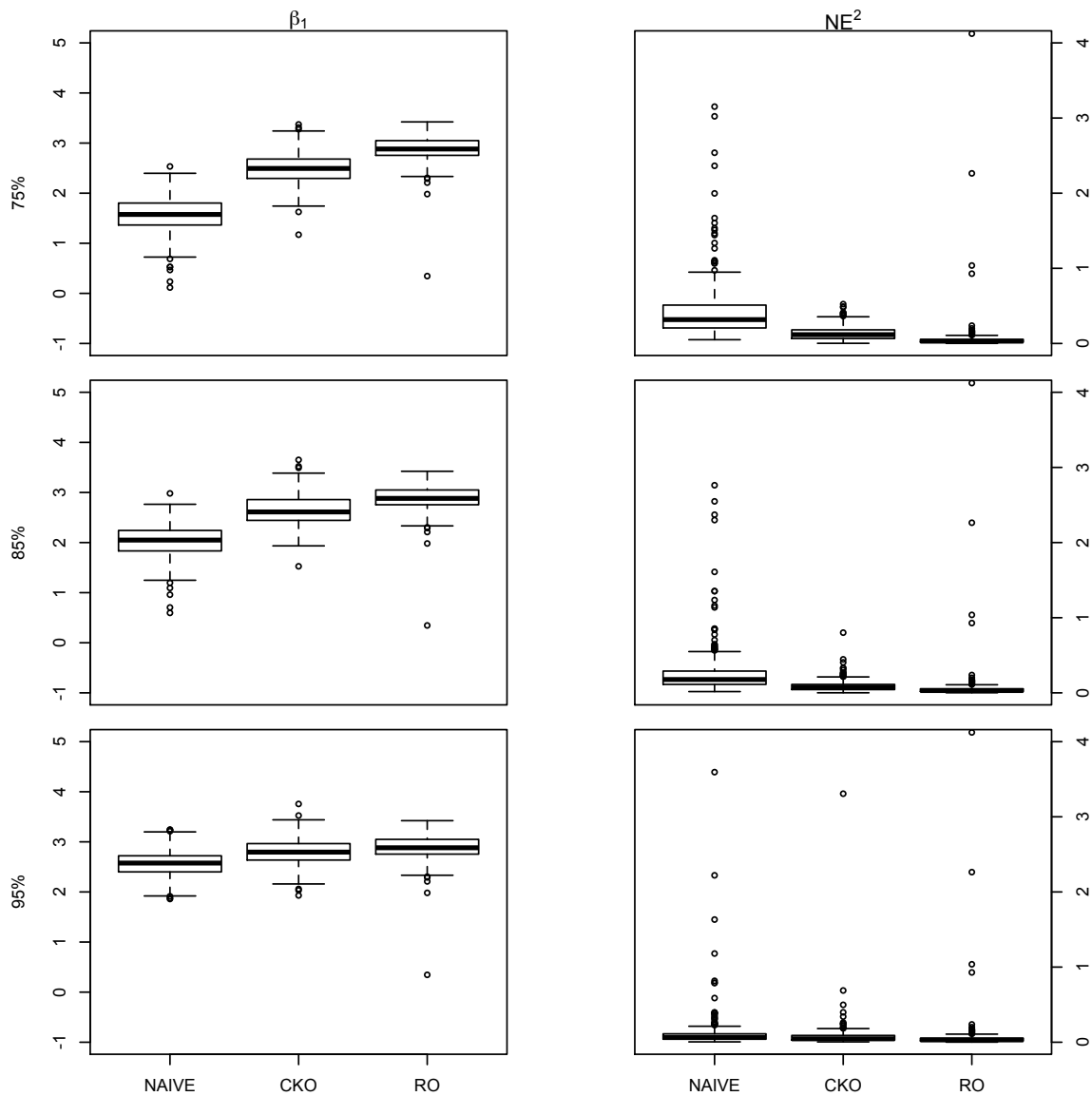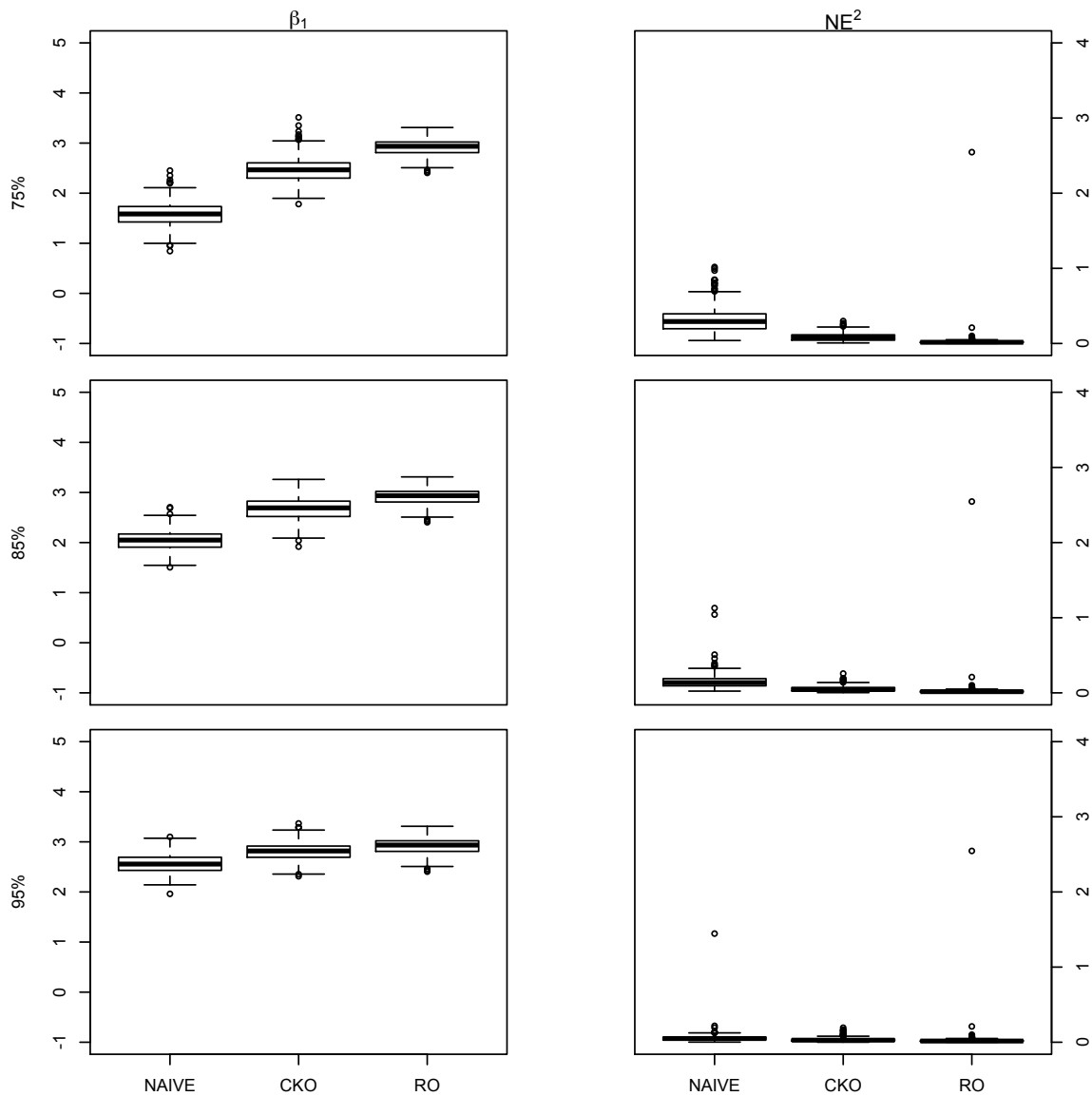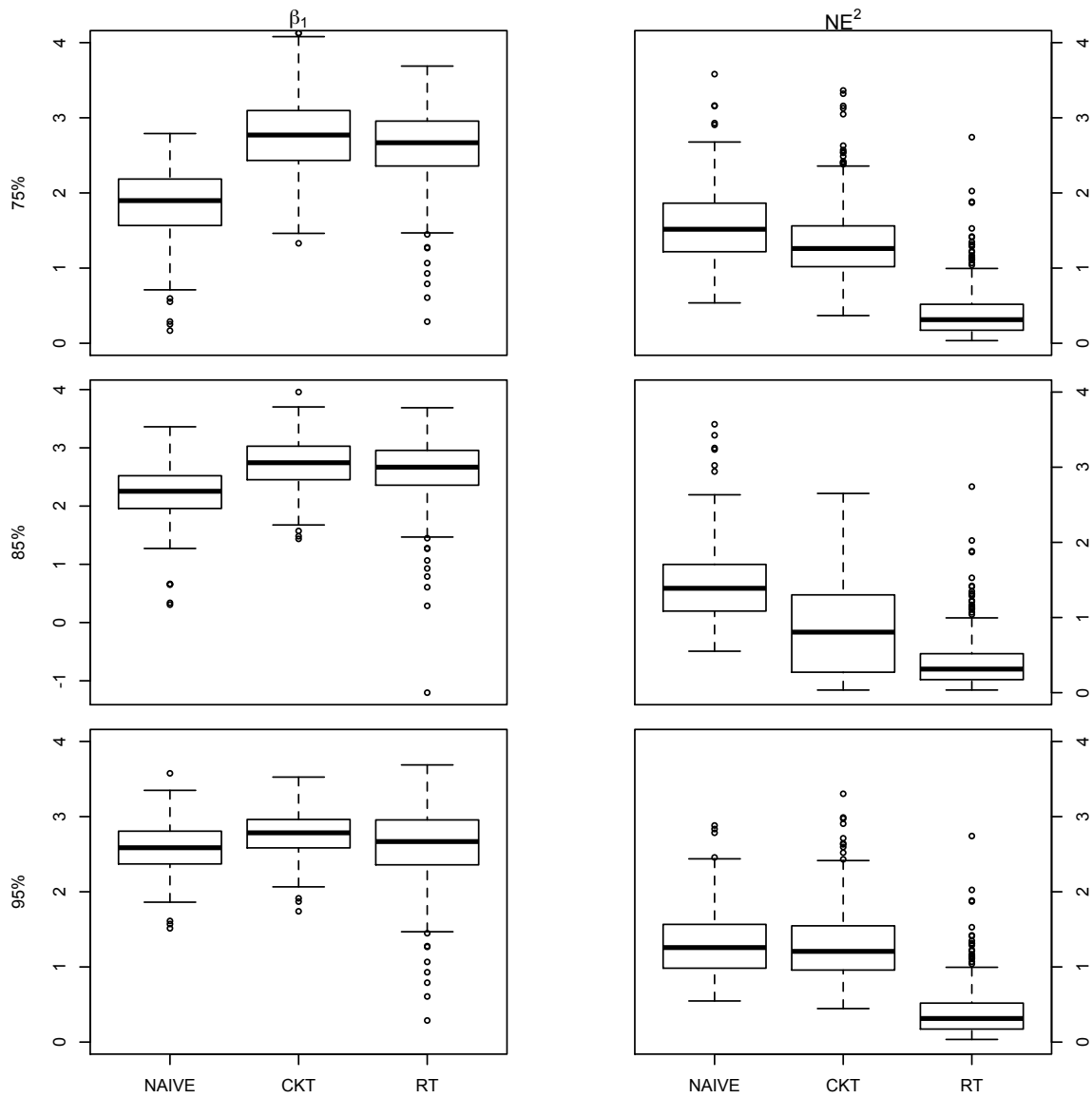
Figure H.2: Two-layer tuning parameter selection. Under the simulation setting (F2) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage method (RO) in the absence of measurement error, respectively.

190
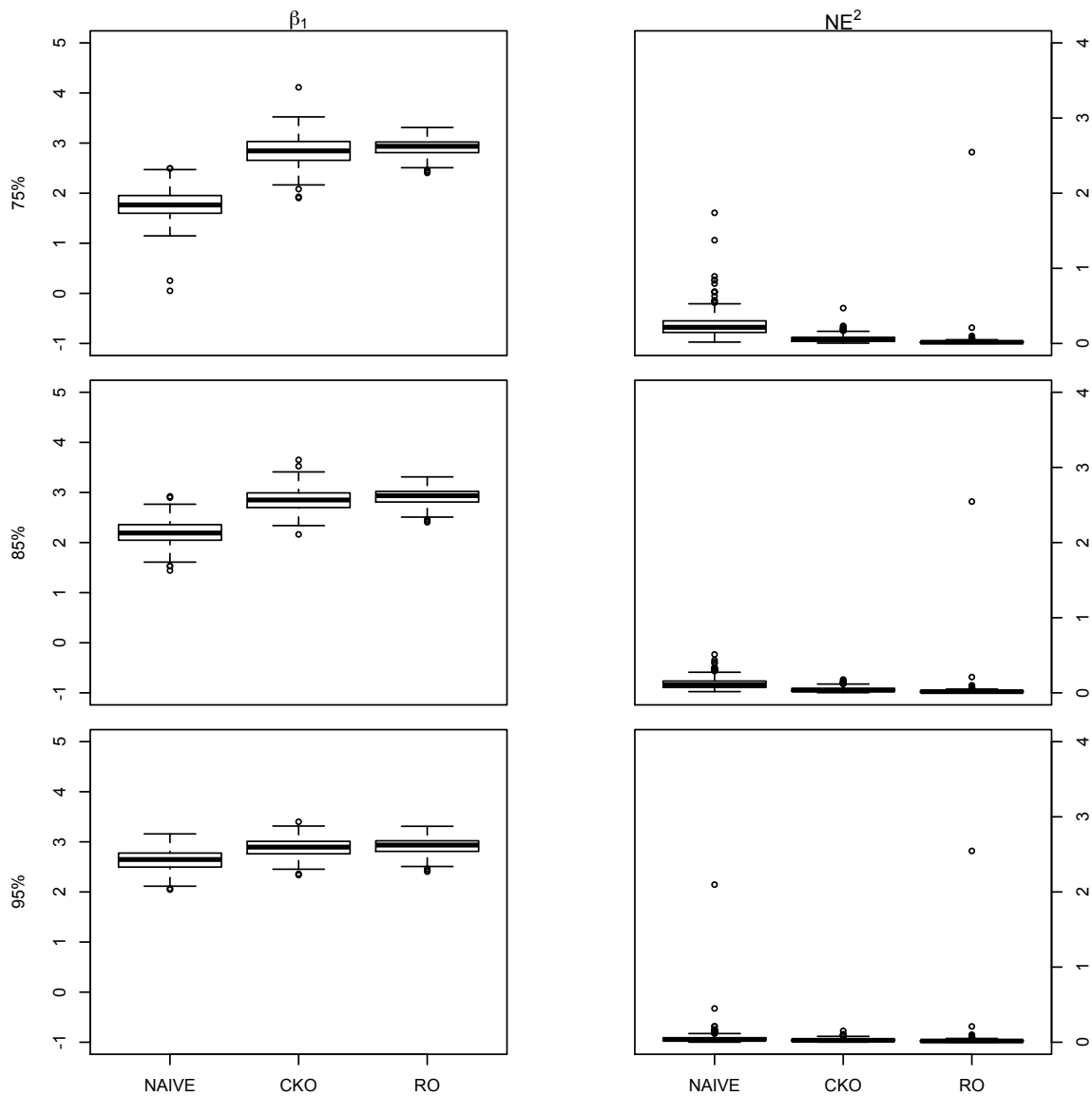
Figure H.3: Two-layer tuning parameter selection. Under the simulation setting (F2) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage (RO) in the absence of measurement error, respectively.
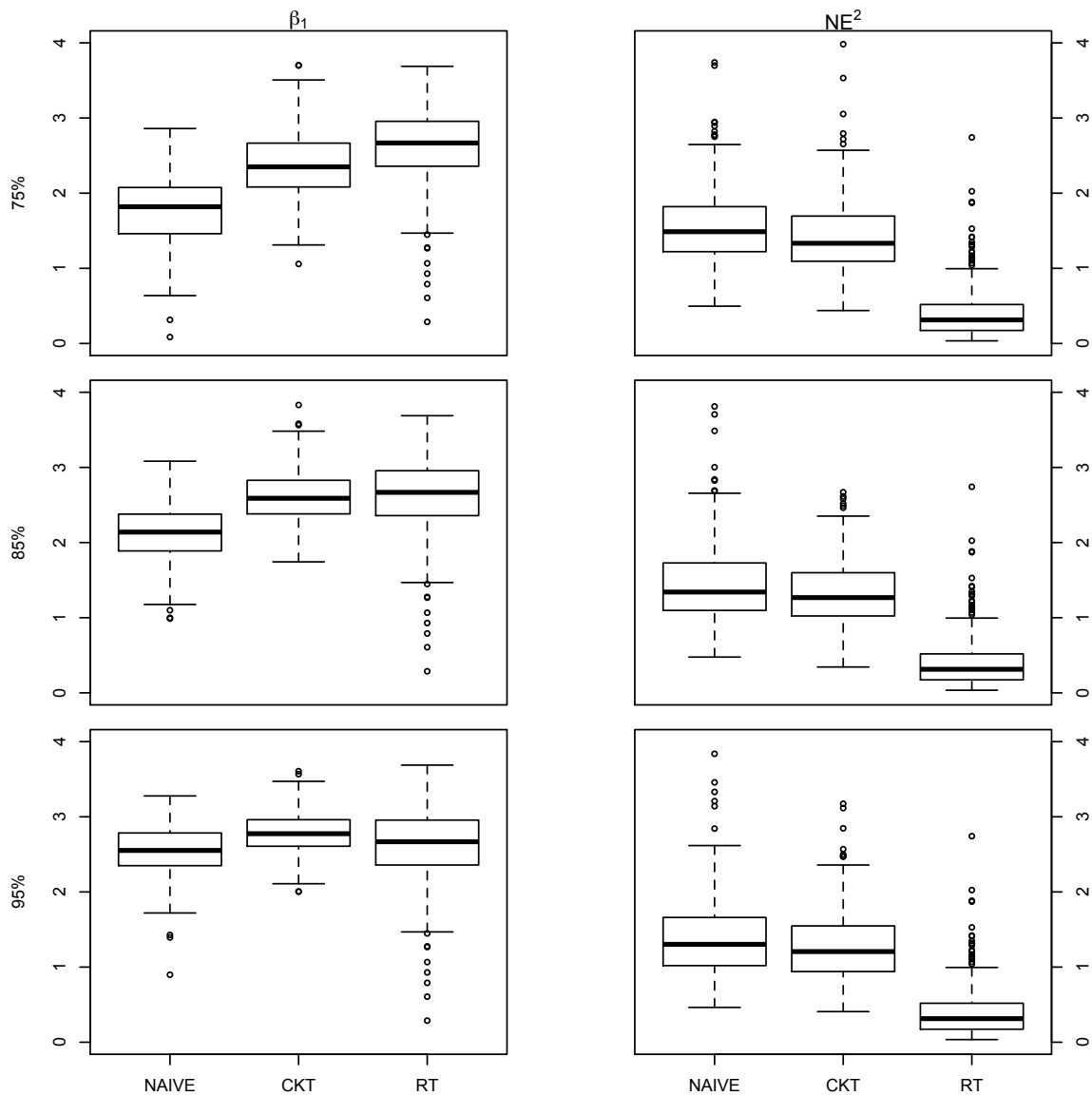
191

Figure H.4: Two-layer tuning parameter selection. Under the simulation setting (F2) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive one-stage method (NAIVE), the corrected kernel one-stage method (CKO), and one-stage method (RO) in the absence of measurement error, respectively.
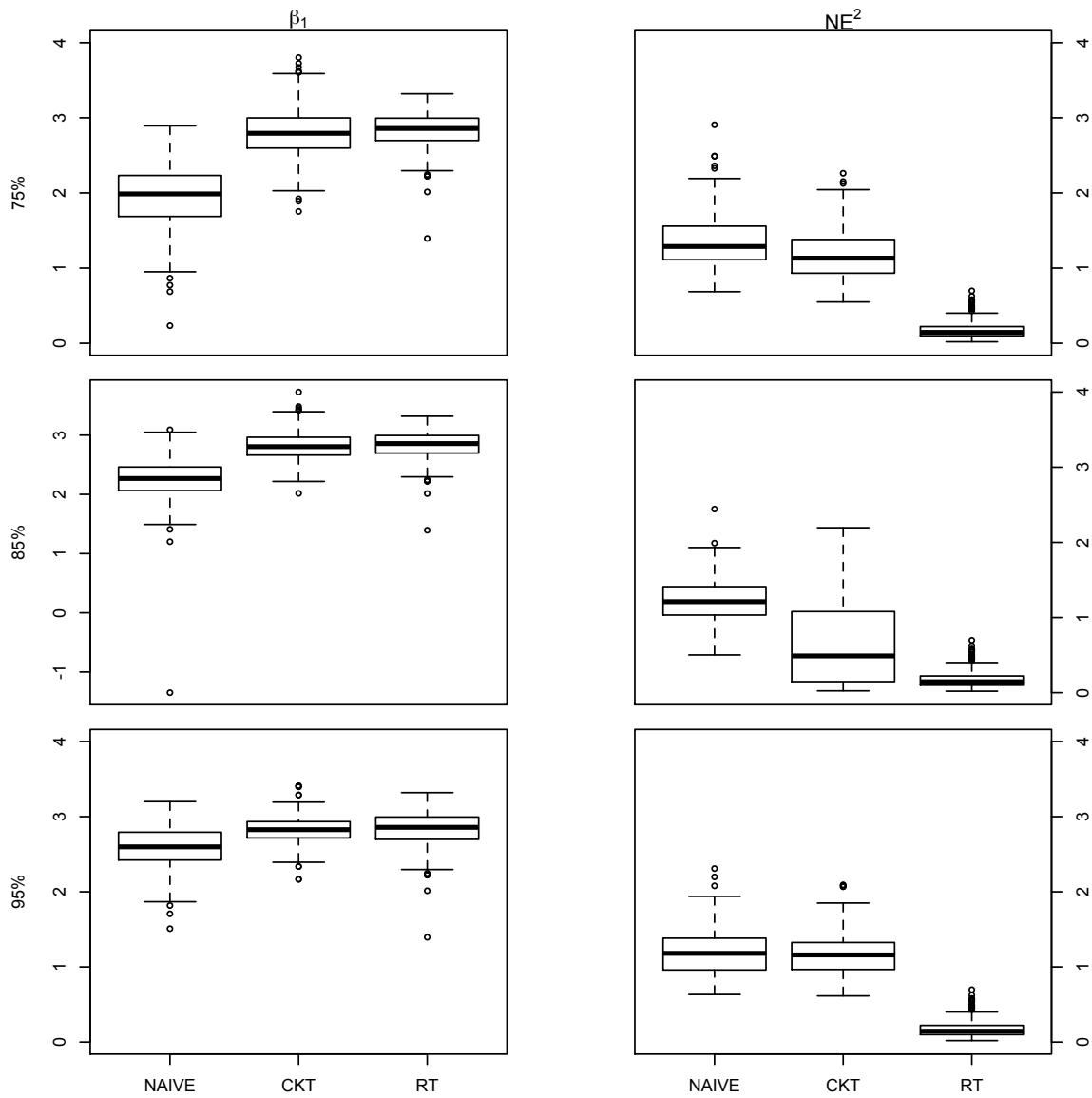
192

Figure H.5: Two dimensional cross validation tuning parameter selection. Under the simulation setting (F4) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage (RT) in the absence of measurement error, respectively.

Figure H.6: Two dimensional cross validation. Under the simulation setting (F4) and the sample size $n = 200$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage method (RT) in the absence of measurement error, respectively.

194

Figure H.7: Two dimensional cross validation tuning parameter selection. Under the simulation setting (F4) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Laplace measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage (RT) in the absence of measurement error, respectively.
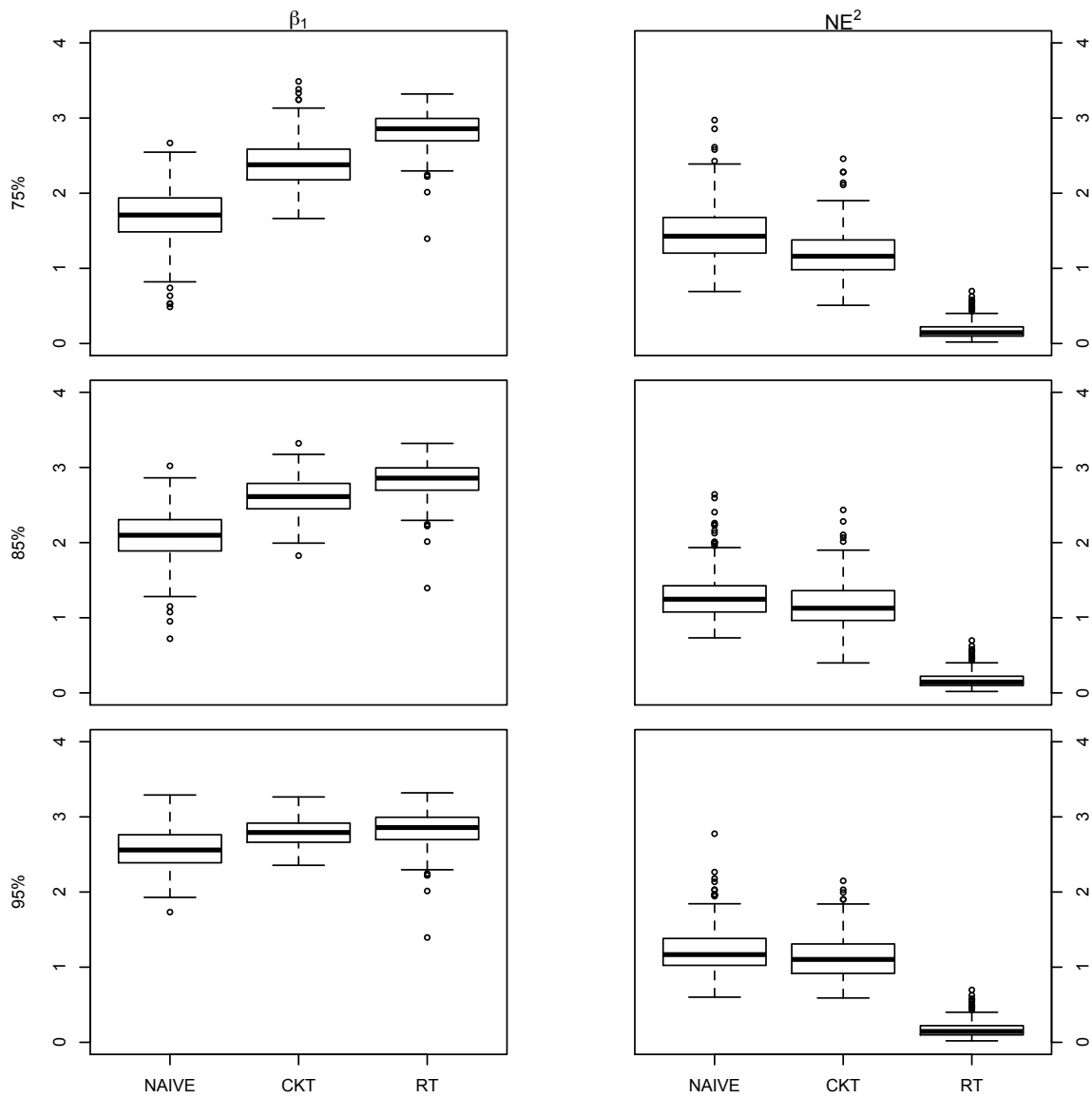
Figure H.8: Two dimensional cross validation. Under the simulation setting (F4) and the sample size $n = 400$. Boxplots of estimates of β_1 (on the left panels) and estimates of NE^2 (on the right panels) when U is Normal measurement error at three levels of reliability ratios (from the top row to the bottom row), $\lambda = 0.75, 0.85, 0.95$. Within each panel, the three estimates (from left to right) result from the naive two-stage method (NAIVE), the corrected kernel two-stage method (CKT), and two-stage method (RT) in the absence of measurement error, respectively.

Table H.3: Averages of parameter estimates from the two-stage estimation method over 300 repetitions under (F4) when $n = 200$. Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truth is $\beta_0 = 1$, $\beta_1 = 3$.

	75%			85%			95%		
	β_0	β_1	NE^2	β_0	β_1	NE^2	β_0	β_1	NE^2
					$U \sim N(0,\sigma^2)$				
Naive	0.56	1.76	1.26	0.71	2.13	1.29	0.83	2.52	1.27
	(0.20)	(0.26)	(0.28)	(0.17)	(0.23)	(0.30)	(0.17)	(0.24)	(0.30)
CK	0.78	2.37	1.28	0.85	2.62	1.23	0.91	2.78	1.22
	(0.17)	(0.26)	(0.30)	(0.16)	(0.21)	(0.27)	(0.14)	(0.16)	(0.28)
					$U \sim \text{Laplace}(0,\sigma^2)$				
Naive	0.59	1.84	1.29	0.70	2.20	1.27	0.82	2.58	1.20
	(0.19)	(0.28)	(0.29)	(0.17)	(0.26)	(0.29)	(0.15)	(0.19)	(0.27)
CK	0.90	2.77	1.24	0.89	2.74	0.76	0.90	2.77	1.22
	(0.19)	(0.29)	(0.29)	(0.16)	(0.24)	(0.35)	(0.15)	(0.16)	(0.28)
TRUE	0.91	2.76	1.27						
	(0.15)	(0.20)	(0.28)						

Table H.4: Averages of parameter estimates from the two-stage estimation method over 300 repetitions under (F4) when $n = 400$. Numbers in parentheses are $(10 \times$ standard errors) associate with the averages. The truth is $\beta_0 = 1$, $\beta_1 = 3$.

	75%			85%			95%		
	β_0	β_1	NE^2	β_0	β_1	NE^2	β_0	β_1	NE^2
					$U \sim N(0,\sigma^2)$				
Naive	0.52	1.71	1.18	0.64	2.08	1.10	0.83	2.56	1.15
	(0.15)	(0.21)	(0.21)	(0.14)	(0.19)	(0.20)	(0.12)	(0.16)	(0.20)
CK	0.80	2.39	1.12	0.87	2.61	1.12	0.92	2.79	1.10
	(0.12)	(0.18)	(0.19)	(0.11)	(0.13)	(0.19)	(0.10)	(0.10)	(0.18)
					$U \sim \text{Laplace}(0,\sigma^2)$				
Naive	0.60	1.94	1.13	0.71	2.26	1.12	0.84	2.59	1.14
	(0.15)	(0.24)	(0.20)	(0.13)	(0.22)	(0.18)	(0.12)	(0.16)	(0.19)
CK	0.89	2.79	1.10	0.91	2.81	0.61	0.94	2.84	1.13
	(0.13)	(0.19)	(0.19)	(0.11)	(0.14)	(0.31)	(0.11)	(0.11)	(0.18)
TRUE	0.94	2.86	1.15						
	(0.11)	(0.14)	(0.19)						

197