Essays in Microeconometrics

Hao Dong

Department of Economics
London School of Economics
and Political Science

A thesis submitted to the Department of Economics
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy

June 9, 2018

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of about 31080 words.

Statement of joint work

I confirm that Chapter 1 and 3 of my thesis are separately joint works with Professor Taisuke Otsu and Professor Qiwei Yao. I contributed 50% of the work in each case.

Hao Dong

Abstract

This thesis consists of three chapters, which are works during my PhD study. In the first two chapters, I investigate the estimation of nonparametric and semiparametric econometric models widely used in empirical studies when the data is mismeasured. In the last chapter, my attention moves to the estimation of the effects of the social interactions.

In Chapter 1, I study the estimation of the nonparametric additive model in the presence of a mismeasured covariate. In such a situation, the conventional method may cause severe bias. Therefore, I propose a new estimator. The estimation procedure is divided into two stages. In the first stage, to adept to the additive structure, I use a series method. And to deal with the ill-posedness brought by the mismeasurement, I introduce a ridge parameter. The convergence rate is then derived for the first stage estimator. For the distributional results required for inference, based on the first stage estimator, I implement the one-step back-fitting with a deconvolution kernel. Asymptotic normality is derived for the second stage estimator.

Chapter 2 investigates the sharp regression-discontinuity (SRD) design when there is a continuously distributed measurement error in the running variable. In such a situation, the discontinuity at the cut-off disappear completely, and using the conventional SRD method cause severe bias. To overcome this, I develop a new estimator of the average treatment effect at the cut-off. Two separate cases characterized by the observability of the treatment status are considered. In the case of observed treatment status, the proposed estimator is the

difference between the deconvolution local linear estimators based on treated and control groups. In the case of unobserved treatment status, the observed running variable cannot be used to divide the sample due to the presence of measurement errors. So, the one-sided kernel functions are implemented, and an additional ridging parameter is introduced for regularization. Asymptotic properties of proposed estimators are derived for both cases.

Chapter 3 develops a new method to estimate the spillover effects using the factor structure of the variables generating the spillovers. Specifically, we find that such a factor structure implies constraints on the spillovers, which can be utilized to improve the performance of the existing estimator, like LASSO, by adding the factor-induced constraints. The L_2 error bound is derived for the proposed estimator. Compared with the unconstrained case, the proposed estimator is more accurate in the sense that it has approximately sharper error bound. Simulation results demonstrate our findings.

Acknowledgements

I am indebted to my supervisors, Taisuke Otsu and Qiwei Yao, for their supervision and continued guidance. They have been extremely patient with me and provided me with the maximum supports throughout my study at LSE.

The faculty at LSE have always been incredibly generous with their time and advice. I am also grateful to the econometricians visiting LSE and providing helpful comments on my work. In particular, I would like to thank Abhimanyu Gupta, Javier Hidalgo, Tatiana Komarova, Peter Robinson, Marcia Schafgans for their help over the years, and thank Zhipeng Liao, Yukitoshi Matsushita, and many others for their time during their visits. I also thank my colleagues and peers at the LSE, in particular the participants at the work in progress seminars, for their comments and discussions on my work and beyond: Karun Adusumilli, Michel Azulai, Shiyu Bo, Qiu Chen, Jiajia Gu, Chao He, Hanwei Huang, Yu-Hsiang Lei, Yan Liang, Yatang Lin, William Matcham, Cheng Qian, Marcus Roel, Xuezhu Shi, Manuel Staab, Luke Taylor, Mengshan Xu, Tianle Zhang, and many others.

Finally, this has been a long journey and it would not even be possible without the support of my family. To Meng, thank you for marring me and being such a great companion. It is your timely encouragement that raise me up time after time and get me through this journey in the most positive way. To my parents and parents-in-law, I owe you everything for your unconditional trust and endless patience despite the long distance between us.

Contents

List of tables **viii**

List of figures **ix**

1 Nonparametric Estimation of Additive Model with Errors-in-Variables **1**

1.1 Introduction . . . 1

1.2 Setup and Estimators . . . 3

1.3 Asymptotic Properties . . . 8

1.3.1 First Stage Estimator . . . 8

1.3.2 Second Stage Estimator . . . 11

1.4 Conclusion . . . 13

1.5 Appendix . . . 14

1.5.1 Proofs of Main Results . . . 14

1.5.2 Proofs of Lemmas . . . 23

2 Sharp Regression-Discontinuity Design with a Mismeasured Running Variable **45**

2.1 Introduction . . . 45

2.2 Setup and Identification . . . 48

2.2.1 Sharp Regression-Discontinuity Design . . . 48

2.2.2 Mismeasured Running Variable . . . 49

2.2.3 Identification . . . 51

2.3 Estimation . . . 52

2.3.1 Case of Observed Treatment . . . 53

2.3.2 Case of Unobserved Treatment . . . 55

2.4 Theoretical Properties . . . 58

2.5 Numerical Properties . . . 62

2.5.1 Simulation Settings . . . 63

2.5.2 Smoothing-parameter Choice . . . 63

2.5.3 Simulation Results . . . 65
2.6 Application: Causal Effects of Being Eligible for Medicaid/CHIP . . . 68
2.6.1 Data and Variables . . . 70
2.6.2 Model and Results . . . 72
2.7 Conclusion . . . 75
2.8 Appendix . . . 77
2.8.1 Proofs of Main Results . . . 77
2.8.2 Proofs of Lemmas . . . 93

3 Estimating Spillovers under Factor-induced Constraints 98
3.1 Introduction . . . 98
3.2 Spillovers and Factor-induced Constraints . . . 100
3.3 Estimation . . . 101
3.3.1 LASSO under Factor-induced Constraints . . . 101
3.3.2 Construction of Constraints . . . 102
3.4 Theoretical Results . . . 103
3.5 Numerical Results . . . 108
3.5.1 Simulation Design . . . 108
3.5.2 Simulation Results . . . 109
3.6 Conclusion . . . 114
3.7 Appendix . . . 115
3.7.1 Proofs of Main Results . . . 115
3.7.2 Proofs of Lemmas . . . 120

Bibliography 122

List of tables

2.1 Rooted MSE of Estimators with Known f_ϵ for DGP1 65

2.2 Rooted MSE of Estimators with known f_ϵ for DGP2 66

2.3 Rooted MSE of Estimators with Unknown f_ϵ for DGP1 68

2.4 Rooted MSE of Estimators with Unknown f_ϵ for DGP2 68

2.5 Descriptive of Health Insurance Participation 72

2.6 Causal Effects of being Eligible for Medicaid 73

2.7 Causal Effects of being Eligible for CHIP 74

3.1 Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 1 with $\nu = 0$. 109

3.2 Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 1 with $\nu = 0.5$. 109

3.3 Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 2.1 and 2.2 with $\nu = 0$. 110

3.4 Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 2.1 and 2.2 with $\nu = 0.5$. . . . 111

List of figures

2.1 SRD Design with a Continuous Error in Running Variable. 50

2.2 Fourier Transform of One-sided Kernel Functions 56

2.3 Ill-posedness and Regularization by Ridging 58

2.4 Simulation Results of DGP1, $N = 1000$ 66

2.5 Simulation Results of DGP2, $N = 1000$ 67

2.6 Total Family Income 71

3.1 Realized L_2 Errors ($\nu = 0$) 112

3.2 Realized L_2 Errors ($\nu = 0.5$) 113

Chapter 1

Nonparametric Estimation of Additive Model with Errors-in-Variables

1.1 Introduction

This chapter investigates the nonparametric estimation of the additive model with a mismeasured covariate as follows.

$$Y = \mu + g(X^*) + m_1(Z_1) + \cdots + m_D(Z_D) + U, \tag{1.1.1}$$

where Y is a response variable, $Z = (Z_1, \ldots, Z_D)$ are observable covariates, U is an error term, and X^* is an error-free but unobservable covariate. Instead of X^*, we observe a mismeasured covariate

$$X = X^* + \epsilon, \tag{1.1.2}$$

where ϵ is a measurement error. As in the literature of nonparametric deconvolution methods[1], we assume that ϵ is independent from X^*, i.e. ϵ is a classical measurement error, and the density of ϵ is known to the researcher. We wish to estimate the unknown functions $g, m_1, \ldots, m_D$ and intercept μ by the observables (Y, X, Z).

[1] See Meister (2009) for the review.

If X^* is observable, it is a standard nonparametric additive model with the identity link function, which has been well studied in the literature; see, e.g., Stone (1985), Stone (1986), Buja et al. (1989), Linton and Nielsen (1995), Linton and Härdle (1996), Opsomer and Ruppert (1997), Fan et al. (1998), Opsomer (2000), and Horowitz and Mammen (2004). However, when X^* is mismeasured, these conventional estimators are in general inconsistent. To the best of our knowledge, there has not been any estimation method for the nonparametric additive model in the presence of measurement errors in covariates.

In this chapter, we develop a nonparametric estimator for the unknown functions $g, m_1, \ldots, m_D$ and intercept μ by extending the two-stage approach of Horowitz and Mammen (2004) to deal with the measurement error based on the deconvolution technique. In the first stage, Horowitz and Mammen (2004) estimated the unknown functions by a series approximation method. In the presence of measurement error, the coefficients in the series approximation are estimated by the ridge-based regularized estimator as in Hall and Meister (2007). In the second stage, Horowitz and Mammen (2004) implemented the one-step backfitting based on local linear regression to achieve asymptotic normality of the estimator. In our case, this stage is implemented by nonparametric deconvolution kernel regression.

There are extensive literature on nonparametric estimation of additive models; see the papers cited above. This chapter contributes to this literature by extending the model and estimation method to the errors-in-variables case. This chapter also contributes to the literature of nonparametric deconvolution methods for measurement error models. In particular, we employ the ridge-based regularization method by Hall and Meister (2007) to estimate moments involving error-free unobservable covatiates. Also for the second stage backfitting, we apply the nonparametric deconvolution kernel regression; see, e.g., Stefanski and Carroll (1990), Carroll and Hall (1988), Fan (1991a), Fan (1991b), Fan and Masry (1992), Fan and Truong (1993), Delaigle et al. (2008), and Hall and Lahiri (2008).

This chapter is organized as follows. Section 1.2 introduces the basic setup and develops our two-stage estimator. Section 1.3 presents main results. In Section 1.3.1, we derive the convergence rate of the first stage estimator. In Section 1.3.2, we establish the limiting distribution of the second stage estimator. Section 1.4 concludes the chapter. All proofs are contained in Section 1.5.

1.2 Setup and Estimators

Before presenting our estimator, we first show that the functions $g, m_1, \ldots, m_D$ and intercept μ can be identified from the observables (Y, X, Z). In this chapter, we consider the following setup.

Assumption 1.

(1) ϵ is independent of (X^*, Z, U).

(2) The distribution of (ϵ, X^*, Z) is absolutely continuous with respect to the Lebesgue measure.

(3) f_ϵ is known.

(4) $f_{X^*,Z}$ is bounded away from zero on $\mathcal{I} \times [-1,1]^D$, where $\mathcal{I}$ is a known compact subset of the support of X^* and $[-1,1]$ is the support of Z_d for $d = 1, \ldots, D$.

(5) $E[U|X^*, Z] = 0$.

(6) $g, m_1, \ldots, m_D$ are normalized as

$$\int_{\mathcal{I}} g(w)dw = \int_{-1}^{1} m_1(w)dw = \cdots = \int_{-1}^{1} m_D(w)dw = 0.$$

Assumption 1 (1) claims that the measurement error discussed in this chapter is classical. Assumption 1 (2) is to guarantee the existence of densities on which the following discussions rely. Assumption 1 (3) is commonly made in the literature of nonparametric estimation with a measurement error, and it can be relaxed by using the auxiliary information such as the repeated measurements, but we will focus on the known error distribution case in this chapter to keep things simple. Assumption 1 (5) and (6) are the normalization for the identification purpose, which are standard in the estimation of nonparametric additive model; see, e.g., Horowitz and Mammen (2004).

Assumption 1 (4) requires all the covariates to be continuously distributed on their support. Since we estimate infinite dimensional objects, it is necessary to work with continuous variables. As in Horowitz and Mammen (2004), we assume that observable covariates Z are supported on $[-1,1]^D$. This is an innocuous assumption because we can always carry out some invertible transformation to achieve this and work with the transformed variables. However, this argument fails when X^* is unobservable. Indeed, such a transformation does not preserve the additive structure in (1.1.2) except when it is linear. Thus, even though the distribution of ϵ is known, it is difficult to recover the distribution of X^* from the transformation of X through deconvolution. Given these considerations, we do not impose any condition on the support of X, X^*, and ϵ, but focus on the estimation of the function g over some known compact set $\mathcal{I}$ of interest. It is assumed that the density of X^* is bounded away from zero on $\mathcal{I}$ so that the conditional expectation function is well defined.

Under this assumption, all unknown objects in the model (1.1.1) are identified, and this result is summarized in Theorem 1 as follows.

Theorem 1. *Under Assumption 1, μ, g, and $m_1, \ldots, m_D$ are identified.*

This theorem follows by an application of the marginal integration argument for the nonparametric additive model combined with the identification of the joint density of (Y, X^*, Z) based on the deconvolution technique. Details of the proof are left to Section 1.5.

We now introduce our estimation strategy. For the expository purpose, we tentatively assume that the error-free covariate X^* is observed. To estimate μ, m_d over $[-1,1]$, and g over the subset $\mathcal{I}$ under the normalization in (**??**), the first stage estimation of Horowitz and Mammen (2004) is implemented by minimizing

$$\sum_{j=1}^{n} \mathbb{I}\{X_j^* \in \mathcal{I}\} \left[Y_j - \mu - \sum_{k=1}^{\kappa} p_k(X_j^*)\theta_k^0 - \sum_{d=1}^{D}\sum_{k=1}^{\kappa} q_k(Z_{d,j})\theta_k^d \right]^2, \tag{1.2.1}$$

with respect to $\theta = (\mu, \theta_1^0, \ldots, \theta_\kappa^0, \theta_1^1, \ldots, \theta_\kappa^1, \ldots, \theta_1^D, \ldots, \theta_\kappa^D)'$, where $\mathbb{I}\{\cdot\}$ is the indicator function, $\{p_k\}_{k=1}^{\infty}$ and $\{q_k\}_{k=1}^{\infty}$ are basis functions supported on $\mathcal{I}$ and $[-1,1]$, respectively, and κ is a tunning parameter characterizing the accuracy of the series approximation. The trimming term $\mathbb{I}\{X_j^* \in \mathcal{I}\}$ appears because we are interested in estimating g over $\mathcal{I}$.

If X^* is mismeasured, this method is obviously infeasible because X^* is unobservable. Also the least square estimation for the above criterion by replacing X_j^* with observable X_j would yield inconsistent estimates in general. In fact, implementing the least square estimation for (1.2.1) by ignoring the measurement error cannot provide the coefficients to construct the estimator of the unknown functions $g, m_1, \dots, m_D$, but a weighted version of them, where the weight is the conditional density $f_{X^*|X,Z}$.

To estimate θ in (1.2.1), we consider the population counterpart of (1.2.1), that is

$$E[\mathbb{I}\{X^* \in \mathcal{I}\}Y^2] + \theta' E[P_\kappa P_\kappa']\theta - 2E[YP_\kappa']\theta,$$

where $P_\kappa = (p_0(X^*), p_1(X^*), \dots, p_\kappa(X^*), q_{01}(Z_1), \dots, q_{0\kappa}(Z_1), \dots, q_{01}(Z_D), \dots, q_{0\kappa}(Z_D))'$ with $p_0(X^*) = \mathbb{I}\{X^* \in \mathcal{I}\}$ and $q_{0k}(Z_d) = p_0(X^*)q_k(Z_d)$ for $k = 1, \dots, \kappa$ and $d = 1, \dots, D$. Thus, once we have estimators for $E[P_\kappa P_\kappa']$ and $E[YP_\kappa']$, denoted by $\hat{E}[P_\kappa P_\kappa']$ and $\hat{E}[YP_\kappa']$ respectively, θ can be estimated by

$$\hat{\theta} = (\mathfrak{Re}\,\hat{E}[P_\kappa P_\kappa'])^{-}\mathfrak{Re}\,\hat{E}[YP_\kappa'], \tag{1.2.2}$$

where $\mathfrak{Re}\,\{\cdot\}$ denotes the real part of a complex-valued matrix or vector, and the inverse here may be the Moore-Penrose inverse. Based on this, the first stage estimators of g and m_d for $d = 1, \dots, D$ are separately given by

$$\hat{g}(x^*) = \sum_{k=1}^{\kappa} p_k(x^*)\hat{\theta}_k^0, \qquad \hat{m}_d(z_d) = \sum_{k=1}^{\kappa} q_k(z_d)\hat{\theta}_k^d. \tag{1.2.3}$$

To implement the estimator in (1.2.3) based on (1.2.2), we need to estimate the expectations $E[P_\kappa P_\kappa']$ and $E[YP_\kappa']$. Any moments that do not involve X^* can be estimated by the method of moments. For the moments depending on X^*, we need to employ a deconvolution technique. For example, consider $E[Yp_k(X^*)]$ that appears in $E[YP_\kappa']$.

To prepare for the discussion, we introduce the following notations. Let $\|f\|_2 = \left(\int |f(w)|^2 dw\right)^{1/2}$ be the L_2-norm of a function $f : \mathbb{R} \to \mathbb{C}$, $L_2(\mathbb{R}) = \{f : \|f\|_2 < \infty\}$ be the L_2-space, and $\langle f_1, f_2 \rangle = \int f_1(w)\overline{f_2(w)}dw$ be the inner product in $L_2(\mathbb{R})$, where $\bar{c}$ denotes the complex conjugate of $c \in \mathbb{C}$. Also let $\mathrm{i} = \sqrt{-1}$ and $f^{\mathrm{ft}}(t) = \int f(x)e^{\mathrm{i}tx}dx$

be the Fourier transform of f. By the Plancherel's isometry (see Lemma 1 (1) in Section 1.5.2), this moment is written as

$$\begin{aligned} E[Yp_k(X^*)] &= \langle mf_{X^*}, p_k \rangle \\ &= \frac{1}{2\pi} \left\langle [mf_{X^*}]^{\mathrm{ft}}, p_k^{\mathrm{ft}} \right\rangle \\ &= \frac{1}{2\pi} \int E[Ye^{\mathrm{i}tX}] \frac{p_k^{\mathrm{ft}}(-t)}{f_\epsilon^{\mathrm{ft}}(t)} dt, \end{aligned}$$

where $m(x^*) = E[Y|X^* = x^*]$, the last equality follows by the law of iterated expectations and independence of ϵ and (Y, X^*). A naive estimator of this moment may be given by replacing $E[Ye^{\mathrm{i}tX}]$ by its sample analogue $n^{-1}\sum_{j=1}^n Y_j e^{\mathrm{i}tX_j}$. However, it is well known that this estimator is not well-behaved due to the fact that $f_\epsilon^{\mathrm{ft}}(t) \to 0$ as $|t| \to \infty$. Intuitively, the estimation error of the sample analogue can be exceptionally amplified in tails, so that the integration transformation is not continuous. Regularization is always required in such a situation. Here we employ the ridge approach in Hall and Meister (2007) and suggest to estimate $E[Yp_k(X^*)]$ by

$$\hat{E}[Yp_k(X^*)] = \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n Y_j e^{\mathrm{i}tX_j} \right) \frac{p_k^{\mathrm{ft}}(-t) f_\epsilon^{\mathrm{ft}}(-t) |f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} dt,$$

where $r \geq 0$ is a tunning parameter to control the smoothness of the integrand and $n^{-\zeta}$ with $\zeta > 0$ is a ridge term to keep the denominator away from zero. Similarly, the moments $E[p_k(X^*)q_{0l}(Z_d)]$ and $E[p_k(X^*)p_l(X^*)]$ appearing in $E[P_\kappa P_\kappa']$ can be estimated by

$$\begin{aligned} \hat{E}[p_k(X^*)q_{0l}(Z_d)] &= \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n q_l(Z_{d,j}) e^{\mathrm{i}tX_j} \right) \frac{[p_0 p_k]^{\mathrm{ft}}(-t) f_\epsilon^{\mathrm{ft}}(-t) |f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} dt, \\ \hat{E}[p_k(X^*)p_l(X^*)] &= \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n e^{\mathrm{i}tX_j} \right) \frac{[p_k p_l]^{\mathrm{ft}}(-t) f_\epsilon^{\mathrm{ft}}(-t) |f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} dt. \end{aligned}$$

By applying these estimators to each element in (1.2.2), we can implement the first stage estimator (1.2.3).

To conduct statistical inference, we construct the second stage estimator. If X^* is observable, we can implement the one-step backfitting as in Horowitz and Mammen (2004). The second stage estimator of g is given by the nonparametric kernel or local polynomial fitting from the residuals $Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})$ by the first stage estimates on the covariate X_j^*. When X^* is mismeasured and unobservable, we modify this second stage estimation by applying the deconvolution kernel regression. In particular, let

$$\mathbb{K}_h(w) = \frac{1}{2\pi} \int e^{-\mathrm{i}tw} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt,$$

be the deconvolution kernel, where K is a kernel function and h is the bandwidth. The second stage estimator of g, denoted as $\tilde{g}$, is defined as

$$\tilde{g}(x^*) = \frac{\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j) \left[Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})\right]}{\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j)}.$$

The second stage estimator of m_d, however, cannot be a direct practice of the deconvolution kernel regression because the unobservable X^* now shows up in the dependent variable $Y_j - \hat{\mu} - \hat{g}(X_j^*) - \sum_{d' \neq d}^{D} \hat{m}_{d'}\left(Z_{d',j}\right)$ instead of the covariate in a non-linear way. One immediate thought could be to first estimate $g(x^*) + m_d(z_d)$ by the deconvolution kernel regression of $Y_j - \hat{\mu} - \sum_{d' \neq d}^{D} \hat{m}_{d'}\left(Z_{d',j}\right)$ on $\left(X_j^*, Z_{d,j}\right)$ then deduct $\hat{g}(x^*)$. This, however, would make the estimator of m_d depend on the choice of x^*, which is unpleasant in practice. Alternatively, we consider the standard kernel regression of $Y_j - \hat{\mu} - \sum_{d' \neq d}^{D} \hat{m}_{d'}(Z_{d',j})$ on $Z_{d,j}$ and deduct the estimator of $E[g(X^*)|Z_d]$. The later is a conditional moment of a function of the mismeasured covariate, and can be estimated based on estimates of g and the joint density of X^* and Z_d. For the joint density of X^* and Z_d, we use the deconvolution density estimator. For the unknown function g, it is natural to consider its first stage estimator $\hat{g}$. However, it is worthy to note that $\hat{g}(x^*)$ is a valid estimator of $g(x^*)$ only when $x^* \in \mathcal{I}$, which is also reflected by the choice of the series in the first stage. This insight implies that the second stage estimation of m_d should be conditional on $X^* \in \mathcal{I}$. In particular, we have

$$\begin{aligned} m_d(z_d) &= E\Big[Y - \mu - g(X^*) - \sum_{d' \neq d} m_{d'}(Z_{d'})|Z_d = z_d, X^* \in \mathcal{I}\Big] \\ &= \frac{\int_{\mathcal{I}} E\left[Y - \mu - g(X^*) - \sum_{d' \neq d} m_{d'}(Z_{d'})|Z^d = z_d, X^* = x^*\right] f_{Z_d,X^*}(z_d, x^*) dx^*}{\int_{\mathcal{I}} f_{Z_d,X^*}(z_d, x^*) dx^*}, \end{aligned}$$

which suggests the following second stage estimator of m_d.

$$\tilde{m}_d(z_d) = \frac{\sum_{j=1}^n \int_{\mathcal{I}} \mathbb{K}_h(x^* - X_j)\left[Y_j - \hat{\mu} - \hat{g}(x^*) - \sum_{d' \neq d} \hat{m}_{d'}(Z_{d',j})\right] dx^* K_h(z_d - Z_{d,j})}{\sum_{j=1}^n \int_{\mathcal{I}} \mathbb{K}_h(x^* - X_j) dx^* K_h(z_d - Z_{d,j})}$$

1.3 Asymptotic Properties

1.3.1 First Stage Estimator

We now study the asymptotic properties of the first stage estimator in (1.2.3). Let $\|A\| = [\text{trace}(A^\dagger A)]^{1/2}$ be the Frobenius norm of a complex matrix A, and $A^\dagger$ be A's conjugate transpose. Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ separately denote the largest and the smallest eigenvalue of a Hermitian positive semi-definite matrix A. Let $\mathcal{F}_{\alpha,c} = \{f \in L_2(\mathbb{R}) : \int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq c\}$ denote the Sobolev class of order $\alpha > 0$ and $c > 0$ [2]. Let $\delta_{k,k'}$ be the Kronecker delta, which equals to 0 if $k \neq k'$ and equals to 1 if $k = k'$. Based on these notations, we impose the following assumptions.

Assumption 2.

(1) $\{Y_j, X_j, Z_j\}_{j=1}^n$ *is i.i.d.*

(2) $E[Y^2|X^*, Z] < \infty$.

(3) $f_{X^*}, f_{X^*|Z_d=z_d}, f_{X^*|Z_d=z_d, Z_{d'}=z_{d'}}, E[Y|X^*]f_{X^*}, E[Y|X^* = \cdot, Z_d = z_d]f_{X^*|Z_d=z_d}$ *belong to* $\mathcal{F}_{\alpha,c_{\text{sob}}}$ *for all* $d, d' = 1, \ldots, D$ *and* $z_d, z_{d'} \in [-1, 1]$.

(4) $\{p_k\}_{k=1}^\infty$ *is a series of basis functions on* $\mathcal{I}$ *such that* $\int_{\mathcal{I}} p_k(w)dw = 0$ *for all* k *and* $\int_{\mathcal{I}} p_k(w)p_{k'}(w)dw = \delta_{k,k'}$ *for all* k, k'.

(5) $\{q_k\}_{k=1}^\infty$ *is a series of basis functions on* $[-1, 1]$ *such that* $\int_{-1}^1 q_k(w)dw = 0$ *for all* k *and* $\int_{-1}^1 q_k(w)q_{k'}(w)dw = \delta_{k,k'}$ *for all* k, k'.

(6) $\lambda_{\min}(E[P_\kappa P_\kappa']) \geq \underline{\lambda} > 0$ *for all* κ.

[2]Even though it looks somewhat different, the Sobolev condition imposed here is essentially equivalent to the one used in (2.30) of Meister (2009), which using our notations requires $\int |f^{\text{ft}}(t)|^2|t|^{2\alpha} < c$. First, it is easy to see $\int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha < c$ implies $\int |f^{\text{ft}}(t)|^2|t|^{2\alpha} < c$. For the other direction, we have $\int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq 2^\alpha \int_{|t|\leq 1} |f^{\text{ft}}(t)|^2 dt + 2^\alpha \int |f^{\text{ft}}(t)|^2|t|^{2\alpha}dt < c'$, where the first inequality follows by $2^\alpha|t|^{2\alpha} \geq (1+|t|^2)^\alpha \Leftrightarrow |t| \geq 1$ and the second inequality follows by $f \in L_2(\mathbb{R})$ and (2.30) of Meister (2009).

(7) $\sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} \|P_\kappa(x^*,z)\| = O(\kappa^{1/2})$ *as* $\kappa \to \infty$.

(8) There exists $\theta_0 = (\mu_0, \theta_0^0, \theta_0^1, \ldots, \theta_0^D)$ *such that*

$$\sup_{x^*\in\mathcal{I}} \left| g(x^*) - P'_{\kappa,0}(x^*)\theta_0^0 \right| = O(\kappa^{-2}),$$

$$\sup_{z_d\in[-1,1]} \left| m_d(z_d) - P'_{\kappa,d}(z_d)\theta_0^d \right| = O(\kappa^{-2}),$$

where $P_{\kappa,0}(x^*) = (p_1(x^*), \ldots, p_\kappa(x^*))$ *and* $P_{\kappa,d}(z_d) = (q_1(z_d), \ldots, q_\kappa(z_d))$ *for* $d = 1, \ldots, D$.

Assumption 2 (3) is the Sobolev condition for various densities and regression functions, which is about the smoothness of the underlying objects, and is imposed to control the size of the bias terms in the estimation. Assumption 2 (4) and (5) contain the conditions about the basis functions $\{p_k\}_{k=1}^\infty$ and $\{q_k\}_{k=1}^\infty$. Similar conditions are adopted by Horowitz and Mammen (2004) for the first-stage estimator with observed X^*. Assumption 2 (6)-(8) are commonly used assumptions for series-based estimation, e.g. Assumption 2 and 3 of Newey (1997).

It is known in the literature that the rate of convergence of a deconvolution-based estimator depends on the smoothness of the error density. Intuitively, since the deconvolution-based estimators always have an error characteristic function in the denominator, the estimation error in the numerator would be greatly amplified in both tails where the error characteristic function is close to zero. Since the smoother the error density corresponds to an error characteristic function decays to zero faster in tails, the smoother the error distribution is, the slower the convergence rate of the estimator will be. Therefore, for the measurement error density f_ϵ, we consider the two categories that are commonly employed in the deconvolution literature in the following discussion.

f_ϵ is said to be ordinary smooth of order β, if there exist some constants $c_{\text{os},1} > c_{\text{os},0} > 0$ and $\beta > 0$ such that

$$c_{\text{os},0}(1+|t|)^{-\beta} \le |f_\epsilon^{\text{ft}}(t)| \le c_{\text{os},1}(1+|t|)^{-\beta} \quad \text{for all } t \in \mathbb{R}.$$

f_ϵ is said to be supersmooth of order β, if there exist some constants $c_{\mathrm{ss},1} > c_{\mathrm{ss},0} > 0$, $\beta_0 > 0$, and $\beta > 0$ such that

$$c_{\mathrm{ss},0} \exp(-\beta_0 |t|^\beta) \le |f_\epsilon^{\mathrm{ft}}(t)| \le c_{\mathrm{ss},1} \exp(-\beta_0 |t|^\beta) \quad \text{for all } t \in \mathbb{R}.$$

In particular, the characteristic function of an ordinary smooth error distribution decays at a polynomial rate, while the characteristic function of a supersmooth error distribution decays at an exponential rate. Typical examples of ordinary smooth densities are the Laplace and gamma densities, and typical examples of supersmooth densities are the normal and Cauchy densities. To facilitate the discussion of the convergence rate of the first stage estimator, we impose the following assumptions to specify the smoothness of the error distribution.

Assumption 3. *f_ϵ is ordinary smooth of order $\beta > 1/2$.*
Assumption 4. *f_ϵ is supersmooth of order $\beta > 0$.*

Under these assumptions, the convergence rate of the first-stage estimator is presented as follows.

Theorem 2. *Suppose that Assumption 1 and 2 hold true.*

(1) Under Assumption 3, we have

$$\|\hat\theta - \theta_0\| = O_p\left(\kappa n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa^{\frac{1}{2}} n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-2}\right).$$

(2) Under Assumption 4, we have

$$\|\hat\theta - \theta_0\| = O_p\left(\kappa^{\frac{1}{2}} (\log n)^{-\frac{\alpha}{\beta}} + \kappa^{-2}\right).$$

Theorem 3. *Suppose that Assumption 1 and 2 hold true.*

(1) Under Assumption 3, we have

$$\sup_{x^* \in \mathcal{I}} |\hat g(x^*) - g(x^*)| = O_p\left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$
$$\sup_{z_d \in [-1,1]} |\hat m_d(z_d) - m_d(z_d)| = O_p\left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$

for $d = 1, \dots, D$.

(2) *Under Assumption 4, we have*

$$\sup_{x^* \in \mathcal{I}} |\hat{g}(x^*) - g(x^*)| = O_p\left(\kappa(\log n)^{-\frac{\alpha}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$
$$\sup_{z_d \in [-1,1]} |\hat{m}_d(z_d) - m_d(z_d)| = O_p\left(\kappa(\log n)^{-\frac{\alpha}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$

for $d = 1, \dots, D$.

1.3.2 Second Stage Estimator

For the distributional results of the second stage estimator $\tilde{g}$, we impose further assumptions as follows.

Assumption 5.

(1) f_{X^*} *is twice continuously differentiable,* $\|f_X\|_\infty < \infty$*, and* g *is three times continuously differentiable.*

(2) $\sup_x E\left[|U|^{2+\eta}|X = x\right] < \infty$ *for some constant* $\eta > 0$.

(3) $\int wK(w)dw = 0$, $\int w^2K(w)dw < \infty$, $\|K^{\mathrm{ft}}\|_\infty < \infty$, $\|K^{\mathrm{ft}'}\|_\infty < \infty$.

(4) $h \to 0$ *as* $n \to \infty$.

Assumption 5 states the regularity conditions used to derive the asymptotic distribution of $\tilde{m}$. Assumption 5 (1) is the smoothness condition about the density f_{X^*} and the regression function g, which is used for the bias reduction in the proof. Assumption 5 (2) is the common assumption used for the Lyapunov central limit theorem. Assumption 5 (3) requires the kernel function K to be symmetric and the second-order moment to exist, which is also commonly used for the bias reduction in the nonparametric estimation problem.

Assumption 6.

(1) $E\left[\left|g(X^*) + U - g(x^*)\right|^2|X = x\right]$ *as a function of* x *is continuous for any* $x^* \in \mathcal{I}$.

(2) $\|f_\epsilon^{\mathrm{ft}'}\|_\infty < \infty$, $|s|^\beta\left|f_\epsilon^{\mathrm{ft}}(s)\right| \to c_\epsilon$*, and* $|s|^{\beta+1}\left|f_\epsilon^{\mathrm{ft}'}(s)\right| \to \beta c_\epsilon$ *for some constant* $c_\epsilon > 0$ *as* $|s| \to \infty$.

(3) $\int |s|^{\beta}\left[\left|K^{\mathrm{ft}}(s)\right| + \left|K^{\mathrm{ft}'}(s)\right|\right] ds < \infty$, $\int |s|^{2\beta}\left|K^{\mathrm{ft}}(s)\right|^{2} ds < \infty$.

(4) $nh^{2\beta+1} \to \infty$ *as* $n \to \infty$.

Assumption 6 (1) is a technical assumption, which would be satisfied if all densities are continuous. Assumption 6 (2) is commonly used in the deconvolution problem with an ordinary smooth error. It goes further than Assumption 3, as Assumption 6 (2) characterizes the exact limit, not the upper and lower bounds, of the error characteristic function and its derivative in tails. Assumption 6 (3) requires the smoothness of the kernel function K to be adapt to that of the measurement error. Assumption 6 (4) requires the bandwith h to decay to zero no faster than $n^{-\frac{1}{2\beta+1}}$, which is due to the large variance brought by the measurement error.

Assumption 7.

(1) K^{ft} *is supported on* $[-1, 1]$.

(2) $nhe^{-2\beta_0 h^{-\beta}} \to \infty$ *as* $n \to \infty$.

(3) $E|G_{1,n,1}|^{2} n^{\frac{\eta}{2+\eta}} h^{\frac{2+2\eta}{2+\eta}} e^{-2\beta_0 h^{-\beta}} \to \infty$ *as* $n \to \infty$, *where* $G_{1,n,1}$ *is defined as in Section 1.5.1.*

Rather than adapting smoothness of the kernel function to the smoothness of measurement error density as in the ordinary smoothness case, Assumption 7 (1) directly assumes the kernel function K is infinite order smooth. Assumption 7 (2) requires the bandwidth h to decay at most in a logarithm rate, which is due to the fact that the error characteristic function in the denominator decays in an exponential rate and is commonly observed in the deconvolution problem with a supersmooth error. Assumption 7 (3) is a technical assumption used to verify the Lyapunov condition (see the proof of Theorem 4 for details), and more primitive conditions, like Condition 3.1 of Fan and Masry (1992), could certainly be imposed here. To keep the simplest notations, following Delaigle et al. (2009), we stick to the current version throughout this chapter.

Theorem 4. *Suppose that Assumption 1, 2, and 5 hold true.*

(1) Under Assumption 3 and 6, we have

$$\frac{\tilde{g}(x^*) - g(x^*) - Bias\left\{\tilde{g}(x^*)\right\}}{\sqrt{Var[\tilde{g}(x^*)]}} \xrightarrow{d} N(0,1).$$

(2) Under Assumption 4 and 7, we have

$$\frac{\tilde{g}(x^*) - g(x^*) - Bias\{\tilde{g}(x^*)\}}{\sqrt{Var[\tilde{g}(x^*)]}} \xrightarrow{d} N(0,1).$$

To conduct statistical inference, the variance of $\tilde{g}$ should be estimated. To this end, we just need to estimate $E|G_{1,n,1}|^2$ where $G_{1,n,1}$ is defined in the proof of Theorem 4, for which we can consider $\frac{1}{n}\sum_{j=1}^{n} G_{1,n,j}$ and substitute the μ, $m_1, \ldots, m_D$, and g by their corresponding estimators.

1.4 Conclusion

In this chapter, we develop a new estimator for the nonparametric additive model in the presence of the mismeasured covariate. The estimation procedure is divided into two stages. In the first stage, to adept to the additive structure, we use a series method, together with a ridge approach to deal with the ill-posedness brought by the mismeasurement. Convergence rate for the first stage estimator is derived. To establish the limiting distribution, we consider the second stage estimator obtained by the one-step backfitting with a deconvolution kernel based on the first stage estimator. The asymptotic normality of the regression function corresponding to the mismeasured covariate is derived.

Further research is needed to explore the asymptotic normality of the proposed second stage estimator for the regression function corresponding to the accurately measured covariates, and the extensions to the case of multiple mismeasured covariates, non-identity link function, and the case when the error distribution is unknown but auxiliary information such as the repeated measurements are available.

1.5 Appendix

1.5.1 Proofs of Main Results

Proof of Theorem 1: Let $z = (z_1, \ldots, z_D)$, $z_{-d} = (z_1, \ldots, z_{d-1}, z_{d+1}, \ldots, z_D)$, $A(\mathcal{I})$ be the length of the set $\mathcal{I}$, and $f^{\mathrm{ft}}_{Y,X,Z}(y, \cdot, z)(t) = \int f_{Y,X,Z}(y, x, z)e^{\mathrm{i}tx}dx$. By Assumption 1 and Lemma 1(2), we have

$$f_{Y,X^*,Z}(y, x^*, z) = \frac{1}{2\pi}\int e^{-\mathrm{i}tx^*}\frac{f^{\mathrm{ft}}_{Y,X,Z}(y, \cdot, z)(t)}{f^{\mathrm{ft}}_{\epsilon}(t)}dt.$$

Since the joint density $f_{Y,X^*,Z}$ is identified, the conditional mean $E[Y|X^*, Z]$ is also identified. Therefore, by Assumption 1, μ, g, $m_1, \ldots, m_D$ are identified as

$$\begin{aligned}
\mu &= 2^{-D}A(\mathcal{I})^{-1}\int_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} E[Y|X^* = x^*, Z = z]dx^*dz, \\
g(x^*) &= 2^{-D}\int_{[-1,1]^D} E[Y|X^* = x^*, Z = z]dz - \mu, \\
m_d(z_d) &= 2^{-(D-1)}\int_{[-1,1]^{D-1}} E[Y|X^* = x^*, Z = z]dz_{-d} - \mu - g(x^*),
\end{aligned}$$

for $d = 1, \ldots, D$. Thus, the conclusion is obtained. ■

Proof of Theorem 2: Let $\hat{M}_\kappa = \mathfrak{Re}\,\hat{E}[P_\kappa P_\kappa']$, $\hat{C}_\kappa = \mathfrak{Re}\,\hat{E}[YP_\kappa']$, $M_\kappa = E[P_\kappa P_\kappa']$, $C_\kappa = E[P_\kappa Y]$, $\theta^* = M_\kappa^{-1}C_\kappa$, and $r_\kappa = E[Y|X^*, Z] - P_\kappa'\theta_0$. First, we have

$$\begin{aligned}
\|\hat{\theta} - \theta^*\|^2 &= \|\hat{M}_\kappa^{-1}\hat{C}_\kappa - M_\kappa^{-1}C_\kappa\|^2 \\
&= \|\hat{M}_\kappa^{-1}(\hat{C}_\kappa - C_\kappa) + \hat{M}_\kappa^{-1}(M_\kappa - \hat{M}_\kappa)\theta^*\|^2 \\
&\leq 2\|\hat{M}_\kappa^{-1}(\hat{C}_\kappa - C_\kappa)\|^2 + 2\|\hat{M}_\kappa^{-1}(M_\kappa - \hat{M}_\kappa)\theta^*\|^2 \\
&\leq 2\lambda_{\max}(\hat{M}_\kappa^{-2})\{\|\hat{C}_\kappa - C_\kappa\|^2 + \|\hat{M}_\kappa - M_\kappa\|^2\|\theta^*\|^2\},
\end{aligned}$$

where the first inequality follows by the Jensen's inequality, and the second inequality follows by $\lambda_{\max}(A) = \sup_{\|\delta\|=1}\delta' A\delta$ and $\lambda_{\max}(A'A) \leq \|A\|^2$.

Note $\|\hat{M}_\kappa - M_\kappa\|^2 \leq \|\hat{E}[P_\kappa P'_\kappa] - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2 \leq \|\hat{E}[P_\kappa Y] - C_\kappa\|^2$. Then, the orders of $\|\hat{M}_\kappa - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2$ follows by Lemma 4 in Section 1.5.2. We also note that $\lambda_{\max}(\hat{M}_\kappa^{-2}) = \lambda_{\min}^{-2}(\hat{M}_\kappa)$, and $\lambda_{\min}(A) = \inf_{\|\delta\|=1} \delta' A \delta$. Thus, the upper bound of $\lambda_{\max}(\hat{M}_\kappa^{-2})$ follows by

$$\inf_{\|\delta\|=1} \delta' \hat{M}_\kappa \delta \geq \inf_{\|\delta\|=1} \delta' (\hat{M}_\kappa - M_\kappa)\delta + \lambda_{\min}(M_\kappa),$$
$$\left(\inf_{\|\delta\|=1} \delta'(\hat{M}_\kappa - M_\kappa)\delta\right)^2 \leq \|\hat{M}_\kappa - M_\kappa\|^2 \xrightarrow{p} 0,$$

and $\lambda_{\min}(M_\kappa) \geq \underline{\lambda} > 0$. Moreover, we note $C_\kappa = E[P_\kappa E[Y|X^*, Z]]$ and

$$\begin{aligned}
\|\theta^*\|^2 &= C'_\kappa M_\kappa^{-2} C_\kappa \\
&\leq \lambda_{\max}(M_\kappa^{-1}) C'_\kappa M_\kappa^{-1} C_\kappa \\
&\leq \underline{\lambda}^{-1} E[E[Y|X^*, Z]^2] \\
&< \infty,
\end{aligned}$$

where the first inequality follows by the property of the maximum eigenvalue, and the second inequality follows by Theorem 1 of Tripathi (1999) and the last inequality is due to the fact that $g, m_1, \cdots, m_D$ are all bounded and are supported on $\mathcal{I}$ and $[-1, 1]$ respectively. Combining these results, we have

$$\|\hat{\theta} - \theta^*\|^2 = \begin{cases} O_p\left(\kappa^2 n^{2\zeta + \frac{\zeta}{\beta} - 1} + \kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O_p\left(\kappa (\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4} \end{cases}.$$

Since $\theta^* = \theta_0 + M_\kappa^{-1} E[P_\kappa r_k]$, we have

$$\begin{aligned}
\|\theta^* - \theta_0\|^2 &= E[P'_\kappa r_k] M_\kappa^{-2} E[P_\kappa r_k] \\
&\leq \lambda_{\max}(M_\kappa^{-1}) E[P'_\kappa r_k] M_\kappa^{-1} E[P_\kappa r_k] \\
&\leq \underline{\lambda}^{-1} E[r_k^2] \\
&= O(\kappa^{-4}),
\end{aligned}$$

where the last equality follows by Assumption 2 (8). Therefore, the convergence rate of $\|\hat{\theta} - \theta_0\|$ follows by the triangle inequality. ■

Proof of Theorem 3: Let $\hat{\theta} = \left(\hat{\mu}, \hat{\theta}^0, \hat{\theta}^1, \ldots, \hat{\theta}^D\right)$, where $\hat{\theta}^0$ is the vector of estimated coefficients corresponding to $P_{\kappa,0}$, and $\hat{\theta}^d$ is the vector of estimated coefficients corresponding to $P_{\kappa,d}$ for $d = 1, \ldots, D$. Let $r_{\kappa,0}(x^*) = g(x^*) - P'_{\kappa,0}\theta_0^0$ and $r_{\kappa,m_d}(z_d) = m_d(z_d) - P'_{\kappa,d}\theta_0^d$.

Note $\sup_{x^*\in\mathcal{I}} \|P_{\kappa,0}(x^*)\| \leq \sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} \|P_\kappa(x^*,z)\|, \sup_{z_d\in[-1,1]} \|P_{\kappa,d}(z_d)\| \leq \sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} \|P_\kappa(x^*,z)\|$, $\|\hat{\theta}^0 - \theta_0^0\| \leq \|\hat{\theta} - \theta_0\|$, and $\|\hat{\theta}^d - \theta_0^d\| \leq \|\hat{\theta} - \theta_0\|$ for $d = 1, \ldots, D$. Then, for the uniform convergence rate of $\hat{g}$, we have

$$
\begin{aligned}
&\sup_{x^*\in\mathcal{I}} \left|\hat{g}(x^*) - g(x^*)\right| \\
&\leq \sup_{x^*\in\mathcal{I}} |P_{\kappa,0}(x^*)'(\hat{\theta}^0 - \theta_0^0)| + \sup_{x^*\in\mathcal{I}} |r_{\kappa,0}(x^*)| \\
&\leq \sup_{x^*\in\mathcal{I}} \|P_{\kappa,g}(x^*)\| \|\hat{\theta}^0 - \theta_0^0\| + O(\kappa^{-2}) \\
&= \begin{cases} O_p\left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right), & \text{under Assumption 3} \\ O_p\left(\kappa(\log n)^{-\frac{\alpha}{\beta}} + \kappa^{-\frac{3}{2}}\right), & \text{under Assumption 4} \end{cases},
\end{aligned}
$$

where the last inequality is obtained by using the Cauchy-Schwartz inequality and Assumption 2 (8), and the last equality follows by Assumption 2 (7) and Theorem 2. Similarly, the uniform convergence rate of $\hat{m}_d$ for $d = 1, \ldots, D$ follows by

$$
\begin{aligned}
&\sup_{z_d\in[-1,1]} \left|\hat{m}_d(z_d) - m_d(z_d)\right| \\
&\leq \sup_{z_d\in[-1,1]} |P_{\kappa,d}(z_d)'(\hat{\theta}^d - \theta_0^d)| + \sup_{z_d\in[-1,1]} |r_{\kappa,m_d}(z_d)| \\
&\leq \sup_{z_d\in[-1,1]} \|P_{\kappa,d}(z_d)\| \left[\|\hat{\theta}^d - \theta_0^d\|\right] + O(\kappa^{-2}) \\
&= \begin{cases} O_p\left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right), & \text{under Assumption 3} \\ O_p\left(\kappa(\log n)^{-\frac{\alpha}{\beta}} + \kappa^{-\frac{3}{2}}\right), & \text{under Assumption 4} \end{cases},
\end{aligned}
$$

where the last inequality is obtained by using the Cauchy-Schwartz inequality and Assumption 2 (8), and the last equality follows by Assumption 2 (7) and Theorem 2.

■

Proof of Theorem 4: To simplify notations, we intentionally suppress the dependence on x^* in the following discussion, at which the function g is valued. Let $a = f_{X^*}(x^*) \int K(w)dw$. Also, let $\mathbb{A}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j)$ and $\mathbb{B}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j)\Big[Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})\Big]$, then $\tilde{g} = \frac{\mathbb{B}_n}{\mathbb{A}_n}$. Also, we note

$$\tilde{g} - g = \frac{1}{n}\sum_{j=1}^{n} G_{n,j},$$

where $G_{n,j} = G_{1,n,j} + G_{2,n,j} + G_{3,n,j} + G_{4,n,j}$ and

$$G_{1,n,j} = \frac{1}{2\pi a}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[Y_j - \mu - \sum_{d=1}^{D} m_d(Z_{d,j}) - g(x^*)\Big]dt,$$

$$G_{2,n,j} = \frac{[\mathbb{A}_n^{-1} - a^{-1}]}{2\pi}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[Y_j - \mu - \sum_{d=1}^{D} m_d(Z_{d,j}) - g(x^*)\Big]dt,$$

$$G_{3,n,j} = \frac{1}{2\pi a}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[\mu + \sum_{d=1}^{D} m_d(Z_{d,j}) - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})\Big]dt,$$

$$G_{4,n,j} = \frac{[\mathbb{A}_n^{-1} - a^{-1}]}{2\pi}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[\mu + \sum_{d=1}^{D} m_d(Z_{d,j}) - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})\Big]dt.$$

The rest of the proof is then divided into three steps.

Step 1:

$$\frac{\sum_{j=1}^{n} G_{1,n,j} - nE[G_{1,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{d} N(0,1) \tag{1.5.1}$$

By Lyapunov central limit theorem, for (1.5.1), it is suffice to show

$$\lim_{n\to\infty} \frac{E\big|G_{1,n,1}\big|^{2+\eta}}{n^{\eta/2}\Big[E\big|G_{1,n,1}\big|^2\Big]^{(2+\eta)/2}} = 0, \tag{1.5.2}$$

for some constant $\eta > 0$.

Let $\mu_{g,2+\eta}(x) = E[\,|g(X^*) + U - g(x^*)|^{2+\eta}|X = x\,]f_X(x)$ for the constant $\eta \geq 0$. Then using the law of iterated expectation, we can write $E\left|G_{1,n,1}\right|^{2+\eta}$ as

$$E\left|G_{1,n,1}\right|^{2+\eta} = \int_x \left|\frac{1}{2\pi a}\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^{2+\eta} \mu_{g,2+\eta}(x)dx. \tag{1.5.3}$$

If $\eta > 0$, we have

$$\begin{aligned} E\left|G_{1,n,1}\right|^{2+\eta} &\leq \frac{h^{-(\beta+1)\eta}}{(2\pi)^\eta a^{(2+\eta)}}\left(h^{\beta+1}\int\frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|}dt\right)^\eta \\ &\quad \times \frac{h^{2\beta+1}}{4\pi^2}\int_x\left|\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^2 \mu_{g,2+\eta}(x)dx \\ &= O\left(h^{-(\beta+1)(\eta+2)+1}\right), \end{aligned} \tag{1.5.4}$$

where the equality follows by Lemma 5 and Lemma 7. If $\eta = 0$, we have

$$\begin{aligned} E\left|G_{1,n,1}\right|^{2} &= \frac{h^{-(2\beta+1)}}{a^2}\left(\frac{h^{2\beta+1}}{4\pi^2}\int_x\left|\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^2 \mu_{g,2+\eta}(x)dx\right) \\ &= \frac{h^{-(2\beta+1)}\mu_{g,2}(x^*)}{2\pi a^2 c_\epsilon^2}\int |s|^{2\beta}\left|K^{\mathrm{ft}}(s)\right|^2 ds\,(1+o_p(1)), \end{aligned} \tag{1.5.5}$$

where the second equality follows by Lemma 7. Thus, (1.5.4) and (1.5.5) together imply that (1.5.1) hold true if $nh \to \infty$ as $n \to \infty$.

Step 2:

$$\frac{\sum_{j=1}^n G_{3,n,j} - nE[G_{3,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{p} 0. \tag{1.5.6}$$

For the numerator, we note

$$\sum_{j=1}^n G_{3,n,j} - nE[G_{3,n,1}] = O_p\left(\sqrt{n\,E\left|G_{3,n,1}\right|^2}\right), \tag{1.5.7}$$

and

$$
\begin{aligned}
E\left|G_{3,n,1}\right|^2 &= \int_x E\left[\left|\mu + \sum_{d=1}^{D} m_d(Z_{d,1}) - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,1})\right|^2 |X = x\right] \\
&\quad \times \left|\int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt\right|^2 f_X(x)\,dx \\
&\le \left(|\hat{\mu} - \mu| + \sum_{d=1}^{D} \sup_{z_d \in [-1,1]} \left|\hat{m}_d(z_d) - m_d(z_d)\right|\right)^2 \\
&\quad \times 4\pi^2 h^{-(2\beta+1)} \left\{\frac{h^{2\beta+1}}{4\pi^2} \int_x \left|\int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt\right|^2 f_X(x)\,dx\right\} \\
&= O_p\left(\kappa^3 n^{2\zeta + \frac{\zeta}{\beta} - 1} h^{-(2\beta+1)} + \kappa^2 n^{-\frac{2\alpha\zeta}{\beta}} h^{-(2\beta+1)} + \kappa^{-3} h^{-(2\beta+1)}\right),
\end{aligned} \tag{1.5.8}
$$

where the last equality follows by Theorem 3 and Lemma 7. For the denominator, we note

$$
\begin{aligned}
aE[G_{1,n,1}] &= \frac{1}{2\pi} \int e^{-\mathrm{i}tx^*} K^{\mathrm{ft}}(th) \left\{E\left[e^{\mathrm{i}tX^*} g(X^*)\right] - E\left[e^{\mathrm{i}tX^*}\right] g(x^*)\right\} dt \\
&= E[K_h(x^* - X^*) g(X^*)] - E[K_h(x^* - X^*)] g(x^*) \\
&= \int K_h(x^* - w) g(w) f_{X^*}(w) dw - g(x^*) \int K_h(x^* - w) f_{X^*}(w) dw \\
&= O\left(h^2\right),
\end{aligned} \tag{1.5.9}
$$

where the last equality follows by the second order differentiability of f_{X^*}, the third order differentiability of g, the symmetry of K, $\int K(w) w^2 dw < \infty$, and the following fact

$$
\begin{aligned}
&\int K_h(x^* - w) g(w) f_{X^*}(w) dw - g(x^*) \int K_h(x^* - w) f_{X^*}(w) dw \\
&\qquad = f_{X^*}(x^*) g''(x^*) \int K(w) w^2 dw\, h^2 + o(h^2).
\end{aligned}
$$

And (1.5.9) together with (1.5.5) imply that $Var[G_{1,n,1}]$ is strictly dominated by $E|G_{1,n,1}|^2$ for large n. Then by (1.5.5), we have

$$
\frac{1}{Var[G_{1,n,1}]} = O\left(h^{(2\beta+1)}\right). \tag{1.5.10}
$$

Therefore, (1.5.6) holding true if $\kappa^{\frac{3}{2}} n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}} \to 0$ as $n \to \infty$, i.e. the first stage estimator is uniformly consistent.

Step 3:

$$\frac{\sum_{j=1}^{n} G_{k,n,j} - nE[G_{k,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{p} 0, \tag{1.5.11}$$

for $k = 2, 4$.

First, we note $G_{2,n,j} = \left(\frac{a-\mathbb{A}_n}{\mathbb{A}_n}\right) G_{1,n,j}$ and $G_{4,n,j} = \left(\frac{a-\mathbb{A}_n}{\mathbb{A}_n}\right) G_{3,n,j}$. Also note

$$\begin{aligned} \mathbb{A}_n = {} & E\left[\frac{1}{2\pi}\int \frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)} e^{-\mathrm{i}t(x^*-X)} dt\right] \\ & + O_p\left(n^{-1/2}\left[E\left|\frac{1}{2\pi}\int \frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)} e^{-\mathrm{i}t(x^*-X)} dt\right|^2\right]^{1/2}\right), \end{aligned} \tag{1.5.12}$$

where $E\left|\frac{1}{2\pi}\int \frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)} e^{-\mathrm{i}t(x^*-X)} dt\right|^2 = O\left(h^{-(2\beta+1)}\right)$ follows by Lemma 7 and

$$\begin{aligned} E\left[\frac{1}{2\pi}\int \frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)} e^{-\mathrm{i}t(x^*-X)} dt\right] &= \frac{1}{2\pi}\int e^{-\mathrm{i}tx^*} K^{\text{ft}}(th) f_{X^*}^{\text{ft}}(t) dt \\ &= E[K_h(x^* - X^*)] \\ &= a + O(h). \end{aligned} \tag{1.5.13}$$

Hence, we have

$$\mathbb{A}_n - a = O(h) + O_p(n^{-1/2} h^{-(\beta+1/2)}), \tag{1.5.14}$$

which implies that (1.5.11) follows by (1.5.1) and (1.5.6) if $h \to 0$ and $nh^{2\beta+1} \to \infty$.

Combining (1.5.1), (1.5.6), and (1.5.11), we have

$$\frac{\tilde{g}(x^*) - g(x^*) - \text{Bias}\left\{\tilde{g}(x^*)\right\}}{\sqrt{Var[G_{1,n,1}]}} \xrightarrow{d} N(0,1),$$

where Bias $\{\tilde{g}(x^*)\} = E[G_{n,1}]$. To conclude for the case of ordinary smooth f_ϵ, note $Var[\tilde{g}(x^*)] = \frac{1}{n}Var\left[\sum_{k=1}^{4} G_{k,n,1}\right]$. Then by Cauchy-Schwartz inequality, the covariance terms are dominated by the variance terms, then for $Var[\tilde{g}(x^*)]/Var[G_{1,n,1}] \xrightarrow{p} 1$, it is sufficient to show $Var[G_{k,n,1}]/Var[G_{1,n,1}] \xrightarrow{p} 0$ for $k = 2,3,4$, which immediately follows by (1.5.8), (1.5.10), and (1.5.14).

The proof for the case of supersmooth f_ϵ follows a similar route as the proof for the case of ordinary smooth f_ϵ. So I only state the difference as follows. First, we update the upper bound results. In step 1 of the proof for the ordinary smooth case, to verify the Lyapunov condition (1.5.2), by (1.5.3), parallel to (1.5.4), for $\eta > 0$, we have

$$\begin{aligned} E\left|G_{1,n,1}\right|^{2+\eta} &\leq \frac{\sup_x \mu_{g,2+\eta}(x)}{(2\pi a)^{2+\eta}} \left(\int \frac{\left|K^{\mathrm{ft}}(th)\right|}{\left|f_\epsilon^{\mathrm{ft}}(t)\right|}dt\right)^{\eta} \\ &\quad \times \int_x \left|\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^2 dx \\ &= O\left(h^{-(1+\eta)}e^{\beta_0(2+\eta)h^{-\beta}}\right), \end{aligned} \tag{1.5.15}$$

where the last equality follows by Lemma 8 and $\sup_x \mu_{g,2+\eta}(x) < \infty$. For the latter, we note $\|g\|_\infty < c_g$ for some $c_g > 0$ and

$$\begin{aligned} \left|g(X^*) + U - g(x^*)\right|^{2+\eta} &\leq \left(|g(X^*)| + |U| + |g(x^*)|\right)^{2+\eta} \\ &\leq \left(2c_g + |U|\right)^{2+\eta} \\ &\leq c_1 + c_2|U|^{2+\eta}, \end{aligned}$$

for constants $c_1 = 2^{1+\eta}(2c_g)^{2+\eta}$ and $c_2 = 2^{1+\eta}$. Hence, $\sup_x \mu_{g,2+\eta}(x) < \infty$ follows by $\|f_X\|_\infty < \infty$ and $\sup_x E[|U|^{2+\eta}|X = x] < \infty$.

By a similar argument as in (1.5.15), we have

$$\begin{aligned} \int_x \left|\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^2 f_X(x)\,dx &\leq \|f_X\|_\infty \int_x \left|\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^2 dx \\ &= O\left(h^{-1}e^{2\beta_0 h^{-\beta}}\right), \end{aligned} \tag{1.5.16}$$

where the equality follows by $\|f_X\|_\infty < \infty$ and Lemma 8. Therefore, for the parallel result to (1.5.8), using (1.5.16), we have

$$E\left|G_{3,n,1}\right|^2 = O_p\left(\kappa^3 n^{2\zeta+\frac{\zeta}{\beta}-1}h^{-1}e^{2\beta_0 h^{-\beta}} + \kappa^2 n^{-\frac{2\alpha\zeta}{\beta}}h^{-1}e^{2\beta_0 h^{-\beta}} + \kappa^{-3}h^{-1}e^{2\beta_0 h^{-\beta}}\right), \tag{1.5.17}$$

For the parallel result to (1.5.14), using (1.5.16), we have

$$\mathbb{A}_n - a = O(h) + O_p\left(n^{-1/2}h^{-1/2}e^{\beta_0 h^{-\beta}}\right), \tag{1.5.18}$$

which implies that (1.5.11) still hold if $h \to 0$ and $nhe^{-2\beta_0 h^{-\beta}} \to \infty$.

To verify the Lyapunov condition (1.5.2), besides (1.5.15), we also need the parallel result to (1.5.5). There is, however, no parallel result to Lemma 7 in the case of supersmooth f_ϵ. Therefore, the lower bound of $E\left|G_{1,n,1}\right|^2$ is commonly derived to verify (1.5.2). Primitive conditions, like Condition 3.1 of Fan and Masry (1992), can be imposed to this end. In this chapter, however, to avoid the unnecessary complication, we directly assume the lower bound of $E\left|G_{1,n,1}\right|^2$ in Assumption 7 (3). Hence, under Assumption 7 (3), the Lyapunov condition (1.5.2) holds true, and the conclusion follows.

■

1.5.2 Proofs of Lemmas

For $\zeta > 0$, let $G_{\epsilon,n,\zeta} = \{t \in \mathbb{R} : |f_\epsilon^{\text{ft}}(t)| < n^{-\zeta}\}$ characterize the region over which the riding regularization is implemented, and $G^c_{\epsilon,n,\zeta} = \mathbb{R} \backslash G_{\epsilon,n,\zeta}$. First, we introduce Lemma 1, Lemma 2, and Lemma 3 to prepare for the proof of Lemma 4, which is used in the proof of Theorem 2.

Lemma 1. *For $f_1, f_2, f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and $c \in \mathbb{R}$, we have*

(1) $\langle f_1, f_2 \rangle = \frac{1}{2\pi}\langle f_1^{\text{ft}}, f_2^{\text{ft}} \rangle$,

(2) $\left(\int f_1(w - w')f_2(w')dw'\right)^{\text{ft}}(t) = f_1^{\text{ft}}(t)f_2^{\text{ft}}(t)$,

(3) $\left(f_1 f_2\right)^{\text{ft}}(t) = \frac{1}{2\pi}\int f_1^{\text{ft}}(t-s)f_2^{\text{ft}}(s)ds$,

(4) $f^{\text{ft}}(t-s) = \{f(w)e^{-\text{i}sw}\}^{\text{ft}}(t)$,

(5) $f^{\text{ft}}(ct) = \left[f(\cdot/c)/c\right]^{\text{ft}}(t)$.

Lemma 1 (1) is known as the Plancherel's isometry and its proof can be found at Theorem A.4. of Meister (2009). One of its useful special case is when $f_1 = f_2 = f$, which gives $\|f\|_2^2 = \frac{1}{2\pi}\|f^{\text{ft}}\|_2^2$ and this is known as the Parseval's identity. Lemma 1 (2) is known as the convolution theorem and its proof can be found at Theorem A.5. of Meister (2009). Lemma 1 (3) can be understood as the convolution theorem with respect to the inverse Fourier transform, which will be used in the following discussion, and its proof is attached as follows. Lemma 1 (4) immediately follows by the definition of the Fourier transform. Lemma 1 (5) is known as the linear stretching property of the Fourier transform, and its proof can be found in Lemma A.1(e) of Meister (2009).

Proof of Lemma 1 (3): Let $\delta(w)$ be the Dirac delta function. Then, we have

$$\begin{aligned}
\frac{1}{2\pi}\int f_1^{\text{ft}}(t-s)f_2^{\text{ft}}(s)ds &= \frac{1}{2\pi}\int_s\int_w f_1(w)e^{\text{i}(t-s)w}dw\int_{w'} f_2(w')e^{\text{i}sw'}dw'ds \\
&= \int_w f_1(w)e^{\text{i}tw}\int_{w'}\left\{\frac{1}{2\pi}\int_s e^{\text{i}s(w'-w)}ds\right\}f_2(w')dw' \\
&= \int_w f_1(w)e^{\text{i}tw}\int_{w'}\delta(w'-w)f_2(w')dw' \\
&= \int f_1(w)f_2(w)e^{\text{i}tw}dw,
\end{aligned}$$

where the third equality follows by $\delta(w) = \frac{1}{2\pi}\int e^{\mathrm{i}tw}dt$ and the last equality follows by the property of Dirac delta function, that is $\int \delta(w'-w)f(w')dw' = f(w)$. ■

Lemma 2. *Suppose Assumption 2 holds true.*

(1) If f_ϵ is ordinary smooth of order $\beta > 0$, we have

$$\int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right),$$

$$\sup_{z_d\in[-1,1]} \int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right),$$

$$\sup_{z_d,z_{d'}\in[-1,1]} \int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right).$$

(2) If f_ϵ is supersmooth of order $\beta > 0$, we have

$$\int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\beta}}\right),$$

$$\sup_{z_d\in[-1,1]} \int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\beta}}\right),$$

$$\sup_{z_d,z_{d'}\in[-1,1]} \int\limits_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\beta}}\right).$$

Proof of Lemma 2: If f_ϵ is ordinary smooth of order β, $c_{\mathrm{os},0}(1+|t|)^{-\beta} < |f^{\mathrm{ft}}_\epsilon(t)|$ for $t \in \mathbb{R}$, and it follows $(1+|t|)^{-\beta} < c_{\mathrm{os},0}^{-1} n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$. Note that Jensen's inequality $(1+|t|) \leq \sqrt{2}(1+|t|^2)^{1/2}$ implies $(1+t^2)^{-\alpha} < 2^\alpha(1+|t|)^{-2\alpha}$, and it follows $(1+t^2)^{-\alpha} < 2^\alpha c_{\mathrm{os},0}^{-\frac{2\alpha}{\beta}} n^{-\frac{2\alpha\zeta}{\beta}}$. Also note that $\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt \leq \int |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt < c_{\mathrm{sob}}$ by $f_{X^*} \in \mathcal{F}_{\alpha,c_{\mathrm{sob}}}$. Then, we have

$$\begin{aligned}\int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_{X^*}(t)|^2 dt &= \int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_{X^*}(t)|^2 (1+t^2)^{\alpha}(1+t^2)^{-\alpha} dt \\ &\leq 2^{\alpha} c_{\text{os},0}^{-\frac{2\alpha}{\beta}} n^{-\frac{2\alpha\zeta}{\beta}} \int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_{X^*}(t)|^2 (1+t^2)^{\alpha} dt \qquad (1.5.19) \\ &= O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right).\end{aligned}$$

By a similar argument, using $f_{X^*|Z_d=z_d} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ and $f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ separately for any $z_d, z_{d'} \in [-1,1]$, we have

$$\sup_{z_d\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_{X^*|Z_d=z_d}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right),$$

$$\sup_{z_d,z_{d'}\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right).$$

If f_ϵ is supersmooth of order β, $c_{\text{ss},0}\exp(-\beta_0|t|^\beta) < n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$, and it implies that there exists some constant $C > 0$ such that $(1+t^2)^{-\alpha} \leq C(\log n)^{-\frac{2\alpha}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$, which follows by

$$\begin{aligned}& c_{\text{ss},0}\exp(-\beta_0|t|^\beta) < n^{-\zeta} \\ \Rightarrow\quad & \exp(\beta_0|t|^\beta) > c_{\text{ss},0} n^{\zeta} \\ \Rightarrow\quad & \beta_0|t|^\beta > \log(c_{\text{ss},0}) + \zeta\log(n) \\ \Rightarrow\quad & |t|^\beta > \beta_0^{-1}\left[\log(c_{\text{ss},0}) + \zeta\log(n)\right] \\ \Rightarrow\quad & 1+|t|^2 > 1+\beta_0^{-\frac{2}{\beta}}\left[\log(c_{\text{ss},0}) + \zeta\log(n)\right]^{\frac{2}{\beta}} \\ \Rightarrow\quad & (1+|t|^2)^{-\alpha} < \left(1+\beta_0^{-\frac{2}{\beta}}\left[\log(c_{\text{ss},0}) + \zeta\log(n)\right]^{\frac{2}{\beta}}\right)^{-\alpha} \quad . \\ & \qquad\qquad \leq \beta_0^{\frac{2\alpha}{\beta}}\left[\log(c_{\text{ss},0}) + \zeta\log(n)\right]^{-\frac{2\alpha}{\beta}} \\ & \qquad\qquad \leq \beta_0^{\frac{2\alpha}{\beta}}\zeta^{-\frac{2\alpha}{\beta}}\left(\log n\right)^{-\frac{2\alpha}{\beta}}\end{aligned}$$

Then, similar to the previous ordinary smooth case, we have

$$
\begin{aligned}
\int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\text{ft}}(t)|^2 dt &= \int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\text{ft}}(t)|^2 (1+t^2)^{\alpha} (1+t^2)^{-\alpha} dt \\
&\leq C(\log n)^{-\frac{2\alpha}{\beta}} \int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\text{ft}}(t)|^2 (1+t^2)^{\alpha} dt \\
&= O\left((\log n)^{-\frac{2\alpha}{\beta}} \right).
\end{aligned}
\tag{1.5.20}
$$

Again, by a similar argument, using $f_{X^*|Z_d=z_d} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ and $f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ separately for any $z_d, z_{d'} \in [-1,1]$, we have

$$
\begin{aligned}
\sup_{z_d\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f_{X^*|Z_d=z_d}^{\text{ft}}(t)|^2 dt &= O\left((\log n)^{-\frac{2\alpha}{\beta}} \right), \\
\sup_{z_d,z_{d'}\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}^{\text{ft}}(t)|^2 dt &= O\left((\log n)^{-\frac{2\alpha}{\beta}} \right).
\end{aligned}
$$

■

Lemma 3. *Suppose Assumption F holds true.*

(1) If f_ϵ is ordinary smooth of order β with $\beta > 1/2(r+1)$, we have

$$
\int \frac{|f_\epsilon^{\text{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\text{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O\left(n^{\frac{\zeta(2\beta+1)}{\beta}} \right).
$$

(2) If f_ϵ is supersmooth of order $\beta > 0$, we have

$$
\int \frac{|f_\epsilon^{\text{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\text{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O\left(n^{2\zeta(r+2)} \right).
$$

Proof of Lemma 3: By the definition of $G_{\epsilon,n,\zeta}$, we have

$$
\int \frac{|f_\epsilon^{\text{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\text{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = n^{2\zeta(r+2)} \int_{G_{\epsilon,n,\zeta}} |f_\epsilon^{\text{ft}}(t)|^{2r+2} dt + \int_{G_{\epsilon,n,\zeta}^c} \frac{1}{|f_\epsilon^{\text{ft}}(t)|^2} dt. \tag{1.5.21}
$$

If f_ϵ is ordinary smooth of order β, $c_{\text{os},0}(1+|t|)^{-\beta} \le |f_\epsilon^{\text{ft}}(t)| \le c_{\text{os},1}(1+|t|)^{-\beta}$ for $t \in \mathbb{R}$. For $t \in G_{\epsilon,n,\zeta}$, we have $c_{\text{os},0}(1+|t|)^{-\beta} \le |f_\epsilon^{\text{ft}}(t)| < n^{-\zeta}$, which implies $(1+|t|)^{-\beta} < c_{\text{os},0}^{-1} n^{-\zeta}$. Thus, there exists some constant $0 < \eta < 2\beta(r+1) - 1$ such that $(1+|t|)^{-2\beta(r+1)+1+\eta} < c_{\text{os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$ if $\beta > 1/2(r+1)$. Also note $\int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-1-\eta} dt \to 0$ as $n \to \infty$ because $1+|t| > c_{\text{os},0}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$ and $\int (1+|t|)^{-1-\eta} dt < \infty$ for any $\eta > 0$. Thus, we have the following result.

$$
\begin{aligned}
\int_{G_{\epsilon,n,\zeta}} |f_\epsilon^{\text{ft}}(t)|^{2r+2} dt &\le c_{\text{os},1}^2 \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-2\beta(r+1)} dt \\
&\le c_{\text{os},1}^2 \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-2\beta(r+1)+1+\eta} (1+|t|)^{-1-\eta} dt \\
&\le c_{\text{os},1} c_{\text{os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}} \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-1-\eta} dt \\
&= O\left(n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}} \right).
\end{aligned}
\tag{1.5.22}
$$

For $t \in G_{\epsilon,n,\zeta}^c$, $|f_\epsilon^{\text{ft}}(t)|^{-2} \le n^{2\zeta}$. Moreover, if f_ϵ is ordinary smooth of order $\beta > 0$, $c_{\text{os},1}(1+|t|)^{-\beta} \ge |f_\epsilon^{\text{ft}}(t)| \ge n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}^c$, which implies $|t| < c_{\text{os},1}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}}$. Then, it follows

$$
\int_{G_{\epsilon,n,\zeta}^c} |f_\epsilon^{\text{ft}}(t)|^{-2} dt \le n^{2\zeta} \int_{G_{\epsilon,n,\zeta}^c} dt \le 2 c_{\text{os},1}^{\frac{1}{\beta}} n^{\frac{\zeta(2\beta+1)}{\beta}} = O\left(n^{\frac{\zeta(2\beta+1)}{\beta}} \right). \tag{1.5.23}
$$

Then, by choosing a sequence of η converges to 0, (1.5.21), (1.5.22), and (1.5.23) together implies

$$
\int \frac{|f_\epsilon^{\text{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\text{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O\left(n^{\frac{\zeta(2\beta+1)}{\beta}} \right).
$$

For $t \in G_{\epsilon,n,\zeta}^c$, $|f_\epsilon^{\text{ft}}(t)| \ge n^{-\zeta}$, which implies $|f_\epsilon^{\text{ft}}(t)|^{-2r-4} \le n^{2\zeta(r+2)}$. Then, we have

$$
\int \frac{|f_\epsilon^{\text{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\text{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt \le n^{2\zeta(r+2)} \int |f_\epsilon^{\text{ft}}(t)|^{2r+2} dt. \tag{1.5.24}
$$

If f_ϵ is supersmooth of order $\beta > 0$, we have

$$\int |f_\epsilon^{\mathrm{ft}}(t)|^{2r+2} dt \leq 2c_{\mathrm{ss},1}^{2r+2} \int_0^{+\infty} \exp\big(-(2r+2)\beta_0|t|^\beta\big)dt, \tag{1.5.25}$$

where the inequality follows by the smoothness of f_ϵ and the symmetry of the integration. Note $t^2\exp(-(2r+2)\beta_0|t|^\beta) \to 0$ as $t \to \infty$, due to the strict monotonicity of t^2 and $\exp(\,(2r+2)\beta_0|t|^\beta)$, there exists an constant δ such that $\exp(\,(2r+2)\beta_0|t|^\beta) > t^2$ for any $t > \delta$. Then, we have

$$\begin{aligned}
\int_0^{+\infty} \exp\big(-(2r+2)\beta_0|t|^\beta\big)dt &= \int_0^{\delta} + \int_\delta^{+\infty} \exp\big(-(2r+2)\beta_0|t|^\beta\big)dt \\
&\leq \delta + \int_\delta^{+\infty} t^{-2}dt \\
&= \delta + \delta^{-1} < \infty.
\end{aligned} \tag{1.5.26}$$

Then, put (1.5.24), (1.5.25), and (1.5.26) together, we have

$$\int \frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O\left(n^{2\zeta(r+2)}\right).$$

■

Let $\mathcal{I}_{M_\kappa} = \{(p, Q) : E[p(X^*)Q]$ is an element of $M_\kappa\}$ be the index set characterizing the components of M, where p is a product of $\{p_0, p_1, \ldots, p_\kappa\}$ and Q is a product of $\{1, q_1(Z_1), \ldots, q_\kappa(Z_D)\}$.

Lemma 4. *Suppose Assumption 1 and 2 hold true.*

(1) Under Assumption 3, we have

$$\begin{aligned}
|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 &= O_p\left(\kappa^2 n^{2\zeta + \frac{\zeta}{\beta} - 1} + \kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), \\
|\hat{E}[P_\kappa Y] - C_\kappa|^2 &= O_p\left(\kappa n^{2\zeta + \frac{\zeta}{\beta} - 1} + n^{-\frac{2\alpha\zeta}{\beta}}\right).
\end{aligned}$$

(2) Under Assumption 4 with $r \geq 0$ and $0 < \zeta < \frac{1}{2(r+2)}$, we have

$$\begin{aligned} |\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 &= O_p\left(\kappa(\log n)^{-\frac{2\alpha}{\beta}}\right), \\ |\hat{E}[P_\kappa Y] - C_\kappa|^2 &= O_p\left((\log n)^{-\frac{2\alpha}{\beta}}\right). \end{aligned}$$

Proof of Lemma 4: Since the proof is similar, we focus on the proof for $|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2$. Let $B_{p,Q} = E\{\hat{E}[p(X^*)Q]\} - E[p(X^*)Q]$ be the bias of the proposed estimator of the element of M_κ characterized by p and Q. And let $V_{p,Q} = \hat{E}[p(X^*)Q] - E\{\hat{E}[p(X^*)Q]\}$, and $V_{p,Q,j}$ be its component associated with the jth observation, i.e. $V_{p,Q} = \frac{1}{n}\sum_{j=1}^n V_{p,Q,j}$.

First, we note

$$\begin{aligned} E|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 &= \frac{1}{n^2}\sum_{j,j'=1}^n \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E\left[(B_{p,Q}+V_{p,Q,j})\overline{(B_{p,Q}+V_{p,Q,j'})}\right] \\ &= \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} |B_{p,Q}|^2 + \frac{1}{n}\sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E|V_{p,Q,1}|^2 \\ &\equiv B + V, \end{aligned}$$

where the second equality follows by Assumption F (1).

For the bias term B, note

$$\begin{aligned} E[p(X^*)Q] &= \langle E[Q|X^*]f_{X^*}, p\rangle \\ &= \frac{1}{2\pi}\int E[Qe^{\mathrm{i}tX^*}]p^{\mathrm{ft}}(-t)dt, \end{aligned}$$

where the second equality follows by Lemma 1 (1) and the law of iterated expectation, and

$$\begin{aligned} E\{\hat{E}[p(X^*)Q]\} &= \frac{1}{2\pi}\int E\left[\frac{1}{n}\sum_{j=1}^n Q_j e^{\mathrm{i}tX_j}\right]\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt \\ &= \frac{1}{2\pi}\int E[Qe^{\mathrm{i}tX^*}]\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}p^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt, \end{aligned}$$

Then, the bias term B can be written as

$$
\begin{aligned}
B &= \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} \left| \frac{1}{2\pi} \int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[Qe^{\mathrm{i}tX^*}] p^{\mathrm{ft}}(-t) dt \right|^2 \\
&\equiv B_1 + \cdots + B_7,
\end{aligned}
$$

where $B_1, \ldots, B_7$ are summations of the terms whose (p, Q) has the form $(p_0, 1)$, $(p_k, 1)$, $(p_k p_l, 1)$, $(p_0, q_k(Z_d))$, $(p_k, q_l(Z_d))$, $(p_0, q_k(Z_d)q_l(Z_d))$, $(p_0, q_k(Z_d)q_l(Z_{d'}))$ separately for $k, l = 1, \ldots, \kappa$ and $d, d' = 1, \ldots, D$ with $d \neq d'$.

Since the proof is similar for B_1, B_2, and B_3, we focus on the proof of B_3. For B_3, we have

$$
\begin{aligned}
B_3 &= \sum_{k,l=1}^{\kappa} \left| \frac{1}{2\pi} \int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t) (p_k p_l)^{\mathrm{ft}}(-t) dt \right|^2 \\
&= \sum_{k,l=1}^{\kappa} \left| \frac{1}{4\pi^2} \int\int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t) p_k^{\mathrm{ft}}(-t-s) p_l^{\mathrm{ft}}(s) ds dt \right|^2 \\
&= \sum_{k,l=1}^{\kappa} \left| \frac{1}{4\pi^2} \int\int \left(\frac{|f_\epsilon^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(u-v)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(u-v) p_k^{\mathrm{ft}}(-u) p_l^{\mathrm{ft}}(v) du dv \right|^2 \\
&\leq \frac{1}{16\pi^4} \int_v \left\{ \sum_{k=1}^{\kappa} \left| \left\langle \left(\frac{|f_\epsilon^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(u-v)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(u-v), p_k^{\mathrm{ft}}(u) \right\rangle_u \right|^2 \right\} \sum_{l=1}^{\kappa} |p_l^{\mathrm{ft}}(v)|^2 dv \\
&\leq \frac{\kappa}{4\pi^2} \int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\mathrm{ft}}(t)|^2 dt \\
&= \begin{cases} O\left(\kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O\left(\kappa (\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4} \end{cases},
\end{aligned}
$$

where the second equality follows by Lemma 1 (2), the third equality follows by the change of variables $(u, v) = (t + s, s)$, the last equality follows by Lemma 2, and the last inequality follows by Lemma 1 (1), the orthonormality of $\{p_l\}_{l=1}^{\kappa}$, and the following fact

$$
\begin{aligned}
&\sum_{k=1}^{\kappa}\left|\left\langle\left(\frac{|f_{\epsilon}^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_{\epsilon}^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}}-1\right)f_{X^*}^{\mathrm{ft}}(u-v),p_k^{\mathrm{ft}}(u)\right\rangle_u\right|^2\\
&=4\pi^2\sum_{k=1}^{\kappa}\left|\left\langle h_1(w)e^{-\mathrm{i}vw},p_k(w)\right\rangle_w\right|^2\\
&\le 4\pi^2\left\|h_1(w)e^{-\mathrm{i}vw}\right\|_2^2\\
&\le 2\pi\left\|\left(\frac{|f_{\epsilon}^{\mathrm{ft}}(t)|^{r+2}}{\{|f_{\epsilon}^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)f_{X^*}^{\mathrm{ft}}(t)\right\|_2^2\\
&=2\pi\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\mathrm{ft}}(t)|^2dt,
\end{aligned}
$$

where h_1 denotes the Fourier inverse of $\left(\frac{|f_{\epsilon}^{\mathrm{ft}}(t)|^{r+2}}{\{|f_{\epsilon}^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)f_{X^*}^{\mathrm{ft}}(t)$, the first equality follows by Lemma 1 (1) and (4), the first inequality follows by the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, the second inequality follows by $|e^{-\mathrm{i}vw}|=1$ and Lemma 1 (1), and the last equality follows by the definition of $G_{\epsilon,n,\zeta}$.

By similar arguments, we have

$$
\begin{aligned}
B_1 &= \left|\frac{1}{2\pi}\int\left(\frac{|f_{\epsilon}^{\mathrm{ft}}(t)|^{r+2}}{\{|f_{\epsilon}^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)f_{X^*}^{\mathrm{ft}}(t)p_0^{\mathrm{ft}}(-t)dt\right|^2\\
&\le \frac{1}{4\pi^2}\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\mathrm{ft}}(t)|^2dt\int|p_0^{\mathrm{ft}}(t)|^2dt\\
&= \frac{A(\mathcal{I})}{2\pi}\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\mathrm{ft}}(t)|^2dt\\
&= \begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3}\\ O\left((\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the inequality follows by Cauchy-Schwartz inequality, the second equality follows by Lemma 1 (1) and the definition of p_0, and the last equality follows by Lemma 2.

$$
\begin{aligned}
B_2 &= 2\sum_{k=1}^{\kappa}\left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)f_{X^*}^{\mathrm{ft}}(t)p_k^{\mathrm{ft}}(-t)dt\right|^2\\
&= 2\sum_{k=1}^{\kappa}|\langle h_1, p_k\rangle|^2\\
&\leq 2\|h_1\|_2^2\\
&\leq \frac{1}{\pi}\int\limits_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\mathrm{ft}}(t)|^2dt\\
&= \begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3}\\ O\left((\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the inequality follows by Lemma 1 (1) and the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, and the equality follows by Lemma 2.

Since the proof is similar for B_4 and B_5, we focus on the proof of B_5. For B_5, we have

$$
\begin{aligned}
B_5 &= 2\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa}\left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_l(Z_d)e^{\mathrm{i}tX^*}]p_k^{\mathrm{ft}}(-t)dt\right|^2\\
&= \frac{1}{2\pi^2}\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa}\left|\left\langle\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_l(Z_d)e^{\mathrm{i}tX^*}], p_k^{\mathrm{ft}}(t)\right\rangle_t\right|^2\\
&\leq \frac{1}{\pi}\sum_{d=1}^{D}\int\limits_{G_{\epsilon,n,\zeta}}\left\{\sum_{l=1}^{\kappa}\left|\int f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_l(z_d)dz_d\right|^2\right\}dt\\
&\leq \frac{1}{\pi}\sum_{d=1}^{D}\int\limits_{G_{\epsilon,n,\zeta}}\left\{\int|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2|f_{Z_d}(z_d)|^2dz_d\right\}dt\\
&\leq \frac{2c_{z,1}^2D}{\pi}\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}\int\limits_{G_{\epsilon,n,\zeta}}|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2dt\\
&= \begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3}\\ O\left((\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the first inequality follows by

$$\begin{aligned}
&\sum_{k=1}^{\kappa}\left|\left\langle\left(\frac{|f_{\epsilon}^{\text{ft}}(t)|^{r+2}}{\{|f_{\epsilon}^{\text{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_l(Z_d)e^{\mathrm{i}tX^*}],p_k^{\text{ft}}(t)\right\rangle_t\right|^2\\
&=4\pi^2\sum_{k=1}^{\kappa}\left|\langle h_{2,l,d},p_k\rangle\right|^2\\
&\leq 4\pi^2\|h_{2,l,d}\|_2^2\\
&=2\pi\int\limits_{G_{\epsilon,n,\zeta}}\left|E[q_l(Z_d)e^{\mathrm{i}tX^*}]\right|^2dt,
\end{aligned}$$

where $h_{2,l,d}$ denotes the Fourier inverse of $\left(\frac{|f_{\epsilon}^{\text{ft}}(t)|^{r+2}}{\{|f_{\epsilon}^{\text{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_l(Z_d)e^{\mathrm{i}tX^*}]$, and

$$\begin{aligned}
E[q_l(Z_d)e^{\mathrm{i}tX^*}] &= \int\limits_{z_d}\int\limits_{x^*}e^{\mathrm{i}tx^*}q_l(z_d)f_{X^*,Z_d}(x^*,z_d)dx^*dz_d\\
&= \int\limits_{z_d}\left\{\int\limits_{x^*}e^{\mathrm{i}tx^*}f_{X^*|Z_d=z_d}(x^*)dx^*\right\}f_{Z_d}(z_d)q_l(z_d)dz_d\\
&= \int\limits_{z_d}f_{X^*|Z_d=z_d}^{\text{ft}}(t)f_{Z_d}(z_d)q_l(z_d)dz_d,
\end{aligned}$$

the second inequality follows by the orthnomality of $\{q_l\}_{l=1}^{\kappa}$, the third inequality follows by that f_{Z_d} is supported on $[-1,1]$ and $\max_{d\in\{1,\ldots,D\}}\sup_{z_d\in[-1,1]}|f_{Z_d}(z_d)|\leq c_{z,1}$, and the last equality follows by Lemma 2.

By a similar argument, we have

$$
\begin{aligned}
B_4 &= 2\sum_{d=1}^{D}\sum_{k=1}^{\kappa}\left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_k(Z_d)e^{\mathrm{i}tX^*}]p_0^{\mathrm{ft}}(-t)dt\right|^2 \\
&= \frac{1}{2\pi^2}\sum_{d=1}^{D}\sum_{k=1}^{\kappa}\left|\left\langle\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_k(Z_d)e^{\mathrm{i}tX^*}],p_0^{\mathrm{ft}}(t)\}\right\rangle_t\right|^2 \\
&\le \frac{A(\mathcal{I})}{\pi}\sum_{d=1}^{D}\int_{G_{\epsilon,n,\zeta}}\left\{\sum_{k=1}^{\kappa}\left|\int f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_k(z_d)dz_d\right|^2\right\}dt \\
&\le \frac{A(\mathcal{I})}{\pi}\sum_{d=1}^{D}\int_{G_{\epsilon,n,\zeta}}\left\{\int|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2|f_{Z_d}(z_d)|^2dz_d\right\}dt \\
&\le \frac{2A(\mathcal{I})c_{z,1}^2D}{\pi}\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}\int_{G_{\epsilon,n,\zeta}}|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2dt \\
&= \begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O\left((\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases}.
\end{aligned}
$$

For B_6, we have

$$
\begin{aligned}
B_6 &= \sum_{d=1}^{D}\sum_{k,l=1}^{\kappa}\left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}]p_0^{\mathrm{ft}}(-t)dt\right|^2 \\
&\le \frac{A(\mathcal{I})}{2\pi}\sum_{d=1}^{D}\int_{G_{\epsilon,n,\zeta}}\sum_{k,l=1}^{\kappa}\left|\left\langle f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_k(z_d),q_l(z_d)\right\rangle_{z_d}\right|^2dt \\
&\le \frac{A(\mathcal{I})}{2\pi}\sum_{d=1}^{D}\int_{G_{\epsilon,n,\zeta}}\sum_{k=1}^{\kappa}\left\{\int|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_k(z_d)|^2dz_d\right\}dt \\
&\le \frac{A(\mathcal{I})c_Z^2D}{2\pi}\int_{G_{\epsilon,n,\zeta}}\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2\left\{\sum_{k=1}^{\kappa}\int|q_k(z_d)|^2dz_d\right\}dt \\
&= \frac{2A(\mathcal{I})c_{z,1}^2D\kappa}{2\pi}\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}\int_{G_{\epsilon,n,\zeta}}|f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2dt \\
&= \begin{cases} O\left(\kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O\left(\kappa(\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the first inequality follows by the Cauchy-Schwarz inequality and

$$
\begin{aligned}
E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}] &= \int\limits_{z_d}\int\limits_{x^*} e^{\mathrm{i}tx^*} q_k(z_d)q_l(z_d)f_{X^*,Z_d}(x^*,z_d)dx^*dz_d \\
&= \int\limits_{z_d}\left\{\int\limits_{x^*} e^{\mathrm{i}tx^*} f_{X^*|Z_d=z_d}(x^*)dx^*\right\} f_{Z_d}(z_d)dz_d \\
&= \int\limits_{z_d} f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)f_{Z_d}(z_d)q_k(z_d)q_l(z_d)dz_d,
\end{aligned}
$$

the second inequality follows by the orthonormality of $\{q_l\}_{l=1}^{\kappa}$, third inequality follows by $\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}|f_{Z_d}(z_d)| \le c_Z$, the second equality follows by the unity of q_k, and the last equality follows by Lemma 2.

For B_7, we have

$$
\begin{aligned}
B_7 &= \sum_{d,d'=1}^{D}\sum_{k,l=1}^{\kappa}\left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_k(Z_d)q_l(Z_{d'})e^{\mathrm{i}tX^*}]p_0^{\mathrm{ft}}(-t)dt\right|^2 \\
&\le \frac{A(\mathcal{I})}{2\pi}\sum_{d,d'=1}^{D}\int\limits_{G_{\epsilon,n,\zeta}}\left\{\sum_{k,l=1}^{\kappa}\left|\int\limits_{z_d,z_{d'}} f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)f_{Z_d,Z_{d'}}(z_d,z_{d'})q_k(z_d)q_l(z_{d'})dz_d dz_{d'}\right|^2\right\}dt \\
&\le \frac{A(\mathcal{I})}{2\pi}\sum_{d,d'=1}^{D}\int\limits_{G_{\epsilon,n,\zeta}}\left\{\int\limits_{z_d,z_{d'}}\left|f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)f_{Z_d,Z_{d'}}(z_d,z_{d'})\right|^2 dz_d dz_{d'}\right\}dt \\
&\le \frac{2A(\mathcal{I})c_{z,2}^2D^2}{\pi}\int\limits_{G_{\epsilon,n,\zeta}}\max_{d,d'\in\{1,\cdots,D\}}\sup_{z_d,z_{d'}\in[-1,1]}\left|f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)\right|^2 dt \\
&= \begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O\left((\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the first inequality follows by the Cauchy-Schwarz inequality and

$$
\begin{aligned}
E[q_k(Z_d)q_l(Z_{d'})e^{\mathrm{i}tX^*}] &= \int\limits_{z_d,z_{d'}} \int\limits_{x^*} e^{\mathrm{i}tx^*} q_k(z_d)q_l(z_{d'}) f_{X^*,Z_d,Z_{d'}}(x^*,z_d,z_{d'})dx^*dz_ddz_{d'} \\
&= \int\limits_{z_d,z_{d'}} \left\{ \int\limits_{x^*} e^{\mathrm{i}tx^*} f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(x^*)dx^* \right\} f_{Z_d,Z_{d'}}(z_d,z_{d'})dz_ddz_{d'} \\
&= \int\limits_{z_d,z_{d'}} f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t) f_{Z_d,Z_{d'}}(z_d,z_{d'}) q_k(z_d)q_l(z_{d'})dz_ddz_{d'},
\end{aligned}
$$

the second inequality follows by the orthonormality of $\{q_k\}_{k=1}^{\kappa}$, the third inequality follows by $\max_{d,d'\in\{1,\cdots,D\}} \sup_{z_d,z_{d'}\in[-1,1]} |f_{Z_d,Z_{d'}}(z_d,z_{d'})| \leq c_{z,2}$, and the last equality follows by Lemma 2.

Combining the results so far, we obtain

$$
B = \begin{cases} O\left(\kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), & \text{under Assumption 3} \\ O\left(\kappa (\log n)^{-\frac{2\alpha}{\beta}}\right), & \text{under Assumption 4} \end{cases}.
$$

We now consider the variance term V. Similarly as the bias term, we have

$$
\begin{aligned}
V &\leq \frac{1}{n} \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E\left| \frac{1}{2\pi} \int Q e^{\mathrm{i}tX} \frac{f^{\mathrm{ft}}_\epsilon(-t)|f^{\mathrm{ft}}_\epsilon(t)|^r p^{\mathrm{ft}}(-t)}{\{|f^{\mathrm{ft}}_\epsilon(t)| \vee n^{-\zeta}\}^{r+2}} \right|^2 \\
&\equiv V_1 + \cdots + V_7,
\end{aligned}
$$

where $V_1,\ldots,V_7$ are summations of non-central second moments terms with (p,Q) in the forms of $(p_0,1)$, $(p_k,1)$, $(p_kp_l,1)$, $(p_0,q_k(Z_d))$, $(p_k,q_l(Z_d))$, $(p_0,q_k(Z_d)q_l(Z_d))$, $(p_0,q_k(Z_d)q_l(Z_{d'}))$ separately for $k,l=1,\ldots,\kappa$ and $d,d'=1,\ldots,D$ with $d\neq d'$.

Since the proof is similar, we focus on the proof of V_3. For V_3, we have

$$
\begin{aligned}
V_3 &= \frac{1}{n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r(p_kp_l)^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int_t\int_s e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}\,p_k^{\mathrm{ft}}(-t-s)p_l^{\mathrm{ft}}(s)dsdt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int_v\int_u e^{\mathrm{i}(u-v)X}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}}\,p_k^{\mathrm{ft}}(-u)p_l^{\mathrm{ft}}(v)dudv\right|^2 \\
&\le \frac{1}{16\pi^4 n}\int_v\int_x\left\{\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}(u-v)x}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}},p_k^{\mathrm{ft}}(u)\right\rangle_u\right|^2\right\}f_X(x)dx\sum_{l=1}^{\kappa}|p_l^{\mathrm{ft}}(v)|^2dv \\
&\le \frac{\kappa}{4\pi^2 n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\
&= \begin{cases} O\left(\kappa n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4}\end{cases},
\end{aligned}
$$

where the second equality follows by Lemma 1 (2), the third equality follows by the change of variables $(u,v)=(t+s,s)$, the last equality follows by Lemma 3, and the last inequality follows by Lemma 1 (1) , the unity of $\{p_l\}_{l=1}^{\kappa}$, and the following fact

$$
\begin{aligned}
&\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}(u-v)x}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}},p_k^{\mathrm{ft}}(u)\right\rangle_u\right|^2 \\
&= 4\pi^2\sum_{k=1}^{\kappa}\left|\left\langle h_{3,x}(w)e^{-\mathrm{i}vw},p_k(w)\right\rangle_w\right|^2 \\
&\le 4\pi^2\|h_{3,x}(w)e^{-\mathrm{i}vw}\|_2^2 \\
&= 2\pi\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt,
\end{aligned}
$$

where $h_{3,x}$ denotes the Fourier inversion of $e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}$ with respect to t for every x in the support of X, the first equality follows by Lemma 1 (1) and (4), the inequality follows by the orthnormality of $\{p_k\}_{k=1}^{\kappa}$, the second equality follows by $|e^{-ivw}|=1$,

$|e^{itx}| = 1$, and Lemma 1 (1). By a similar argument, we have

$$V_1, V_2 \quad = \quad \begin{cases} O\left(n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4} \end{cases}.$$

Since the proof is similar for V_4 and V_5, we focus on the proof of V_5. For V_5, we have

$$\begin{aligned} V_5 &= 2\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int q_l(Z_d)e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p_k^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\ &= \frac{1}{2\pi^2 n}\sum_{d=1}^{D}\sum_{l=1}^{\kappa}\int_{z_d}\int_x |q_l(z_d)|^2\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}, p_k^{\mathrm{ft}}(t)\right\rangle_t\right|^2 f_{Z_d,X}(z_d,x)dxdz_d \\ &\leq \frac{c_{z,1}D\kappa}{\pi n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\ &= \begin{cases} O\left(\kappa n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4} \end{cases}, \end{aligned}$$

By a similar argument, we have

$$V_4 \quad = \quad \begin{cases} O\left(\kappa n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4} \end{cases},$$

For V_6, we have

$$
\begin{aligned}
V_6 &= \sum_{d=1}^{D}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p_0^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa}\int_{z_d,x}\left\{\begin{array}{c}\left|\int e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}p_0^{\mathrm{ft}}(-t)dt\right|^2 \\ \times|q_k(z_d)q_l(z_d)|^2 f_{Z_d,X}(z_d,x)\end{array}\right\}dxdz_d \\
&= \frac{A(\mathcal{I})}{2\pi n}\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa}\int_{z_d,x}\left\{\begin{array}{c}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\ \times|q_k(z_d)q_l(z_d)|^2 f_{Z_d,X}(z_d,x)\end{array}\right\}dxdz_d \\
&\le \frac{A(\mathcal{I})c_{z,2}D\kappa^2}{2\pi n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\
&= \begin{cases} O\left(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa^2 n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4}\end{cases}.
\end{aligned}
$$

For V_7, we have

$$
\begin{aligned}
V_7 &= \sum_{d,d'=1}^{D}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int q_k(Z_d)q_l(Z_{d'})e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p_0^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{d,d'=1}^{D}\sum_{k,l=1}^{\kappa}\int_{z_d,z_{d'},x}\left\{\begin{array}{c}\left|\int e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}p_0^{\mathrm{ft}}(-t)dt\right|^2 \\ \times|q_k(z_d)q_l(z_{d'})|^2 f_{Z_d,Z_{d'},X}(z_d,z_{d'},x)\end{array}\right\}dxdz_d d_{z_{d'}} \\
&= \frac{A(\mathcal{I})}{2\pi n}\sum_{d,d'=1}^{D}\sum_{k,l=1}^{\kappa}\int_{z_d,z_{d'},x}\left\{\begin{array}{c}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\ \times|q_k(z_d)q_l(z_{d'})|^2 f_{Z_d,Z_{d'},X}(z_d,z_{d'},x)\end{array}\right\}dxdz_d d_{z_{d'}} \\
&\le \frac{A(\mathcal{I})c_{z,2}D^2\kappa^2}{2\pi n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt \\
&= \begin{cases} O\left(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa^2 n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4}\end{cases}.
\end{aligned}
$$

Combining these results, we obtain

$$V = \begin{cases} O\left(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1}\right), & \text{under Assumption 3} \\ O\left(\kappa^2 n^{2\zeta(r+2)-1}\right), & \text{under Assumption 4} \end{cases}.$$

Under Assumption 4, κ can only diverge in a logarithm rate so that $\kappa(\log n)^{-\frac{2\alpha}{\beta}}$ converges to zero. Therefore $\kappa^2 n^{2\zeta(r+2)-1} \ll \kappa(\log n)^{-\frac{2\alpha}{\beta}}$ for $0 < \zeta < \frac{1}{2(r+2)}$ and n large enough. Combining these results, the conclusion follows. ∎

Lemma 5. *Suppose Assumption 3 and 6 hold true. There exist a function $\psi \in L_1(\mathbb{R})$ such that*

$$\sup_n h^\beta \frac{|K^{\mathrm{ft}}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|} \leq \psi(s),$$

which implies that there exist constants $c > 0$ such that

$$h^{\beta+1} \int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt \leq c.$$

Proof of Lemma 5: Since $\lim_{|t|\to\infty} |t|^\beta |f_\epsilon^{\mathrm{ft}}(t)| = c_\epsilon$, there exists a constant c_F such that $|t|^\beta |f_\epsilon^{\mathrm{ft}}(t)| > c_\epsilon/2$ for all $t \geq c_F$. Then, for constants $c_1 > 0$ such that $c_1 > h^\beta$ and $c_2 > 0$ such that $c_2 > c_F h$ for all $n = 1, 2, \ldots$, we have

$$\begin{aligned} h^\beta \frac{|K^{\mathrm{ft}}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|} &\leq h^\beta \frac{\max_{|s|\leq c_F h} |K^{\mathrm{ft}}(s)|}{\min_{|s|\leq c_F} |f_\epsilon^{\mathrm{ft}}(s)|} \mathbb{1}\{|s| \leq c_F h\} \\ &\quad + \frac{|K^{\mathrm{ft}}(s)||s|^\beta}{(|s|/h)^\beta |f_\epsilon^{\mathrm{ft}}(s/h)|} \mathbb{1}\{|s| > c_F h\} \\ &\leq c_1 c_{\mathrm{os},0}^{-1} (1 + c_F)^\beta \|K^{\mathrm{ft}}\|_\infty \mathbb{1}\{|s| \leq c_2\} \\ &\quad + \frac{2|K^{\mathrm{ft}}(s)||s|^\beta}{c_\epsilon} \\ &\equiv \psi(s), \end{aligned}$$

where the integrability of $\psi(s)$ follows by $\|K^{\mathrm{ft}}\|_\infty < \infty$, the ordinary smoothness of f_ϵ, and $\int |K^{\mathrm{ft}}(s)||s|^\beta ds < \infty$. And the second statement immediately follows by the change of variable $t = s/h$. ■

Lemma 6. *Suppose K_n is a sequence of Borel functions satisfying*

$$K_n(x) \to K(x) \text{ and } \sup_n |K_n(x)| \leq K^*(x),$$

where K^* satisfies

$$\int K^*(x)dx < \infty \text{ and } \lim_{x\to\infty} |xK^*(x)| = 0.$$

If c is a continuity point of f, then for any sequence $h \to 0$ as $n \to \infty$,

$$\int h^{-1}K_n\big(h^{-1}(c-x)\big)f(x)dx = f(c)\int K(x)dx + o(1).$$

Proof of Lemma 6: See Lemma 2.1 of Fan (1991a). ■

Lemma 7. *Suppose f is continuous at x^*, f_ϵ is ordinary smooth of order β, $\|f_\epsilon^{\mathrm{ft}'}\|_\infty < \infty$, $|s|^\beta \big|f_\epsilon^{\mathrm{ft}}(s)\big| \to c_\epsilon$, and $|s|^{\beta+1}\big|f_\epsilon^{\mathrm{ft}'}(s)\big| \to \beta c_\epsilon$, $\|K^{\mathrm{ft}}\|_\infty < \infty$, $\|K^{\mathrm{ft}'}\|_\infty < \infty$, $\int |s|^\beta |K^{\mathrm{ft}}(s)|ds < \infty$, and $\int |s|^\beta|K^{\mathrm{ft}'}(s)|ds < \infty$. Then, we have*

$$\lim_{n\to\infty} h^{2\beta+1}\int_x \frac{1}{4\pi^2}\left|\int_t \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}e^{-\mathrm{i}t(x^*-x)}dt\right|^2 f(x)\,dx = \frac{f(x^*)}{2\pi c_\epsilon^2}\int_s |s|^{2\beta}\big|K^{\mathrm{ft}}(s)\big|^2 ds$$

Proof of Lemma 7: First, we note

$$\begin{aligned}
&\lim_{n\to\infty}\frac{h^\beta}{2\pi}\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}\,e^{-\mathrm{i}sx}ds \\
=\ &\lim_{n\to\infty}\frac{1}{2\pi}\int \frac{K^{\mathrm{ft}}(s)|s|^\beta}{(|s|/h)^\beta f_\epsilon^{\mathrm{ft}}(s/h)}\,e^{-\mathrm{i}sx}ds \\
=\ &\frac{1}{2\pi}\int\left\{\lim_{n\to\infty}\frac{K^{\mathrm{ft}}(s)|s|^\beta}{(|s|/h)^\beta f_\epsilon^{\mathrm{ft}}(s/h)}\mathbb{1}\{|s| > c_F h\}\right\}e^{-\mathrm{i}sx}ds \\
=\ &\frac{1}{2\pi c_\epsilon}\int K^{\mathrm{ft}}(s)|s|^\beta\,e^{-\mathrm{i}sx}ds,
\end{aligned}$$

where the second equality follows by the dominant convergence theorem and Lemma 5. And it follows

$$\frac{h^{2\beta}}{4\pi^2}\left|\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}\, e^{-\mathrm{i}sx}ds\right|^2 \to \frac{1}{4\pi^2 c_\epsilon^2}\left|\int K^{\mathrm{ft}}(s)|s|^\beta\, e^{-\mathrm{i}sx}ds\right|^2. \tag{1.5.27}$$

Moreover, by the integration by parts, we have

$$\begin{aligned}\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx}ds = &\ \frac{1}{\mathrm{i}x}\int \frac{K^{\mathrm{ft}'}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx}ds \\ &+ \frac{1}{\mathrm{i}xh}\int \frac{K^{\mathrm{ft}}(s) f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}^2}(s/h)} e^{-\mathrm{i}sx}ds.\end{aligned} \tag{1.5.28}$$

Since $|s|^\beta \left|f_\epsilon^{\mathrm{ft}}(s)\right| \to c_\epsilon$ and $|s|^{\beta+1}\left|f_\epsilon^{\mathrm{ft}'}(s)\right| \to \beta c_\epsilon$ as $s \to \infty$, there exists an constant $c_F > 0$ be a constant such that $|s|^\beta \left|f_\epsilon^{\mathrm{ft}}(s)\right| > c_\epsilon/2$ and $|s|^{\beta+1}\left|f_\epsilon^{\mathrm{ft}'}(s)\right| < 5\beta c_\epsilon/4$ for any s satisfying $|s| > c_F$. Then, we have

$$\begin{aligned}&\left|\frac{1}{\mathrm{i}x}\int \frac{K^{\mathrm{ft}'}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx}ds\right| \\ &\leq \ \frac{1}{|x|}\int \frac{|K^{\mathrm{ft}'}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|}ds \\ &\leq \ \frac{h}{|x|}\left(\frac{2c_F \max_{|s|\leq c_F h}|K^{\mathrm{ft}'}(s)|}{\min_{|s|\leq c_F}|f_\epsilon^{\mathrm{ft}}(s)|}\right) + \frac{h^{-\beta}}{|x|}\int\limits_{|s|>c_F h} \frac{|K^{\mathrm{ft}'}(s)||s|^\beta}{(|s|/h)^\beta |f_\epsilon^{\mathrm{ft}}(s/h)|}ds \\ &\leq \ \frac{h}{|x|}\, 2c_F c_{\mathrm{os},0}^{-1}(1+c_F)^\beta \|K^{\mathrm{ft}'}\|_\infty + \frac{h^{-\beta}}{|x|}\left(\frac{2}{c_\epsilon}\right)\int |K^{\mathrm{ft}'}(s)||s|^\beta ds \\ &= O\left(h^{-\beta}|x|^{-1}\right),\end{aligned} \tag{1.5.29}$$

and

$$
\begin{aligned}
\left|\frac{1}{\mathrm{i}xh}\int \frac{K^{\mathrm{ft}}(s)f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}2}(s/h)}ds\right| &\leq \frac{h^{-1}}{|x|}\int \frac{|K^{\mathrm{ft}}(s)||f_\epsilon^{\mathrm{ft}'}(s/h)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|^2}ds \\
&\leq \frac{1}{|x|}\left(\frac{2c_F \max_{|s|\leq c_F h}|K^{\mathrm{ft}}(s)| \max_{|s|\leq c_F}|f_\epsilon^{\mathrm{ft}'}(s)|}{\min_{|s|\leq c_F}|f_\epsilon^{\mathrm{ft}}(s)|^2}\right) \\
&+ \frac{h^{-\beta}}{|x|}\int\limits_{|s|>c_F h} \frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1}(|s|/h)^{\beta+1}|f_\epsilon^{\mathrm{ft}'}(s/h)|}{(|s|/h)^{2\beta}|f_\epsilon^{\mathrm{ft}}(s/h)|^2}ds \\
&\leq \frac{h}{|x|}2c_F c_{\mathrm{os},0}^{-2}(1+c_F)^{2\beta}\|K^{\mathrm{ft}}\|_\infty \|f_\epsilon^{\mathrm{ft}'}\|_\infty \\
&+ \frac{h^{-\beta}}{|x|}\left(\frac{5\beta}{c_\epsilon}\right)\int |K^{\mathrm{ft}}(s)||s|^{\beta-1}ds \\
&= O\left(h^{-\beta}|x|^{-1}\right).
\end{aligned}
\tag{1.5.30}
$$

Thus, Lemma 5, (1.5.28), (1.5.29), and (1.5.30) imply that there are a pair of constants $c_1, c_2 > 0$ such that

$$
\sup_n h^{2\beta}\left|\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds\right|^2 \leq \min\{c_1, c_2|x|^{-2}\} \tag{1.5.31}
$$

Therefore, the conclusion follows by

$$
\begin{aligned}
&\lim_{n\to\infty} h^{2\beta+1}\int\limits_x \frac{1}{4\pi^2}\left|\int\limits_t \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}e^{-\mathrm{i}t(x^*-x)}dt\right|^2 f(x)\,dx \\
&= \lim_{n\to\infty}\int\limits_x \frac{h^{2\beta-1}}{4\pi^2}\left|\int\limits_s \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\frac{\mathrm{i}s(x^*-x)}{h}}ds\right|^2 f(x)\,dx \\
&= \frac{f(x^*)}{c_\epsilon^2}\int\limits_x \left|\frac{1}{2\pi}\int\limits_s K^{\mathrm{ft}}(s)|s|^\beta e^{-\mathrm{i}sx}ds\right|^2 dx \\
&= \frac{f(x^*)}{2\pi c_\epsilon^2}\int \left|K^{\mathrm{ft}}(s)\right|^2 |s|^{2\beta}ds,
\end{aligned}
\tag{1.5.32}
$$

where the first equality follows by the change of variable $s = th$, the second equality follows by Lemma 6 with $K_n(x) = \frac{h^{2\beta}}{4\pi^2}\left|\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds\right|^2$ and $K^*(x) = \min\{c_1, c_2|x|^{-2}\}$, and the third equality follows by Lemma 1 (1). ■

Lemma 8. *Suppose Assumption 4 and 7 hold true. There exists a constant $c > 0$ such that*

$$he^{-\beta_0 h^{-\beta}} \int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt \leq c,$$

$$he^{-2\beta_0 h^{-\beta}} \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 dx \leq c.$$

Proof of Lemma 8: The first statement follows by

$$\begin{aligned}
\int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt &= h^{-1} \int \frac{|K^{\mathrm{ft}}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|} ds \\
&\leq c_{\mathrm{ss},0}^{-1} h^{-1} \int_{|s| \leq 1} |K^{\mathrm{ft}}(s)| \, e^{\beta_0 (|s|/h)^\beta} ds \\
&= O\left(h^{-1} e^{\beta_0 h^{-\beta}}\right),
\end{aligned}$$

where the first equality follows by the change of variable $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that K^{ft} is supported on $[-1, 1]$, and the last equality uses $\|K^{\mathrm{ft}}\|_\infty < \infty$.

The second statement follows by

$$\begin{aligned}
\int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 dx &= 2\pi \int \frac{|K^{\mathrm{ft}}(th)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} dt \\
&= 2\pi h^{-1} \int \frac{|K^{\mathrm{ft}}(s)|^2}{|f_\epsilon^{\mathrm{ft}}(s/h)|^2} ds \\
&\leq 2\pi c_{\mathrm{ss},0}^{-2} h^{-1} \int_{|s| \leq 1} |K^{\mathrm{ft}}(s)|^2 \, e^{2\beta_0 (|s|/h)^\beta} ds \\
&= O\left(h^{-1} e^{2\beta_0 h^{-\beta}}\right),
\end{aligned}$$

where the first equality follows by Lemma 1 (1), the second equality follows by the change of variable $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that K^{ft} is supported on $[-1, 1]$, and the last equality uses $\|K^{\mathrm{ft}}\|_\infty < \infty$. ■

Chapter 2

Sharp Regression-Discontinuity Design with a Mismeasured Running Variable

2.1 Introduction

Regression discontinuity (RD) design is a method using non-experimental data to estimate the causal effect of exogenous policy intervention. Compared with other non-experimental approaches, it exploits discreteness in the rules used to assign individuals to receive a treatment. Specifically, in the RD design, the treatment is assigned based on an observed variable, known as the running variable in the literature, while the probability of receiving treatment changes abruptly at a known cut-off point in the support of the running variable. Since the late 1990s, RD design has been widely applied to research questions in various fields of economics and other social sciences. Van der Klaauw (2008), Imbens and Lemieux (2008), Lee and Lemieuxa (2010), and DiNardo and Lee (2011) provide reviews of related works. An appealing feature of RD design is that its identification assumptions are relatively weaker than other non-experimental designs.

The validity of the RD estimate critically depends on the accuracy of the running variable. Empirical works utilizing the RD design mainly assume that the running variable is error-free. Such an assumption, however, is driven more by convenience than by reality, especially when survey data is in use. In fact, the measurement problem of the running variable has been recognized in practice, for example by Battistin et al. (2009) and Hullegie and Klein (2010). Battistin et al. (2009) studied the causal effect of retirement on consumption using the RD approach. They found that a significant number of respondents were reported to be retired while they were still ineligible. Likewise, Hullegie and Klein (2010) studied the causal effect of private medical insurance on health and medical usage. They found that a sizable number of observations were reported to be privately insured while they were ineligible according to their reported income. The mismeasured running variable in RD design also has attracted the attention of econometricians, and there has been a recent wave of theoretical works on this topic. Dong (2015) examined the rounding error in the running variable, Pei and Shen (2016) studied the case where both the true underlying running variable and the measurement error were discrete, and Davezies and Le Barbanchon (2017) considered a continuous measurement error in the running variable.

In this chapter, I consider a classical measurement error in the running variable, where both the true underlying running variable and the measurement error are continuous, thus making Davezies and Le Barbanchon (2017) the most relevant work. Davezies and Le Barbanchon (2017)'s identification strategy requires the design to be two-sided fuzzy, in that the probability of receiving treatment is strictly positive on both side of the cut-off point. Distinct from their work, I study the sharp RD (SRD) design in this chapter, which is a special case of the RD design where the probability of receiving treatment changes from zero to one at the cut-off point. The SRD design is the simplest example of RD design when there is no measurement error. In the presence of measurement error, however, the treatment status may be unobservable together with the true underlying running variable, which could bring extra difficulty to the estimation. Even though I focus on the SRD design throughout this chapter, the proposed method could be easily extended to the RD design with other types of discreteness that was not covered by Davezies and Le Barbanchon (2017), for example the one-sided fuzzy design and the kink design.

Nonparametric methods are commonly used in the theoretical literature of RD analysis to allow for flexible functional forms, which take the parametric approach as a special case [1]. For the error-free RD design, Hahn et al. (2001) and Porter (2003) pro-

[1] The linear regression using observations nearby the cut-off point, which is known as the parametric RD approach in practice, is in fact the same as the local linear regression using the rectangular kernel and a specific bandwidth.

posed a nonparametric estimator for the average treatment effect (ATE) at the cut-off. Specifically, the ATE at the assignment threshold could be estimated by the difference between the nonparametric estimates of regression functions based on observations on each side of the cut-off.

The nonparametric estimation of a regression function in the presence of a classical measurement error in a regressor is studied in the deconvolution literature. This literature began with early works on the kernel density estimation with known error distribution, which include Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991b), and Fan (1991a). Several subsequent works relaxed the known error distribution assumption, including Diggle and Hall (1993), Horowitz and Markatou (1996), Neumann and Hössjer (1997), Efromovich (1997), Li and Vuong (1998), Delaigle et al. (2008), Johannes (2009), and Comte and Lacour (2011). Following the kernel density estimation, earlier works on the estimation of regression function in the presence of measurement error in regressors also focused on the case of known error distribution, including Fan and Truong (1993), Fan and Masry (1992), Delaigle and Meister (2007), Delaigle et al. (2009), Delaigle et al. (2015). Delaigle et al. (2009) specifically proposed a local polynomial estimator when the error distribution is known. For the case of unknown error distribution, Schennach (2004) substituted the estimated error characteristic function in the deconvolution local constant estimator by Fan and Truong (1993) and derived its asymptotic distribution.

Estimators of the ATE at the cut-off in an SRD design with a continuous measurement error in the running variable are developed in this chapter. Two separate cases characterized by the observability of the treatment status are considered for this study. In the case of observed treatment, the sample can be divided into two groups according to the treatment staus. The proposed estimator is then the difference of the deconvolution local linear estimators by Delaigle et al. (2009) based on observations of each group. In the case of unobserved treatment, the sample can no longer be divided explicitly. To reflect the one-sided property of the estimation problem, I adjust the standard deconvolution local linear estimator by using one-sided kernel functions. The Fourier transform of kernel function is usually required to be compactly supported in the deconvolution estimation due to the ill-poseness of the problem; see Horowitz (2014). Since the Fourier transform of the one-sided kernel function cannot be compactly supported except when the original kernel function is constantly zero, an additional ridge parameter is introduced for the regularization purpose.

The remainder of this chapter is organized as follows. Section 2.2 introduces the error-free SRD design framework and the measurement error in the running variable, shows the failure of the classical identification strategy, and provides the identification results with a contaminated sample. In Section 2.3, the construction of the estimators is discussed for two separate cases characterized by the observability of the treatment status. Section 2.4 derives the asymptotic properties of the proposed estimators, and Section 2.5 investigates the finite sample performance of proposed estimators by simulation. In Section 2.6, as an application, the proposed estimators are used to estimate the causal effects of being eligible for Medicaid and CHIP with an income threshold using a two-year (2012-13) longitudinal file released by the Current Population Survey (CPS).

2.2 Setup and Identification

2.2.1 Sharp Regression-Discontinuity Design

Let $D \in \{0,1\}$ be the indicator of treatment status, where $D = 1$ if the treatment is received and $D = 0$ otherwise, and Y_d be the potential outcome corresponding to the treatment status $d \in \{0,1\}$. Then, the observed outcome $Y = Y_D$. Let R^* be the true underlying running variable which determines the treatment assignment with a known cut-off point r_0. Throughout this chapter, let $r_0 = 0$ [2]. In the SRD design, the treatment status sharply changes at the cut-off without uncertainty, which implies $D = \mathbb{1}\{R^* \geq 0\}$. The object of interest is the ATE at the cut-off

$$\theta = E[Y_1 - Y_0 | R^* = 0].$$

In the error-free case, R^* is observed. Then, under Assumption 8 (see Assumption A1-2 of Hahn et al. (2001)) below, θ can be identified by

$$E[Y|R^* = 0+] - E[Y|R^* = 0-]. \tag{2.2.1}$$

[2] If r_0 is known, setting $r_0 = 0$ does not lose generality because the normalized variable $R^* - r_0$ can be used as the running variable rather than R^*.

Assumption 8. *$f_{R^*}(0) > 0$ and $E[Y_d|R^* = r]$ is continuous at $r = 0$ for $d = 0, 1$.*

2.2.2 Mismeasured Running Variable

In the presence of a measurement error in running variable, instead of R^*, researchers observe

$$R = R^* + \epsilon,$$

where ϵ is the measurement error. In this chapter, ϵ is assumed to be independent of the true underlying running variable R^* and the observed outcome Y, and both R^* and ϵ are continuous. These are summarized in Assumption 9 below.

Assumption 9. *R^* and ϵ are continuous, and $\epsilon \perp\!\!\!\perp (Y, R^*)$.*

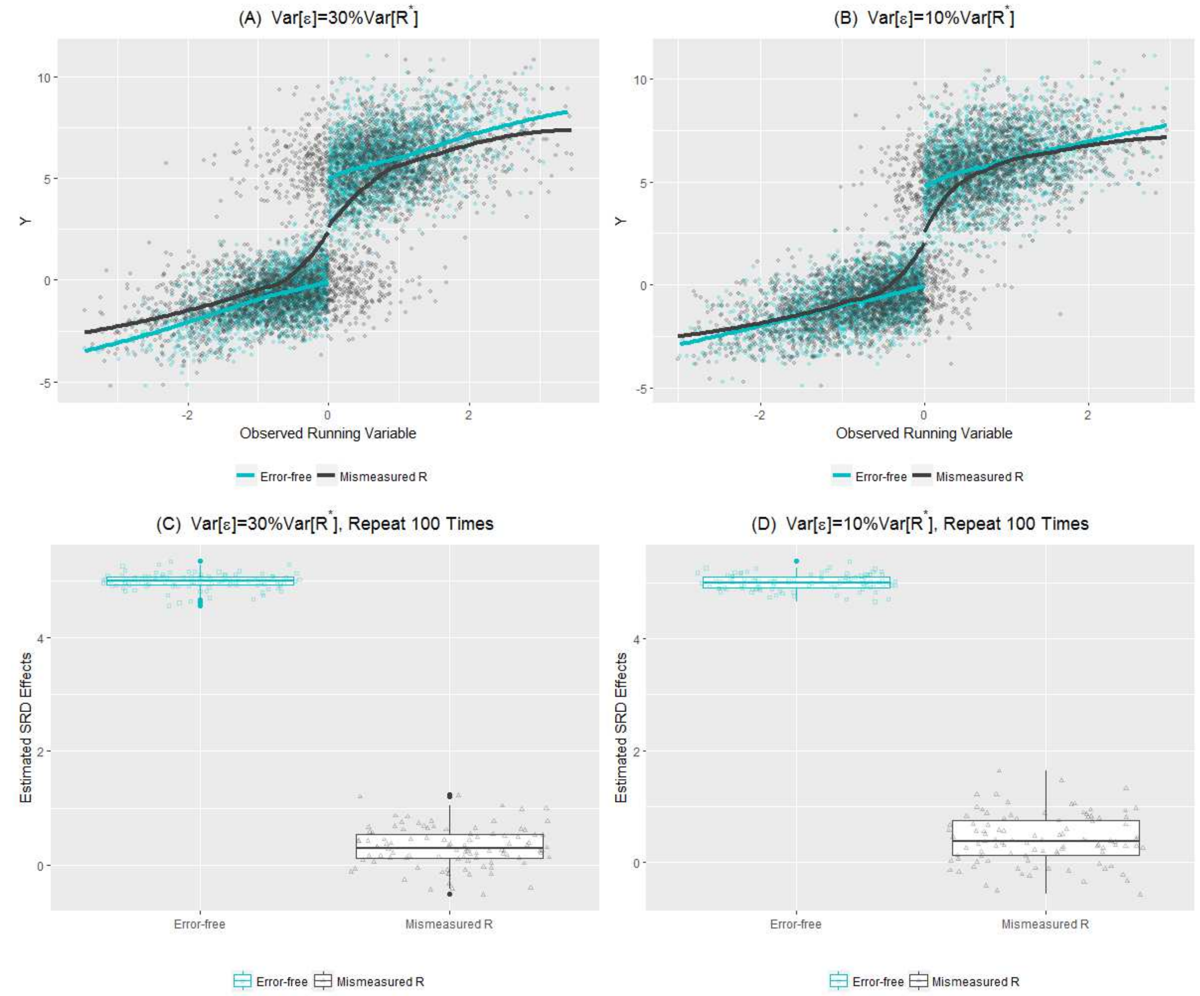
Similar to Proposition 1 of Davezies and Le Barbanchon (2017), I have Proposition 1 as follows.

Proposition 1. *Under Assumption 9, if the density of ϵ is continuous, $E[Y|R]$ is continuous on the interior of the support of R.*

Proposition 1 implies that $E[Y|R = 0-] = E[Y|R = 0+]$ if the measurement error has a continuous density and 0 is an interior point of the support of R. Then, ignoring a continuous measurement error in the running variable fails the identification strategy (2.2.1) in general. The proof of Proposition 1 is an application of the dominant convergence theorem, and is left to Section 2.8.1.

Figure 2.1 visualizes the effect of a continuous measurement error in the running variable using datasets generated by process (a) in Section 2.5 with different signal-to-noise ratios. In Figure 2.1 (A) and (B), the black points depict a sample of the error-free case, where R^* is observed; the green points depict a sample of the contaminated case, where only R is observed; the black and green curves are corresponding local linear fittings by Fan and Gijbels (1996). Figure 2.1 (A) implies that $E[Y|R = r]$ is continuous at $r = 0$ and $E[Y|R^* = 0+] - E[Y|R^* = 0+] = 5$. A comparison between Figure 2.1(A) and Figure 2.1(B) shows that this continuity of $E[Y|R = r]$ at the cut-off does not depend on the magnitude of measurement error, i.e. even a small measurement error

Figure 2.1: SRD Design with a Continuous Error in Running Variable.



in the running variable is enough to clear any discontinuity of regression function at the cut-off. Figure 2.1 (C) and (D) provide further evidences by simulation.

As shown in Figure 2.1, the contaminated sample is a random shift of the error-free sample. So a measurement error in the running variable seems to make an SRD design fuzzy around the cut-off point. However, it is worthy to note that the SRD design with a continuous measurement error in running variable cannot be treated as a fuzzy design without measurement error in general. There are two key differences between an SRD design with a continuous measurement error in running variable and an error-free fuzzy RD design: (1) the identification of a fuzzy RD design requires an observed treatment status, while the treatment status could be unobservable in an SRD design with a mismeasured running variable; (2) even if the treatment status is observed, a continuous measurement error in the running variable implies $E[D|R = 0-] = E[D|R = 0+]$ (see Proposition 1 of Davezies and Le Barbanchon (2017)), while the identification of a fuzzy RD design requires the discontinuity of the probability of receiving treatment at the cut-off.

2.2.3 Identification

Let $\mathrm{i} = \sqrt{-1}$ and $\phi_f(t) = \int e^{\mathrm{i}tx} g(x)\, dx$ be the Fourier transform of a function f. For a random variable X, let f_X denote its density function, and $\phi_X(t)$ denote its characteristic function.

If f_ϵ is known, under Assumption 9, the identification of θ follows by (2.2.1) and

$$E[Y|R^* = r] = \frac{\int \frac{\phi_{E[Y|R]f_R}(t)}{\phi_\epsilon(t)} \exp(-\mathrm{i}tr)\, dt}{\int \frac{\phi_R(t)}{\phi_\epsilon(t)} \exp(-\mathrm{i}tr)\, dt}. \tag{2.2.2}$$

and its proof is left to Section 2.8.1.

Even though it is often assumed in the deconvolution literature, the known error distribution assumption may still be restrictive in practice and cause severe misspecification bias. To identify f_ϵ, auxiliary information is needed in general. Let an extra noisy measure of R^* be

$$R_2 = R^* + \epsilon_2$$

where ϵ_2 is the corresponding measurement error.

Given a pair of noisy measures, there are two typical methods used in the literature to identify the error distribution. One assumes that the densities of both measurement errors are identical and symmetric around zero. Specifically, it requires the following.

Assumption 10.

(1) $(R^*, \epsilon, \epsilon_2)$ *are mutually independent.*

(2) (ϵ, ϵ_2) *are identically distributed and* f_ϵ *is symmetric around zero.*

Under Assumption 10, the identification of f_ϵ follows by

$$\phi_\epsilon(t) = [\phi_{R-R_2}(t)]^{1/2}. \tag{2.2.3}$$

The other method based on the Kotlarski's identity requires weaker assumptions, and allows for cases of non-identical and asymmetric error distributions. However, the corresponding estimation procedure is more technically involved than that of (2.2.3). Further details can be found in Li and Vuong (1998) and Schennach (2004).

2.3 Estimation

For the expository purpose, the true running variable R^* is tentatively assumed to be observable. To estimate θ in this error-free case, it is common to apply some nonparametric techniques, such as Hahn et al. (2001) and Porter (2003). For instance, $E[Y|R^* = 0-]$ and $E[Y|R^* = 0+]$ can be estimated by the local linear estimators $\hat{\mu}_o^+$ and $\hat{\mu}_o^-$, which are defined as solutions to minimization problems (2.3.1) with respect to b_0.

$$\min_{b_0,b_1}\sum_{j=1}^{n}\left(Y_j-b_0-b_1R_j^*\right)^2K_h(R_j^*)D_j^{\pm}, \tag{2.3.1}$$

where $D^+=D$, $D^-=1-D$, K is the kernel function, h is the smoothing bandwidth, and $K_h(x)=K(x/h)/h$. The identification result (2.2.1) thus implies that θ can be estimated as follows

$$\hat{\theta}_0=\hat{\mu}_o^+-\hat{\mu}_o^-. \tag{2.3.2}$$

To construct the estimator for the contaminated case when R is observed instead of R^*, I introduce the following notations. Let $K_{h,k}(x)=(x/h)^kK_h(x)$ for $k=0,1,2$. Let the unit step functions (also known as the heaviside functions) be $u^+(x)=\mathbb{1}\{x>0\}$ and $u^-(x)=\mathbb{1}\{x<0\}$. For a generic function $f(x)$, let $f^{\pm}(x)=f(x)u^{\pm}(x)$.

2.3.1 Case of Observed Treatment

Let us begin with the simplest contaminated case, in which the treatment status D is observed. In this case, the sample can still be divided into two groups according to the value of D as in the error-free case. The deconvolution techniques, as in Delaigle et al. (2009), can be employed to estimate $E[Y|R^*=0-]$ and $E[Y|R^*=0+]$. The identification result (2.2.1) hence implies that θ can be estimated by

$$\hat{\theta}_{d,\epsilon}=\hat{\mu}_{d,\epsilon}^+-\hat{\mu}_{d,\epsilon}^-, \tag{2.3.3}$$

where

$$\hat{\mu}_{d,\epsilon}^{\pm}=\frac{\mathbb{A}_{n,2}^{\pm}\mathbb{B}_{n,0}^{\pm}-\mathbb{A}_{n,1}^{\pm}\mathbb{B}_{n,1}^{\pm}}{\mathbb{A}_{n,2}^{\pm}\mathbb{A}_{n,0}^{\pm}-\left[\mathbb{A}_{n,1}^{\pm}\right]^2},$$

$$\mathbb{A}_{n,k}^{\pm}=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi\mathrm{i}^k}\int\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)D_j^{\pm}\,dt,$$

$$\mathbb{B}_{n,k}^{\pm}=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi\mathrm{i}^k}\int\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)Y_jD_j^{\pm}\,dt.$$

To understand $\hat{\theta}_{d,\epsilon}$, notice that the local linear estimator in the error-free case can be expressed as

$$\hat{\mu}_o^{\pm} = \frac{A_{n,2}^{\pm} B_{n,0}^{\pm} - A_{n,1}^{\pm} B_{n,1}^{\pm}}{A_{n,2}^{\pm} A_{n,0}^{\pm} - \left[A_{n,1}^{\pm}\right]^2},$$

where $A_{n,k}^{\pm} = \frac{1}{n}\sum_{j=1}^{n} K_{h,k}(R_j^*)D_j^{\pm}$ and $B_{n,k}^{\pm} = \frac{1}{n}\sum_{j=1}^{n} K_{h,k}(R_j^*)Y_j D_j^{\pm}$. Heuristically, at a constant bandwidth h, the probability limit of $A_{n,k}^{\pm}$ is $E[K_{h,k}(R^*)D^{\pm}]$. The Plancherel's isometry (see Lemma 1 (1)) then implies

$$\begin{aligned} E\left[K_{h,k}(R^*)D^{\pm}\right] &= \int K_{h,k}(x) f_{R^*}^{\pm}(x)dx \\ &= \frac{1}{2\pi}\int \phi_{K_{h,k}}(t)\phi_{f_{R^*}^{\pm}}(-t)dt \\ &= \frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)} E[\exp(-\mathrm{i}tR)D^{\pm}]\, dt, \end{aligned} \tag{2.3.4}$$

where the last equality follows by $\phi_{K_{h,k}}(t) = \int x^k K(x) e^{\mathrm{i}thx}dx = \mathrm{i}^{-k}\phi_K^{(k)}(th)$, which is an immediate result by using Lemma 1 (5) and Lemma 12. Notice $E[\exp(-\mathrm{i}tR)D^{\pm}]$ is a moment of the observables only and can be estimated by its sample analogue $\frac{1}{n}\sum_{j=1}^{n}\exp(-\mathrm{i}tR_j)D_j^{\pm}$. $\mathbb{A}_{n,k}^{\pm}$ is then obtained by replacing $E[\exp(-\mathrm{i}tR)D^{\pm}]$ on the right hand side of (2.3.4) by $\frac{1}{n}\sum_{j=1}^{n}\exp(-\mathrm{i}tR_j)D_j^{\pm}$. Similar argument can be extended to motivate $\mathbb{B}_{n,k}^{\pm}$ from $B_{n,k}^{\pm}$.

An additional concern arises when using $\mathbb{A}_{n,k}^{\pm}$ and $\mathbb{B}_{n,k}^{\pm}$ as estimators of $E\left[K_{h,k}(R^*)D^{\pm}\right]$ and $E\left[K_{h,k}(R^*)YD^{\pm}\right]$: the error characteristic function $\phi_\epsilon(-t)$ in the denominator would be close to zero when $|t|$ is large. Intuitively, estimation errors of the sample analogues of $E[\exp(-\mathrm{i}tR)D^{\pm}]$ and $E[\exp(-\mathrm{i}tR)YD^{\pm}]$ would be exceptionally amplified when ϕ_ϵ is close to zero in tails, as shown in Figure 2.3 (A). Rigorously, it makes the identifying mapping discontinuous in a way that prevents consistent estimation of the parameter of interest by replacing the population distribution of the data with a consistent sample analogue, and the estimation problem is thus known as ill-posed according to Horowitz (2014). An ill-posed estimation problem thus should always employ certain regularization method. In the present case when D is observed, K could be chosen such that the support of $\phi_K^{(k)}$ is compact. Then, the compactly supported $\phi_K^{(k)}$ can be used to truncate the ill-behaved part of the integrands in tails.

2.3.2 Case of Unobserved Treatment

A more challenging case comes when D is unobserved. In this case, the sample cannot be divided explicitly by the treatment status. Then how to estimate $E[Y|R^* = 0\pm]$ separately is not clear. To overcome this problem, notice $D^{\pm}$ are known functions of R^* in an SRD design. Specifically, at the current setting, $D^{\pm} = u^{\pm}(R^*)$, which imply $K_{h,k}(R^*)D^{\pm} = K^{\pm}_{h,k}(R^*)$ and $K_{h,k}(R^*)YD^{\pm} = K^{\pm}_{h,k}(R^*)Y$. As an alternative of (2.3.4), the Plancherel's isometry implies

$$\begin{aligned} E\left[K_{h,k}(R^*)D^{\pm}\right] &= \int K^{\pm}_{h,k}(x) f_{R^*}(x)dx \\ &= \frac{1}{2\pi}\int \phi_{K^{\pm}_{h,k}}(t)\phi_{f_{R^*}}(-t)dt \\ &= \frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi^{(k)}_{K^{\pm}}(th)}{\phi_{\epsilon}(-t)} E[\exp(-\mathrm{i}tR)]\,dt. \end{aligned} \tag{2.3.5}$$

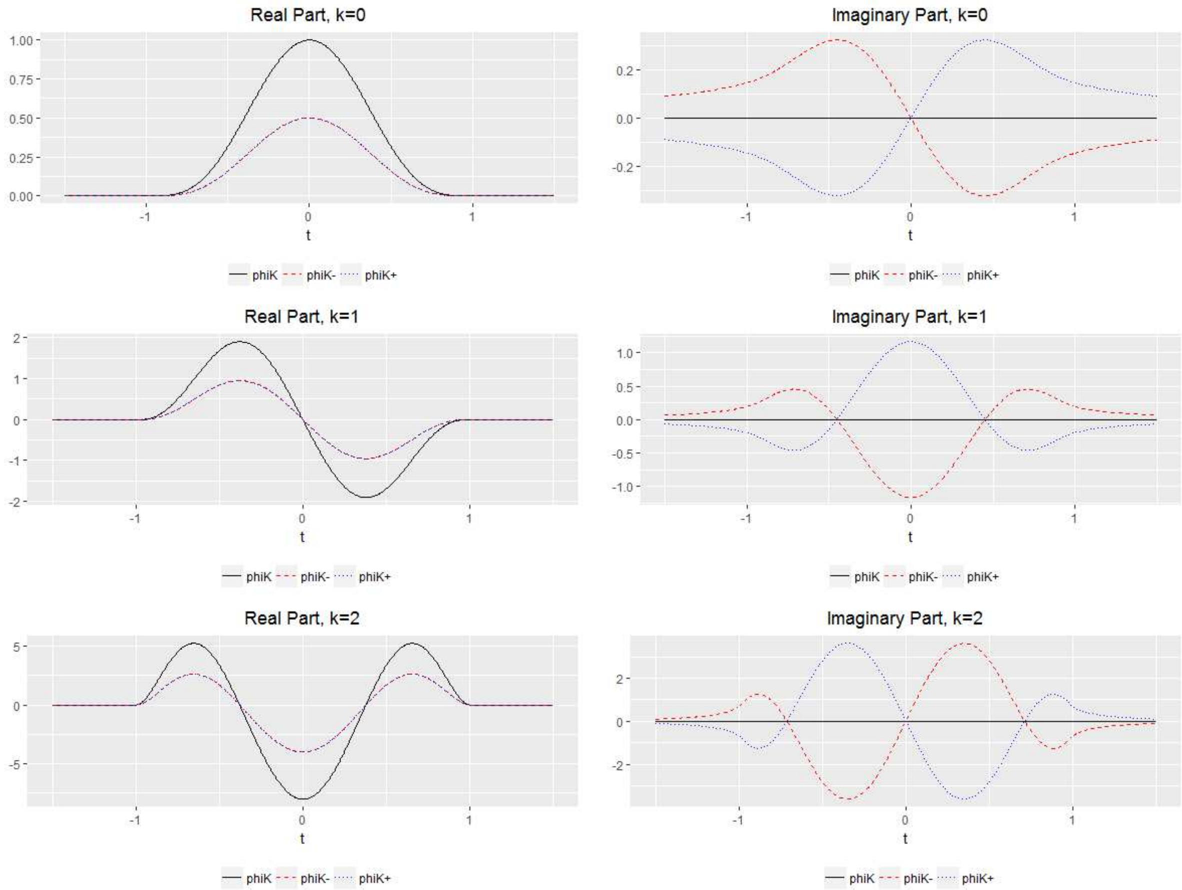
$E[\exp(-\mathrm{i}tR)]$ is a moment of R only, and can be estimated by $\frac{1}{n}\sum_{j=1}^{n}\exp(-\mathrm{i}tR_j)$. $E\left[K_{h,k}(R^*)D^{\pm}\right]$ can then be estimated by replacing $E[\exp(-\mathrm{i}tR)]$ in (2.3.5) by its sample analogue $\frac{1}{n}\sum_{j=1}^{n}\exp(-\mathrm{i}tR_j)$. Similar argument can be extended for $E\left[K_{h,k}(R^*)YD^{\pm}\right]$. Notice that the estimation of $E[\exp(-\mathrm{i}tR)]$ does not require any knowledge of D. Then, the one-sided property of the estimation problem can be reflected by using a pair of truncated or one-sided kernel functions $K^{\pm}$.

Similar to the case of observed D, regularization is required as the estimation problem is ill-posed. Different from the case of observed D when I can choose K with compactly supported Fourier transform to truncate the ill-behaved parts of the integrand in tails, $\phi_{K^{\pm}}$ can never be compactly supported despite the choice of K, except for the trivial case when K is constantly zero. This result is summarized in Proposition 2, and the proof is left to Section 2.8.1.

Proposition 2. *If $\phi_{K^{\pm}}$ is compactly supported, $K = 0$.*

As an example, I consider the kernel function K with Fourier transform $\phi_K(t) = (1-t^2)^4\mathbb{1}\{|t| < 1\}$, and compare $\phi^{(k)}_K$ with $\phi^{(k)}_{K^+}$ and $\phi^{(k)}_{K^-}$ for $k = 0, 1, 2$. As shown in Figure 2.2, even when $\phi^{(k)}_K$ is supported on $[-1, 1]$, the imaginary part of $\phi^{(k)}_{K^{\pm}}$ do not vanish everywhere. The real part of $\phi^{(k)}_{K^+}$ and $\phi^{(k)}_{K^-}$ equal to one half of $\phi^{(k)}_K$. The

Figure 2.2: Fourier Transform of One-sided Kernel Functions



imaginary part of $\phi^{(k)}_{K^+}$ is opposite to $\phi^{(k)}_{K^-}$.

Therefore, the one-sided kernel function can not be used for regularization. Hall and Meister (2007) introduced a kernel-free deconvolution estimator for the density estimation [3] using a ridge-parameter approach. Following their paper, I introduce an additional ridge parameter $\zeta > 0$ to regularize the behaviors of the integrands. Different from their paper, however, I still have kernel functions in my estimators, because the one-sided kernel function is still needed to reflect the one-sided property of the estimation problem when the explicitly division is not available.

If f_ϵ is known, I suggest the estimator of θ as follows

$$\hat{\theta}_\epsilon = \hat{\mu}^+_\epsilon - \hat{\mu}^-_\epsilon, \tag{2.3.6}$$

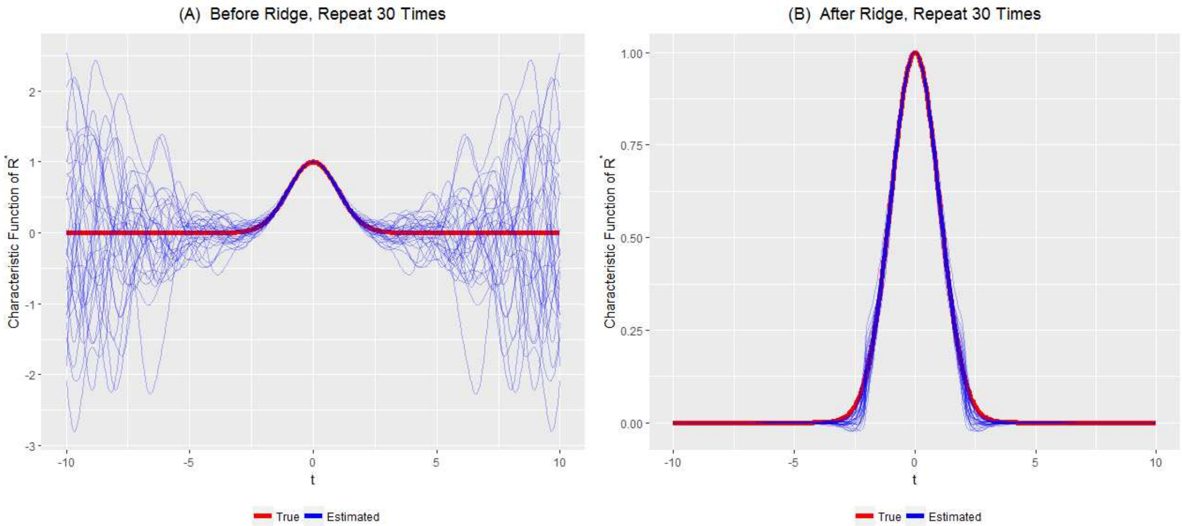
where

$$\begin{aligned}
\hat{\mu}^{\pm}_\epsilon &= \frac{\mathbb{S}^{\pm}_{n,2}\mathbb{T}^{\pm}_{n,0} - \mathbb{S}^{\pm}_{n,1}\mathbb{T}^{\pm}_{n,1}}{\mathbb{S}^{\pm}_{n,2}\mathbb{S}^{\pm}_{n,0} - \left[\mathbb{S}^{\pm}_{n,1}\right]^2}, \\
\mathbb{S}^{\pm}_{n,k} &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi^{(k)}_{K^\pm}(th)\,\phi_\epsilon(t)}{\left\{|\phi_\epsilon(t)| \vee n^{-\zeta}\right\}^2}\exp(-\mathrm{i}tR_j)dt, \\
\mathbb{T}^{\pm}_{n,k} &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi^{(k)}_{K^\pm}(th)\,\phi_\epsilon(t)}{\left\{|\phi_\epsilon(t)| \vee n^{-\zeta}\right\}^2}\exp(-\mathrm{i}tR_j)Y_j dt.
\end{aligned}$$

To see the effectiveness of the ridge-parameter approach, I compare the performance of two estimates of $\phi_{R^*}(-t)$: $\frac{1}{n}\sum_{j=1}^{n}\frac{\exp(-\mathrm{i}tR_j)}{\phi_\epsilon(-t)}$ and $\frac{1}{n}\sum_{j=1}^{n}\frac{\exp(-\mathrm{i}tR_j))\phi_\epsilon(t)}{\{|\phi_\epsilon(t)| \vee n^{-\zeta}\}^2}$. Both estimates are based on $\phi_{R^*}(-t) = \frac{\phi_R(-t)}{\phi_\epsilon(-t)}$. The former is obtained by replacing $\phi_R(-t)$ by its sample analogue and the latter adds an extra ridging term in the denominator. The data is generated by the process (a) in Section 2.5 with $n = 1000$ and $\zeta = 0.25$. As shown in Figure 2.3 (A), the ill-posedness is reflected by the horrendous performance of the estimate in tails. Choosing the kernel function with a compactly supported Fourier transform can solve the problem by restricting the integration to the central

[3] Hall and Meister (2007) also proposed an estimator of the regression function in their paper without deriving the asymptotic property. Their estimator was a Nadaraya-Watson type estimator, which is known to suffer from a bias problem when the point of interest is at the boundary. In this chapter, I focus on the deconvolution variant of the local linear estimator.

Figure 2.3: Ill-posedness and Regularization by Ridging



part, which makes it the natural option for the regularization when D is observed. Figure 2.3 (B) shows that the ridging term $n^{-\zeta}$ iron out the ill-behaved parts in tails. By Proposition 2, the ridge-parameter approach is the automatic option for the regularization when D is unobserved.

2.4 Theoretical Properties

This section discusses the asymptotic properties of the proposed estimators. The estimators based on different regularization methods have different properties. The discussion is thus divided into two parts. First, I derive the asymptotic distribution of the estimators when the treatment D is observed, where the regularization is fulfilled by the kernel function whose fourier transform is compactly supported. I then derive the rate of convergence of the estimators when the treatment D is unobserved, where the one-sided kernel is used and the regularization is fulfilled by the ridge parameter.

I use $\|v\| = \sqrt{v'v}$ and $\|v\|_\infty = \sup_j |v_j|$ to denote the L_2-norm and the L_∞-norm of a vector v respectively, and $\|A\| = \sqrt{\lambda_{\max}(A^\dagger A)}$ and $\|A\|_F = \sqrt{\text{trace}(A^\dagger A)}$ to denote the spectral norm and the Frobenius norm of a complex matrix A, where $A^\dagger$ is A's conjugate transpose. Let $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ separately denote the largest the smallest eigenvalue of a Hermitian positive semi-definite matrix A. Let $m(x) = E[Y|R^* = x] - \theta \mathbb{1}\{x \geq 0\}$ denote the continuous part of the regression function $E[Y|R^* = x]$. Let $\mathcal{N}_0$ denote the interior of a compact interval containing 0. For the asymptotic distribution of $\hat{\theta}_{d,\epsilon}$, the estimator of θ when D is observed and f_ϵ is known, I impose Assumption 11 as follows.

Assumption 11.

(1) $\{Y_j, R_j^*, \epsilon_j\}_{j=1}^n$ *are i.i.d.,* $\epsilon \perp\!\!\!\perp (Y, R^*)$*,* f_ϵ *is known.*

(2) $\int |\phi_{R^*}(t)|dt < \infty$*,* $\phi_\epsilon(t) \neq 0$ *for all* $t \in \mathbb{R}$*,* $\phi_K^{(k)}$ *is not identically zero, and* $\int |\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(t)}|dt < \infty$ *for all* $h > 0$ *and* $k = 0, 1, 2$*.*

(3) For $x \in \mathcal{N}_0$ *,* $f_{R^*}(x)$ *is twice continuously differentiable and bounded away from zero,* $m(x)$ *is twice continuously differentiable for* $x \in \mathcal{N}_0 \backslash \{0\}$*, and* $m(x)$ *is continuous at* $x = 0$ *with finite right and left-hand derivatives to order 2.*

(4) K *is a real and symmetric kernel such that* $\int K(x)dx = 1$ *and has finite moments of order 3.*

Assumption 11 (1) is common in the literature of classical measurement error. Assumption 11 (2) is standard in deconvolution literature. Assumption 11 (3) is Assumption 2 (a) in Porter (2003) when the order of smoothness is 2, which imposes the Hölder type smoothness restrictions on f_{R^*} and m for bias reduction. Assumption 11 (4) assumes the existence of moments of kernel functions which is standard in the nonparametric literature.

It is known in the deconvolution literature that the asymptotic properties of a deconvolution estimator depend on the smoothness of the error distribution. Following the literature, I consider two separate cases characterized by the smoothness of the measurement error distribution: ordinary smooth and supersmooth. f_ϵ is said to be ordinary smooth of order β if

$$c_{\text{os},0}(1 + |t|)^{-\beta} \leq |\phi_\epsilon(t)| \leq c_{\text{os},1}(1 + |t|)^{-\beta} \quad \text{for all } t \in \mathbb{R}, \tag{2.4.1}$$

for some constants $c_{\text{os},1} > c_{\text{os},0} > 0$ and $\beta > 1$. Specifically, the characteristic function of an ordinary smooth distribution decays in tails to zero at a polynomial rate. Typical examples of ordinary smooth densities are the Laplace and gamma density. Different conditions on K and ϕ_ϵ are needed for measurement error with different smoothness. In the case of ordinary smooth f_ϵ, I impose the following assumption.

Assumption 12.

(1) f_ϵ *is ordinary smooth of order* β *with* $\beta > 1/2$.

(2) $\|\phi_\epsilon^{(1)}\|_\infty < \infty$, $\lim_{t\to\infty} |t|^\beta \phi_\epsilon(t) = c_\epsilon$ *and* $\lim_{t\to\infty} |t|^{\beta+1}\phi_\epsilon^{(1)}(t) = -c_\epsilon \beta$.

(3) $\|\phi_K^{(k)}\|_\infty < \infty$, $\int \left[|t|^\beta + |t|^{\beta-1}\right] |\phi_K^{(k)}(t)| dt < \infty$, *and* $\int |t|^{2\beta} |\phi_K^{(k)}(t)||\phi_K^{(k')}(t)| dt < \infty$ *for* $k, k' = 0, 1, 2$.

Assumption 12 (1) requires the measurement error to be ordinary smooth of order $\beta > 1/2$. In the literature, $\beta > 0$ is usually required for an ordinary smooth error, which is sufficient to derive the result, such as Theorem 5 , when D is observed. However, when D is unobserved, the ridge-parameter approach has further requirement on the smoothness order of the measurement error. To make this assumption to be adaptable for both cases, I require $\beta > 1/2$. Assumption 12 (2) is a further restriction on tail behaviors of $\phi_\epsilon(t)$ required for asymptotic normality, see Fan (1991a). Rather than setting bounds, specific limits of $\phi_\epsilon(t)$ and $\phi_\epsilon^{(1)}(t)$ are required as $t \to \infty$. Assumption 12 (3) provides extra conditions needed for an ordinary smooth f_ϵ, which would be automatically satisfied if $\phi_K^{(k)}$ is compactly supported. Let $\mu_{K,k} = \int_0^{+\infty} u^k K(u) du$, $\sigma_{k,k'} = \frac{1}{2\pi c_\epsilon^2} \int |t|^{2\beta} \phi_K^{(k)}(t) \phi_K^{(k')}(t) dt$, and $v_\epsilon(x) = \int Var[Y|R^* = u] f_{R^*}(u) f_\epsilon(x-u)\, du$.

Theorem 5. *Under Assumption 8, 9, 11 and 12, if* $nh^{2\beta+1} \to \infty$ *as* $n \to \infty$,

$$\frac{\hat{\theta}_{d,\epsilon} - \theta - Bias[\hat{\theta}_{d,\epsilon}]}{\sqrt{Var[\hat{\theta}_{d,\epsilon}]}} \xrightarrow{d} N(0,1),$$

where

$$Bias[\hat{\theta}_{d,\epsilon}] = \frac{(\mu_{K,2}^2 - \mu_{K,1}\mu_{K,3})[m^{(2)+}(0) - m^{(2)-}(0)]}{2\,(\mu_{K,0}\mu_{K,2} - \mu_{K,1}^2)} h^2 + o(h^2),$$

$$Var[\hat{\theta}_{d,\epsilon}] = \frac{(\sigma_{0,0}\mu_{K,2}^2 + \sigma_{1,1}\mu_{K,1}^2) v_\epsilon(0)}{(\mu_{K,0}\mu_{K,2} - \mu_{K,1}^2)^2 f_{R^*}^2(0) n h^{2\beta+1}} + o\left(\frac{1}{nh^{2\beta+1}}\right).$$

f_ϵ is said to be supersmooth of order β if

$$c_{\text{ss},0} \exp(-\beta_0 |t|^\beta) \le |\phi_\epsilon(t)| \le c_{\text{ss},1} \exp(-\beta_0 |t|^\beta) \quad \text{for all } t \in \mathbb{R}, \tag{2.4.2}$$

for some constants $c_{\text{ss},1} > c_{\text{ss},0} > 0$, $\beta_0 > 0$, and $\beta > 0$. Specifically, the characteristic function of an supersmooth distribution decays in tails to zero at an exponential rate. Typical examples of supersmooth densities are the Cauchy and Gaussian densities. In the case of supersmooth f_ϵ, I impose the following assumption.

Assumption 13.

(1) f_ϵ is supersmooth of order β with $\beta > 0$.

(2) ϕ_K is supported on $[-1,1]$ and $\|\phi_K^{(k)}\|_\infty < \infty$ for $k = 0,1,2$.

(3) $E|P_{1,n,1}^{\pm}|^2 n^{\frac{\eta}{2+\eta}} h^{\frac{2+2\eta}{2+\eta}} e^{-2\beta_0 h^{-\beta}} \to \infty$ as $n \to \infty$, where $P_{1,n,1}^{\pm}$ is defined as in Section 2.8.1.

Assumption 13 (1) requires the measurement error to be supersmooth of order $\beta > 0$, which is standard in the deconvolution literature. Assumption 13 (2) is imposed to make $\mathbb{A}_{n,k}^{\pm}$ and $\mathbb{B}_{n,k}^{\pm}$ well-defined and meanwhile allow the smoothness of the kernel function K to adapt to that of the measurement error density. Assumption 13 (3) is a refined version of the Lyapounov condition, and it imposes a lower bound on the speed of the bandwidth converging to zero. It could be more primitive, such as Assumption $B_{m,1}$ of Fan (1991a) or Condition 3.1 of Fan and Masry (1992), but it would be technically involved and thus will be omitted here.

Theorem 6. *Under Assumption 8, 9, 11 and 13, if $h = (b\beta_0)^{1/\beta} (\log n)^{-1/\beta}$ for $b > 2$,*

$$\frac{\hat\theta_{d,\epsilon} - \theta - Bias[\hat\theta_{d,\epsilon}]}{\sqrt{Var[\hat\theta_{d,\epsilon}]}} \xrightarrow{d} N(0,1),$$

where $Bias[\hat\theta_{d,\epsilon}]$ is the same as in Theorem 5.

In Theorem 5 and 6 stated above, the asymptotic normality is obtained after normalization. To conduct statistical inference, the variance of the proposed estimator should be estimated. In the case when f_ϵ is ordinary smooth, an explicit form of the dominant term of the variance is derived in Theorem 5. So to estimate this dominant term, I just need to estimate $f_{R^*}(0)$ and $Var[Y|R^* = 0]$, for which the corresponding deconvolution estimators can be considered. In the case when f_ϵ is supersmooth, however, since the variance does not have an explicit form, it is not clear how to estimate.

To derive the convergence rate of $\hat{\theta}_\epsilon$, the estimator of θ when the treatment status D is unobserved, I need further smoothness condition on f_{R^*} and m. This is because the ridging approach here, even though regularize the ill-poseness brought by the mismeasurement, like the ridging approach in many other situations introduces additional bias compared to the estimators in the case of observed treatment status D. Specifically, given $\alpha > 0$ and $c > 0$, denote $\mathcal{F}_{\alpha,c} = \left\{ f \in L_2(\mathbb{R}) : \int |\phi_f(t)|^2(1+|t|^2)^\alpha dt \le C \right\}$ as the Sobolev class of order α, where $L_2(\mathbb{R}) = \{f : \int |f(x)|^2 dx < \infty\}$. I impose the Sobolev condition as follows.

Assumption 14.

(1) $f_{R^*}, mf_{R^*} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ *with* $\alpha > 1/2$ *and* $c_{\text{sob}} > 0$.

(2) $0 < \zeta < 1/4$.

Theorem 7. *Under Assumption 11 and 14,*

(1) if f_ϵ *is ordinary smooth of order* β, $h = n^{-\frac{2\alpha\zeta}{3\beta}}$, *and* $\zeta = \frac{3\beta}{4\alpha+6\beta+3}$,

$$\|\hat{\theta}_\epsilon - \theta\| = O_p\left(n^{-\frac{2\alpha}{4\alpha+6\beta+3}}\right).$$

(2) if f_ϵ *is supersmooth of order* β *and* $h = (\log n)^{-\frac{2\alpha}{3\beta}}$,

$$\|\hat{\theta}_\epsilon - \theta\| = O_p\left((\log n)^{-\frac{2\alpha}{3\beta}}\right).$$

2.5 Numerical Properties

In this section, I provide the simulation results of the estimators proposed in Section 2.3 to investigate their finite sample performances.

2.5.1 Simulation Settings

I consider two different data generating processes.

(1) $R^* \sim N(0,1)$, $Y_0 = R^* + e_0$, $Y_1 = 3 + Y_0 - e_1$ with $e_0 \sim N(0,1)$, $e_1 \sim N(0,1)$, (e_0, e_1, R^*) are mutually independent, , $R = R^* + \epsilon$, and $R_2 = R^* + \epsilon_2$.

(2) $R^* \sim U[-3,3]$, $Y_0 = R^* \exp(-R^*/10) + e_0$, $Y_1 = -(R^* + 3^2)^{1/2} + e_1$, with $e_0 \sim N(0,2)$, $e_1 \sim N(0,1/2)$, (e_0, e_1, R^*) are mutually independent, $R = R^* + \epsilon$, and $R_2 = R^* + \epsilon_2$.

In DGP (1), the regression function is linear on both sides of the cut-off, and the true underlying running variable R^* is infinitely supported. In DGP(2), the regression function has different curvatures on each side of the cut-off, and the true underlying running variable is supported on $[-3,3]$.

For measurement errors, both the ordinary smooth error and the supersmooth error are considered, where Laplace$(0,1)$ is used as an example of ordinary smooth error, and $N(0,1)$ is used as an example of supersmooth error. For the case of known f_ϵ, I consider two different cases characterized by smoothness of f_ϵ. For the case of unknown f_ϵ, I consider two different cases characterized by the smoothness of f_ϵ and f_{ϵ_2}. For the kernel function, I use K such that $\phi_K(t) = (1-t^2)^4 \mathbb{1}\{|t| < 1\}$. In all DGPs, I consider four different sample sizes: $N = 200, 400, 800, 1500$.

2.5.2 Smoothing-parameter Choice

To implement proposed estimators, it requires choosing the bandwidth h in the case of observed treatment status, and both the bandwidth h and the ridge parameter ζ in the case of unobserved treatment status.

In the case of observed treatment, I use SIMEX-type method by Delaigle and Hall (2008) to choose h. Precisely, let $\epsilon_{j,b}^{(k)}$ for $j = 1, \cdots, n$, $b = 1, \cdots, B$, $k = 1,2$ be a collection of independent random variables, which are drawn from the same distribution as ϵ and are independent of data $\{Y_j, D_j, R_j\}_{j=1}^n$. Denote $R_{j,b}^{(1)} = R_j + \epsilon_{j,b}^{(1)}$ and $R_{j,b}^{(2)} = R_j + \epsilon_{j,b}^{(1)} + \epsilon_{j,b}^{(2)}$ for $j = 1, \cdots, n$ and $b = 1, \cdots, B$. Then, SIMEX bandwidth

$\hat{h}$ is given by

$$\hat{h} = \left[\hat{h}_1\right]^2 / \hat{h}_2$$

where

$$\hat{h}_1 = \arg\min_h \sum_{b=1}^{B} \sum_{j=1}^{n} \left(Y_j - \hat{\mu}^{(1)}_{-j,b}(R_j)\right)^2$$
$$\hat{h}_2 = \arg\min_h \sum_{b=1}^{B} \sum_{j=1}^{n} \left(Y_j - \hat{\mu}^{(2)}_{-j,b}(R^{(1)}_{b,j})\right)^2$$

and $\hat{\mu}^{(k)}_{-j,b}$ denote the deconvolution local linear estimator used in Section 2.3 using the sample $\{Y_{j'}, R^{(k)}_{j',b}\}_{j' \neq j}$ for $k = 1, 2$.

In the case of unobserved treatment, both h and ζ, or equivalently $\xi = n^{-\zeta}$, requires to be chosen. Instead of choosing ξ and h simultaneously, I propose a two-stage method. First, I implement the cross-validation procedures by Hall and Meister (2007) to choose ξ. Since $\mathbb{S}^{\pm}_{n,k}$ and $\mathbb{T}^{\pm}_{n,k}$ estimate object of different smoothness, I use different ξ. Specifically, denote ξ^0 as the ridge parameter associated with $\mathbb{S}^{\pm}_{n,k}$ and ξ^1 as the ridge parameter associated with $\mathbb{T}^{\pm}_{n,k}$. Then, $\hat{\xi}^k$ is chosen by

$$\hat{\xi}^k = \arg\min_{\xi} J_k(\xi) - 2\mathfrak{Re}\, \hat{I}_k(\xi)$$

where

$$J_k(\xi) = \frac{1}{2\pi} \int \frac{|\phi_\epsilon(t)|^2}{\{|\phi_\epsilon(t)| \vee \xi\}^4} \left(\frac{1}{n}\sum_{j=1}^{n} Y_j^k \exp(\mathrm{i}tR_j)\right)^2 dt,$$
$$\hat{I}_k(\xi) = \frac{1}{2\pi n(n-1)} \int \frac{1}{\{|\phi_\epsilon(t)| \vee \xi\}^2} \sum_{j \neq j'} Y_j^k Y_{j'}^k \exp(\mathrm{i}t(R_j - R_{j'}))\, dt.$$

Once $\hat{\xi}^k$ is obtained, the SIMEX-type method similar to the case of observed treatment can be implemented to choose h.

2.5.3 Simulation Results

Table 2.1 and 2.2 report the performance of the proposed estimators for DGP1 and DGP2 when the distribution of the measurement error is known. *Oracle* stands for the estimator by Hahn et al. (2001) when R^* is observed; *Naive* stands for the same estimator but R is used in the place of R^*; *OT* stands for the proposed estimator when the treatment status is observed, i.e. $\hat{\theta}_{d,\epsilon}$; *UT* stands for the proposed estimator when the treatment status is unobserved, i.e. $\hat{\theta}_{\epsilon}$. The bandwidth h used in *Oracle* and *Naive* is chosen by Imbens and Kalyanaraman (2012). To ease the computational burden, SIMEX-type method is not used in simulation, and the theoretical optimal choice is implemented.

Table 2.1: Rooted MSE of Estimators with Known f_ϵ for DGP1

		$N = 200$	$N = 500$	$N = 1000$	$N = 2000$
Error-free	*Oracle*	0.509	0.361	0.262	0.154
OS Error	*Naive*	2.990	2.959	2.881	2.879
	OT	1.149	0.546	0.368	0.286
	UT	1.608	0.692	0.475	0.340
SS Error	*Naive*	3.149	3.160	3.063	3.024
	OT	0.918	0.487	0.334	0.213
	UT	1.491	0.799	0.556	0.313

The *Oracle* estimator performs best for all sample sizes, but it is infeasible in the presence of a measurement error. The *Naive* estimator suffers from severe bias, and the rooted mean square errors are approximately the same as the true value of θ, which is due to the failure of identification shown in Proposition 1. The *OT* estimator and *UT* estimator do not outperform the *Oracle* estimator because of the measurement error, but they perform well when the sample size is reasonable, and will converge as the sample size grows. Comparison between Table 2.1 and 2.2 indicates that *UT* estimator is more sensitive to the change of the curvature at the cut-off than the *OT* estimator in the finite sample.

Figure 2.4: Simulation Results of DGP1, $N = 1000$

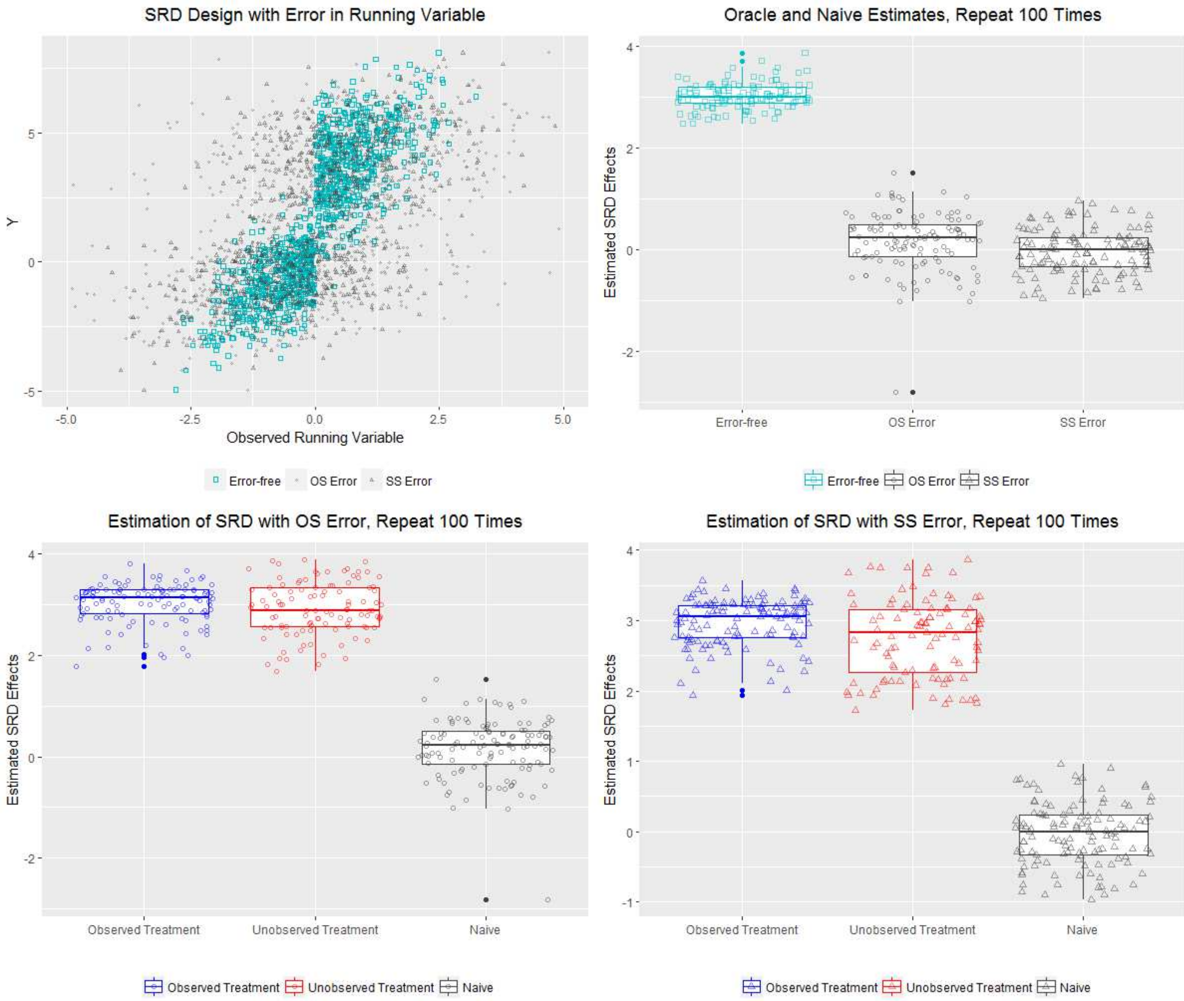


Table 2.2: Rooted MSE of Estimators with known f_ϵ for DGP2

		$N = 200$	$N = 500$	$N = 1000$	$N = 2000$
Error-free	*Oracle*	0.491	0.301	0.223	0.161
OS Error	*Naive*	2.856	2.834	2.818	2.851
	OT	0.560	0.355	0.296	0.206
	UT	1.157	0.757	0.675	0.450
SS Error	*Naive*	3.055	2.834	3.061	2.942
	OT	0.465	0.338	0.326	0.272
	UT	0.791	0.625	0.502	0.399

Figure 2.5: Simulation Results of DGP2, $N = 1000$

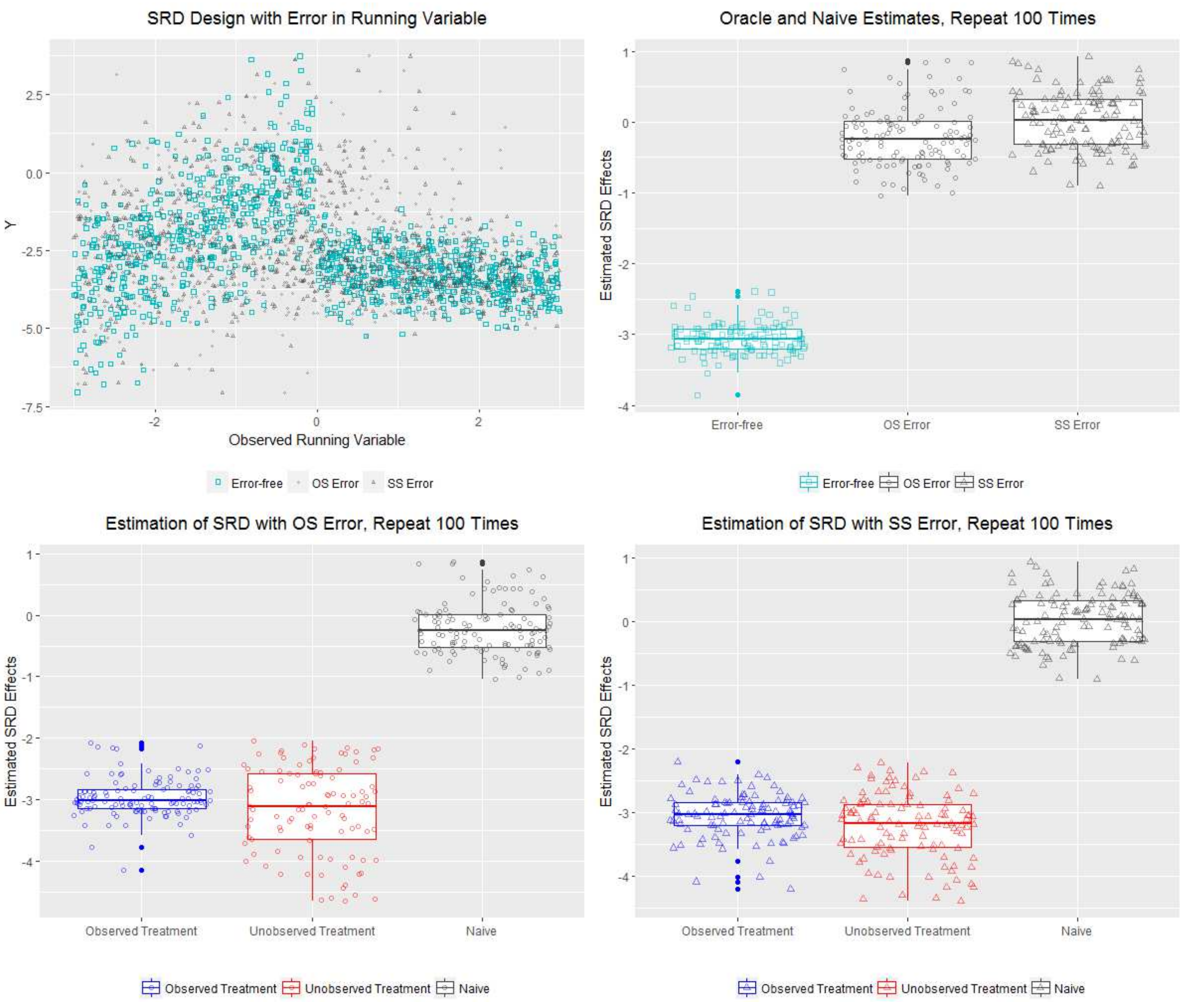


Table 2.3 and 2.4 report the performance of the proposed estimators for DGP1 and DGP2 when the distribution of measurement error is unknown and an extra noisy measure R_2 is used to estimate the error distribution by the sample analogue of (2.2.3). OT2 stands for the proposed estimator when the treatment status is observed; UT2 stands for the proposed estimator when the treatment status is unobserved.

Table 2.3: Rooted MSE of Estimators with Unknown f_ϵ for DGP1

		$N = 200$	$N = 500$	$N = 1000$	$N = 2000$
OS Errors	$OT2$	1.074	0.554	0.393	0.319
	$UT2$	1.627	0.879	0.668	0.472
SS Errors	$OT2$	1.042	0.582	0.328	0.265
	$UT2$	1.960	0.928	0.686	0.468

Table 2.4: Rooted MSE of Estimators with Unknown f_ϵ for DGP2

		$N = 200$	$N = 500$	$N = 1000$	$N = 2000$
OS Errors	$OT2$	0.504	0.241	0.202	0.158
	$UT2$	1.381	0.994	0.689	0.402
SS Errors	$OT2$	0.428	0.186	0.137	0.085
	$UT2$	1.719	1.019	0.804	0.538

2.6 Application: Causal Effects of Being Eligible for Medicaid/CHIP

The United States has taken a continuous effort to reduce the proportion of the uninsured population, with particular emphasis on children and other disadvantaged groups. Medicaid, a government insurance program for people from low-income families, has come to effect since 1965. It is targeting eligible low-income adults, children, pregnant women, elderly adults and people with disabilities. The CHIP was then introduced in 1997, providing health insurance coverage to children from families whose incomes are too high to be eligible for Medicaid but are not able to afford private insurance. The program has expanded over the years, and almost all the states are covering children whose family income are at least 200% of the Federal Poverty Level (FPL).

Given the magnitude of the insurance coverage expansion program targeting children, accurate evaluation of the program effects is not only valuable to researchers but also of significant practical meaning to policymakers. There has been an extensive literature providing insights on the program effects of Medicaid and CHIP on insurance participation, health care utilization, health outcomes and labor market outcomes with special focus on children, parents, and the elderly. Cutler and Gruber (1996) estimated the crowding-out effects on private insurance participation following an expansion of eligibility for the first time, ensuing extensive debates from various literature. The methods adopted by these studies were also varied. To address policy endogeneity, Cutler and Gruber (1996) used instrument variable (IV) approach to estimate the substitution of private insurance after an expansion of public insurance. Blumberg et al. (2000) examined the expansions of public programs by comparing newly eligible and ineligible children using the panel data structure. With increasing popularity of RD design in evaluating program effects in non-experimental settings, several previous studies have used RD approach to estimate the program effect of private health insurance crowding-out, and also on other health related outcomes (Card and Shore-Sheppard (2004), De La Mata (2012), and Koch (2013)). Card and Shore-Sheppard (2004) used age as the running variable to compare cohorts of children immediately on each side of the expansion threshold. De La Mata (2012) and Koch (2013) used the income eligibility as the threshold to compare children who were eligible to those who were immediately ineligible.

However, as I have mentioned in Section 2.1, most empirical works assume the running variable in the RD design is error-free, which is not a reliable assumption given the nature of the survey data used in most empirical literature. In this section, I am going to implement my new estimator to examine the program effect of Medicaid and CHIP, with the consideration of a continuous measurement error in the family income measurements. The identification relies on the income eligibility rules, following De La Mata (2012) and Koch (2013). However, I will adopt a different dataset from what they used, to demonstrate the performance of my proposed estimator in a large sample.

2.6.1 Data and Variables

My empirical implementation requires data on the income of family with children (0-19 years old), eligibility rules for low-cost health insurance, and health insurance participations. I constructed a panel dataset by merging March supplement to Current Population Survey (March CPS) prepared by the CEPR with Medicaid/CHIP eligibility data from the Henry J. Kaiser Family Foundation reports[4]. To avoid questionnaire inconsistency, I restricted the sample to year 2012 and 2013 as the 2014 March CPS included redesigned health insurance and income questions. CPS is conducted on a rotating basis, that is: households from all 50 states and the District of Columbia are in the survey for 4 consecutive months, out for 8, and then return for 4 consecutive months before leaving the sample [5]. Therefore, after merging two years' March CPS, only half of the sample remained from each year. As new respondents are sampled on a probability basis each month, this attrition will not affect the sample distribution.

The family income question from March CPS is about the previous calendar year. The family income reported in 2012 March CPS is, in fact, the income of 2011, and the family income variable in 2013 reports the income of 2012. Therefore, I will use insurance participation variables from 2012 as the outcome variables, and family income variable from 2013 as the primary noisy measure of the assignment variable. To implement my estimator, I introduce income variable from the 2012 dataset as a second noisy measure of the assignment variable. Taking into account of inflation between years, I adjust the income from the 2012 survey with the CPI multiplier between March 2011 to March 2012 by Bureau of Labor Statistics. The poverty level is thus calculated using the ratio of total family income to Federal Poverty Level (FPL) based on family size and year. In this study, I will use 2012 FPL as the denominator.

As RD design focuses only on local effect, I dropped observations whose poverty levels are in the lowest 10 percentiles and the highest 25 percentiles. The remaining sample includes families whose poverty levels are between 65% and 510% of the federal poverty line if using my primary income measure. The upper limit is higher than De La Mata (2012) and Koch (2013) because I am also interested in the effects of CHIP eligibility, and the threshold of which could be as high as 300% to 400% (e.g.,

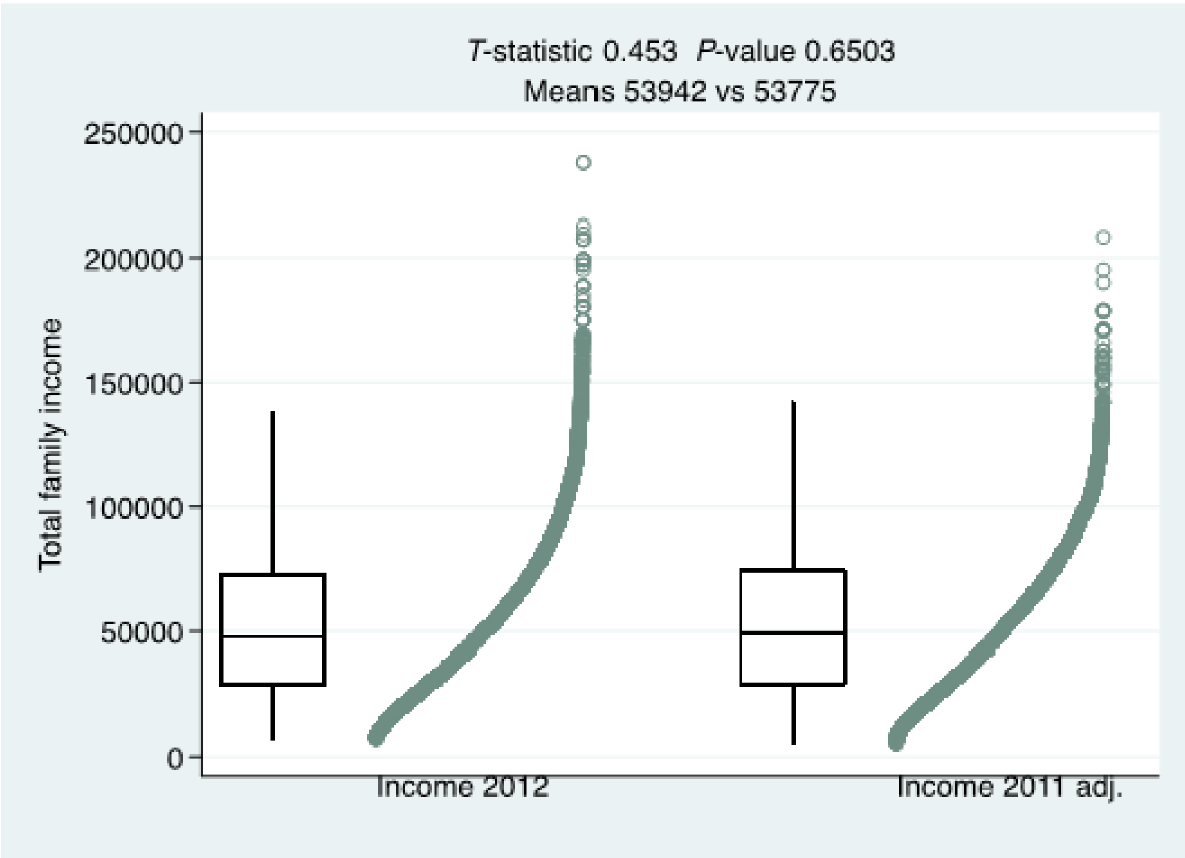
[4]https://www.kff.org/data-collection/trends-in-medicaid-income-eligibility-limits/

[5]Current Population Survey Techinical Paper 66, https://www.census.gov/prod/2006pubs/tp-66.pdf

New York and New Jersey).

Figure 2.6 presents the distribution of the two income measures. The box chart indicates that the two income measures have similar means and distribution, and there is also no significant difference based on the t-test result provided above the figure.

Figure 2.6: Total Family Income



Due to the limited scope of this paper, I only exploit the causal effect of Medicaid/CHIP eligibility on the participation of any insurance, public insurance, Medicaid (take-up effect), and private insurance (crowd-out effect). The CEPR version of March CPS 2012 has four dummy variables indicating the four outcomes mentioned above. Descriptive statistics of the outcome variables by poverty levels are presented in Table 2.5.

Table 2.5: Descriptive of Health Insurance Participation

	Full Sample	100%>PVL	300%>PVL≥ 100%	PVL≥300%
Any insurance	88.43	84.97	87.04	93.49
Public insurance	41.10	71.71	43.89	16.01
Private insurance	56.00	19.51	52.61	86.05
Medicaid	37.51	70.23	39.91	11.87
N	13863	2368	7747	3748

The average participation rate of any insurance of March CPS sample in 2012 is 88%, and the groups with higher income level have higher participation rate. Meanwhile, the highest income group has the lowest participation rate in public insurance, but the highest participation rate in private insurance plans. There are still some children from relatively higher income (≥300%) families participating in Medicaid, as some of the states (e.g., Hawaii, Maryland, New Hampshire and Colombia) have very high-income eligibility thresholds.

2.6.2 Model and Results

Our main analysis is based on a regression discontinuity model of the following form:

$$\text{HIP}_{j,2012} = \theta \mathbb{1}\{\text{Distance}^*_{j,2012} \leq 0\} + m(\text{Distance}^*_{j,2012}) + e_j,$$

for $j = 1, \ldots, n$, where $\text{HIP}_{j,2012}$ is health insurance participation (any health insurance, public health insurance, private health insurance, or Medicaid), $\text{Distance}^*_{j,2012}$ is the normalized family poverty level (ratio of family income to FPL) with respect to income eligibility threshold (% FPL) for Medicaid or CHIP varying by state and time, m is an unknown continuous function, and θ is the causal effects of being eligible to Medicaid or CHIP. The March CPS family income is a survey estimate with nonresponse, and hence the derived family poverty level will inevitably suffer from measurement problem. Instead of the true family poverty level, the constructed poverty level variable can only be regarded as a noisy measure of the true underlying variable: $\text{Distance}_{j,2012} = \text{Distance}^*_{j,2012} + \epsilon_{j,2012}$. As I have mentioned in the variable description part, the 4-8-4 structure of March CPS allows construction of a panel from two consecutive years. In this paper, I use family poverty level in 2011 as another noisy measure for $\text{Distance}^*_{j,2012}$. Considering the possibility of income level affected by inflation, the second noisy measure is adjusted according to CPI index: $\text{Distance}^{\text{CPIadj}}_{j,2012} = \text{Distance}_{j,2011} \times \text{CPI}_{11-12}$.

Table 2.6: Causal Effects of being Eligible for Medicaid

		Medicaid	Private	Public	Any
Total	*Naive*	0.005	0.017	-0.014	0.001
	*UT*2	0.133	-0.189	0.145	0.136
	N	13341	13341	13341	13341
Age 0-1	*Naive*	0.044	0.041	0.05	0.061
	*UT*2	0.253	-0.109	0.281	0.243
	N	643	643	643	643
Age 2-5	*Naive*	0.026	0.014	0.035	0.045
	*UT*2	0.256	-0.118	0.283	0.184
	N	2804	2804	2804	2804
Age 6-19	*Naive*	-0.018	0.01	-0.029	-0.003
	*UT*2	0.143	-0.2	0.155	0.127
	N	9894	9894	9894	9894
Male	*Naive*	-0.024	0.021	-0.03	-0.009
	*UT*2	0.148	-0.054	0.162	0.376
	N	6957	6957	6957	6957
Female	*Naive*	0.01	0.019	0.001	0.015
	*UT*2	0.186	-0.286	0.204	0.055
	N	6384	6384	6384	6384

Table 2.7: Causal Effects of being Eligible for CHIP

		Private	Public	All
Total	*Naive*	0.017	0.013	0.006
	UT2	-0.142	0.284	0.185
	N	11165	11165	11165
Age 0-1	*Naive*	-0.034	0.133	0.028
	UT2	0.005	0.132	0.244
	N	502	502	502
Age 2-5	*Naive*	0.081	-0.028	0.017
	UT2	-0.287	0.05	0.096
	N	2253	2253	2253
Age 6-19	*Naive*	0.013	0.024	0.01
	UT2	-0.055	0.111	0.25
	N	7993	7993	7993
Male	*Naive*	0.021	0.006	0.01
	UT2	-0.056	0.004	0.442
	N	5326	5326	5326
Female	*Naive*	0.032	-0.013	0.007
	UT2	-0.122	0.119	0.256
	N	5839	5839	5839

Table 2.6 and 2.7 show a comparison of estimation results using the naive estimator and the proposed estimator (*UT2*) based on an estimated error characteristic function by using the sample analogue of (2.2.3) and the repeated noisy measures of the normalized family poverty levels constructed above. Table 2.6 uses Medicaid eligibility, and Table 2.7 uses CHIP eligibility as treatment. I first estimate the effect of income eligibility using the full sample, then I restrict the sample based on age groups and gender to address possible heterogeneity. The proposed estimator almost always provides estimates with larger magnitude, which is consistent with the simulation results. The sign of the proposed estimator is also consistent with literature. For example, the naive estimations of the crowding-out effect on private insurance are mostly positive and close to zero, but the proposed estimator provides large negative estimates except for infants in the CHIP eligibility evaluation.

2.7 Conclusion

This chapter develops, to the best of my knowledge, the first nonparametric RD estimator for the sharp design with a continuous measurement error in the running variable. In particular, the measurement error does not affect the outcome variable, and may only impact the running variable. Estimators are proposed for two separate cases characterized by the observability of the treatment. In the case of observed treatment, the proposed estimator is the difference between a pair of deconvolution local linear estimators using observations with different treatment. In the case of unobserved treatment, the sample cannot be explicitly divided. To overcome this, I employ the one-sided kernel functions, and introduce an additional ridge parameter for regularization. Asymptotic properties of the proposed estimators are derived for both cases, and simulation results demonstrate their validity. As a real data example, I consider the estimation of the causal effects of being eligible for Medicaid and CHIP. The results show that ignoring the measurement error in the reported family income may cause severe bias in the conventional RD estimates, and the proposed estimator provides superior results on various outcome variables controlling for demographic heterogeneity.

There are a number of natural avenues for future work stemming from this chapter. First, this chapter focuses on the classical measurement error, which may be overly restrictive in many situations. It would therefore be useful to develop an equivalent estimator that is able to accommodate nonclassical error. Moreover, this chapter derives the convergence rate of the estimator in the case of unobserved treatment, and it would be helpful to further investigate the distributional results and bootstrap methods for inference. Finally, auxiliary information such as the repeated measurements may not always be available to identify the measurement error distribution in practice, and it would be interesting to study the partial identification of the ATE at the cut-off when the measurement error distribution is not point identified.

2.8 Appendix

2.8.1 Proofs of Main Results

Proof of Proposition 1: First, notice that $R^* \perp\!\!\!\perp \epsilon$ implies

$$f_R(r) = \int f_{R^*}(u) f_\epsilon(r-u)du. \tag{2.8.1}$$

By the dominant convergence theorem, the continuity of f_ϵ suggests that f_R is continuous. Moreover, $(Y, R^*) \perp\!\!\!\perp \epsilon$ implies that

$$\begin{aligned} E[Y|R=r]f_R(r) &= \int E[Y|R=r, R^*=a] f_{R^*|R}(a|r)da\, f_R(r) \\ &= \int E[Y|R^*=a] f_{R^*}(a) f_\epsilon(r-a)da, \end{aligned} \tag{2.8.2}$$

where the second equality follows by

$$E[Y|R=r, R^*=a] = E[Y|\epsilon = r-a, R^*=a] = E[Y|R^*=a],$$
$$f_{R^*|R}(a|r)f_R(r) = f_{R^*,R}(a,r) = f_{R^*,\epsilon}(a, r-a) = f_{R^*}(a) f_\epsilon(r-a).$$

The continuity of f_R and f_ϵ then suggest that $E[Y|R=r]$ is continuous by the dominant convergence theorem. ■

Proof of (2.2.2): By the convolution theorem, (2.8.1) and (2.8.2) imply $\phi_{E[Y|R]f_R}(t) = \phi_{E[Y|R^*]f_{R^*}}(t)\phi_\epsilon(t)$ and $\phi_R(t) = \phi_{R^*}(t)\phi_\epsilon(t)$. Then (2.2.2) follows by

$$\begin{aligned} E[Y|R^*=r]f_{R^*}(r) &= \frac{1}{2\pi}\int e^{-itr}\frac{\phi_{E[Y|R]f_R}(t)}{\phi_\epsilon(t)}dt \\ f_{R^*}(r) &= \frac{1}{2\pi}\int e^{-itr}\frac{\phi_R(t)}{\phi_\epsilon(t)}dt \end{aligned}$$

■

Proof of (2.2.3): First, notice $E[\exp\big(\mathrm{i}t(R-R_2)\big)] = E[\exp\big(\mathrm{i}t(\epsilon-\epsilon_2)\big)]$. (2.2.3) then follows by

$$\begin{aligned}E\big[\exp\big(\mathrm{i}t(\epsilon-\epsilon_2)\big)\big] &= E[\exp(\mathrm{i}t\epsilon)]E[\exp(-\mathrm{i}t\epsilon_2)]\\ &= E[\exp(\mathrm{i}t\epsilon)]E[\exp(-\mathrm{i}t\epsilon)]\\ &= \big|E[\exp(\mathrm{i}t\epsilon)]\big|^2,\end{aligned}$$

where the first equality follows by $\epsilon \perp\!\!\!\perp \epsilon_2$, the second equality follows by the fact that ϵ and ϵ_2 are identically distributed, and the last equality follows by the fact that f_ϵ is symmetric around zero. ■

In the complex analysis, a function $g : \mathbb{C} \to \mathbb{C}$ is said to be analytic on an open set $G \subset \mathbb{C}$ if it is complex differentiable at every point of G. Moreover, g is called analytic at x_0 if it is analytic on a neighbourhood of x_0, and is called entire if it is analytic on $\mathbb{C}$.

Lemma 9. *Zeros of a non-constant analytic function are isolated.*

By the isolation of zeros, I mean that if $x_0 \in \mathbb{C}$ is one zero of $g(x)$, there exists an $\epsilon > 0$ such that $f(x_1) \neq 0$ for any $x_1 \in \{x \in \mathbb{C} : 0 < |x - x_0| < \epsilon\}$. An immediate corollary of Lemma 9 is: if a function g is analytic at x_0 and $g(x_0) = 0$, either g is constantly zero in a neighborhood of x_0 or there is no other zero in a neighborhood of x_0.

Lemma 10. *Given that functions g_1 and g_2 are analytic on a connected open set $G \subset \mathbb{C}$, if $g_1 = g_2$ on some non-empty open subset of G, then $g_1 = g_2$ on G.*

Lemma 10 is known as the identity theorem for analytic function, and it says that a analytic function is completely determined by its values on a neighborhood. Denote the holomorphic Fourier transform of $f : \mathbb{R} \to \mathbb{C}$ as

$$A_f(z) = \frac{1}{2\pi}\int e^{\mathrm{i}xz} f(x)dx$$

where $z = z_1 + z_2\mathrm{i}$. A_f extends the domain of the ordinary Fourier transform ϕ_f from $\mathbb{R}$ to $\mathbb{C}$. Specifically,

$$\begin{aligned} A_f(z) &= \frac{1}{2\pi}\int \exp(\mathrm{i}x(z_1+z_2\mathrm{i}))f(x)dx \\ &= \frac{1}{2\pi}\int \exp(\mathrm{i}xz_1)\left[\exp(-xz_2)f(x)\right]dx \\ &= \phi_{\exp(-xz_2)f(x)}(z_1). \end{aligned} \tag{2.8.3}$$

(2.8.3) shows that the holomorphic Fourier transform is the ordinary Fourier transform of the function $\exp(-xz_2)f(x)$. As $\exp(-xz_2)$ grows rapidly at infinity, A_f may not be well-defined for a generic function f. For the proof of Proposition 2, however, it is enough to focus on the function f that has a compact support. A_f is therefore well-defined because $\exp(-xz_2)$ is bounded on a compact set. The support information about f can be characterized by the analyticity conditions of A_f, and the result for the case of a compactly supported f is given in Lemma 11 below.

Lemma 11. *If f is supported in $[-b, b]$ for some $b > 0$ and satisfy $\int_{-b}^{b} |f(x)|^2 dx < \infty$, A_f is an entire function.*

Lemma 9 and Lemma 10 can be found in Theorem 1.2 of Lang (2013), and Lemma 11 is known as the Paley-Wiener theorem (see Theorem 7.2.1 of Strichartz (2003)).

Proof of Proposition 2: If ϕ_{K^+} is compactly supported, $A_{\phi_{K^+}}$ is entire by Lemma 11. Notice $A_{\phi_{K^+}}(x) = K^+(x)$ for $x \in \mathbb{R}$. $K^+(x) = 0$ for any $x < 0$ then implies that $A_{\phi_{K^+}}(x) = 0$ for any $x < 0$. Lemma 9 then suggests that $A_{\phi_{K^+}}(x_0) = 0$ in a neighborhood of a specific x_0 for any $x_0 < 0$. Lemma 10 therefore claims that $A_f(z) = 0$ for any $z \in \mathbb{C}$, which indicates $K^+ \equiv 0$. ■

Proof of Theorem 5: Let $U = Y - E[Y|R^*]$. Then, we have

$$\hat{\theta}_{d,\epsilon} - \theta = \hat{m}^+_{d,\epsilon} - \hat{m}^-_{d,\epsilon},$$

where

$$\hat{m}^{\pm}_{d,\epsilon} = \frac{\mathbb{A}^{\pm}_{n,2}\mathbb{B}^{\pm}_{n,0,m} - \mathbb{A}^{\pm}_{n,1}\mathbb{B}^{\pm}_{n,1,m}}{\mathbb{A}^{\pm}_{n,2}\mathbb{A}^{\pm}_{n,0} - [\mathbb{A}^{\pm}_{n,1}]^2},$$

$$\mathbb{B}^{\pm}_{n,k,m} = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi^{(k)}_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)[m(R^*_j) + U_j]D^{\pm}_j.$$

Notice that $\hat{m}^{+}_{d,\epsilon}$ and $\hat{m}^{-}_{d,\epsilon}$ are independent with each other because they are separately based on observations from each side of the cut-off point. For the asymptotic distribution of $\hat{\theta}_{d,\epsilon}$, I then proceed by deriving the asymptotic distribution of $\hat{m}^{+}_{d,\epsilon}$. The distributional result of $\hat{m}^{-}_{d,\epsilon}$ can be obtained in a similar way.

Let $a^{+}_k = f_{R^*}(0)\mu^{+}_{K,k}$ for $k = 0, 1, 2$. Then, it follows

$$\hat{m}^{+}_{d,\epsilon} - m(0) = \frac{1}{n}\sum_{j=1}^{n} P^{+}_{n,j}, \tag{2.8.4}$$

where $P^{+}_{n,j} = P^{+}_{1,n,j} + P^{+}_{2,n,j}$ and

$$\begin{aligned}
P^{+}_{1,n,j} &= \left(\frac{a^{+}_2}{a^{+}_0 a^{+}_2 - \left[a^{+}_1\right]^2}\right)\frac{1}{2\pi}\int \frac{\phi_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)[m(R^*_j) + U_j - m(0)]D^{+}_j\, dt \\
&- \left(\frac{a^{+}_1}{a^{+}_0 a^{+}_2 - \left[a^{+}_1\right]^2}\right)\frac{1}{2\pi\mathrm{i}}\int \frac{\phi^{(1)}_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)[m(R^*_j) + U_j - m(0)]D^{+}_j\, dt, \\
P^{+}_{2,n,j} &= \left(\frac{\mathbb{A}^{+}_{n,2}}{\mathbb{A}^{+}_{n,2}\mathbb{A}^{+}_{n,0} - \left[\mathbb{A}^{+}_{n,1}\right]^2} - \frac{a^{+}_2}{a^{+}_0 a^{+}_2 - \left[a^{+}_1\right]^2}\right) \\
&\quad \times \frac{1}{2\pi}\int \frac{\phi_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)[m(R^*_j) + U_j - m(0)]D^{+}_j\, dt \\
&- \left(\frac{\mathbb{A}^{+}_{n,1}}{\mathbb{A}^{+}_{n,2}\mathbb{A}^{+}_{n,0} - \left[\mathbb{A}^{+}_{n,1}\right]^2} - \frac{a^{+}_1}{a^{+}_0 a^{+}_2 - \left[a^{+}_1\right]^2}\right) \\
&\quad \times \frac{1}{2\pi\mathrm{i}}\int \frac{\phi^{(1)}_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)[m(R^*_j) + U_j - m(0)]D^{+}_j\, dt.
\end{aligned}$$

Step 1:

$$\frac{\sum_{j=1}^{n} P_{1,n,j}^{+} - nE[P_{1,n,1}^{+}]}{\sqrt{nVar[P_{1,n,1}^{+}]}} \xrightarrow{d} N(0,1) \tag{2.8.5}$$

By Lyapunov central limit theorem, for (2.8.5), it is suffice to show

$$\lim_{n\to\infty} \frac{E\left|P_{1,n,1}^{+}\right|^{2+\eta}}{n^{\eta/2}\left[E\left|P_{1,n,1}^{+}\right|^{2}\right]^{(2+\eta)/2}} = 0 \tag{2.8.6}$$

First, by Jensen inequality, we have

$$\begin{aligned} E\left|P_{1,n,1}^{+}\right|^{2+\eta} \leq\ & 2^{1+\eta}\left(\frac{a_2^{+}}{a_0^{+}a_2^{+} - [a_1^{+}]^2}\right)^{2+\eta} \\ & \times E\left|\frac{1}{2\pi}\int \frac{\phi_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*) + U_1 - m(0)]D_1^{+}\,dt\right|^{2+\eta} \\ & + 2^{1+\eta}\left(\frac{a_1^{+}}{a_0^{+}a_2^{+} - [a_1^{+}]^2}\right)^{2+\eta} \\ & \times E\left|\frac{1}{2\pi\mathrm{i}}\int \frac{\phi_K^{(1)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*) + U_1 - m(0)]D_1^{+}\,dt\right|^{2+\eta} \end{aligned} \tag{2.8.7}$$

Let $\mu_{m,2+\eta}^{+}(x) = E[|m(R^*) + U - m(0)|^{2+\eta}D^{+}|R = x]f_R(x)$ for the constant $\eta \geq 0$. Then using the law of iterated expectation, for $k = 0, 1$, we can write

$$\begin{aligned} & E\left|\frac{1}{2\pi\mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*) + U_1 - m(0)]D_1^{+}\,dt\right|^{2+\eta} \\ & \quad = \int_x \left|\frac{1}{2\pi\mathrm{i}^k}\int_t \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tx)\right|^{2+\eta} \mu_{m,2+\eta}^{+}(x)dx. \end{aligned}$$

If $\eta > 0$, we have

$$
\begin{aligned}
& E\left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right|^{2+\eta} \\
& \leq h^{-(2+\eta)(1+\beta)+1}\left(\frac{h^{\beta+1}}{2\pi}\int \frac{|\phi_K^{(k)}(th)|}{|\phi_\epsilon(-t)|}\,dt\right)^{\eta} \\
& \quad \times h^{2\beta+1}\int\left|\frac{1}{2\pi}\int\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp\left(-\mathrm{i}tx\right)\,dt\right|^2 \mu^+_{m,2+\eta}(x)dx \\
& = O\left(h^{-(2+\eta)(1+\beta)+1}\right),
\end{aligned}
\tag{2.8.8}
$$

where the last equality follows by Lemma 13 and Lemma 14. Then, (2.8.7) and (2.8.8) together imply

$$
E\left|P^+_{1,n,1}\right|^{2+\eta} = O\left(h^{-(2+\eta)(1+\beta)+1}\right). \tag{2.8.9}
$$

Also, we note

$$
\begin{aligned}
& E\left|P^+_{1,n,1}\right|^2 \\
= & \left(\frac{a_2^+}{a_0^+a_2^+-[a_1^+]^2}\right)^2 E\left|\frac{1}{2\pi}\int\frac{\phi_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right|^2 \\
+ & \left(\frac{a_1^+}{a_0^+a_2^+-[a_1^+]^2}\right)^2 E\left|\frac{1}{2\pi\mathrm{i}}\int\frac{\phi_K^{(1)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right|^2 \\
- & \frac{a_2^+a_1^+}{[a_0^+a_2^+-[a_1^+]^2]^2}\left\{E\left(\frac{\mathrm{i}}{4\pi^2}\int\frac{\phi_K(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right.\right. \\
& \quad\left.\times\int\frac{\phi_K^{(1)}(-th)}{\phi_\epsilon(t)}\exp(\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right) \\
& \quad + E\left(\frac{1}{4\pi^2\mathrm{i}}\int\frac{\phi_K^{(1)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right. \\
& \quad\left.\left.\times\int\frac{\phi_K(-th)}{\phi_\epsilon(t)}\exp(\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right)\right\}.
\end{aligned}
\tag{2.8.10}
$$

And by Lemma 14, we have

$$\begin{aligned}
& E\left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right|^2 \\
&= h^{-2\beta-1}\left\{h^{2\beta+1}\int\left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp\left(-\mathrm{i}tx\right)\,dt\right|^2 \mu_{m,2}^+(x)dx\right\} \qquad (2.8.11)\\
&= \frac{h^{-2\beta-1}\mu_{m,2}^+(0)}{2\pi c_\epsilon^2}\int |s|^{2\beta}|\phi_K^{(k)}(s)|^2 ds(1+o(1)),
\end{aligned}$$

and

$$\begin{aligned}
& E\left(\frac{1}{4\pi^2 \mathrm{i}^{k-k'}}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right.\\
&\qquad \left.\times\int \frac{\phi_K^{(k')}(-th)}{\phi_\epsilon(t)}\exp(\mathrm{i}tR_1)[m(R_1^*)+U_1-m(0)]D_1^+\,dt\right)\\
&= \frac{h^{-2\beta-1}}{i^{k-k'}}\left\{h^{2\beta+1}\int\left(\frac{1}{2\pi}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tx)dt\right)\right. \qquad (2.8.12)\\
&\qquad \left.\times\left(\frac{1}{2\pi}\int \frac{\phi_K^{(k')}(-th)}{\phi_\epsilon(t)}\exp(\mathrm{i}tx)dt\right)\mu_{m,2}^+(x)dx\right\}\\
&= \frac{h^{-2\beta-1}\mu_{m,2}^+(0)}{2\pi c_\epsilon^2 i^{k-k'}}\int |s|^{2\beta}\phi_K^{(k)}(s)\phi_K^{(k')}(s)ds(1+o(1))
\end{aligned}$$

Note that (2.8.12) implies that sum of two expected cross-product terms of (2.8.10) vanishes as $n\to\infty$. Then, (2.8.10), (2.8.11), and (2.8.12) together imply

$$E|P_{1,n,1}^+|^2 = \frac{(\sigma_{0,0}\mu_{K,2}^2+\sigma_{1,1}\mu_{K,1}^2)v_\epsilon^+(0)}{(\mu_{K,0}\mu_{K,2}-\mu_{K,1}^2)^2 f_{R^*}^2(0)h^{2\beta+1}} + o\left(\frac{1}{h^{2\beta+1}}\right) \qquad (2.8.13)$$

where $\sigma_{k,k'} = \frac{1}{2\pi c_\epsilon^2}\int |t|^{2\beta}\phi_K^{(k)}(t)\phi_K^{(k')}(t)dt$ and $v_\epsilon^+(x) = \int_0^{+\infty} Var[Y|R^*=u]f_{R^*}(u)f_\epsilon(x-u)\,du$. Hence, the Lyapunov condition (2.8.6) holds if $nh\to\infty$ as $n\to\infty$.

Step 2:

$$\frac{\sum_{j=1}^{n} P_{2,n,j}^{+} - nE[P_{2,n,1}^{+}]}{\sqrt{nVar[P_{1,n,j}^{+}]}} = o_p(1) \tag{2.8.14}$$

First, we note

$$\begin{aligned}
\mathbb{A}_{n,k}^{+} &= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_j)D_j^{+}dt \\
&= E\left[\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)D_1^{+}dt\right] \\
&+ O_p\left(n^{-1/2}\left[E\left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)D_1^{+}dt\right|^2\right]^{1/2}\right),
\end{aligned} \tag{2.8.15}$$

where

$$E\left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)D_1^{+}\,dt\right|^2 = O(h^{-(2\beta+1)}) \tag{2.8.16}$$

follows by Lemma 14, and

$$\begin{aligned}
&E\left[\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-\mathrm{i}tR_1)D_1^{+}dt\right] \\
&= \frac{1}{2\pi \mathrm{i}^k}\int \phi_K^{(k)}(th)\phi_{f_{R^*}^{+}}(-t)dt \\
&= \int_0^{+\infty} K_{h,k}(x)f_{R^*}(x)dx \\
&= f_{R^*}(0)\int_0^{+\infty} K(z)z^k dz + hf_{R^*}^{(1)}(0)\int_0^{+\infty} K(z)z^{k+1}dz + o(h) \\
&= a_k^{+} + O(h),
\end{aligned} \tag{2.8.17}$$

Hence, we have

$$\mathbb{A}_{n,k}^{+} - a_k^{+} = O(h) + O_p\left(\frac{1}{n^{1/2}h^{\beta+1/2}}\right) \tag{2.8.18}$$

Moreover, let $M_a^+ = \begin{pmatrix} a_0^+ & a_1^+ \\ a_1^+ & a_2^+ \end{pmatrix}$ and $M_{\mathbb{A}}^+ = \begin{pmatrix} \mathbb{A}_{n,0}^+ & \mathbb{A}_{n,1}^+ \\ \mathbb{A}_{n,1}^+ & \mathbb{A}_{n,2}^+ \end{pmatrix}$. Then, for $k = 0, 1, 2$, we have

$$
\begin{aligned}
&\left| \frac{\mathbb{A}_{n,k}^+}{\mathbb{A}_{n,2}^+ \mathbb{A}_{n,0}^+ - \left[\mathbb{A}_{n,1}^+\right]^2} - \frac{a_k^+}{a_0^+ a_2^+ - \left[a_1^+\right]^2} \right| \\
&\leq \| [M_{\mathbb{A}}^+]^{-1} - [M_a^+]^{-1} \| \\
&\leq \| [M_{\mathbb{A}}^+]^{-1} \| \| [M_a^+]^{-1} \| \| M_{\mathbb{A}}^+ - M_a^+ \| \\
&\leq \lambda_{\min}^{-1}(M_a^+) \big(\lambda_{\min}(M_a^+) + O_p(\| M_{\mathbb{A}}^+ - M_a^+ \|) \big)^{-1} \| M_{\mathbb{A}}^+ - M_a^+ \|,
\end{aligned}
\tag{2.8.19}
$$

where the first inequality follows by the fact that $|m_{j,j'}| \leq \sqrt{\max_{1 \leq j \leq k} \sum_{j'=1}^{k} |m_{j,j'}|^2} \leq \sqrt{\lambda_{\max}(M^\dagger M)} = \|M\|$ for a matrix $M = (m_{j,j'}) \in \mathbb{C}^{k \times k}$, and last inequality follows by $\lambda_{\min}(M_a^+) > 0$, $\lambda_{\min}(M_{\mathbb{A}}^+) \geq \inf_{|\delta|=1} \delta'(M_{\mathbb{A}}^+ - M_a^+)\delta + \lambda_{\min}(M_a^+)$, and $|\inf_{|\delta|=1} \delta'(M_{\mathbb{A}}^+ - M_a^+)\delta \leq \|M_{\mathbb{A}}^+ - M_a^+\|$. Note $\|M\|^2 = \lambda_{\max}(M^\dagger M) \leq \operatorname{tr}(M^\dagger M) = \sum_{j,j'=1}^{k} |m_{j,j'}|^2$ implies

$$
\|M_{\mathbb{A}}^+ - M_a^+\| = O_p\left(\max_{k=0,1,2} \left| \mathbb{A}_{n,k}^+ - a_k^+ \right| \right). \tag{2.8.20}
$$

Then, if $nh^{2\beta+1} \to \infty$, by (2.8.18), (2.8.19), and (2.8.20), we have

$$
\frac{\mathbb{A}_{n,k}^+}{\mathbb{A}_{n,2}^+ \mathbb{A}_{n,0}^+ - \left[\mathbb{A}_{n,1}^+\right]^2} - \frac{a_k^+}{a_0^+ a_2^+ - \left[a_1^+\right]^2} = o_p(1), \tag{2.8.21}
$$

which implies $P_{2,n,j}^+ - E[P_{2,n,j}^+] = o_p(1)[P_{1,n,j}^+ - E[P_{1,n,j}^+]]$ and then (2.8.5) implies (2.8.14) holds.

Combining (2.8.5) and (2.8.14) together, we have

$$
\frac{\hat{m}_{d,\epsilon}^+ - m(0) - E[P_{n,1}^+]}{\sqrt{Var[P_{1,n,1}^+]/n}} \xrightarrow{d} N(0,1). \tag{2.8.22}
$$

Let $\delta_m^+(x) = [m(x) - m(0)]f_{R^*}(x)\mathbb{1}\{x \geq 0\}$. For the bias term $E[P_{1,n,1}^+]$, we have

$$
\begin{aligned}
E[P_{1,n,1}^+] &= \frac{1}{a_0^+ a_2^+ - \left[a_1^+\right]^2} \left\{ \frac{a_2^+}{2\pi} \int \phi_K(th)\phi_{\delta_m^+}(-t)\,dt - \frac{a_1^+}{2\pi \mathrm{i}} \int \phi_K^{(1)}(th)\phi_{\delta_m^+}(-t)\,dt \right\} \\
&= \frac{1}{a_0^+ a_2^+ - \left[a_1^+\right]^2} \left\{ a_2^+ \int K_h(x)\delta_m^+(x)\,dx - a_1^+ \int K_{h,1}(x)\delta_m^+(x)\,dx \right\} \\
&= \frac{f_{R^*}(0)}{a_0^+ a_2^+ - \left[a_1^+\right]^2} \left\{ f_{R^*}(0) \int_0^{+\infty} x^2 K(x)dx \int_0^{+\infty} K(z)[m(zh) - m(0)]\,dz \right. \\
&\quad - f_{R^*}(0) \int_0^{+\infty} xK(x)dx \int_0^{+\infty} zK(z)[m(zh) - m(0)]\,dz \\
&\quad + \int_0^{+\infty} x^2 K(x)dx \int_0^{+\infty} K(z)[m(zh) - m(0)][f_{R^*}(zh) - f_{R^*}(0)]\,dz \\
&\quad \left. - \int_0^{+\infty} xK(x)dx \int_0^{+\infty} zK(z)[m(zh) - m(0)][f_{R^*}(zh) - f_{R^*}(0)]\,dz \right\} \\
&\equiv \sum_{k=1}^{4} BP_k^+,
\end{aligned}
\tag{2.8.23}
$$

where

$$
\begin{aligned}
BP_1^+ &= \frac{1}{\mu_{K,0}^+ \mu_{K,2}^+ - [\mu_{K,1}^+]^2} \int_0^{+\infty} x^2 K(x)dx \int_0^{+\infty} K(z)[m(zh) - m(0)]dz \\
&= \frac{1}{\mu_{K,0}^+ \mu_{K,2}^+ - [\mu_{K,1}^+]^2} \int_0^{+\infty} x^2 K(x)dx \int_0^{+\infty} K(z) \left[m^{(1)}(0)zh + \frac{1}{2} m^{(2)}(\xi_{zh}^1)(zh)^2 \right] dz,
\end{aligned}
\tag{2.8.24}
$$

and

$$
\begin{aligned}
BP_2 &= -\frac{1}{\mu_{K,0}^+ \mu_{K,2}^+ - [\mu_{K,1}^+]^2} \int_0^{+\infty} xK(x)dx \int_0^{+\infty} zK(z)[m(zh) - m(0)]dz \\
&= \frac{1}{\mu_{K,0}^+ \mu_{K,2}^+ - [\mu_{K,1}^+]^2} \int_0^{+\infty} xK(x)dx \int_0^{+\infty} zK(z) \left[m^{(1)}(0)zh + \frac{1}{2} m^{(2)}(\xi_{zh}^2)(zh)^2 \right] dz.
\end{aligned}
\tag{2.8.25}
$$

Then, it follows

$$
BP_1 + BP_2 \to \frac{(\mu_{K,2}^+)^2 - \mu_{K,1}^+ \mu_{K,3}^+}{2[\mu_{K,0}^+ \mu_{K,2}^+ - (\mu_{K,1}^+)^2]} m^{(2)+}(0)h^2. \tag{2.8.26}
$$

For BP_3^+ and BP_4^+, it is easy to show that they are of order $o(h^2)$. Moreover, (2.8.21) implies $E[P_{2,n,1}^+] = o(E[P_{1,n,1}^+])$, which gives

$$E[P_{n,1}^+] = \frac{(\mu_{K,2}^+)^2 - \mu_{K,1}^+\mu_{K,3}^+}{2[\mu_{K,0}^+\mu_{K,2}^+ - (\mu_{K,1}^+)^2]} m^{(2)+}(0)h^2 + o(h^2). \tag{2.8.27}$$

Results as (2.8.13), (2.8.22), and (2.8.27) can be obtained for $\hat{m}_{d,\epsilon}^-$ following a similar argument, which together with symmetry of K imply

$$\frac{\hat{\theta}_{d,\epsilon} - \theta - \text{Bias}[\hat{\theta}_{d,\epsilon}]}{Var[\hat{\theta}_{d,\epsilon}]} \xrightarrow{d} N(0,1). \tag{2.8.28}$$

■

Proof of Theorem 6: The proof for the case of supersmooth f_ϵ follows a similar route as that for the case of the ordinary smooth case, as in Theorem 5. So I only focus on the differences in the following discussion. First, we update the upper bound results.

In step 1 of the proof for the ordinary smooth case, to verify the Lyapunov condition (2.8.6), parallel to (2.8.8), for $\eta > 0$, we have

$$\begin{aligned}
& E\left|\frac{1}{2\pi i^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-itR_1)[m(R^*)+U-m(0)]D^+\,dt\right|^{2+\eta} \\
\leq\; & \exp(\beta_0(2+\eta)h^{-\beta})h^{-(1+\eta)}\left(\frac{h\exp(-\beta_0 h^{-\beta})}{2\pi}\int \frac{|\phi_K^{(k)}(th)|}{|\phi_\epsilon(-t)|}\,dt\right)^{\eta} \\
& \times h\exp(-2\beta_0 h^{-\beta})\int\left|\frac{1}{2\pi}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp\left(-itx\right)\,dt\right|^2 \mu_{m,2+\eta}^+(x)dx \\
=\; & O\big(\exp(\beta_0(2+\eta)h^{-\beta})h^{-(1+\eta)}\big),
\end{aligned} \tag{2.8.29}$$

where the last equality follows by Lemma 15 and $\sup_x \mu_{m,2+\eta}^+(x) < \infty$.

To verify the Lyapunov condition (2.8.6), besides (2.8.29), we also need the parallel result to (2.8.13). There is, however, no parallel result to Lemma 14 in the case of supersmooth f_ϵ. Therefore, the lower bound of $E\left|P^{+}_{1,n,1}\right|^2$ is commonly derived to verify (2.8.6). Primitive conditions, like Condition 3.1 of Fan and Masry (1992), can be imposed to this end. In this chapter, however, to avoid the unnecessary complication, we directly assume the lower bound of $E\left|P^{+}_{1,n,1}\right|^2$ in Assumption 13. Hence, under Assumption 13, the Lyapunov condition (2.8.6) holds true, and the conclusion follows.

By a similar argument as in (2.8.29), parallel to (2.8.16), we have

$$\begin{aligned} & E\left|\frac{1}{2\pi i^k}\int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp(-itR_1)D_1^{+}\,dt\right|^2 \\ \leq\ & C\int\left|\frac{1}{2\pi i^k}\int\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)}\exp\left(-itx\right)\,dt\right|^2 dx \\ =\ & O\big(\exp(2\beta_0 h^{-\beta})h^{-1}\big), \end{aligned} \tag{2.8.30}$$

where the equality follows by $\|f_X\|_\infty < \infty$ and Lemma 15. Therefore, for the parallel result to (2.8.18), we have

$$\mathbb{A}^{+}_{n,k} - a^{+}_k = O(h) + O_p\left((nh)^{-1/2}\exp(\beta_0 h^{-\beta})\right). \tag{2.8.31}$$

So, if $nh\exp(-2\beta_0 h^{-\beta}) \to \infty$, (2.8.21) holds true. Therefore, the conclusion follows by a similar argument as in Theorem 5.

For $Var[\hat{\theta}_{d,\epsilon}] = o\big(\text{Bias}^2[\hat{\theta}_{d,\epsilon}]\big)$, first note $\text{Bias}[\hat{\theta}_{d,\epsilon}]$ does not depend on error distribution and is of order h^2. Moreover, we note $Var[\hat{\theta}_{d,\epsilon}] \sim E\left|P^{+}_{1,n,1}\right|^2/n$ and (2.8.29) implies

$$E\left|P^{+}_{1,n,1}\right|^2 \leq C\exp(2\beta_0 h^{-\beta})h^{-1}.$$

Then, the conclusion follows by

$$h = (b\beta_0)^{1/\beta}(\log n)^{-1/\beta},$$

where $b > 2$. ∎

Proof of Theorem 7: Following a similar argument to the proof of Theorem 5 and 6, we have $\hat{\theta}_\epsilon - \theta = \hat{m}_\epsilon^+ - \hat{m}_\epsilon^-$, where

$$\hat{m}_\epsilon^\pm = \frac{\mathbb{S}_{n,2}^\pm \mathbb{T}_{n,0,m}^\pm - \mathbb{S}_{n,1}^\pm \mathbb{T}_{n,1,m}^\pm}{\mathbb{S}_{n,2}^\pm \mathbb{S}_{n,0}^\pm - \mathbb{S}_{n,1}^\pm \mathbb{S}_{n,1}^\pm},$$

$$\mathbb{T}_{n,k,m}^\pm = \frac{1}{n}\sum_{j=1}^n \frac{1}{2\pi \mathrm{i}^k} \int \frac{\phi_{K^\pm}^{(k)}(th)\phi_\epsilon(t)}{\left\{|\phi_\epsilon(t)| \vee n^{-\zeta}\right\}^2} \exp(-\mathrm{i}tR_j)[m(R_j^*) + U_j]dt.$$

For convergence rate of $\hat{\theta}_\epsilon$, it is suffice to get separate convergence rate of $\hat{m}_\epsilon^+$ and $\hat{m}_\epsilon^-$. In the following discussion, I focus on $\hat{m}_\epsilon^+$, while the result for $\hat{m}_\epsilon^-$ can be obtained in a similar way.

Let $e_1 = (1,0)'$, $b_{k,m}^+ = m(0) f_{R^*}(0) \int_0^{+\infty} z^k K(z)dz$, and

$$M_b^+ = \begin{pmatrix} b_{0,m}^+ \\ b_{1,m}^+ \end{pmatrix}, \quad M_\mathbb{S}^+ = \begin{pmatrix} \mathbb{S}_{n,0}^+ & \mathbb{S}_{n,1}^+ \\ \mathbb{S}_{n,1}^+ & \mathbb{S}_{n,2}^+ \end{pmatrix}, \quad M_\mathbb{T}^+ = \begin{pmatrix} \mathbb{T}_{n,0,m}^+ \\ \mathbb{T}_{n,1,m}^+ \end{pmatrix}.$$

Then, we have

$$\begin{aligned} \hat{m}_\epsilon^+ - m(0) &= e_1' \left[M_\mathbb{S}^+\right]^{-1} M_\mathbb{T}^+ - m(0) \\ &= e_1' \left(\left[M_\mathbb{S}^+\right]^{-1} M_\mathbb{T}^+ - \left[M_a^+\right]^{-1} M_b^+\right) \\ &\quad + \left(e_1' \left[M_a^+\right]^{-1} M_b^+ - m(0)\right) \\ &= T_{n,1} + T_{n,2}, \end{aligned} \tag{2.8.32}$$

where

$$\begin{aligned} |T_{n,1}|^2 &\le \left\| \left[M_\mathbb{S}^+\right]^{-1}\left[M_\mathbb{T}^+ - M_b^+\right] + \left\{\left[M_\mathbb{S}^+\right]^{-1} - \left[M_a^+\right]^{-1}\right\} M_b^+ \right\|^2 \\ &\le 2\left\| \left[M_\mathbb{S}^+\right]^{-1}\left[M_\mathbb{T}^+ - M_b^+\right]\right\|^2 + 2\left\| \left[M_\mathbb{S}^+\right]^{-1}\left[M_a^+ - M_\mathbb{S}^+\right]\left[M_a^+\right]^{-1} M_b^+\right\|^2 \\ &\le 2\lambda_{\max}^2\left(\left[M_\mathbb{S}^+\right]^{-1}\right)\left\{\left\|M_\mathbb{T}^+ - M_b^+\right\|^2 + \left\|M_\mathbb{S}^+ - M_a^+\right\|^2 \left\|\left[M_a^+\right]^{-1} M_b^+\right\|^2\right\}. \end{aligned} \tag{2.8.33}$$

Notice $\lambda_{\max}\left(\left[M_{\mathbb{S}}^{+}\right]^{-1}\right)=\lambda_{\min}^{-1}\left(M_{\mathbb{S}}^{+}\right)$, and the upper bound of $\lambda_{\max}\left(\left[M_{\mathbb{S}}^{+}\right]^{-1}\right)$ follows by

$$\inf_{|\delta|=1}\delta' M_{\mathbb{S}}^{+}\delta \geq \inf_{|\delta|=1}\delta'\left(M_{\mathbb{S}}^{+}-M_{a}^{+}\right)\delta+\lambda_{\min}\left(M_{a}^{+}\right)$$

and $\lambda_{\min}\left(M_{a}^{+}\right)>0$. By $\|\left[M_{a}^{+}\right]^{-1}M_{b}^{+}\|^{2}<\infty$, which follows by $\lambda_{\min}\left(M_{a}^{+}\right)>0$ and $\|M_{b}^{+}\|<\infty$, for the upper bound of $|T_{n,1}|$, it is suffice to establish upper bounds for $\|M_{\mathbb{S}}^{+}-M_{a}^{+}\|^{2}$ and $\|M_{\mathbb{T}}^{+}-M_{b}^{+}\|^{2}$ respectively.

To find these upper bounds, first, we note $\|M_{\mathbb{S}}^{+}-M_{a}^{+}\|^{2}\leq\|\hat{M}_{a}^{+}-M_{a}^{+}\|_{F}^{2}=\sum_{k,k'=0}^{1}\left|\mathbb{S}_{n,k+k'}^{+}-a_{k+k'}^{+}\right|^{2}$, where $\mathbb{S}_{n,k}^{+}=\frac{1}{n}\sum_{j=1}^{n}Q_{n,k,j}^{s+}$ with $Q_{n,k,j}^{s+}=\frac{1}{2\pi\mathrm{i}^{k}}\int\frac{\phi_{K^{+}}^{(k)}(th)\phi_{\epsilon}(t)}{\{|\phi_{\epsilon}(t)|\vee n^{-\zeta}\}^{2}}\exp(-\mathrm{i}tR_{j})dt$. Also, we note

$$\left|\mathbb{S}_{n,k}^{+}-a_{k}^{+}\right|^{2}\leq 3\left(I_{n,1}^{a}+I_{n,2}^{a}+I_{n,3}^{a}\right), \tag{2.8.34}$$

where

$$\begin{aligned}
I_{n,1}^{a}&=\left|\frac{1}{n}\sum_{j=1}^{n}Q_{n,k,j}^{s+}-E[Q_{n,k,j}^{s+}]\right|^{2},\\
I_{n,2}^{a}&=\left|E[Q_{n,k,j}^{s+}]-\frac{1}{2\pi\mathrm{i}^{k}}\int\phi_{K^{+}}^{(k)}(th)\phi_{R^{*}}(-t)dt\right|^{2},\\
I_{n,3}^{a}&=\left|\frac{1}{2\pi\mathrm{i}^{k}}\int\phi_{K^{+}}^{(k)}(th)\phi_{R^{*}}(-t)dt-a_{k}^{+}\right|^{2}.
\end{aligned}$$

For $I_{n,1}^{a}$, we have

$$\begin{aligned}
E[I_{n,1}^{a}]&=\frac{1}{n}Var[Q_{n,k,j}^{s+}]\\
&\leq\frac{1}{n}E\left|\frac{1}{2\pi\mathrm{i}^{k}}\int\frac{\phi_{K^{+}}^{(k)}(th)\phi_{\epsilon}(t)}{\left\{|\phi_{\epsilon}(t)|\vee n^{-\zeta}\right\}^{2}}\exp(-\mathrm{i}tR_{j})dt\right|^{2}\\
&\leq\frac{1}{2\pi n}\int\frac{|\phi_{K^{+}}^{(k)}(th)|^{2}|\phi_{\epsilon}(t)|^{2}}{\left\{|\phi_{\epsilon}(t)|\vee n^{-\zeta}\right\}^{4}}dt\\
&=O\left(\frac{1}{n}\int\frac{|\phi_{\epsilon}(t)|^{2}}{\left\{|\phi_{\epsilon}(t)|\vee n^{-\zeta}\right\}^{4}}dt\right).
\end{aligned} \tag{2.8.35}$$

Then, (2.8.35) and Lemma 3 imply

$$I_{n,1}^{a}=\begin{cases} O_p\left(n^{\frac{\zeta(2\beta+1)}{\beta}-1}\right), & \text{if } f_\epsilon \text{ is ordinary smooth of order } \beta \\ O_p(n^{4\zeta-1}), & \text{if } f_\epsilon \text{ is supersmooth of order } \beta \end{cases}. \tag{2.8.36}$$

For $I_{n,2}^{a}$, we have

$$\begin{aligned} I_{n,2}^{a} &= \left|\frac{1}{2\pi \mathrm{i}^k}\int \frac{\phi_{K^+}^{(k)}(th)\phi_\epsilon(t)\phi_R(-t)}{|\phi_\epsilon(t)|^2}\left\{\frac{|\phi_\epsilon(t)|^2}{\{|\phi_\epsilon(t)|\vee n^{-\zeta}\}^2}-1\right\}dt\right|^2 \\ &= \left|\frac{1}{2\pi \mathrm{i}^k}\int\limits_{G_{\epsilon,n,\zeta}} \frac{\phi_{K^+}^{(k)}(th)\phi_\epsilon(t)\phi_R(-t)}{|\phi_\epsilon(t)|^2}dt\right|^2 \\ &\le \frac{1}{2\pi}\int\left|\phi_{K^+}^{(k)}(th)\right|^2 dt\,\frac{1}{2\pi}\int\limits_{G_{\epsilon,n,\zeta}}|\phi_{R^*}(t)|^2\,dt, \\ &= O\left(h^{-1}\int\limits_{G_{\epsilon,n,\zeta}}|\phi_{R^*}(t)|^2\,dt\right), \end{aligned} \tag{2.8.37}$$

where the last equality follows by $\frac{1}{2\pi}\int\left|\phi_{K^+}^{(k)}(th)\right|^2 dt = h^{-1}\int_0^{+\infty} z^{2k}K^2(z)dz$. Then Lemma 2 implies

$$I_{n,2}^{a}=\begin{cases} O\left(n^{-\frac{2\alpha\zeta}{\beta}}h^{-1}\right), & \text{if } f_\epsilon \text{ is ordinary smooth of order } \beta \\ O\left((\log n)^{-\frac{2\alpha}{\beta}}h^{-1}\right), & \text{if } f_\epsilon \text{ is supersmooth of order } \beta \end{cases}. \tag{2.8.38}$$

For $I_{n,3}^{a}$, we have

$$\begin{aligned} I_{n,3}^{a} &= \left|\int_0^{+\infty} K_{h,k}(x)f_{R^*}(x)dx - f_{R^*}(0)\int_0^{+\infty} u^k K(u)du\right|^2 \\ &= \left|\int_0^{+\infty} z^k K(z)[f_{R^*}(zh)-f_{R^*}(0)]dz\right|^2 \\ &= O(h^2), \end{aligned} \tag{2.8.39}$$

where the last equality follows by the twice continuously differentiability of f_{R^*} and $\int_0^{+\infty} z^{k+1}K(z)dz < \infty$. Combining (2.8.36), (2.8.38), and (2.8.39) together, we have

$$\|M_{\mathbb{S}}^+ - M_a^+\|^2 = \begin{cases} O_p\left(n^{\frac{\zeta(2\beta+1)}{\beta}-1} + n^{-\frac{2\alpha\zeta}{\beta}}h^{-1} + h^2\right), & \text{if } f_\epsilon \text{ is ordinary smooth of order } \beta \\ O_p\left(n^{4\zeta-1} + (\log n)^{-\frac{2\alpha}{\beta}}h^{-1} + h^2\right), & \text{if } f_\epsilon \text{ is supersmooth of order } \beta \end{cases}. \tag{2.8.40}$$

Following a similar argument, it is easy to show

$$\|M_{\mathbb{T}}^+ - M_b^+\|^2 = \begin{cases} O_p\left(n^{\frac{\zeta(2\beta+1)}{\beta}-1} + n^{-\frac{2\alpha\zeta}{\beta}}h^{-1} + h^2\right), & \text{if } f_\epsilon \text{ is ordinary smooth of order } \beta \\ O_p\left(n^{4\zeta-1} + (\log n)^{-\frac{2\alpha}{\beta}}h^{-1} + h^2\right), & \text{if } f_\epsilon \text{ is supersmooth of order } \beta \end{cases}. \tag{2.8.41}$$

Also we note $|T_{n,2}|^2 = O(h^4) = o(h^2)$. Then by (2.8.40) and (2.8.41), $|T_{n,1}|$ is the dominant term, and we have

$$\|\hat{\theta}_\epsilon - \theta\| = \begin{cases} O_p\left(n^{\frac{\zeta(2\beta+1)}{2\beta}-\frac{1}{2}} + n^{-\frac{\alpha\zeta}{\beta}}h^{-1/2} + h\right), & \text{if } f_\epsilon \text{ is ordinary smooth of order } \beta \\ O_p\left(n^{2\zeta-\frac{1}{2}} + (\log n)^{-\frac{\alpha}{\beta}}h^{-1/2} + h\right), & \text{if } f_\epsilon \text{ is supersmooth of order } \beta \end{cases}.$$

If f_ϵ is ordinary smooth of order β, the trade-off between $n^{-\frac{\alpha\zeta}{\beta}}h^{-1/2}$ and h implies $h = n^{-\frac{2\alpha\zeta}{3\beta}}$, and the trade-off between $n^{\frac{\zeta(2\beta+1)}{2\beta}-\frac{1}{2}}$ and $n^{-\frac{\alpha\zeta}{\beta}}h^{-1/2}$ with $h = n^{-\frac{2\alpha\zeta}{3\beta}}$ implies $\zeta = \frac{3\beta}{4\alpha+6\beta+3}$, which implies

$$\|\hat{\theta}_\epsilon - \theta\| = O_p\left(n^{-\frac{2\alpha}{4\alpha+6\beta+3}}\right).$$

If f_ϵ is supersmooth of order β, consistency of $\hat{\theta}_\epsilon$ implies $h(\log n)^{\frac{2\alpha}{\beta}} \to \infty$ as $n \to \infty$. Then, $\zeta < 1/4$ implies $n^{2\zeta-1/2} = o(h)$. Then, the trade-off between $(\log n)^{-\frac{\alpha}{\beta}}h^{-1/2}$ and h implies $h = (\log n)^{-\frac{2\alpha}{3\beta}}$, which implies

$$\|\hat{\theta}_\epsilon - \theta\| = O_p\left((\log n)^{-\frac{2\alpha}{3\beta}}\right).$$

■

2.8.2 Proofs of Lemmas

To establish (2.3.4), we introduce the following properties of Fourier transform.

Lemma 12. *If $f(x), x^k f(x) \in L_1(\mathbb{R})$ for a positive integer k, $\phi_f^{(k)} = \mathrm{i}^k \int x^k f(x) e^{\mathrm{i}tx} dx$.*

Proof of Lemma 12: Use Leibniz rule. ■

Lemma 13. *Suppose Assumption 11 and 12 hold true. There exist a function $\psi \in L_1(\mathbb{R})$ such that*

$$\sup_n h^\beta \frac{|\phi_K^{(k)}(s)|}{\left|\phi_\epsilon(-s/h)\right|} \leq \psi_k(s),$$

which implies that there exist constants $c > 0$ such that

$$h^{\beta+1} \int \frac{|\phi_K^{(k)}(th)|}{|\phi_\epsilon(-t)|} dt \leq c.$$

Proof of Lemma 13: Since $\lim_{|t|\to\infty} |t|^\beta \left|\phi_\epsilon(-t)\right| = c_\epsilon$, there exists a constant c_F such that $|t|^\beta \left|\phi_\epsilon(-t)\right| > c_\epsilon/2$ for all $t \geq c_F$. Then, for constants $c_1 > 0$ such that $c_1 > h^\beta$ and $c_2 > 0$ such that $c_2 > c_F h$ for all $n = 1, 2, \ldots$, we have

$$\begin{aligned}
h^\beta \frac{|\phi_K^{(k)}(s)|}{\left|\phi_\epsilon(-s/h)\right|} &\leq h^\beta \frac{\max_{|s|\leq c_F h} |\phi_K^{(k)}(s)|}{\min_{|s|\leq c_F} |f_\epsilon^{\mathrm{ft}}(s)|} \mathbb{1}\{|s| \leq c_F h\} \\
&\quad + \frac{|\phi_K^{(k)}(s)||s|^\beta}{(|s|/h)^\beta |\phi_\epsilon(-s/h)|} \mathbb{1}\{|s| > c_F h\} \\
&\leq c_1 c_{\mathrm{os},0}^{-1} (1 + c_F)^\beta \|\phi_K^{(k)}\|_\infty \mathbb{1}\{|s| \leq c_2\} \\
&\quad + \frac{2|\phi_K^{(k)}(s)||s|^\beta}{c_\epsilon} \\
&\equiv \psi_k(s),
\end{aligned}$$

where the integrability of $\psi(s)$ follows by $\|\phi_K^{(k)}\|_\infty < \infty$, the ordinary smoothness of f_ϵ, and $\int |\phi_K^{(k)}(s)||s|^\beta ds < \infty$. And the second statement immediately follows by the change of variable $t = s/h$. ■

Lemma 14. *Suppose f is continuous at 0, ϕ_ϵ does not vanish everywhere, $\|\phi_\epsilon\|_\infty < \infty$, $\|\phi_\epsilon^{(1)}\|_\infty < \infty$, $|s|^\beta \phi_\epsilon(s) \to c_\epsilon$, and $|s|^{\beta+1}\phi_\epsilon^{(1)}(s) \to -\beta c_\epsilon$. For a pair of positive integers k and k', $\|\phi_K^{(k)}\|_\infty < \infty$, $\|\phi_K^{(k')}\|_\infty < \infty$, $\int |s|^\beta |\phi_K^{(k)}(s)|ds < \infty$, and $\int |s|^\beta |\phi_K^{(k')}(s)|ds < \infty$.*

$$\begin{aligned}\lim_{n\to\infty} h^{2\beta+1} \int \left(\frac{1}{4\pi^2 i^{k-k'}} \int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)} \exp(-\mathrm{i}tx)dt \int \frac{\phi_K^{(k')}(-th)}{\phi_\epsilon(t)} \exp(\mathrm{i}tx)dt \right) f(x)dx \\ = \frac{f(0)}{2\pi c_\epsilon^2 i^{k-k'}} \int |s|^{2\beta} \phi_K^{(k)}(s)\phi_K^{(k')}(s)ds\end{aligned}$$

Proof of Lemma 14: First, we note

$$\begin{aligned}
&\lim_{n\to\infty} \frac{h^\beta}{2\pi \mathrm{i}^k} \int \frac{\phi_K^{(k)}(s)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}sx)ds \\
=& \lim_{n\to\infty} \frac{1}{2\pi \mathrm{i}^k} \int \frac{\phi_K^{(k)}(s)|s|^\beta}{(|s|/h)^\beta \phi_\epsilon(-s/h)} \exp\left(-\mathrm{i}sx\right) ds \\
=& \frac{1}{2\pi \mathrm{i}^k} \int \left\{ \lim_{n\to\infty} \frac{\phi_K^{(k)}(s)|s|^\beta}{(|s|/h)^\beta \phi_\epsilon(-s/h)} \mathbb{1}\{|s| > Mh\} \right\} \exp\left(-\mathrm{i}sx\right) ds \\
=& \frac{1}{2\pi \mathrm{i}^k c_\epsilon} \int \phi_K^{(k)}(s)|s|^\beta \exp(-\mathrm{i}sx)ds,
\end{aligned} \tag{2.8.42}$$

where the second equality follows by the dominant convergence theorem and Lemma 13.

And it follows

$$\begin{aligned}
&\frac{h^{2\beta}}{4\pi^2 i^{k-k'}} \int \frac{\phi_K^{(k)}(s/h)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}tx)dt \int \frac{\phi_K^{(k')}(-s)}{\phi_\epsilon(s/h)} \exp(\mathrm{i}tx)dt \\
\to& \frac{1}{4\pi^2 \mathrm{i}^{k-k'} c_\epsilon^2} \int \phi_K^{(k)}(s)|s|^\beta \exp(-\mathrm{i}sx)ds \int \phi_K^{(k')}(-s)|s|^\beta \exp(\mathrm{i}sx)ds
\end{aligned} \tag{2.8.43}$$

Moreover, by the integration by parts, we have

$$\begin{aligned}
\int \frac{\phi_K^{(k)}(s)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}sx)ds =& \frac{1}{\mathrm{i}x} \int \frac{\phi_K^{(k+1)}(s)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}sx)ds \\
&+ \frac{1}{\mathrm{i}xh} \int \frac{\phi_K^{(k)}(s)\phi_\epsilon^{(1)}(-s/h)}{\phi_\epsilon^2(-s/h)} \exp(-\mathrm{i}sx)ds.
\end{aligned} \tag{2.8.44}$$

Since $|s|^{\beta}\big|\phi_\epsilon(s)\big| \to c_\epsilon$ and $|s|^{\beta+1}\big|\phi_\epsilon^{(1)}(s)\big| \to \beta c_\epsilon$ as $s \to \infty$, there exists an constant $c_F > 0$ be a constant such that $|s|^{\beta}\big|\phi_\epsilon(s)\big| > c_\epsilon/2$ and $|s|^{\beta+1}\big|\phi_\epsilon^{(1)}(s)\big| < 5\beta c_\epsilon/4$ for any s satisfying $|s| > c_F$. Then, we have

$$\begin{aligned}
&\left|\frac{1}{ix}\int \frac{\phi_K^{(k+1)}(s)}{\phi_\epsilon(-s/h)}\exp(-\mathrm{i}sx)ds\right| \\
\leq\ & \frac{1}{|x|}\int \frac{|\phi_K^{(k+1)}(s)|}{|\phi_\epsilon(-s/h)|}ds \\
\leq\ & \frac{h}{|x|}\left(\frac{2M\max_{|s|\leq Mh}|\phi_K^{(k+1)}(s)|}{\min_{|s|\leq M}|\phi_\epsilon(s)|}\right) + \frac{h^{-\beta}}{|x|}\int_{|s|>Mh}\frac{|\phi_K^{(k+1)}(s)||s|^{\beta}}{(|s|/h|)^{\beta}|\phi_\epsilon(-s/h)|}ds \qquad (2.8.45)\\
\leq\ & \frac{h}{|x|}\left(\frac{2M\max_{|s|\leq M}|\phi_K^{(k+1)}(s)|}{\min_{|s|\leq M}|\phi_\epsilon(s)|}\right) + \frac{h^{-\beta}}{|x|}\left(\frac{2}{c_\epsilon}\right)\int |\phi_K^{(k+1)}(s)||s|^{\beta}ds \\
=\ & O\big(h^{-\beta}|x|^{-1}\big),
\end{aligned}$$

and

$$\begin{aligned}
&\left|\frac{1}{\mathrm{i}xh}\int \frac{\phi_K^{(k)}(s)\phi_\epsilon^{(1)}(-s/h)}{\phi_\epsilon^2(-s/h)}\right| \\
\leq\ & \frac{h^{-1}}{|x|}\int \frac{|\phi_K^{(k)}(s)||\phi_\epsilon^{(1)}(-s/h)|}{|\phi_\epsilon(-s/h)|^2}ds \\
\leq\ & \frac{1}{|x|}\left(\frac{2M\max_{|s|\leq Mh}|\phi_K^{(k)}(s)|\max_{|s|\leq M}|\phi_\epsilon^{(1)}(s)|}{\min_{|s|\leq M}|\phi_\epsilon(s)|}\right) \\
&+\frac{h^{-\beta}}{|x|}\int_{|s|>Mh}\frac{|\phi_K^{(k)}(s)||s|^{\beta-1}(|s|/h|)^{\beta+1}|\phi_\epsilon^{(1)}(-s/h)|}{(|s|/h|)^{2\beta}|\phi_\epsilon(-s/h)|^2}ds \qquad (2.8.46)\\
\leq\ & \frac{h}{|x|}\left(\frac{2M\max_{|s|\leq M}|\phi_K^{(k+1)}(s)|\max_{|s|\leq M}|\phi_\epsilon^{(1)}(s)|}{\min_{|s|\leq M}|\phi_\epsilon(s)|}\right) \\
&+\frac{h^{-\beta}}{|x|}\left(\frac{5\beta}{c_\epsilon}\right)\int |\phi_K^{(k)}(s)||s|^{\beta-1}ds \\
=\ & O\big(h^{-\beta}|x|^{-1}\big).
\end{aligned}$$

Thus, Lemma 13, (2.8.44), (2.8.46), and (2.8.45) imply that there are a pair of constants $c_1, c_2 > 0$ such that

$$\sup_n h^{2\beta} \left| \frac{1}{4\pi^2 i^{k-k'}} \int \frac{\phi_K^{(k)}(s/h)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}tx)dt \int \frac{\phi_K^{(k')}(-s)}{\phi_\epsilon(s/h)} \exp(\mathrm{i}tx)dt \right| \leq \min\{c_1, c_2|x|^{-2}\} \tag{2.8.47}$$

Therefore, the conclusion follows by

$$\begin{aligned}
&\lim_{n\to\infty} h^{2\beta+1} \int \left\{ \frac{1}{2\pi \mathrm{i}^k} \int \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(-t)} \exp(-\mathrm{i}tx)dt \right\} \left\{ \frac{1}{2\pi i^{-k'}} \int \frac{\phi_K^{(k')}(-th)}{\phi_\epsilon(t)} \exp(\mathrm{i}tx)dt \right\} f(x)dx \\
=& \lim_{n\to\infty} h^{2\beta} \int \frac{1}{4\pi^2 \mathrm{i}^{k-k'}} \int \frac{\phi_K^{(k)}(s)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}sz)ds \int \frac{\phi_K^{(k')}(-s)}{\phi_\epsilon(s/h)} \exp(\mathrm{i}sz)ds f(zh)dz \\
=& \frac{f(0)}{4\pi^2 c_\epsilon^2 \mathrm{i}^{k-k'}} \int \left\{ \int |s|^\beta \phi_K^{(k)}(s) \exp(-\mathrm{i}sz)ds \right\} \left\{ \int |s|^\beta \phi_K^{(k')}(-s) \exp(\mathrm{i}sz)ds \right\} dz \\
=& \frac{f(0)}{2\pi c_\epsilon^2 \mathrm{i}^{k-k'}} \int |s|^{2\beta} \phi_K^{(k)}(s) \phi_K^{(k')}(s)ds,
\end{aligned} \tag{2.8.48}$$

where the first equality follows by the change of variable $s = th$, the second equality follows by Lemma 6 with $K_n(x) = \frac{h^{2\beta}}{4\pi^2 \mathrm{i}^{k-k'}} \int \frac{\phi_K^{(k)}(s)}{\phi_\epsilon(-s/h)} \exp(-\mathrm{i}sx)ds \int \frac{\phi_K^{(k')}(-s)}{\phi_\epsilon(s/h)} \exp(\mathrm{i}sx)ds$ and $K^*(x) = \min\{c_1, c_2|x|^{-2}\}$, and the third equality follows by Lemma 1 (1). ■

Lemma 15. *Suppose Assumption 11 and 13 hold true. There exists a constant $c > 0$ such that*

$$h \exp\left(-\beta_0 h^{-\beta}\right) \int \frac{|\phi_K^{(k)}(th)|}{|\phi_\epsilon(t)|} dt \leq c,$$

$$h \exp\left(-2\beta_0 h^{-\beta}\right) \int_x \left| \int_t \exp\left(-\mathrm{i}t(x^* - x)\right) \frac{\phi_K^{(k)}(th)}{\phi_\epsilon(t)} dt \right|^2 dx \leq c.$$

Proof of Lemma 15: The first statement follows by

$$\begin{aligned}
\int \frac{\left|\phi_K^{(k)}(th)\right|}{\left|\phi_\epsilon(t)\right|}dt &= h^{-1}\int \frac{\left|\phi_K^{(k)}(s)\right|}{\left|\phi_\epsilon(s/h)\right|}ds \\
&\leq c_{\mathrm{ss},0}^{-1}h^{-1}\int\limits_{|s|\leq 1} \left|\phi_K^{(k)}(s)\right| \exp\left(\beta_0(|s|/h)^{\beta}\right)ds \\
&= O\left(h^{-1}e^{\beta_0 h^{-\beta}}\right),
\end{aligned}$$

where the first equality follows by the change of variable $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that $\phi_K^{(k)}$ is supported on $[-1, 1]$, and the last equality uses $\|\phi_K^{(k)}\|_\infty < \infty$.

The second statement follows by

$$\begin{aligned}
\int\limits_x \left|\int\limits_t \exp\left(-\mathrm{i}t(x^* - x)\right)\frac{\phi_K^{(k)}(th)}{\phi_\epsilon(t)}dt\right|^2 dx &= 2\pi\int \frac{\left|\phi_K^{(k)}(th)\right|^2}{\left|\phi_\epsilon(t)\right|^2}dt \\
&= 2\pi h^{-1}\int \frac{\left|\phi_K^{(k)}(s)\right|^2}{\left|\phi_\epsilon(s/h)\right|^2}ds \\
&\leq 2\pi c_{\mathrm{ss},0}^{-2}h^{-1}\int\limits_{|s|\leq 1} \left|\phi_K^{(k)}(s)\right|^2 \exp\left(2\beta_0(|s|/h)^{\beta}\right)ds \\
&= O\left(h^{-1}\exp\left(2\beta_0 h^{-\beta}\right)\right),
\end{aligned}$$

where the first equality follows by Lemma 1 (1), the second equality follows by the change of variable $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that $\phi_K^{(k)}$ is supported on $[-1, 1]$, and the last equality uses $\|\phi_K^{(k)}\|_\infty < \infty$. ■

Chapter 3

Estimating Spillovers under Factor-induced Constraints

3.1 Introduction

The spillover effect, also known as the externality, is the effect of a treatment assigned to one individual of a group on the outcomes of other members of the same group. It presents in almost all contexts with a strategic setup in economics, examples include technological adoptions, the peer effects of education, and many others. Ignoring the spillover effects could cause severe bias in the estimation of the treatment effect. Besides, the structure of the social interactions that drive the spillovers might also be of its own interest.

In the estimation of spillover effects, one of the challenges comes from the relative large number of pair-specific parameters compared to the limited sample size. In such a situation, a panel data structure can relieve the problem because the variations across the additional time dimension can be used as one source of the knowledge on the spillover structure. However, given the prevalence of the short panel, which is the case when there are more cross-sectional units than the number of periods, the spillover effects could still be under-identified even with a panel dataset. Therefore, as pointed out in Blume et al. (2015), prior knowledge on the spillover structure is necessary to study the spillovers.

Most existing works focus on the case when the spillover structure is observed; see De Paula (2016) for the review. There is a recent wave of researches focusing on the case when the spillover structure is unobserved but known to be sparse, for example Manresa (2016) and Lam and Souza (2013) among others. Specifically, they require each individual only has a relatively limited number of connections relative to the size of the population. Compared to the observed spillover structure assumption, the sparsity assumption is more restrictive on the number of spillovers but less restrictive on the identity and intensity. Under the sparsity assumption, certain penalized regression method, such as LASSO by Tibshirani (1996) and the adaptive LASSO by Zou (2006), can then be used.

To the best of our knowledge, the prior knowledge about the spillovers considered in the exsiting literature are all direct knowledge. In this chapter, in advance of the existing works, we are interested in the case when there is a factor structure under the variables that generate the spillovers. For example, when estimating the technological spillovers from the R&D investments to the productivities of firms in the same market, the variation of the R&D investments of different firms may be driven by the same macro factors of the market. And we show that such a factor structure is in fact an indirect knowledge on the spillover structure. In particular, it implies the linear constraints on the parameters characterizing the spillovers. Therefore, we can improve the performance of existing estimators by adding these factor-induced constraints.

The remainder of this chapter is organized as follows. Section 3.2 introduces the model used to characterize the spillovers and the factor structure, and provides the intuition on how the factor structure could be treated as a knowledge about the spillover structure. In Section 3.3 discusses the general construction of the estimator based on the factor-induced constraints. Section 3.4 derives the properties of the proposed estimator, Section 3.5 investigates the finite sample performance, and Section 3.6 concludes this chapter.

3.2 Spillovers and Factor-induced Constraints

In this chapter, we study the estimation of the linear regression model

$$y_t \quad = \quad \beta' x_t + \gamma' z_t + \epsilon_t, \tag{3.2.1}$$

where $t = 1, \cdots, T$, $x_t = (x_{1t}, \cdots, x_{Nt})'$ is a N-dimensional vector of the variables generating the spillovers, and $z_t = (z_{1t}, \cdots, z_{kt})$ is a k-dimensional vector of the additional controls. In particular, we focus on the case when $N > T$, which makes (3.2.1) a high-dimensional problem in the sense that the number of unknown parameters is larger than the number of the observations. For the estimation of spillover effects using a panel dataset, β in (3.2.1) can be interpreted as capturing the spillovers from a specific individual, that is $(y_t, x_t, z_t, \epsilon_t)$ and (β, γ) in (3.2.1) should all be indexed by i when using a panel dataset. Throughout this chapter, instead of the general panel data setup, the discussion will focus on (3.2.1) for simplicity.

For the factor structure of x_t, we consider the model as follows.

$$x_t = A f_t + u_t, \tag{3.2.2}$$

where A is a $N \times r$ matrix of the factor loadings, f_t is a r-dimensional vector of the factors, and u_t is a N-dimensional vector of the idiosyncratic shocks. To avoid the potential bias caused by omitting relevant factors, I treat all factors f_t as latent in this chapter. In case when f_t is partly observed, x_t then could be treated as the residuals from the regressions on the observed factors.

To understand how (3.2.2) could be treated as the indirect knowledge about β, note that substituting (3.2.2) into (3.2.1) gives a reduced form as follows.

$$y_t = \psi' f_t + \gamma' z_t + e_t, \tag{3.2.3}$$

where $\psi = A'\beta$ and $e_t = \beta' u_t + \epsilon_t$. If A and ψ were both observed, $\psi = A'\beta$ are the linear constraints on β, which are equally informative as those direct knowledge on β. Even though A and ψ cannot be directly observed in practice, both of them can be estimated with a minor cost, and the constraints can then be established based on the estimates of A and ψ.

3.3 Estimation

In this section, we discuss the construction of the spillover estimator making use of the latent factor structure. Let $\hat{A}$, $\hat{\psi}$, and $\hat{\gamma}$ separately denote the estimate of A, ψ, and γ. For the expository purpose, we first treat them as if they had been defined in Section 3.3.1, and the details of their construction are discussed later in Section 3.3.2.

3.3.1 LASSO under Factor-induced Constraints

If the spillover structure is sparse, the penalized regression methods can be used to estimate β in (3.2.1) when $N > T$. One of the most popular among these methods is the LASSO, which places an L_1 penalty on β. Specifically, let $\tilde{\beta}$ denote the LASSO estimator of β, which is defined as the solution to the following minimization problem.

$$\min_b \|Y - Xb - Z\hat{\gamma}\|_2^2/2 + \lambda\|\hat{D}b\|_1, \tag{3.3.1}$$

where $Y = (y_1, \cdots, y_T)'$, $X = (x_1, \cdots, x_T)'$, $Z = (z_1, \cdots, z_T)'$, and $\hat{D}$ is a diagonal matrix introduced to normalize X. Compared with other penalized regression methods, the LASSO is featured by its ability of shrinking certain regression coefficients to exact zero.

In the presence of the latent factor structure (3.2.2) in x_t, since the factor structure implies linear constraints on β, we adjust the LASSO estimator by adding the factor-induced constraints. The constrained LASSO estimator, denoted as $\hat{\beta}$, is then defined as the solution to the following minimization problem.

$$\min_b \|Y - Xb - Z\hat{\gamma}\|_2^2/2 + \lambda\|\hat{D}b\|_1 \quad \text{s.t.} \quad \hat{\psi} = \hat{A}'b. \tag{3.3.2}$$

In the situation when the true parameters satisfy the constraints, James et al. (2012) shows that the constrained LASSO outperform the unconstrained one in the sense that it has a sharper error bound. This implies that the infeasible constrained LASSO when A and ψ were observed, denoted as $\check{\beta}$, outperform the unconstrained LASSO, denoted as $\tilde{\beta}$. In our setting, even though the true value of β may not exactly satisfies the constraints in (3.3.2) due to the estimation error of $\hat{A}$ and $\hat{\psi}$, the gap would decrease as the sample size grows. In particular, if $\hat{A}$ and $\hat{\psi}$ converge sufficiently fast, we show that the error bound of $\hat{\beta}$ is close to that of $\check{\beta}$, which is strictly sharper than that of $\tilde{\beta}$.

3.3.2 Construction of Constraints

In the rest of this section, we provide details on $\hat{A}$, $\hat{\psi}$, and $\hat{\gamma}$, which are required in the construction of the constrained LASSO (3.3.2). Specifically, we will separately discuss the estimation of the factor model (3.2.2), which gives $\hat{A}$, and the corresponding factor-augmented regression, which gives $\hat{\psi}$ and $\hat{\gamma}$.

The estimation methods of the factor model (3.2.2), which is known as the large factor model due to the fact that $N > T$, are well documented in the literature. Early attempts include the principle components method (PC) in Bai and Ng (2002) and the generalized principle components method (GPC) in Choi (2012) using estimated covariance matrix suggested by Bai and Liao (2013) as weights. If data exhibits nonzero serial correlation, Lam et al. (2011) develops a new approach based on the information from the autocovariance matrix at non-zero lags, instead of the covariance matrix as in PC and GPC. Compared with PC and GPC, Lam et al. (2011) has better performance when there exist strong cross-correlation in the idiosyncratic shocks. In this chapter, we focus on the situation where x_t has non-zero serial correlation, which makes Lam et al. (2011) an automatic choice for the estimation of the factor model (3.2.2).

In particular, when the number of factors r is known, Lam et al. (2011) suggests the estimators of A and f_t as follows.

$$\hat{A} = (\hat{s}_1, \cdots, \hat{s}_r), \text{ and } \hat{f}_t = \hat{A}' x_t, \tag{3.3.3}$$

where $(\hat{s}_1, \cdots, \hat{s}_r)$ be the orthonormal eigenvectors of $\hat{H}_x$ corresponding to its largest r eigenvalues and

$$\hat{H}_x = \sum_{k=1}^{k_0} \widehat{\Sigma}_x(k) \widehat{\Sigma}_x(k)', \widehat{\Sigma}_x(k) = \frac{1}{T} \sum_{j=1}^{T-k} (x_{t+j} - \bar{x})(x_t - \bar{x})', \tag{3.3.4}$$

with $k_0 \geq 1$ be a prespecified integer and $\bar{x} = \frac{1}{T} \sum_{t=1}^{T} x_t$.

Due to the rotational indeterminacy of the factor model [1], instead of the estimator of a specific loading matrix, $\hat{A}$ is an estimator for the factor loading space $\mathcal{M}(A)$, the r-dimensional linear space spanned by the columns of A. However, it is worthy to note that the factor-induced constraints used in (3.3.2) is independent of the choice of A and f_t, in that different choices of A and f_t lead to equivalent $\psi = A'\beta$.

In practice, we need to first estimate the number of factors r before implementing Lam et al. (2011). There are two main types of existing methods to estimate r. One is based on information criteria, for example Bai and Ng (2002). The other is based on the distribution of eigenvalues, for examples Onatski (2009), Lam and Yao (2012), and Ahn and Horenstein (2013). In this chapter, following Lam and Yao (2012), we estimate the number of factor r by the relative magnitude of ratios of eigenvalues, which is defined as follows.

$$\hat{r} = \arg\min_{1\le j\le R} \frac{\hat{\lambda}_{j+1}}{\hat{\lambda}_j}, \tag{3.3.5}$$

where $r < R < N$ and $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_N$ are the decreasingly ordered eigenvalues of $\hat{H}_x$.

To estimate ψ and γ, we consider the factor-augmented regression as follows.

$$y_t \quad = \quad \psi'\hat{f}_t + \gamma' z_t + e_t. \tag{3.3.6}$$

Notice that the number of unknown parameters in (3.3.6) is $r + k < T$. $\hat{\psi}$ and $\hat{\gamma}$ can thus be defined as the solution to minimization problem (3.3.7) with respect to b_1 and b_2 as follows

$$\min_{b_1, b_2} \sum_{t=1}^{T} (y_t - b_1'\hat{f}_t - b_2' z_t)^2. \tag{3.3.7}$$

3.4 Theoretical Results

This section discusses the theoretical properties of the proposed estimator. To facilitate the discussion, we introduce the following notations. For a d-dimensional vector $v = (v_1, \cdots, v_d)$, let $\|v\|_p$ be its L_p-norm, and $\text{supp}(v) = \{j : v_j \neq 0\}$. Given a set

[1] $Af_t = AHH'f_t$ for any orthnormal matrix $H \in \mathbb{R}^{r\times r}$.

of indices $I \subset \{1, \cdots, d\}$, let v_I be the d-dimensional vector whose jth coordiante $(v_I)_j = v_j \mathbb{1}\{j \in I\}$ for $j = 1, \ldots, d$. For a $d_1 \times d_2$ matrix W, let $\|W\|_2 = \sqrt{\lambda_{\max}(W'W)}$ be its spectral norm and $\|W\|_{\min}$ be the square root of the smallest non-zero eigenvalue of WW', where $\lambda_{\max}$ and $\lambda_{\min}$ separately denote the largest the smallest eigenvalue of the matrix. Given a set of indices $I_2 \subset \{1, \cdots, d_2\}$, let W_{I_2} be the rows of W associated with I_2, and $W_{I_2^c}$ be the remaining rows of W. We use the notation $a \asymp b$ to denote the situation when $a = O(b)$ and $b = O(a)$ hold simultaneously. Also, to keep things clean, we simplify (3.2.1) through this section by setting $z_t = 1$ for $t = 1, \ldots, T$, and the general results for (3.2.1) could be obtained by a very similar argument.

The construction of $\hat{\beta}$ and its performance critically depend on the accuracy of $\hat{A}$, $\hat{\psi}$, and $\hat{\gamma}$. Intuitively, a more accurate $\hat{\gamma}$ would reduce the noise of the regression and increase the precision of the estimates. Moreover, the closer $\hat{A}$ and $\hat{\psi}$ are to A and ψ, the closer the constraints in (3.3.2) would be to the infeasible constraints $\psi = A'b$, under which the infeasible constrained LASSO estimator $\check{\beta}$ has sharper error bound than the unconstrained LASSO estimator $\tilde{\beta}$. To quantify the accuracy, we first study the convergence rate of $\hat{A}$, $\hat{\psi}$, $\hat{\gamma}$ separately.

The convergence rate of $\hat{A}$ is shown in Theorem 1 and 2 of Lam et al. (2011). We thus investigate the convergence rate for $\hat{\psi}$ and $\hat{\gamma}$. To this end, we impose assumptions as follows.

Assumption 15.

(1) $A'A = I_r$, f_t *is weakly stationary,* u_t *is a white noise with zero mean and variance* Σ_u, *and* $\operatorname{cov}(f_t, u_s) = 0$ *for* $t \leq s$.

(2) $\|\Sigma_f(k)\|_2 \asymp N^{1-\nu} \asymp \|\Sigma_f(k)\|_{\min}$ *and* $\|\Sigma_{fu}(k)\|_2 = o\left(N^{1-\nu}\right)$ *for some* $\nu \in [0,1]$ *and* $k = 0, 1, \cdots, k_0$, *where* $\Sigma_f(k) = \operatorname{cov}[f_{t+k}, f_t]$ *and* $\Sigma_{fu}(k) = \operatorname{cov}[f_{t+k}, u_t]$*; Moreover,* $\|\Sigma_{fu}(0)\|_2 = o\left(N^{1-v}\right)$ *and* $\lambda_{\min}\left(A\Sigma_{fu}(0)\right) = o\left(N^{1-v}\right)$.

(3) $\{z_t', \epsilon_t'\}$ *is a stationary* α*-mixing process with* $E\|(z_t', \epsilon_t')\|^{2+\gamma} < \infty$ *elementwisely for some* $\gamma > 0$, *and the mixing coefficients* $\alpha(t)$ *satisfying* $\sum_{t\geq 1}^{\infty} \alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty$.

(4) $\|\Sigma_{f\epsilon}\|_2 = O\left(N^{\frac{1}{2}} T^{-\frac{1}{2}}\right)$ *and* $\|\Sigma_{u\epsilon}\|_2 = O\left(N^{\frac{1}{2}} T^{-\frac{1}{2}}\right)$, *where* $\Sigma_{f\epsilon} = cov[f_t, \epsilon_{it}]$ *and* $\Sigma_{u\epsilon} = cov[u_t, \epsilon_{it}]$.

Assumption 15 (1) contains a normalization of the factor loadings A. Due to the rotational indeterminacy of factor model (3.2.2), even under this normalization, only the linear space spanned by the columns of A, denoted as $\mathcal{M}(A)$, is identified instead of the original A [2]. We can then choose any $N \times r$ matrix $\tilde{A}$ such that $\mathcal{M}(\tilde{A}) = \mathcal{M}(A)$ as the target of estimation. Following Lam et al. (2011), we specify $\tilde{A}$ as QV, where Q coming from the thin Q-R decomposition of A is a $N \times r$ matrix satisfying $Q'Q = I_r$, and V is a r-dimensional orthonormal matrix that comes from the spectral decomposition $\sum_{k=1}^{k_0}\{\Sigma_f(k)Q' + \Sigma_{f,u}(k)\}\{\Sigma_f(k)Q' + \Sigma_{f,u}(k)\}' = VDV'$ [3]. Once we specify the factor loading matrix as above, the corresponding factor process is uniquely defined as $V'Rf_t$. To distinguish the objects based on the original factors from those based on the chosen factors, we index the objects based on the original factors by "o" in the following discussion.

Assumption 15 (2) follows by $\|\Sigma^o_f(k)\|_2 \asymp \|\Sigma^o_f(k)\|_{\min} \asymp 1$, $\Sigma^o_{f,u}(k) = O(1)$ elementwise, and $\|a_i\|_2^2 = N^{1-\nu}$ for $i = 1, \cdots, r$ and $0 \leq \nu \leq 1$, where a_i is the ith column of A; see Lemma 1 of Lam et al. (2011). Assumption 15 (3) is prepared to capture the upper bounds of $\|\hat{\Sigma}^o_f(k) - \Sigma^o_f(k)\|_2$ and $\|\hat{\Sigma}^o_{f,u}(k) - \Sigma^o_{f,u}(k)\|_2$, which follows by using the Frobenius norm as upper bound of the spectral norm, and using the central limit theorem of α-mixing process elementwisely, for example Theorem 0 of Bradley (1985). Since we want to discuss the otherwise normal linear regression model, cutting off the correlation between z_t and ϵ is necessary, and this is specified in Assumption 15 (4).

Under Assumption 15, the convergence rate of $\hat{A}$ in spectral norm is provided by Lam et al. (2011), and is stated in Proposition 3 as follows.

Proposition 3. *Under Assumption 15,* $\|\hat{A} - A\|_2 = O_p(N^\nu T^{-1/2})$.

The linear regression with the estimated factors such as (3.3.6) has been studied by Stock and Watson (2002) and Bai and Ng (2006). Both works rely on the factors estimated by the variance-covariance based method such as Bai and Ng (2002). In this chapter, since we use a different approach to estimate the factor model (3.2.2), it is then necessary to investigate the property of $\hat{\psi}$. Specifically, under Assumption 15, the

[2] It is worth noting that even though $A'A = I_r$ is restrictive, it still cannot pin down A because $A'H'HA = I_r$ for any $r \times r$ orthonormal matrix H.

[3] Since the spectral decomposition does not pinned down the sign of V, we need to allow $\hat{A}$ to adjust sign to adept to the sign of chosen V

convergence rate of $\hat{\psi}$ is given by Theorem 8 below. The proof of Theorem 8 is left to Section 3.7.1.

Theorem 8. *Under Assumption 15, if* $N^{\nu}T^{-1/2} = o(1)$*, we have*

$$\|\hat{\psi} - \psi\|_2 = O_p(N^{\nu}T^{-\frac{1}{2}}).$$

Theorem 8 provides the general result on the convergence rate of $\hat{\psi}$ when both N and T go to infinity. According to Theorem 8, the convergence rate of $\hat{\psi}$ depends on the strength of factors ν. If $\nu = 0$, which is the case when the factors are strong, $\hat{\psi}$ would converge in the root-T rate, which is the same as if f_t is directly observed. If $\nu > 0$, which is the case when the factors are weak, the convergence rate of $\hat{\psi}$ would be slow down by N^{ν}. In this sense, the behavior of $\hat{\psi}$ is very similar to that of $\hat{A}$.

To derive the error bound of $\hat{\beta}$, we need further assumptions. The key is to regularize the Gram matrix $M_x = X'X/T$. M_x is singular when $N > T$, and it is impossible to require its smallest eigenvalue is bounded off zero. In such situation, instead of the standard eigenvalue, following the literature of the penalized regression, we impose the lower bound assumption on the restricted eigenvalue of M_x. To characterize the restricted eigenvalue, let $\mathcal{T}$ be a subset of $\{1, \cdots, N\}$ such that $A'_{\mathcal{T}}$ is invertible, $\mathcal{T}^{c} = \{1, \cdots, N\} \backslash \mathcal{T}$, and $\mathcal{S} = \text{supp}(\beta_{\mathcal{T}^c})$. Then, the restricted eigenvalue of M_x is defined in (3.4.1) as follows.

$$\kappa_{\zeta}(M_x) = \min_{\delta \in \Delta_{\zeta}} \frac{\delta' M_x \delta}{\|\delta\|_2^2}, \tag{3.4.1}$$

where

$$\Delta_{\zeta} = \left\{ \delta \in \mathbb{R}^N : \|\delta_{\mathcal{S}^c \cap \mathcal{T}^c}\|_1 \leq \frac{\zeta+1}{\zeta-1} \left(\|\delta_{\mathcal{T}}\|_1 + \|\delta_{\mathcal{S} \cap \mathcal{T}^c}\|_1 \right) + \frac{\zeta}{\zeta-1} \|\delta_f\|_1 \right\}, \tag{3.4.2}$$

and

$$\delta_f = (\hat{A}'_{\mathcal{T}})^{-1}\hat{\psi} - (A'_{\mathcal{T}})^{-1}\psi - [(\hat{A}'_{\mathcal{T}})^{-1}\hat{A}'_{\mathcal{T}^c} - (A'_{\mathcal{T}})^{-1}A'_{\mathcal{T}^c}]\beta_{\mathcal{T}^c}.$$

In Section 3.7.1, we show that the estimation errors of $\hat{\beta}$ belongs to Δ_ζ, which justifies the sufficiency of $\kappa_\zeta(M_x) > 0$ in the study of the error bound of $\hat{\beta}$. Compared to James et al. (2012), our restricted set Δ_ζ is larger due to estimation error of $\hat{A}$ and $\hat{\psi}$. If we know A and ψ exactly, $\delta_f = 0$ and Δ_ζ coincides with the restricted set defined in James et al. (2012). Moreover, if there is no restriction, $\mathcal{T} = \emptyset$ and $\zeta = 2$, which implies that Δ_ζ degenerate to the restricted set of standard LASSO; see Bickel et al. (2009).

Assumption 16.

(1) *$r \leq s = o(T)$, where $r = \text{rank}(A)$ and $s = \|\phi^*\|_0$.*

(2) *$A'_{\mathcal{T}}$ is non-singular for some $\mathcal{T} \subset \text{supp}(\beta)$.*

(3) *For any $\zeta > 0$, there exists a finite constrant $\kappa > 0$, which does not depend on T but may depend on ζ, such that $\kappa_\zeta(M_x) \geq \kappa$ with probability approaching 1 as $T \to \infty$.*

(4) *$\{x_{jt}, \epsilon_t\}_{t=1}^T$ is a strong mixing sequence with $E[x_{jt}\epsilon_t] = 0$ for $j = 1, \cdots, N$ with mixing coefficient $\alpha(t)$ satisfying $\alpha(t) \leq \exp(-ct^{\eta_1})$ for some $\eta_1 > 0$ and $c > 0$; $\sup_{1\leq j\leq r}\sup_t P(|x_{jt}\epsilon_t| > x) \leq \exp(1 - x^{\eta_2})$ for some $\eta_2 > 0$; $0 < l \leq \min_{1\leq j\leq N} V_j \leq \max_{1\leq j\leq N} V_j \leq u < \infty$, where $V_j = \sup_{1\leq t\leq T}\left(Ex_{jt}^2\epsilon_t^2 + 2\sum_{s>t}|E(x_{js}x_{jt}\epsilon_s\epsilon_t)|\right)$.*

(5) *$\log N = o(T^{1/3})$.*

Assumption 16 (1) requires the number of the factors cannot be larger than the number of essentially relevant regressors, which is commonly satisfied in practice because the number of latent factors is usually very limited. Assumption 16 (2) requires the loadings of the essentially relevant regressors are linearly independent with each other. Assumption 16 (3) is a technical assumption preparing for the Fuk-Nagaev inequality, which is used to control the probability when the maximum score is beyond a specific penalty level. Assumption 16 (5) requires that the cross-sectional dimension cannot be too large in the sense that it cannot be larger than the exponential of the time-dimension, which is commonly needed for the penalized regression method such as LASSO.

Theorem 9. *Under Assumption 15 and 16, if $\lambda = K\sqrt{T\log N}$ with $K > 4\zeta\sqrt{\frac{2u}{\log 2}}$ and some constant $\zeta > 1$, we have*

$$\|\hat{\beta} - \beta\|_2^2 \leq T_{n,1} + \sqrt{T_{n,1}^2 + T_{n,2}},$$

where

$$T_{n,1} = \frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r},\sqrt{r}\}\lambda}{2\kappa^2\zeta T},$$
$$T_{n,2} = \frac{\kappa\|\delta_f\|_2^2}{\kappa T} - \frac{(2\zeta+1)\sqrt{r}\lambda\|\delta_f\|_2}{2\kappa^2\zeta T}.$$

When $\nu = 0$, which is the case of strong factors, $T_{n,2}$ is of order $\frac{\sqrt{\log N}}{T}$, and is dominated by $T_{n,1}^2$ whose order is $\frac{\log N}{T}$. The error bound of our constrained LASSO estimator $\hat{\beta}$ is then close to the error bound of the infeasible constrained LASSO estimator $\check{\beta}$; see Theorem 1 of James et al. (2012), which is sharper than that of the standard LASSO estimator $\tilde{\beta}$; see Theorem 1 of Negahban et al. (2009). When $\nu > 0$, which is the case of weak factors, $T_{n,2}$ is of order $(N/T)^{2\nu} + (N/T)^{\nu}\sqrt{\log N}$. Then, $T_{n,2}$ dominates $T_{n,1}$, and the error bound of our constrained LASSO estimator $\hat{\beta}$ thus is not close to that of the infeasible constrained LASSO estimator $\check{\beta}$ even when T is large.

3.5 Numerical Results

In this section, I provide the simulation results of the estimators proposed in Section 3.3 to investigate their finite sample performances.

3.5.1 Simulation Design

Throughout this section, we consider three data generation processes as follows.
Example 1: For the factor model (3.2.2), let $r = 1$, $A = (a_1, \cdots, a_N)'$ with $a_i = 2\cos(2\pi i/N)$, $f_t = 0.4 f_{t-1} + \omega_t$ with $\omega_t \sim i.i.N(0,1)$, and $u_t \sim i.i.N(0, I_N)$. For (3.2.1), we set $\beta_j = 5\mathbb{1}_{\{j=1,2,3\}}$, $z_t \sim i.i.N(0,5)$, $\gamma = 5$, and $\epsilon_t \sim i.i.N(0, 1/\sqrt{2})$.
Example 2.1: For the factor model (3.2.2), let $r = 3$. The elements of A is generated randomly from the $U(-5,5)$ distribution. $f_{1t} = v_t$, $f_{2t} = v_{t-1}$, $f_{3t} = v_{t-2}$ for $v_t = 0.5\omega_{t-1} + \omega_t$ with $\omega_t \sim i.i.N(0,1)$, and $u_t \sim i.i.N(0, I_N)$.

Example 2.2: $u_t \sim i.i.N(0, \Sigma_u)$, where the (i, j) element of Σ_u is defined as

$$\sigma_{i,j} = \begin{cases} \frac{1}{2}\left\{(|i-j|+1)^{2H} - 2|i-j|^{2H} + (|i-j|-1)^{2H}\right\}, & i \neq j; \\ 1, & i = j. \end{cases}$$

with $H = 0.9$ is the Hurst parameter. Everything else is the same as Example 2.1.

3.5.2 Simulation Results

First, we provide results for $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ in Table 3.1, 3.2, 3.3, and 3.4 as follows. Let $N \in \{400, 500\}$ and $T \in \{100, 200, 400\}$. For each example, the rooted mean square error are reported in two situations: $\nu = 0$ and $\nu = 0.5$. For each setting, we replicate the experiment 100 times.

Table 3.1: Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 1 with $\nu = 0$.

	$\|\hat{A} - A\|_2$		$\|\hat{\gamma} - \gamma\|_2$		$\|\hat{\psi} - \psi\|_2$	
	$N = 400$	$N = 500$	$N = 400$	$N = 500$	$N = 400$	$N = 500$
$T = 100$	176_{53}	186_{60}	126_{94}	155_{114}	27_{22}	24_{19}
$T = 200$	120_{24}	113_{20}	90_{78}	107_{87}	16_{12}	16_{11}
$T = 400$	84_{9}	83_{13}	67_{56}	77_{64}	12_{9}	11_{9}

Note: Means and standard deviations are reported for each case, and the reported values are actual values multiplied by 1000.

Table 3.2: Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 1 with $\nu = 0.5$.

	$\|\hat{A} - A\|_2$		$\|\hat{\gamma} - \gamma\|_2$		$\|\hat{\psi} - \psi\|_2$	
	$N = 400$	$N = 500$	$N = 400$	$N = 500$	$N = 400$	$N = 500$
$T = 100$	530_{73}	555_{70}	127_{95}	156_{113}	126_{102}	105_{87}
$T = 200$	425_{56}	426_{46}	91_{77}	108_{87}	84_{67}	88_{60}
$T = 400$	330_{31}	340_{40}	67_{56}	76_{64}	59_{46}	61_{46}

Note: Means and standard deviations are reported for each case, and the reported values are actual values multiplied by 1000.

Table 3.3: Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 2.1 and 2.2 with $\nu = 0$.

		$\Vert\hat{A} - A\Vert_2$		$\Vert\hat{\gamma} - \gamma\Vert_2$		$\Vert\hat{\psi} - \psi\Vert_2$	
		$N = 400$	$N = 500$	$N = 400$	$N = 500$	$N = 400$	$N = 500$
DGP2.1	$T = 100$	408_{176}	450_{256}	203_{148}	169_{129}	108_{83}	101_{91}
	$T = 200$	354_{196}	349_{223}	126_{91}	147_{114}	71_{58}	84_{149}
	$T = 400$	234_{106}	266_{116}	95_{67}	87_{70}	49_{39}	41_{33}
DGP2.2	$T = 100$	438_{305}	402_{249}	363_{280}	298_{241}	181_{282}	121_{150}
	$T = 200$	343_{211}	331_{188}	212_{155}	284_{201}	87_{60}	79_{61}
	$T = 400$	229_{121}	265_{125}	189_{137}	166_{117}	57_{35}	49_{35}

Note: Means and standard deviations are reported for each case, and the reported values are actual values multiplied by 1000.

Table 3.4: Rooted MSE of $\hat{A}$, $\hat{\gamma}$, and $\hat{\psi}$ for Example 2.1 and 2.2 with $\nu = 0.5$.

		$\|\hat{A} - A\|_2$		$\|\hat{\gamma} - \gamma\|_2$		$\|\hat{\psi} - \psi\|_2$	
		$N = 400$	$N = 500$	$N = 400$	$N = 500$	$N = 400$	$N = 500$
	$T = 100$	653_{220}	709_{255}	205_{150}	168_{130}	216_{123}	217_{166}
DGP2.1	$T = 200$	626_{164}	676_{270}	125_{91}	148_{114}	168_{119}	186_{191}
	$T = 400$	552_{161}	575_{92}	95_{67}	87_{70}	127_{120}	107_{64}
	$T = 100$	849_{370}	930_{400}	339_{269}	292_{228}	740_{428}	716_{408}
DGP2.2	$T = 200$	801_{338}	824_{357}	193_{148}	278_{201}	745_{460}	639_{377}
	$T = 400$	646_{285}	784_{367}	176_{136}	164_{112}	561_{370}	571_{335}

Note: Means and standard deviations are reported for each case, and the reported values are actual values multiplied by 1000.

We next provide the results for the constrained LASSO estimator $\hat{\beta}$ for Example 1 in Figure 3.1 and 3.2 as follows. For $T \in \{400, 800\}$ and $N \in \{1.2T, 2.4T\}$, the realized L_2 error of the standard LASSO estimator $\tilde{\beta}$, infeasible constrained LASSO estimator $\check{\beta}$, and the proposed constrained LASSO estimator $\hat{\beta}$ are separately visualized along the path of the tunning parameter λ. In each case, we replicate the simulation 100 times and report the average of the realized L_2 error. To compute the constrained LASSO estimator, we employ the alternating direction method of multipliers (ADMM) by Gaines and Zhou (2016).

Figure 3.1: Realized L_2 Errors ($\nu = 0$)

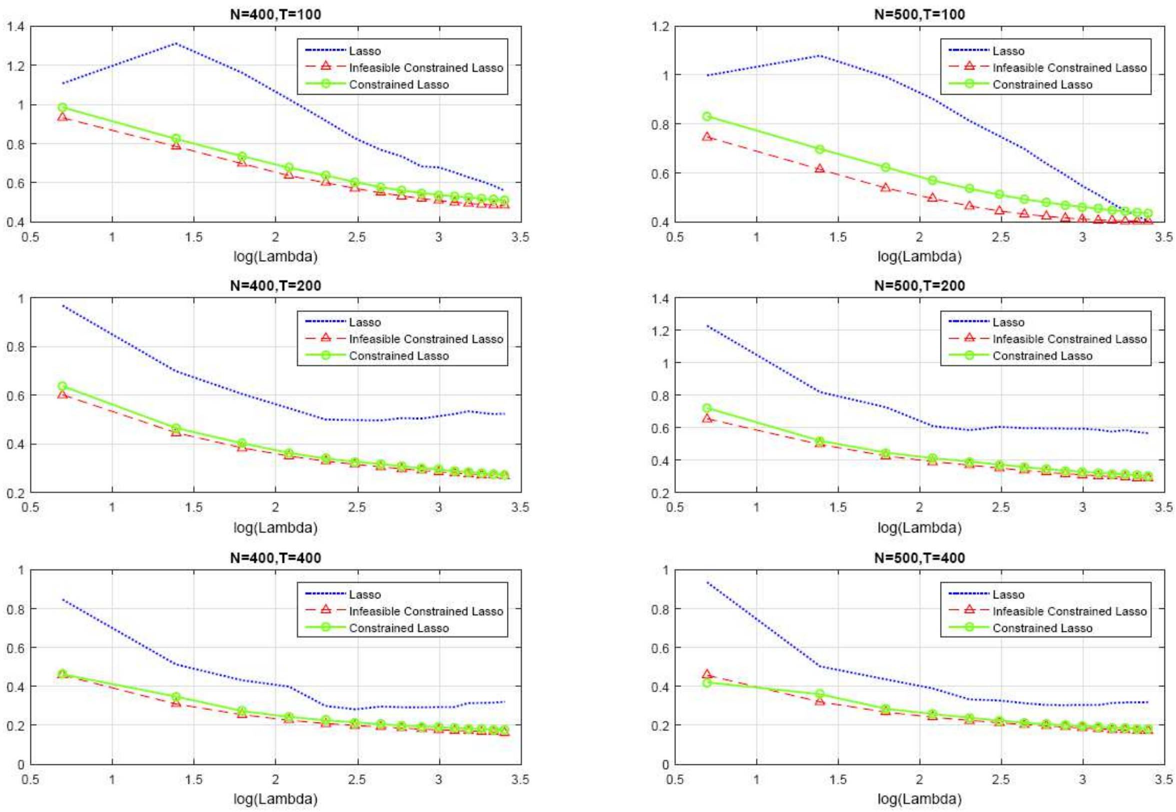
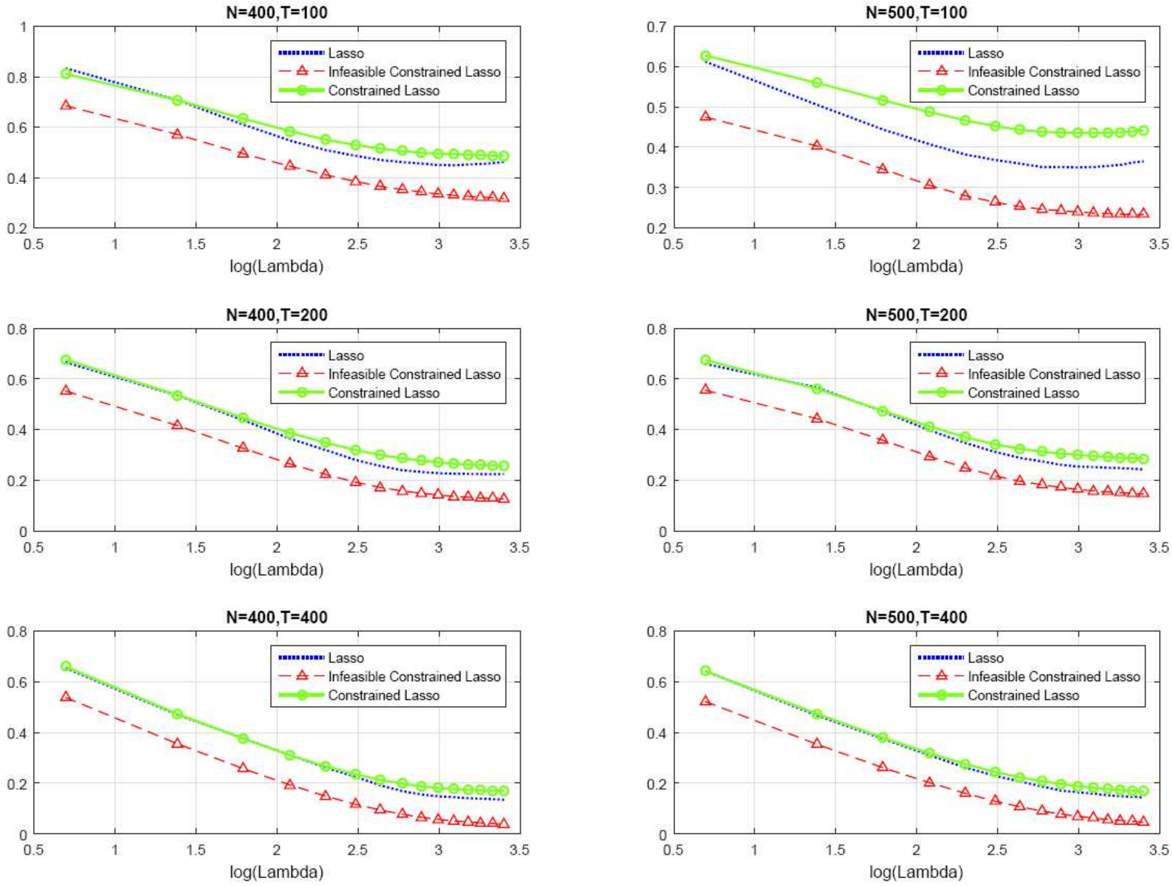


Figure 3.2: Realized L_2 Errors ($\nu = 0.5$)



3.6 Conclusion

This chapter proposes a method to improve the performance of existing spillover estimators by using a latent factor structure in the variables that generate the spillovers. Specifically, we find that such a latent factor structure implies linear constraints on the parameters characterizing the spillovers, and the better performance of existing estimators would be available by adding these factor-induced constraints. The L_2 error bound of the proposed estimator is derived. Compared with the unconstrained estimator, the proposed estimator is more accurate in the sense that it has approximately sharper error bound. Also, we note that the strength of the latent factors is critical for the performance of the factor-induced constraints. In particular, strong factors always provide better constraints than weak factors, which is due to the fact that the estimated factor loadings and corresponding factor process have faster convergence rate when the factors are strong. Our findings are demonstrated by the simulations.

3.7 Appendix

3.7.1 Proofs of Main Results

Let $Y = (y_1, \cdots, y_T)' \in \mathbb{R}^T$, $\epsilon = (\epsilon_1, \cdots, \epsilon_T)' \in \mathbb{R}^T$, $X = (x_1, \cdots, x_T)' \in \mathbb{R}^{T\times N}$, $F = (f_1, \cdots, f_T)' \in \mathbb{R}^{T\times r}$, $U = (u_1, \cdots, u_T)' \in \mathbb{R}^{T\times N}$, and $e = U\beta + \epsilon \in \mathbb{R}^T$. We can then rewrite (3.2.1) as $X = FA' + U$ and (3.2.3) as $Y = i_T\gamma + F\psi + e$. Moreover, let $\hat{F} = (\hat{f}_1, \cdots, \hat{f}_T)' \in \mathbb{R}^{T\times r}$, $M_T = I_T - T^{-1} i_T i_T'$, where I_T is the $T \times T$ identity matrix and i_T is T-dimensional vector whose components are all one.

Proof of Theorem 8: By $Y = i_T\gamma + \hat{F}\psi + [(F - \hat{F})\psi + U\beta + \epsilon]$, we have

$$\hat{\psi} - \psi \quad = \quad \left(\hat{F}' M_T \hat{F}\right)^{-1} \hat{F}' M_T [(F - \hat{F})\psi + U\beta + \epsilon].$$

Given the results in Lemma 15 (2), it is sufficient for us to investigate the order of $\|\hat{F}' M_T (F - \hat{F})\|_2$. To this end, we note $\hat{F}' M_T (F - \hat{F}) = -A'X'M_T UA + (\hat{A} - A)'X'M_T F - R_1$, where $R_1 = (\hat{A} - A)'X'M_T XA + A'X'M_T X(\hat{A} - A) + (\hat{A} - A)'X'M_T X(\hat{A} - A)$. Then, following a similar argument as in the proof of Lemma 15, for some positive constant $\rho < 1 - v$, we have

$$\|A'X'M_T UA\|_2 = O_p(TN^\rho + \sqrt{T}N),$$
$$\|(\hat{A} - A)'X'M_T F - R_1\|_2 = O_p(\sqrt{T}N).$$

Then,we have $\|\hat{F}' M_T (F - \hat{F})\|_2 = O_p(TN^\rho + \sqrt{T}N)$, and the conclusion follows by Lemma 16 (1). ■

Proof of Theorem 9: Let $\hat{\delta} \equiv \hat{\beta} - \beta$. Also, for a d-dimensional vector v and an index set $I \subset \{1, \cdots, d\}$, let $\Pi_I(v) = v_I$.

Step 1: Under Assumption 16, $\hat{\delta} \in \Delta_\zeta$ if $\lambda \geq \zeta \|X'\epsilon\|_\infty$ for some constant $\zeta \geq 1$.

Since $\|\hat{A} - A\|_2 \xrightarrow{p} 0$, $\hat{A}'_{\mathcal{T}}$ is invertible w.p.a.1, which implies

$$b_{\mathcal{T}} = (\hat{A}'_{\mathcal{T}})^{-1}(\hat{\psi} - \hat{A}'_{\mathcal{T}^c} b_{\mathcal{T}^c})$$

hold with w.p.a.1 for any $b \in \mathbb{R}^N$ satisfying $\hat{\psi} = \hat{A}'b$. We then have the constrained problem (3.3.2) with respect to $b \in \mathbb{R}^N$ is equivalent (w.p.a.1) to the following unconstrained problem with respect to $b_2 \in \mathbb{R}^{N-r}$.

$$\min_{b_2 \in \mathbb{R}^{N-r}} \frac{1}{2}\|\tilde{y} - \tilde{X}_{\mathcal{T}^c} b_2\|_2^2 + \lambda\left(\|b_{\mathcal{T}}(b_2)\|_1 + \|b_2\|_1\right), \tag{3.7.1}$$

where $b_{\mathcal{T}}(b_2) = (\hat{A}'_{\mathcal{T}})^{-1}\hat{\psi} - (\hat{A}'_{\mathcal{T}})^{-1}\hat{A}'_{\mathcal{T}^c} b_2$, $\tilde{y} = y - X_{\mathcal{T}}(\hat{A}'_{\mathcal{T}})^{-1}\hat{\psi}$, and $\tilde{X}_{\mathcal{T}^c} = X_{\mathcal{T}^c} - X_{\mathcal{T}}(\hat{A}'_{\mathcal{T}})^{-1}(\hat{A}'_{\mathcal{T}^c})$. Let $\hat{b}_2$ be the solution to (3.7.1) and $\beta_{\mathcal{T}^c}$ be the corresponding true value. Let $\hat{b} = (b_{\mathcal{T}}(\hat{b}_2), \hat{b}_2)$ and $\beta_{\mathcal{T}}$ be the true value of $b_{\mathcal{T}}(\hat{b}_2)$. By the optimality of $\hat{b}_2$, we have

$$\begin{aligned}
0 \geq_{(1)} & -(\tilde{y} - \tilde{X}_{\mathcal{T}^c}\beta_{\mathcal{T}^c})'\tilde{X}_{\mathcal{T}^c}(\hat{b}_2 - \beta_{\mathcal{T}^c}) + \|\tilde{X}_{\mathcal{T}^c}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_2^2 \\
& + \lambda\left(\|b_{\mathcal{T}}(\hat{b}_2)\|_1 - \|b_{\mathcal{T}}(\beta_{\mathcal{T}^c})\|_1 + \|\hat{b}_2\|_1 - \|\beta_{\mathcal{T}^c}\|_1\right) \\
\geq_{(2)} & -\epsilon' X(\hat{b} - \beta) + 3\delta_f' X_{\mathcal{T}}' X(\hat{b} - \beta) - \epsilon' X_{\mathcal{T}}\delta_f + 2\|X_{\mathcal{T}}\delta_f\|_2^2 + \|X(\hat{b} - \beta)\|_2^2 \\
& + \lambda\left(\|\hat{b}\|_1 - \|\beta\|_1 - \|\delta_f\|_1\right),
\end{aligned} \tag{3.7.2}$$

where the first inequality follows by the optimality of $\hat{b}_2$ and the second inequality follows by

$$\begin{aligned}
\tilde{y} - \tilde{X}_{\mathcal{T}^c}\beta_{\mathcal{T}^c} &= \epsilon - X_{\mathcal{T}}\delta_f, \\
\tilde{X}_{\mathcal{T}^c}(\hat{b}_2 - \beta_{\mathcal{T}^c}) &= X(\hat{b} - \beta) + X_{\mathcal{T}}\delta_f, \\
b_{\mathcal{T}}(\beta_{\mathcal{T}^c}) &= \beta_{\mathcal{T}} + \delta_f,
\end{aligned}$$

with $\delta_f = (\hat{A}'_{\mathcal{T}})^{-1}\hat{\psi} - (A'_{\mathcal{T}})^{-1}\psi - [(\hat{A}'_{\mathcal{T}})^{-1}\hat{A}'_{\mathcal{T}^c} - (A'_{\mathcal{T}})^{-1}A'_{\mathcal{T}^c}]\gamma_{\mathcal{T}^c}$.

Then, using $\|X_{\mathcal{T}}\delta_f\|_2^2 \geq 0$, $\|X(\hat{b} - \gamma)\|_2^2 \geq 0$, and

$$\begin{aligned}
-\epsilon' X(\hat{b} - \beta) &\geq -|\epsilon' X(\hat{b} - \beta)| \geq -\|\epsilon' X\|_\infty \|\hat{b} - \beta\|_1 \geq -\frac{\lambda}{\zeta}\|\hat{b} - \beta\|_1, \\
\delta_f' X_{\mathcal{T}}' X(\hat{b} - \beta) &\geq -|\delta_f' X_{\mathcal{T}}' X(\hat{b} - \beta)| \geq -\|\delta_f' X_{\mathcal{T}}' X\|_\infty \|\hat{b} - \beta\|_1 \geq -\frac{\lambda}{\zeta}\|\hat{b} - \beta\|_1, \\
-\epsilon' X_{\mathcal{T}}\delta_f &\geq -|\epsilon' X_{\mathcal{T}}\delta_f| \geq -\|\epsilon' X_{\mathcal{T}}\|_\infty \|\delta_f\|_1 \geq -\frac{\lambda}{\zeta}\|\delta_f\|_1,
\end{aligned}$$

(3.7.2) implies

$$0 \geq \lambda \left(\|\hat{b}\|_1 - \|\beta\|_1 - \frac{1}{\zeta}\|\hat{b} - \beta\|_1 - \|\delta_f\|_1 \right). \tag{3.7.3}$$

Let $\text{supp}(\beta_{\mathcal{T}^c}) = \mathcal{S}$. By Lemma 5 of James et al. (2012), we have

$$\begin{aligned}
&\|\hat{b}\|_1 - \|\beta\|_1 - \frac{1}{\zeta}\|\hat{b} - \beta\|_1 - \|\delta_f\|_1 \\
&= \left(\|b_{\mathcal{T}}(\hat{b}_2)\|_1 - \|\beta_{\mathcal{T}}\|_1 - \frac{1}{\zeta}\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_1 \right) + \left(\|\hat{b}_2\|_1 - \|\beta_{\mathcal{T}^c}\|_1 - \frac{1}{\zeta}\|\hat{b}_2 - \beta_{\mathcal{T}^c}\|_1 \right) - \|\delta_f\|_1 \\
&\geq -\frac{\zeta+1}{\zeta}\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_1 - \frac{\zeta+1}{\zeta}\|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_1 + \frac{\zeta-1}{\zeta}\|\Pi_{\mathcal{S}^c}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_1 - \|\delta_f\|_1.
\end{aligned} \tag{3.7.4}$$

The conclusion thus follows by (3.7.3), (3.7.4), and $\lambda > 0$ and $\zeta > 1$.

Step 2: Under Assumption 16, $\lambda > \zeta \max\{\|\epsilon' X\|_\infty, \|\delta_f' X_{\mathcal{T}}' X\|_\infty\}$ w.p.a.1.

For $\|\epsilon' X\|_\infty$, by Lemma 17, we have

$$\begin{aligned}
P\left(\lambda < \zeta\|\epsilon' X\|_\infty\right) \quad &\leq \quad N \max_{1\leq j\leq N} P\left(\left| \sum_{t=1}^{T} \epsilon_t X_{t,j} \right| > \frac{\lambda}{\zeta} \right) \\
&\leq \quad 4N \max_{1\leq j\leq N} \left\{ \exp\left[-\frac{\lambda^2 \log 2}{32\zeta^2 T V_j} \right] + 4C\zeta T\lambda^{-1} \exp\left[-\frac{c^2(4\zeta T V_j)^\phi}{\lambda^\phi} \right] \right\} \\
&\leq \quad 4N \exp\left[-\frac{\lambda^2 \log 2}{32\zeta^2 T u} \right] + 16C\zeta N T\lambda^{-1} \exp\left[-\frac{c^2(4\zeta T l)^\phi}{\lambda^\phi} \right] \\
&= \quad 4\epsilon_p + C_1 N \sqrt{\frac{T}{\log(\frac{N}{\epsilon_p})}} \exp\left(-C_2 \left[\frac{T}{\log(\frac{N}{\epsilon_p})} \right]^{\frac{\phi}{2}} \right) \\
&\to 0,
\end{aligned}$$

where the first inequality comes from the union bound, the second inequality follows by the quoted theorem and Assumption 16 (4), the third inequality follows by $l \leq \min_{1\leq j\leq N} V_j \leq \max_{1\leq j\leq N} V_j < u$ from Assumption 16 (4), and the equality follows by choosing $\lambda = K\sqrt{T\log(\frac{N}{\epsilon_p})}$ with $K = 4\zeta\sqrt{\frac{2u}{\log 2}}$, $\log(\frac{N}{\epsilon_p}) = o(T)$, and $\epsilon_p \to 0$, and

$\log(\frac{N}{\epsilon_p}) = O(\log N)$, $\log N = o(T^{1/3})$ and $\phi > 1$ from Assumption 16.

For $\|\delta_f' X_{\mathcal{T}}' X\|_\infty$, we note

$$\begin{aligned} \|\delta_f' X_{\mathcal{T}}' X\|_\infty &= \sup_{1\le k\le N} \left| \sum_{j=1}^{r} (\delta_f)_j (X_{\mathcal{T}}' X)_{j,k} \right| \\ &\le \sqrt{r}\|\delta_f\|_2 \max_{\substack{1\le k\le N \\ 1\le j\le r}} \left|(X_{\mathcal{T}}' X)_{j,k}\right|. \end{aligned} \tag{3.7.5}$$

Again, by Lemma 17, we have

$$\begin{aligned} &P\left(\lambda < \sqrt{r}\|\delta_f\|_2 \max_{\substack{1\le k\le N \\ 1\le j\le r}} \left|(X_{\mathcal{T}}' X)_{j,k}\right|\right) - P\left(\sqrt{r}\|\delta_f\|_2 > M_f T^{-1/2}\right) \\ &\le P\left(\max_{\substack{1\le k\le N \\ 1\le j\le r}} \left|(X_{\mathcal{T}}' X)_{j,k}\right| > \frac{\sqrt{T}\lambda}{M_f}\right) \\ &\le Nr \max_{\substack{1\le k\le N \\ 1\le j\le r}} P\left(\left|\sum_{t=1}^{T} (X_{\mathcal{T}})_{t,j} X_{t,k}\right| > \frac{\lambda}{M_f}\right) \\ &\le 4Nr \max_{1\le j\le N} \left\{\exp\left[-\frac{\lambda^2 \log 2}{32 M_f^2 T V_{j,k}'}\right] + 4C M_f T \lambda^{-1} \exp\left[-\frac{c^2 (4 M_f T V_{j,k})^{\phi}}{\lambda^{\phi}}\right]\right\} \\ &\le 4Nr \exp\left[-\frac{\lambda^2 \log 2}{32 M_f^2 T u'}\right] + 16 C M_f N r T \lambda^{-1} \exp\left[-\frac{c^2 (4 M_f T\, l')^{\phi}}{\lambda^{\phi}}\right] \\ &= 4r\epsilon_p' + C_1 N \sqrt{\frac{T}{\log(\frac{N}{\epsilon_p'})}} \exp\left(-C_2 \left[\frac{T}{\log(\frac{N}{\epsilon_p'})}\right]^{\frac{\phi}{2}}\right) \\ &\to 0. \end{aligned} \tag{3.7.6}$$

By Theorem 8, $P\left(\sqrt{r}\|\delta_f\|_2 > M_f T^{-1/2}\right)$ can be arbitrarily small, which together with (3.7.5) implies

$$P\left(\lambda < \|\delta_f' X_{\mathcal{T}}' X\|_\infty\right) \to 0.$$

Step 3:

$$
\begin{aligned}
&\|\hat{b}\|_1 - \|\beta\|_1 - \frac{1}{\zeta}\|\hat{b} - \beta\|_1 \\
&\geq \frac{\zeta+1}{\zeta}\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_1 - \frac{\zeta+1}{\zeta}\|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_1 \\
&\geq -\frac{(\zeta+1)\max\{\sqrt{s-r}, \sqrt{r}\}}{\zeta}(\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_2 + \|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_2) \\
&\geq -\frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r}, \sqrt{r}\}}{\zeta}\|\hat{b} - \beta\|_2,
\end{aligned}
\tag{3.7.7}
$$

where the first inequality follows by (3.7.4) and $\zeta > 1$, the second inequality follows by $\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_1 \leq \sqrt{r}\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_2$ and $\|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_1 \leq \sqrt{s-r}\|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_2$, and the third inequality follows by $\|b_{\mathcal{T}}(\hat{b}_2) - \beta_{\mathcal{T}}\|_2 + \|\Pi_{\mathcal{S}}(\hat{b}_2 - \beta_{\mathcal{T}^c})\|_2 \leq \sqrt{2}\|\hat{b} - \beta\|_2$.

Then, (3.7.2) together with (3.7.7) implies

$$
\begin{aligned}
0 \;&\geq\; \|X(\hat{b} - \beta)\|_2^2 - \frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r}, \sqrt{r}\}\lambda}{\zeta}\|\hat{b} - \beta\|_2 + \|X_{\mathcal{T}}\delta_f\|_2^2 - \frac{(2\zeta+1)\lambda}{2\zeta}\|\delta_f\|_1 \\
&\geq\; T\kappa^2\|\hat{b} - \beta\|_2^2 - \frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r}, \sqrt{r}\}\lambda}{\zeta}\|\hat{b} - \beta\|_2 + \underline{\kappa}^2\|\delta_f\|_2^2 - \frac{(2\zeta+1)\sqrt{r}\lambda}{2\zeta}\|\delta_f\|_2,
\end{aligned}
$$

where the last inequality follows by Assumption 16 (3), $\|X_{\mathcal{T}}\delta_f\|_2^2 \geq \underline{\kappa}^2\|\delta_f\|_2^2$, and $\sqrt{r}\|\delta_f\|_2 \geq \|\delta_f\|_1$.

Then, the conclusion follows by

$$
\begin{aligned}
\|\hat{b} - \beta\|_2^2 \leq &\frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r}, \sqrt{r}\}\lambda}{2\kappa^2\zeta T} \\
&+ \sqrt{\left[\frac{(\zeta+1)\sqrt{2}\max\{\sqrt{s-r}, \sqrt{r}\}\lambda}{2\kappa^2\zeta T}\right]^2 + \frac{\underline{\kappa}\|\delta_f\|_2^2}{\kappa T} - \frac{(2\zeta+1)\sqrt{r}\lambda\|\delta_f\|_2}{2\kappa^2\zeta T}}
\end{aligned}
\tag{3.7.8}
$$

■

3.7.2 Proofs of Lemmas

Lemma 16. *Under Assumption 15,*

(1) $T^{-1}\hat{F}'M_T\hat{F} \xrightarrow{p} M_F$*, where* $M_F \in \mathbb{R}^{r\times r}$ *is a positive definite matrix with* $\lambda_{\min}(M_F) \geq cN^{1-\nu}$ *for some* $c > 0$.

(2) $\|\hat{F}'M_TU\|_2 = O_p(TN^\rho + \sqrt{T}N)$ *for a positive constant* $\rho < 1 - v$*, and* $\|\hat{F}'M_T\epsilon\| = O_p(\sqrt{NT})$.

Proof of Lemma 16: For (1), let $R_1 = (\hat{A} - A)'X'M_TXA + A'X'M_TX(\hat{A} - A) + (\hat{A} - A)'X'M_TX(\hat{A} - A)$. Then by $\hat{F} = X\hat{A}$, it follows

$$\hat{F}'M_T\hat{F} = A'X'M_TXA + R_1.$$

Since $\|A\|_2 = 1$ and $\|\hat{A} - A\|_2 = O_p(N^\nu T^{-1/2}) = o_p(1)$, R_1 is dominated by $A'X'M_TXA$ in the limit. By $X = FA' + U$ and $A'A = I_r$, we have

$$\begin{aligned} T^{-1}A'X'M_TXA &= T^{-1}F'M_TF + T^{-1}F'M_TUA + T^{-1}A'U'M_TF + T^{-1}A'U'M_TUA \\ &\xrightarrow{p} \Sigma_f + \Sigma_{fu}A + A'\Sigma_{uf} + A'\Sigma_uA \\ &\equiv M_F, \end{aligned}$$

where the convergence follows by $T^{-1}F'M_TF \xrightarrow{p} \Sigma_f$, $T^{-1}F'M_TU \xrightarrow{p} \Sigma_{fu}$, $T^{-1}U'M_TU \xrightarrow{p} \Sigma_u$, which come from Assumption 15 (3). Then, (1) follows by

$$\begin{aligned} \lambda_{\min}(M_F) &\geq \lambda_{\min}\left(\Sigma_f\right) + \lambda_{\min}\left(A'\Sigma_uA\right) + 2\lambda_{\min}\left(\Sigma_{fu}A\right) \\ &\geq \|\Sigma_f\|_{\min} + \|\Sigma_u\|_{\min} + o(N^{1-v}) \\ &\asymp N^{1-v}, \end{aligned}$$

where the first inequality follows by $\lambda_{\min}(G_1 + G_2) \geq \lambda_{\min}(G_1) + \lambda_{\min}(G_2)$, and the second inequality follows by $\lambda_{\min}(G) = \|G\|_{\min}$ for real symmetric matrix G, $\|G_1G_2\|_{\min} \geq \|G_1\|_{\min}\|G_2\|_{\min}$, $\|A\|_{\min} = 1$, and $\lambda_{\min}(A\Sigma_{fu}) = o(N^{1-v})$ from Assumption 15 (2), and the equality follows by $\|\Sigma_f\|_{\min} \asymp N^{1-\nu}$ and $\|\Sigma_u\|_{\min} = O(1)$.

For (2), let $R_2 = (\hat{A} - A)'X'M_TU$, and we have $\hat{F}'M_TU = A'X'M_TU + R_2$. Since R_2 is dominated by $A'X'M_TU$, it is suffice to focus on the bounds of $A'X'M_TU$, which follows by

$$\begin{aligned}T^{-1}\|A'X'M_TU\|_2 &\leq \|T^{-1}F'M_TU\|_2 + \|T^{-1}U'M_TU\|_2\\ &\leq \|\Sigma_{fu}\|_2 + \|\Sigma_u\|_2 + \|T^{-1}F'M_TU - \Sigma_{fu}\|_2 + \|T^{-1}U'M_TU - \Sigma_u\|_2\\ &= O(N^\rho) + O(N^{\frac{1}{2}}) + O_p(N^{1-\frac{\nu}{2}}T^{-\frac{1}{2}}) + O_p(NT^{-\frac{1}{2}})\\ &= O_p(N^\rho + NT^{-\frac{1}{2}}),\end{aligned}$$

where the first inequality follows by $X = FA' + U$, $A'A = I_r$, $\|G_1G_2\|_2 \leq \|G_1\|_2\|G_2\|_2$, $\|A\|_2 = 1$, and the triangular inequality, the first equality follows by $\|\Sigma_{fu}\|_2 \asymp N^\rho$, $\|\Sigma_u\|_2 \leq \|\Sigma_u\|_F$, Σ_u is bounded elementwisely, and Lemma 2 of Lam et al. (2011).

Similarly, for the remaining part of (2), let $R_3 = (\hat{A} - A)'X'M_T\epsilon$, and we have $\hat{F}'M_T\epsilon = A'X'M_T\epsilon + R_3$. Since R_3 is dominated by $A'X'M_T\epsilon$, it is suffice to focus on the bounds of $A'X'M_T\epsilon$, which follows by

$$\begin{aligned}T^{-1}\|A'X'M_T\epsilon\|_2 &\leq \|T^{-1}F'M_T\epsilon\|_2 + \|T^{-1}U'M_T\epsilon\|_2\\ &= O_p(N^{\frac{1-\nu}{2}}T^{-\frac{1}{2}}) + O_p(N^{\frac{1}{2}}T^{-\frac{1}{2}})\\ &= O_p(N^{\frac{1}{2}}T^{-\frac{1}{2}}),\end{aligned}$$

where the first equality follows by $\Sigma_{f\epsilon} = O\big(N^{\frac{1}{2}}T^{-\frac{1}{2}}\big)$ and $\Sigma_{u\epsilon} = 0$ and similar argument as in Lemma 2 of Lam et al. (2011). ■

Lemma 17. *Let $\{U_t\}_{t=1}^T$ be a strongly mixing sequence of real-valued and centered random variables with mixing coefficient $\alpha(t)$. There are constants ϕ_1 and $c > 0$ such that $\alpha(t) \leq \exp(-ct^{\phi_1})$, and there is a constant $\phi_2 > 0$ such that $\sup_t P(|U_t| > u) \leq \exp(1 - u^{\phi_2})$. Then, for any $\lambda \geq (TV)^{1/2}$, we have*

$$P\left(\left|\sum_{t=1}^T U_t\right| \geq 4\lambda\right) \leq 4\exp\left(-\frac{\lambda^2\log 2}{2TV}\right) + 4CT\lambda^{-1}\exp\left(-\frac{c^2(TV)^\phi}{\lambda^\phi}\right)$$

, where $\phi = \frac{\phi_1\phi_2}{\phi_1+\phi_2}$ and $V = \sup_{1\leq t\leq T}\left(EU_t^2 + 2\sum_{s>t}|E(U_sU_t)|\right)$.

Proof of Lemma 17: This is an immediate corollary of Theorem 6.2 of Rio (1999). ■

Bibliography

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Bai, J. and Liao, Y. (2013). Statistical inferences using large estimated covariances for panel data and factor models. *Working Paper*.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Battistin, E., Brugiavini, A., Rettore, E., and Weber, G. (2009). The retirement consumption puzzle: evidence from a regression discontinuity approach. *American Economic Review*, 99(5):2209–2226.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732.

Blumberg, L. J., Dubay, L., and Norton, S. A. (2000). Did the medicaid expansions for children displace private insurance? an analysis using the sipp. *Journal of Health Economics*, 19(1):33–60.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2):444–496.

Bradley, R. C. (1985). On the central limit question under absolute regularity. *The Annals of Probability*, pages 1314–1325.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, pages 453–510.

Card, D. and Shore-Sheppard, L. D. (2004). Using discontinuous eligibility rules to identify the effects of the federal medicaid expansions on low-income children. *The Review of Economics and Statistics*, 86(3):752–766.

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

Choi, I. (2012). Efficient estimation of factor models. *Econometric Theory*, 28(02):274–308.

Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):601–627.

Cutler, D. M. and Gruber, J. (1996). Does public insurance crowd out private insurance? *The Quarterly Journal of Economics*, 111(2):391–430.

Davezies, L. and Le Barbanchon, T. (2017). *Regression Discontinuity Design with Continuous Measurement Error in the Running Variable*. CREST.

De La Mata, D. (2012). The effect of medicaid eligibility on coverage, utilization, and children's health. *Health Economics*, 21(9):1061–1079.

De Paula, A. (2016). Econometrics of network models.

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

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

Delaigle, A., Hall, P., and Jamshidi, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):149–169.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, 36(2):665–685.

Delaigle, A. and Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association*, 102(480):1416–1426.

Diggle, P. J. and Hall, P. (1993). A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 523–531.

DiNardo, J. and Lee, D. S. (2011). Program evaluation and research designs. *Handbook of Labor Economics*, 4:463–536.

Dong, Y. (2015). Regression discontinuity applications with rounding errors in the running variable. *Journal of Applied Econometrics*, 30(3):422–446.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526–535.

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

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

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

Fan, J., Härdle, W., and Mammen, E. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26(3):943–971.

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

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

Gaines, B. R. and Zhou, H. (2016). Algorithms for fitting the constrained lasso. *arXiv preprint, arXiv:1611.01511*.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hall, P. and Lahiri, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *The Annals of Statistics*, pages 2110–2134.

Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *The Annals of Statistics*, 35(4):1535–1558.

Horowitz, J. L. (2014). Ill-posed inverse problems in economics. *Annual Review of Economics*, 6(1):21–51.

Horowitz, J. L. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *The Annals of Statistics*, 32(6):2412–2443.

Horowitz, J. L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *The Review of Economic Studies*, 63(1):145–168.

Hullegie, P. and Klein, T. J. (2010). The effect of private health insurance on medical care utilization and self-assessed health in germany. *Health Economics*, 19(9):1048–1062.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

James, G. M., Paulson, C., and Rusmevichientong, P. (2012). The constrained lasso. Technical report, Citeseer.

Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301–2323.

Koch, T. G. (2013). Using rd design to understand heterogeneity in health insurance crowd-out. *Journal of Health Economics*, 32(3):599–611.

Lam, C. and Souza, P. (2013). Regularization for high-dimensional spatial models using the adaptive lasso. Technical report.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Lang, S. (2013). *Complex analysis*, volume 103. Springer Science & Business Media.

Lee, D. S. and Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2):139–165.

Linton, O. B. and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika*, 83(3):529–540.

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, pages 93–100.

Manresa, E. (2016). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*, 71:165–174.

Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer Berlin Heidelberg.

Negahban, S., Yu, B., Wainwright, M. J., and Ravikumar, P. K. (2009). A unified framework for high-dimensional analysis of *m*-estimators with decomposable regularizers. In *Advances in Neural Information Processing Systems*, pages 1348–1356.

Neumann, M. H. and Hössjer, O. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479.

Opsomer, J. D. (2000). Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73(2):166–179.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, pages 186–211.

Pei, Z. and Shen, Y. (2016). The devil is in the tails: Regression discontinuity design with measurement error in the assignment variable. *Working Paper*.

Porter, J. (2003). Estimation in the regression discontinuity model. *Working Paper*, pages 5–19.

Rio, E. (1999). *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media.

Schennach, S. M. (2004). Nonparametric regression in the presence of measurement error. *Econometric Theory*, 20(06):1046–1093.

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

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, pages 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, pages 590–606.

Strichartz, R. S. (2003). *A guide to distribution theory and Fourier transforms*. World Scientific Publishing.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Tripathi, G. (1999). A matrix extension of the cauchy-schwarz inequality. *Economics Letters*, 63(1):1–3.

Van der Klaauw, W. (2008). Regression–discontinuity analysis: a survey of recent developments in economics. *Labour*, 22(2):219–245.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.