2013

JAMES-STEIN TYPE COMPOUND ESTIMATION OF MULTIPLE MEAN RESPONSE FUNCTIONS AND THEIR DERIVATIVES

Limin Feng
University of Kentucky, limin.feng@uky.edu

Limin Feng, Student

Dr. Richard Chainigo, Major Professor

Constance Wood, Director of Graduate Studies

JAMES-STEIN TYPE COMPOUND ESTIMATION OF MULTIPLE MEAN RESPONSE FUNCTIONS AND THEIR DERIVATIVES

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Limin Feng

Lexington, Kentucky

Director: Dr. Richard Charnigo, Professor of Biostatistics and Statistics at the University of Kentucky

Lexington, Kentucky

2013

ABSTRACT OF DISSERTATION

JAMES-STEIN TYPE COMPOUND ESTIMATION OF MULTIPLE MEAN RESPONSE FUNCTIONS AND THEIR DERIVATIVES

Charnigo and Srinivasan originally developed compound estimators to nonparametrically estimate mean response functions and their derivatives simultaneously when there is one response variable and one covariate. The compound estimator maintains self-consistency, infinite differentiability and almost optimal convergence rate. This dissertation studies, in part, compound estimation with multiple responses and/or covariates. An empirical comparison of compound estimation, local regression and spline smoothing is included, and near optimal convergence rates are established in the presence of multiple covariates.

James and Stein proposed an estimator of the mean vector of a p- dimensional multivariate normal distribution, which produces a smaller risk than the maximum likelihood estimator if $p \geq 3$. In this dissertation, we also extend their idea to a nonparametric regression setting. More specifically, we present Steinized local regression estimators of p mean response functions and their derivatives. We consider different covariance structures for the error terms, and whether or not a known upper bound for the estimation bias is assumed.

We also apply Steinization to compound estimation, considering the application of Steinization to both pointwise estimators (for example, as obtained through local regression) and weight functions.

Finally, the new methodology introduced in this dissertation will be demonstrated on numerical data illustrating the outcomes of a laboratory experiment in which radiation induces nanoparticles to scatter evanescent waves. The patterns of scattering, as represented by derivatives of multiple mean response functions, may be used to classify nanoparticles on their sizes and structures.

KEYWORDS: Nonparametric Regression, Compound Estimation, James-Stein Type Estimation, Multple Mean Response Functions, Multiple Covariate

Limin Feng

Student's Signature

07/30/2013

Date

JAMES-STEIN TYPE COMPOUND ESTIMATION OF MULTIPLE MEAN RESPONSE FUNCTIONS AND THEIR DERIVATIVES

By
Limin Feng

Richard Charnigo

Director of Dissertation

Constance Wood

Director of Graduate Studies

07/30/2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

1.1 Nonparametric regression

Suppose we wish to estimate the mean response function $\mu(x)$ in model

$$Y_i = \mu(x_i) + \epsilon_i \tag{1}$$

for $i \in \{1, \ldots, n\}$, where the values of x_i belong to a compact interval $\mathcal{X} \subset \mathbb{R}$, $\mu(x)$ is an unknown, real-valued function which has at least $(J+1)$ continuous derivatives on $\mathcal{X}$ for some positive integer J, the error terms ϵ_i are independent, normally distributed with mean zero and finite variance $\sigma^2 \in (0, \infty)$.

We may apply parametric regression analysis to estimate $\mu(x)$ by assuming that $\mu(x)$ has a specific form, for example, linear functions. However, if people do not want to assume any specific forms of $\mu(x)$, or when it is difficult to make any assumptions, we need to seek nonparametric approaches. The following are two practical applications in which nonparametric regression is used:

Example 1. Loader (1999) studied the ethanol dataset which measures engine exhaust from burning ethanol in a single cylinder engine. The response variable is NOx, the concentration of nitric oxide (NO) and nitrogen dioxide (NO2) and the explanatory variable is E, the equivalence ratio measuring the richness of the air and fuel mix. Suppose we wish to predict the amount of NOx at different values of E. In Figure 1, we observe that there are two peaks around $E = 0.85$ and $E = 0.95$, therefore, none of the parametric models would work well.

Example 2. Suppose we are interested in when a person grows the fastest between the ages of 1 and 18. Figure 2 depicts the first girl's ages and corresponding heights in the Berkeley Growth Study (Ramsay and Silverman (2002)). To find out at which age the girl had the fastest growth, we need to get the local maximum of the first order derivative estimates. Assuming a specific parametric form would place an unrealistic constraint on the first derivative and thus render meaningless the maximization of its estimate, and a nonparametric regression will be used.

Figure 1: Ethanol example

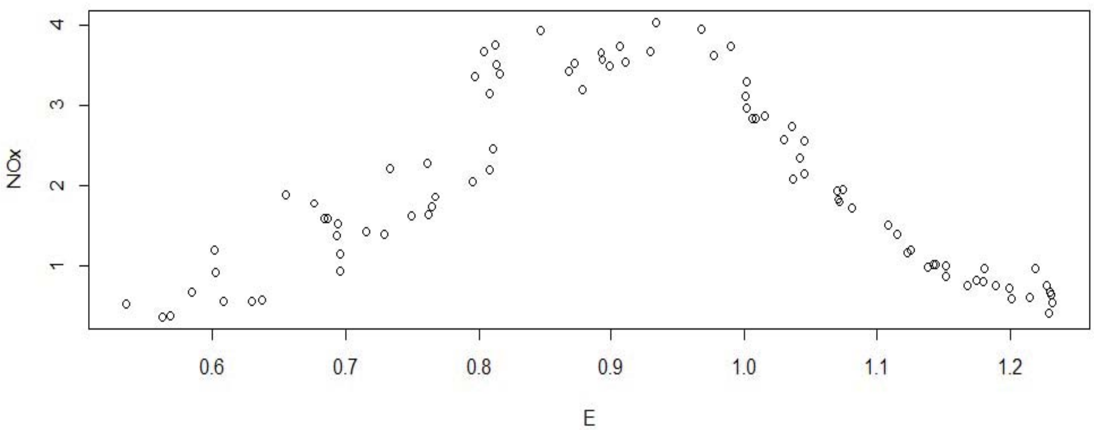
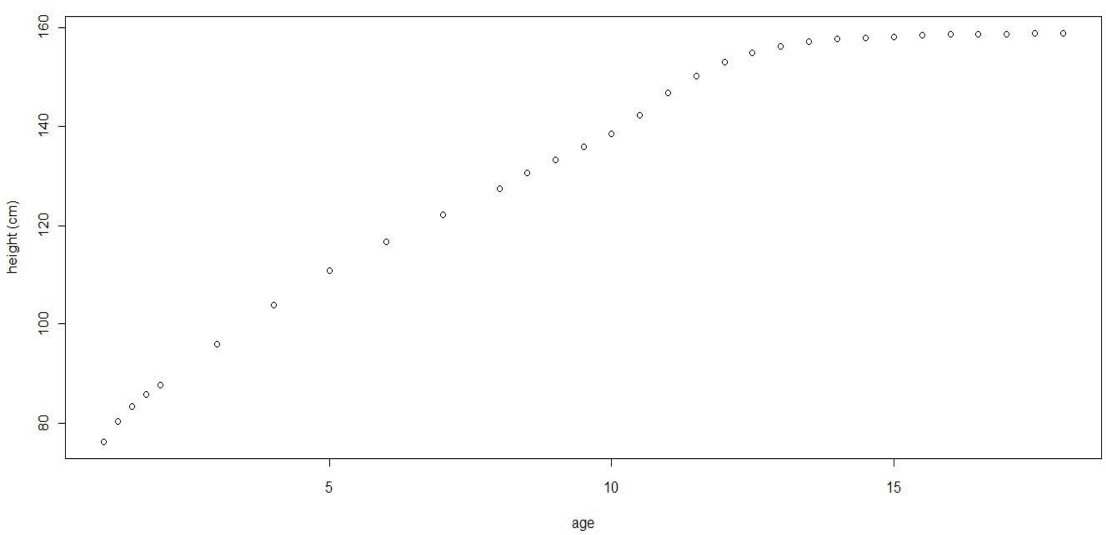


Figure 2: Human growth



Several nonparametric regression methods have been proposed. One of the first and most well studied is kernel smoothing. The basic idea behind kernel smoothing is to use observations close to the fitting point x. The "closeness" is defined by a bandwidth $h > 0$. So that observations within the interval $[x-h, x+h]$ receive more weight when estimating $\mu(x)$. Nadaraya (1964) and Watson (1964) proposed the well-known Nadaraya-Watson estimator of $\mu(x)$

$$\hat{\mu}(x) = \frac{\sum_{i=1}^{n} K(\frac{x-x_i}{h})Y_i}{\sum_{i=1}^{n} K(\frac{x-x_i}{h})} \tag{2}$$

where $K(u)$ is the weight function which assigns the most weight to the observations closest to x. Examples of weight functions are rectangular $W(u) = I_{u\in[-1,1]}$, tricube $W(u) = (1-|u|^3)^3 I_{u\in[-1,1]}$ and Gaussian $W(u) = \frac{1}{\sqrt{(2\pi)}}e^{-u^2/2}$.

A key factor in kernel smoothing is bandwidth selection. If the bandwidth h is too big, a large number of observations will be used and we will get an oversmoothed fitted curve, which means small variance and a large bias. On the other hand, a bandwidth that is too small will result in large variance and a small bias (undersmoothing). The most commonly used criterion for bandwidth selection is mean integrated squared error (MISE)

$$MISE = \int_{\mathcal{X}} E[\hat{\mu}(x) - \mu(x)]^2 dx. \tag{3}$$

It can be approximated by asymptotic mean integrated squared error (AMISE) as shown in Jones (1990). A cross-validation criterion is proposed in Sarda (1993).

If the weight function in the kernel estimator is chosen to be differentiable, which is true for the Gaussian weight function among many others, we can define the estimator of the derivatives of $\mu(x)$ as $\widehat{\mu}'(x) := \frac{\sum_{i=1}^{n} K'(\frac{x-x_i}{h})Y_i}{h\sum_{i=1}^{n} K(\frac{x-x_i}{h})}$, where K' is the derivative of K. Alternatively, if $\hat{\mu}(x)$ were defined as $\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h})Y_i$ to begin with, then $\hat{\mu}(x)$ could be directly differentiated to estimate $\hat{\mu}(x)$. We would then get $\hat{\mu}(x) := \frac{1}{nh^2}\sum_{i=1}^{n} K'(\frac{x-x_i}{h})Y_i$. Higher orders of estimators of derivatives are defined similarly. However, we may not be able to estimate the derivatives of $\mu(x)$ in this manner if K is not differentiable.

Moreover, a severe problem with the kernel smoother is bias at boundary points as observed by Hastie and Loader (1993). Realizing the weaknesses of kernel smoothing, Stone (1977), Cleveland (1979) and Katkovnik (1979) introduced local regression, which is to minimize the locally weighted sum of squares

$$\sum_{i=1}^{n} \omega_i(x)[Y_i - a_0 - a_1(x_i - x) - \frac{a_2}{2!}(x_i - x)^2 - \ldots - \frac{a_J}{J!}(x_i - x)^J]^2 \qquad (4)$$

with respect to a_0, a_1, ..., a_J, where $\omega_i(x) = W(\frac{x_i - x}{h})$ for some weight function W that is typically maximized at 0 and that approaches 0 as its argument becomes large in absolute value. Then we can define $\hat{\mu}(x) := \hat{a}_0$, $\hat{\mu}'(x) := \hat{a}_1$ and so on.

Remark 1. In local regression, estimates of derivatives of $\mu(x)$ exist up to order J even if the weight function $W(u)$ is not differentiable, which is an advantage over kernel smoothing discussed above. On the other hand, one does not have, for example, $\hat{\mu}'(x) = \frac{d}{dx}\hat{\mu}(x)$.

Remark 2. If J is chosen to be zero, it is easy to show that $\hat{\mu}(x) = \frac{\sum_{i=1}^{n} \omega_i(x) Y_i}{\sum_{i=1}^{n} \omega_i(x)}$, which is to say, kernel smoothing is a special case of local regression when the local polynomial degree is zero.

Remark 3. As shown in Hastie and Loader (1993), a big advantage of local regression over kernel smoothing is that local regression deals better with boundary effects, especially when there are multiple predictors. Indeed, formulation (4) can be readily extended to accommodate multiple predictors.

Compared to kernel smoothing, local regression has one more parameter: the local polynomial degree. As it gets higher, the estimate $\hat{\mu}(x)$ produces smaller bias but larger variance. When one is not particularly concerned about estimating derivatives, it is recommended to select a low polynomial degree (local linear or local quadratic) and then choose a bandwidth having fixed the polynomial degree (Loader (1999)). Several criteria for bandwidth selection are introduced, for example, leave-one-out cross validation and generalized cross validation criteria as well as CP criterion (Cleveland and Devlin (1988)). On the other hand, if derivative estimation is of concern, then a higher degree local polynomial is warranted and in fact one may prefer to use a method other than local regression to ensure that, for example, $\hat{\mu}'(x) = \frac{d}{dx}\hat{\mu}(x)$.

A completely different approach to estimate $\mu(x)$ is through minimization of the sum of squares with penalty

$$\sum_{i=1}^{n}[Y_i - \mu(x_i)]^2 + \lambda \int_{\mathcal{X}} \mu''(x)^2 dx \tag{5}$$

where λ is a smoothing parameter. Note that when $\lambda = 0$, there are no restrictions and $\mu(x)$ will be an interpolating function. When $\lambda = \infty$, it returns to ordinary least square estimation. Therefore λ controls the tradeoff between the two cases above. It seems difficult to solve for $\mu(x)$ in (5), however, the solution turns out to be simple, as shown in Wahba and Wold (1975), which depends on a class of basis functions called splines.

The basic idea behind spline smoothing is to partition the range of x into several intervals by choosing a set of ordered points called knots, then fit the basis functions within each interval. Fewer knots or lower polynomial degree of basis functions will result in large bias and small variance. On the other hand, small bias and large variance will be caused with choices of more knots or higher polynomial degrees.

Several criteria have been proposed for selection of tuning parameter λ, examples are leave-one-out cross validation criterion (Silverman (1985)) and generalized cross validation criterion (Craven and Wahba (1979)).

Charnigo and Srinivasan (2011) introduced a new approach to estimate $\mu(x)$ and its derivatives up to the J-th order simultaneously. The "compound estimators" have the so-called self-consistency property which can be described as $\frac{d^j}{dx^j}\hat{\mu}(x) = \widehat{\frac{d^j}{dx^j}\mu(x)}$, for $0 \leq j \leq J$. As observed in the same article, local regression estimators are not self-consistent, for indeed $\frac{d^j}{dx^j}\hat{\mu}(x)$ may not even exist.

To define the compound estimator of $\mu(x)$, let I be a compact subinterval of $(-1, 1)$ (take $\mathcal{X} = [-1, 1]$ without loss of generality). Let M_n be a nondecreasing sequence of nonnegative integers and β_n be a nondecreasing sequence of positive numbers. Partition $[-1, 1]$ into $L_n = 3^{M_n}$ subintervals with equal length, and let I_n be the set of interval midpoints falling inside $(-1, 1)$. Then the compound estimator of $\mu(x)$ is defined as

$$\mu^*(x) = \sum_{a \in I_n} W_{a,n}(x) \tilde{\mu}_{J;a}(x) \tag{6}$$

where $W_{a,n}(x) = \frac{e^{-\beta_n (x-a)^2}}{\sum_{c \in I_n} e^{-\beta_n (x-c)^2}}$, and $\tilde{\mu}_{J;a}(x) = \sum_{j=0}^{J} \tilde{c}_{j;a}(x-a)^j$ for $a \in I_n$, in which $\tilde{c}_{j;a}$, for $0 \leq j \leq J$, are the pointwise estimators of $c_{j;a} = \frac{\mu^{(j)}(a)}{j!}$. There are several options for pointwise estimation, for example, kernel smoothing, spline smoothing as well as local regression. Charnigo and Srinivasan (2011) showed that the compound estimator defined in (6) is self-consistent. Moreover, the compound estimator and its derivatives achieve near optimal convergence rates.

Let (X, Y) be a pair of random variables, where X takes value $x = (x_1, \cdots, x_d) \in \mathcal{R}^d$ for d in a set of positive integers, and Y is real valued. Assume the regression function of Y on X belongs to a collection of functions on a fixed subset of $\mathcal{R}^d$, Θ. Let $T(\theta)$ be a real valued function, where $\theta \in \Theta$, and $\hat{T}_n$ be an estimator of $T(\theta)$. Denote α as a positive number and α is called the optimal convergence rate if

$$\liminf_{n} \sup_{\theta \in \Theta} P_\theta(|\hat{T}_n - T(\theta)| > cn^{-\alpha}) > 0, \forall c > 0 \tag{7}$$

$$\lim_{c \to 0} \liminf_{n} \sup_{\theta \in \Theta} P_\theta(|\hat{T}_n - T(\theta)| > cn^{-\alpha}) = 1 \tag{8}$$

$$\lim_{c \to \infty} \limsup_{n} \sup_{\theta \in \Theta} P_\theta(|\hat{T}_n - T(\theta)| > cn^{-\alpha}) = 0. \tag{9}$$

Stone (1980) showed that under certain assumptions, the optimal convergence rate is $\alpha = \frac{p-m}{2p+d}$, where m and p index the order of the derivative being estimated and the degree of differentiability for the mean response function, respectively.

The simulation study in Charnigo and Srinivasan (2011) compared the performances of three different nonparametric regression methods in estimating $\mu(x)$ and its derivatives: compound estimation using local regression pointwise estimators, spline smoothing, and local regression. The simulation results have shown that compound estimation produces more accurate estimation than the other two methods in many cases.

To select the tuning parameters in compound estimation, which are M_n, β_n and the bandwidth in pointwise estimation via local regression, Charnigo et al (2011) proposed a generalized C_p criterion that attempts to optimize estimation of a specified derivative, which can be described as a residual sum of squares for the fitted j-th order derivative plus a penalty term such that $E\left(GC_p\right) = E\left(\sum_{i=1}^{n} s_i \left(\frac{d^j}{dx^j}\mu(x_i) - \widehat{\frac{d^j}{dx^j}\mu_\lambda(x_i)}\right)^2\right)$ up to an asymptotically negligible remainder related to estimation bias, where $0 \leq s_i \leq 1$ are weights specified by the data analyst. Let $T_n(\lambda) = E\left(\sum_{i=1}^n s_i \left(\frac{d^j}{dx^j}\mu(x_i) - \widehat{\frac{d^j}{dx^j}\mu_\lambda(x_i)}\right)^2\right)$, where λ denotes the runing parameter(s), and suppose $\hat{\lambda}_n$ and ${\lambda_n}^*$ minimize $E\left(GC_p\right)$ and $T_n(\lambda)$, respectively. Charnigo et al (2011) showed that under regularity condistions, $\frac{T_n(\hat{\lambda}_n)}{T_n({\lambda_n}^*)} \to 1$, as $n \to \infty$.

1.2 The James-Stein Type Estimator

If one observes a random vector $X = (X_1, X_2, \ldots, X_p)'$ from a normal distribution with unknown mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_p)'$ and covariance matrix equal to the identity matrix, the usual estimator of μ would be $(X_1, X_2, \ldots, X_p)'$. However, Stein (1956) showed the inadmissibility of the usual estimator when $p \geq 3$. Furthermore, he showed the estimator $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_p)'$

$$\hat{\mu}_j := \left(1 - \frac{b}{a + \sum_{k=1}^p {X_k}^2}\right) X_j, j = 1, 2, \ldots, p \tag{10}$$

for $a, b \geq 0$ and $p \geq 3$, had a smaller expected value of the loss function $E\left[\sum_{j=1}^p (\hat{\mu}_j - \mu_j)^2\right]$ than the usual estimator.

The subsequent work in James and Stein (1961) showed the following estimators

$$\hat{\mu}_j := \left(1 - \frac{p-2}{\sum_{k=1}^p {X_k}^2}\right) X_j, j = 1, 2, \ldots, p \tag{11}$$

had a risk $E\left[\sum_{j=1}^p (\hat{\mu}_j - \mu_j)^2\right] = p - E\frac{(p-2)^2}{p-2+2K}$, where K follows a Poisson distribution with mean $\sum_{j=1}^p {\mu_j}^2/2$, and $p \geq 3$, which is smaller than the risk of the usual estimators, p.

The estimators defined in formula (11) assume $X_1, X_2, \ldots, X_p$ are from a normal distribution with covariance matrix equal to the identity matrix, however, in practical applications, the covariance matrix Σ may be unknown. James and Stein (1961) provided estimator of the form $\hat{\mu} = \left(1 - \frac{(p-2)/(n-p+3)}{X'S^{-1}X}\right) X$, where S is a Wishart matrix with n degrees of freedom and scale matrix Σ. It is shown that the risk function is $p - \frac{n-p+1}{n-p+3}(p-2)^2 E\frac{1}{p-2+2K}$, where K has a Poisson distribution with mean $\mu'\Sigma^{-1}\mu/2$. Other versions of estimators when the covariance matrix is unknown can be found in, for example, Baranchik (1970),

Table 1: Batting average data

j	Name	hits/AB	X_j	$\hat{\mu_j}^{JS}$	μ_j	$\lvert X_j - \mu_j \rvert$	$\lvert \hat{\mu_j}^{JS} - \mu_j \rvert$
1	Miguel Cabrera	31/94	0.330	0.295	0.344	0.014	0.049
2	Adrian Gonzalez	33/105	0.314	0.297	0.338	0.024	0.041
3	Michael Young	38/111	0.342	0.293	0.338	0.004	0.045
4	Victor Martinez	14/56	0.250	0.305	0.330	0.080	0.025
5	Jacoby Ellsbury	25/94	0.266	0.303	0.321	0.055	0.018
6	David Ortiz	23/86	0.267	0.303	0.309	0.042	0.006
7	Dustin Pedroia	25/98	0.255	0.305	0.307	0.052	0.002
8	Casey Kotchman	15/44	0.341	0.293	0.306	0.035	0.013
9	Melky Cabrera	31/116	0.267	0.303	0.305	0.038	0.002
10	Alex Gordon	37/104	0.356	0.291	0.303	0.053	0.012

Lin and Tsai (1973) and Efron and Morris (1976).

Instead of shrinking $X_1, X_2, \ldots, X_p$ toward zero in formula (11), Efron and Morris (1973) proposed estimators shrinking $X_1, X_2, \ldots, X_p$ toward the grand mean $\bar{X} = \sum_{j=1}^{p} X_j/p$, which are defined as

$$\hat{\mu}_j := \bar{X} + \left(1 - \frac{p-3}{\sum_{k=1}^{p}(X_k - \bar{X})^2}\right)(X_j - \bar{X}), j = 1, 2, \ldots, p. \tag{12}$$

Notice that formula (12) uses $p-3$ instead of $p-2$ as appearing in (11), which is due to the dimension reduction from p to $p-1$ when applying (11) on the residuals $X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_p - \bar{X}$.

Note that if the covariance matrix of $X = (X_1, X_2, \ldots, X_p)'$ is $I\sigma^2$ where σ^2 is known, (11) and (12) can be modified by multiplying $p-2$ or $p-3$ by σ^2.

Table 1 is an illustrating example of steinization. It shows the batting averages of 10 baseball players in April of the 2011 regular season. The data can be found on *http://espn.go.com/mlb/stats/batting/_/league/al*. The true values of μ_j are computed as the averages over the whole 2011 regular season and X_j denotes the batting average of the j-th person in April. Let $\hat{\mu_j}^{JS}$ be the James-Stein estimates.

The maximum likelihood estimate of μ_j is X_j, and the James-Stein estimate based on (12) of μ_j is $\bar{X} - 0.131(X_j - \bar{X})$, where -0.131 is calculated as $1 - \frac{(p-3)\sigma^2}{\sum_{k=1}^{p}(X_k - \bar{X})^2}$. σ^2 is approximated by taking the average of estimated variances of $X_1, X_2, \ldots, X_p$ and the variance of X_j is estimated by $X_j(1-X_j)/n_j$ where n_j is the number of at bats of the j-th person.

To see which estimates perform better, we look at the distance between the estimated and our targets μ_j. As shown in Table 1, the maximum likelihood estimates for the first three players are closer to the target values. However, the James-Stein estimates for the rest are better. If we look at all ten players simultaneously, the James-Stein estimates produce a much smaller sum of squared errors $\sum_{j=1}^{p}\left(\hat{\mu}_j^{JS}-\mu_j\right)^2 = 0.007$ than the maximum likelihood estimates $\sum_{j=1}^{p}(X_j-\mu_j)^2 = 0.020$.

Efron and Morris (1971, 1972, 1973, 1975, 1976) studied James-Stein estimators via Bayesian or empirical Bayesian approaches treating the unknown parameters μ_j as random variables with their own distributions. Suppose μ_j follows a normal prior distribution with mean θ and variance τ^2. Suppose also that the conditional distribution of X_j given μ_j is normal with mean μ_j and variance σ^2. It is easy to show that the posterior distribution of μ_j is normal with mean $\frac{X_j+\frac{\sigma^2}{\tau^2}\theta}{1+\frac{\sigma^2}{\tau^2}}$ and variance $\frac{1}{1/\sigma^2+1/\tau^2}$. Therefore, the Bayes estimator is

$$\hat{\mu}_j^{Bayes} = \frac{X_j+\frac{\sigma^2}{\tau^2}\theta}{1+\frac{\sigma^2}{\tau^2}}. \tag{13}$$

On the other hand, the marginal distribution of X_j is $N(\theta,\tau^2+\sigma^2)$. We may estimate θ and τ^2 with $\bar{X}$ and $\sum_{j=1}^{p}(X_j-\bar{X})^2/(p-3)-\sigma^2$, respectively. Substituting them into (13) gives (12).

If we assume $\theta = 0$ and $\sigma^2 = 1$, then (13) simplifies to

$$\hat{\mu}_j^{Bayes} = \frac{X_j}{1+1/\tau^2}. \tag{14}$$

Since the marginal distribution of X_j is normal with mean 0 and variance $1+\tau^2$, we have $\sum_{j=1}^{p} {X_j}^2$ following a rescaled chi-square distribution with p degrees of freedom, $\sum_{j=1}^{p} {X_j}^2 \sim (1+\tau^2){\chi_p}^2$. Therefore, $E\left(\frac{p-2}{\sum_{j=1}^{p} {X_j}^2}\right) = \frac{1}{1+\tau^2}$. Substituting $\frac{1}{1+\tau^2}$ with $\frac{p-2}{\sum_{j=1}^{p} {X_j}^2}$ in (14) yields formula (11).

2 Compound estimation in multiple covariates

2.1 Motivation

Suppose $\mu(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \ldots, x_D)^T$ with D a positive integer, is the mean response function for the nonparametric regression model

$$Y_i = \mu(\mathbf{x}_i) + \epsilon_i \quad \text{for } i \in \{1, \ldots, n\}, \tag{15}$$

where $\mu(\mathbf{x})$ is an unknown, real-valued function which has continuous derivatives of order $(J+1)$, the vectors of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ belong to a compact set $\mathcal{X} \subset \mathbb{R}^D$, and ϵ_i's are independent error terms with mean zero and variance bounded above by a positive constant M.

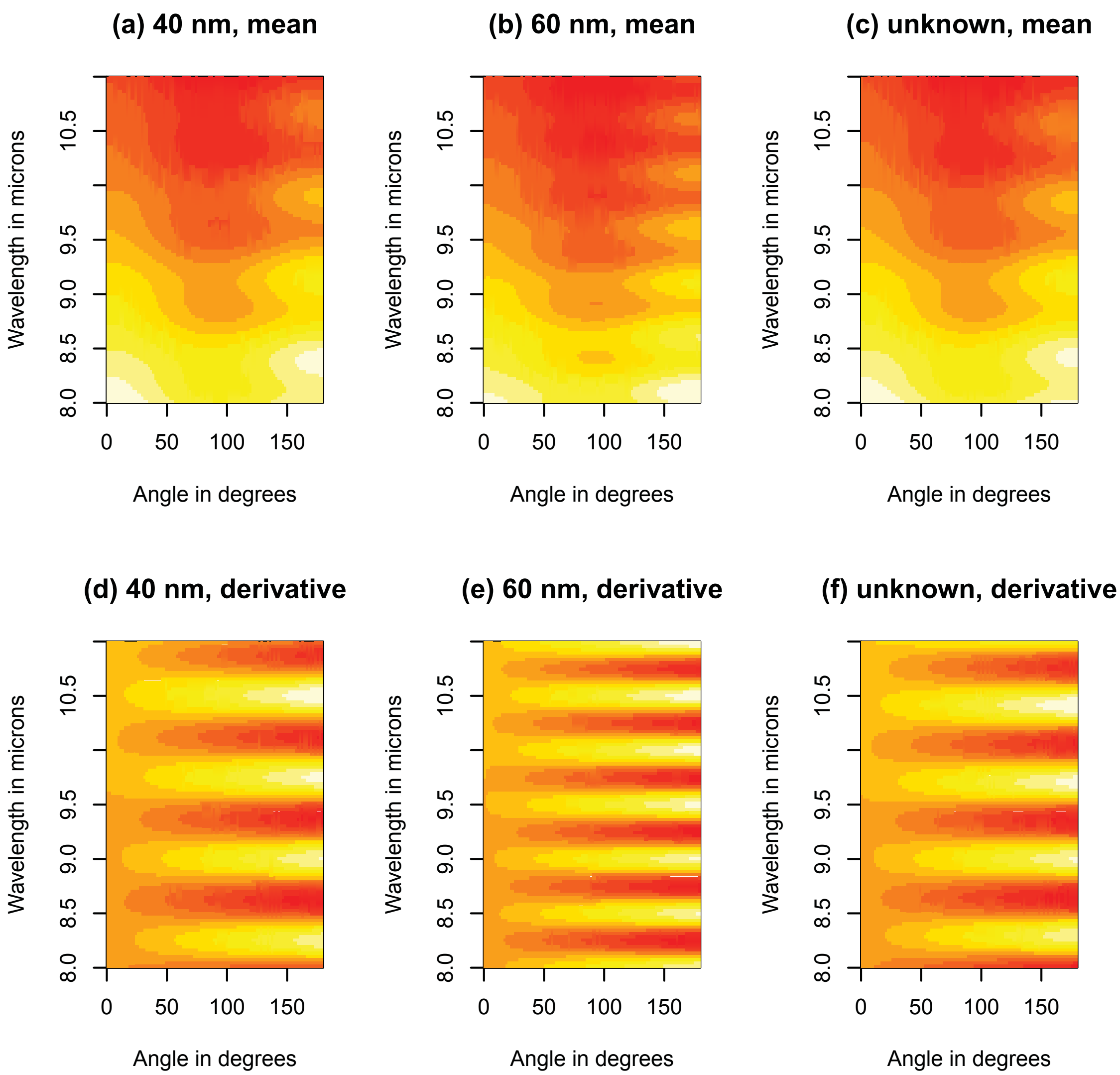
We are interested in estimating the mean response function $\mu(\mathbf{x})$ and its derivatives of up to J-th order simultaneously. Its practical applications can be, for example, predicting concentrations of certain pollutants (Loader, 1999), modeling human growth (Ramsay and Silverman, 2002), characterizing nanoparticles in light scattering experiments (Charnigo et al, 2007) and Parkinson's disease study (Little et al, 2009; Tsanas et al, 2010). Chapter 1 described several nonparametric methods for the case of $D = 1$. A more general case with generic D is also worth studying because of its important applications. The following is an illustrating example with $D = 2$.

Light scattering experiment. Panels a through c of Figure 3 describe the values of a Mueller scattering matrix (defined in Bohren and Huffman, 1998) entry, a function of radiation wavelength and scattering angle from a light scattering experiment, when the diameters of nanoparticles are 40 nm, 60 nm and unknown, respectively. A description of the light scattering experiment can be found in Charnigo et al (2007). The top three panels may be considered as estimated mean response functions with $D = 2$ on $\mathcal{X} := [0, 180] \times [8, 11]$. When comparing the three estimated mean response functions in Figure 3, we may find out if the unknown diameter is closer to 40 nm or 60 nm (shown in panels a and b).

Since it is difficult to distinguish which panel is closer to the unknown mean response function, we take the first order derivative of the mean response function with respect to wavelength (shown in panels d through f). And it shows that f is closer to d, therefore, the unkown diameter is closer to 40 nm.

However, if we fix the wavelength at 9 microns, the means and derivatives visually stay consistent showing no distinguishable difference.

Figure 3: Light scattering: two covariates



Several nonparametric regression methods have been studied to estimate $\mu(\mathbf{x})$ and its derivatives (Nadaraya 1964; Watson 1964; Stone 1977, Cleveland 1979 and Katkovnik 1979; Wahba and Wold 1975; Charnigo and Srinivasan 2011), one of which is compound estimation. One key advantage over other methods is the self-consistency property defined in Charnigo and Srinivasan (2011), which can be described as the estimated derivatives are equal to the derivatives of the estimated mean response function. Indeed, if $\widehat{\mu}(x)$ is the local regression estimator of $\mu(x)$, the derivatives may not even exist. To show the importance of self-consistency in practical applications, next we take a look at the Parkinson's telemonitoring data which can be found on $http://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring$.

Parkinson's telemonitoring study. The Parkinson's telemonitoring study (Little et al, 2009; Tsanas et al, 2010) is a six-month trial in which the voices of 42 subjects, with early-stage Parkinson's disease, were recorded to help predict the patient's symptom. The response variable is the total UPDRS score, an assessment of Parkinson's symptom. In this example, we foucus on two biomedical voice measures, DFA and PPE.

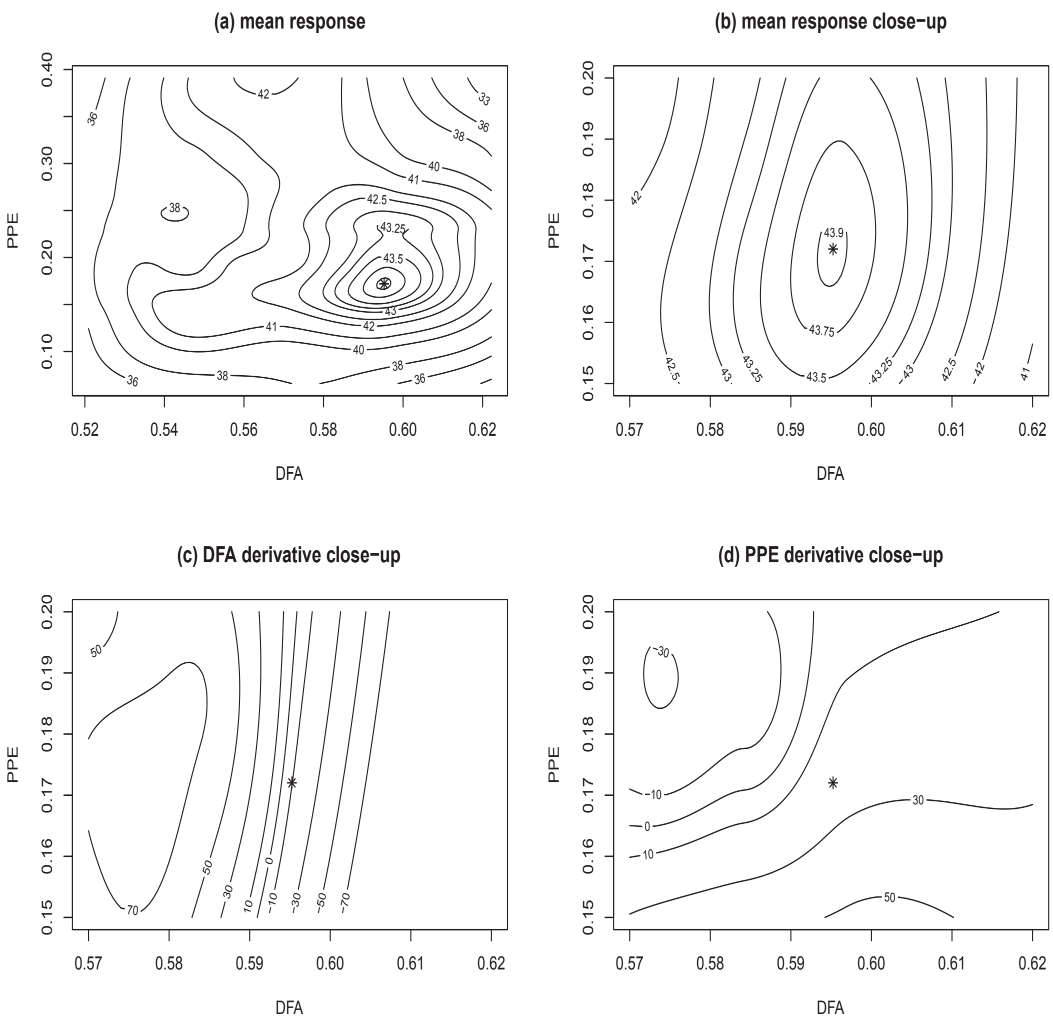
Panels a and b of Figure 4 depict a local regression estimate of the mean total UPDRS score. Panel b shows that $(0.595, 0.172)^T$ maximizes the estimated mean response function. Since $(0.595, 0.172)^T$ is not a boundary point, we would expect both of the estimated first partial derivatives with respect to DFA and PPE to be zero. However, panels c and d show that the estimated partial derivatives are not even close to zero at the maximizer. Moreover, there is no point in $[0.57, 0.62] \times [0.16, 0.19]$ at which both the estimated partial derivatives are zero. The "self-inconsistency" of local regression creates an ambiguity in identifying the optimal point.

2.2 Construction of compound estimator

Define $\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}) := \frac{\partial^{j_1+\cdots+j_D}}{\partial x_1^{j_1}\cdots\partial x_D^{j_D}}\mu(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_D)^T \in \mathcal{X}$, $\mathbf{j} = (j_1, \ldots, j_D)^T$ and $|\mathbf{j}| := \sum_{k=1}^{D} j_k$. Except for the Parkinson's telemonitoring study, we take $\mathcal{X} = [-1, 1]^D$ without loss of generality. In this section, we show how to construct a compound estimator of $\mu(\mathbf{x})$ and its derivatives $\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})$ for any $\mathbf{j}$ with $0 < |\mathbf{j}| \leq J$.

The first step of constructing the compound estimator is to obtain the pointwise estimators of $\mu(\mathbf{x})$ and its derivatives up to order J. Let $\xi_n \in [0, 1/2]$ be a nonincreasing sequence, L_n be a nondecreasing sequence of positive integers and define $I_n := [-(1-\xi_n), (1-\xi_n)]^D \subset [-1, 1]^D$. Partition $[-1, 1]$ into L_n intervals with equal width, and let R_n be the set of all interval midpoints that fall inside I_n. Next, we define the set of centering points $I_{0,n}$ as $I_{0,n} := \{\mathbf{x} = (x_1, \ldots, x_D)^T : x_1 \in R_n, \ldots, x_D \in R_n\} \subset I_n$. Let $\mathbf{j}! := \prod_{k=1}^{D} j_k!$,

Figure 4: Parkinson's study: local regression



$(\mathbf{x}-\mathbf{a})^{\mathbf{j}} := \prod_{k=1}^{D}(x_k - a_k)^{j_k}$. Put $c_{\mathbf{j},\mathbf{a}} := \{\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})|/\mathbf{j}!\}|_{\mathbf{x}=\mathbf{a}}$ for any $\mathbf{j}$ with $0 \le |\mathbf{j}| \le J$, and let $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ be the pointwise estimator. Define

$$\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) := \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J} \widetilde{c}_{\mathbf{j},\mathbf{a}}(\mathbf{x}-\mathbf{a})^{\mathbf{j}}. \tag{16}$$

Assuming $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ is a consistent estimator of $c_{\mathbf{j},\mathbf{a}}$, we observe that

$$\sup_{\mathbf{x}\in[-1,1]^D:|\mathbf{x}-\mathbf{a}|\le\zeta_n} |\mathcal{D}_{\mathbf{j}}\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})| \xrightarrow{P} 0 \quad \text{as } n\to\infty \tag{17}$$

for any $o(1)$ sequence $\zeta_n \in (0,\infty)$.

Proof. Applying Taylor's theorem for multivariate functions, we have

$$\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}) = \sum_{0\le|\mathbf{p}|\le J-|\mathbf{j}|} \mathcal{D}_{\mathbf{p}}\mathcal{D}_{\mathbf{j}}\mu(\mathbf{a})\frac{(\mathbf{x}-\mathbf{a})^{\mathbf{p}}}{\mathbf{p}!} + \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} R_q(\mathbf{x})(\mathbf{x}-\mathbf{a})^{\mathbf{q}}.$$

where $R_q(\mathbf{x}) = \frac{|\mathbf{q}|}{\mathbf{q}!}\int_0^1 (1-t)^{|\mathbf{q}|-1}\mathcal{D}_{\mathbf{q}+\mathbf{j}}\mu(\mathbf{a}+\mathbf{t}(\mathbf{x}-\mathbf{a}))\mathbf{dt}$.

On the other hand, the derivative of $\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x})$ is

$$\mathcal{D}_{\mathbf{j}}\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) = \sum_{0\le|\mathbf{p}|\le J-|\mathbf{j}|} \widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}}\frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!}(\mathbf{x}-\mathbf{a})^{\mathbf{p}}$$

Therefore, we have

$$\begin{aligned} &|\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x})| \\ = &\left|\sum_{0\le|\mathbf{p}|\le J-|\mathbf{j}|} \left[\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{a}) - \widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}}(\mathbf{p}+\mathbf{j})!\right]\frac{(\mathbf{x}-\mathbf{a})^{\mathbf{p}}}{\mathbf{p}!} + \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} R_q(\mathbf{x})(\mathbf{x}-\mathbf{a})^{\mathbf{q}}\right|. \end{aligned}$$

Since $|R_{\mathbf{q}}(\mathbf{x})| \leq \frac{|\mathbf{q}|}{\mathbf{q}!} \max_{|\mathbf{q}|=|\mathbf{p}|} \sup_{\mathbf{y}\in[-\mathbf{1},\mathbf{1}]^{\mathbf{D}}} |\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{y})|$,

$$
\begin{aligned}
&\sup_{\mathbf{x}\in[-1,1]^D:|\mathbf{x}-\mathbf{a}|\leq\zeta_n} |\mathcal{D}_{\mathbf{j}}\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})| \\
= &\sup_{\mathbf{x}\in[-1,1]^D:|\mathbf{x}-\mathbf{a}|\leq\zeta_n} \left| \sum_{0\leq|\mathbf{p}|\leq J-|\mathbf{j}|} [\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{a}) - \widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}}(\mathbf{p}+\mathbf{j})!] \frac{(\mathbf{x}-\mathbf{a})^{\mathbf{p}}}{\mathbf{p}!} \right. \\
&\left. + \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} R_q(\mathbf{x})(\mathbf{x}-\mathbf{a})^{\mathbf{q}} \right| \\
\leq &\sup_{\mathbf{x}\in[-1,1]^D:|\mathbf{x}-\mathbf{a}|\leq\zeta_n} \left\{ \sum_{0\leq|\mathbf{p}|\leq J-|\mathbf{j}|} |[\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{a}) - \widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}}(\mathbf{p}+\mathbf{j})!]| \frac{|\mathbf{x}-\mathbf{a}|^{\mathbf{p}}}{\mathbf{p}!} \right. \\
&\left. + \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} |R_q(\mathbf{x})|\,|\mathbf{x}-\mathbf{a}|^{\mathbf{q}} \right\} \\
\leq &\sum_{0\leq|\mathbf{p}|\leq J-|\mathbf{j}|} \sup_{\mathbf{x}\in[-1,1]^D:|\mathbf{x}-\mathbf{a}|\leq\zeta_n} |[\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{a}) - \widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}}(\mathbf{p}+\mathbf{j})!]| \frac{|\mathbf{x}-\mathbf{a}|^{\mathbf{p}}}{\mathbf{p}!} \\
&+ \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} \frac{|\mathbf{q}|}{\mathbf{q}!} \max_{|\mathbf{q}|=|\mathbf{p}|} \sup_{\mathbf{y}\in[-1,1]^D} |\mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{y})| \sup_{\mathbf{x}\in[-\mathbf{1},\mathbf{1}]^{\mathbf{D}}:|\mathbf{x}-\mathbf{a}|\leq\zeta_{\mathbf{n}}} |\mathbf{x}-\mathbf{a}|^{\mathbf{q}} \\
\xrightarrow{P} &\; 0 \quad \text{as } n\to\infty
\end{aligned}
$$

providing $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ consistently estimates $c_{\mathbf{j},\mathbf{a}}$ and ζ_n has limit 0.

The second step of compound estimator construction is to combine $\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x})$ defined in step 1 smoothly such that the compound estimator is infinitely differentiable and consistent. Relation (17) indicates that $\mathcal{D}_{\mathbf{j}}\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x})$ estimates $\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})$ well when $\mathbf{x}$ is close to $\mathbf{a}$, which motivates us to define the compound estimator as a weighted average of $\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x})$ and directly differentiate the result to estimate $\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})$.

Therefore, let $\mathbf{B}_n$ be a positive definite matrix, refered as convolution matrix, and let

$$W_{\mathbf{a},n}(\mathbf{x}) := \exp[-(\mathbf{x}-\mathbf{a})^T\mathbf{B}_n(\mathbf{x}-\mathbf{a})] / \sum_{\mathbf{c}\in I_{0,n}} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})],$$

the compound estimator is defined as

$$\widehat{\mu}(\mathbf{x}) := \sum_{\mathbf{a}\in I_{0,n}} W_{\mathbf{a},n}(\mathbf{x})\widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}). \tag{18}$$

And the derivatives of $\mu(\mathbf{x})$ can be estimated by

$$\widehat{\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})} := \mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) \tag{19}$$

for any $\mathbf{j}$ such that $0 < |\mathbf{j}| \leq J$.

Remark 1. We can use local regression in pointwise estimation. In fact, any nonparametric regression method that satisfies (21) or (22) can be used to estimate $c_{\mathbf{j},\mathbf{a}}$.

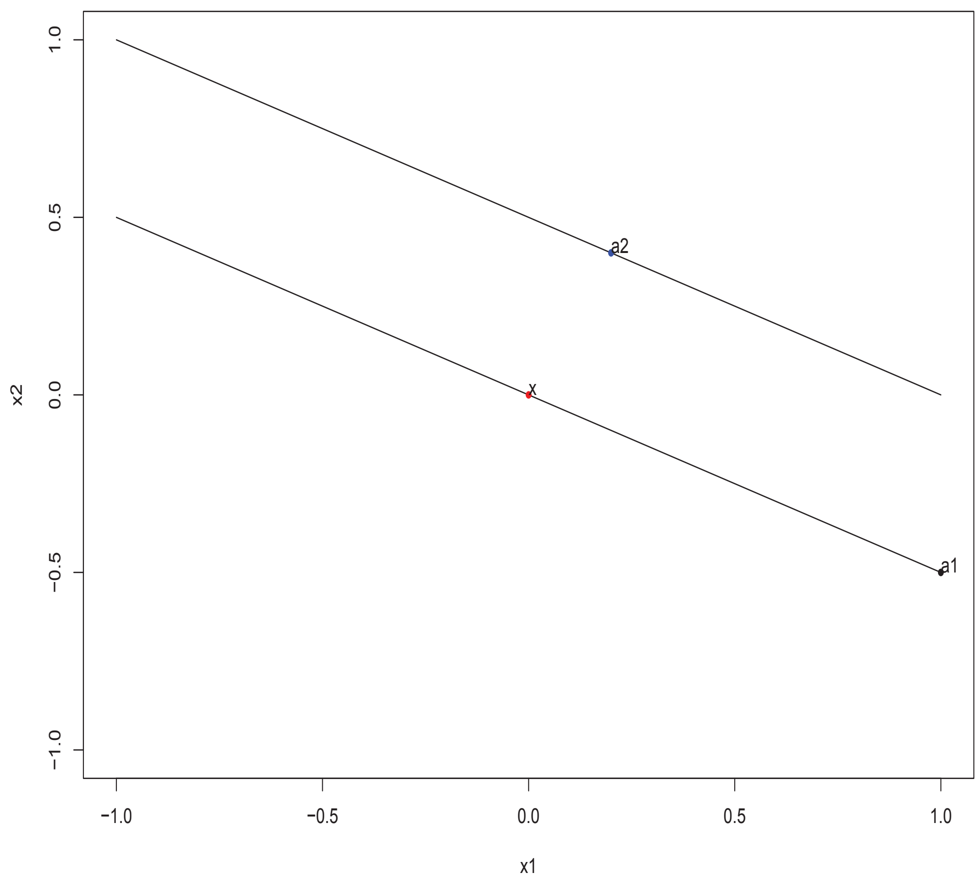
Remark 2. $I_{0,n}$ is restricted to be a subset of I_n, a compact set contained in $[-1,1]^D$, to avoid boundary effects in pointwise estimation at centering points. Consider the one dimensional case where $D = 1$. As shown in Charnigo et al (2011), bias in estimating $\mu^{(J)}(a)$ via kernel smoothing is of order $O(h)$, where h is the bandwidth. This is based on approximating an integral $\int_{\frac{a-1}{h}}^{\frac{a+1}{h}} K^{(J)}(v)\mu(a-hv)dv$ by $\int_{-\infty}^{+\infty} K^{(J)}(v)\mu(a-hv)dv$ (notations can be found in Charnigo et al 2011). So we need $\lim_{n\to\infty} \frac{a-1}{h} = -\infty$ and $\lim_{n\to\infty} \frac{a+1}{h} = +\infty$ to justify this approximation, where a is a centering point. For a uniform result across all centering points, we need $\lim_{n\to\infty} \sup_{a\in I_{0;n}} \frac{a-1}{h} = -\infty$ and $\lim_{n\to\infty} \inf_{a\in I_{0;n}} \frac{a+1}{h} = +\infty$. If we make no restriction on a, then $\lim_{n\to\infty} \sup_{a\in[-1,1]} \frac{a-1}{h} = 0 \neq -\infty$ and $\lim_{n\to\infty} \inf_{a\in[-1,1]} \frac{a+1}{h} = 0 \neq +\infty$. In summary, the pointwise estimators may not achieve condition (21) or (22) without such a restriction.

Remark 3. The proof of Theorem 2.1 shows that $\mathbf{B}_n$ can be selected as $n^\kappa \mathbf{B}_0$, where $\mathbf{B}_0$ is a positive definite matrix and κ is a positive number.

One choice of $\mathbf{B}_0$ is a diagonal matrix with the diagonall elements being a constant. Note that a larger constant provides a less smooth fit. However, we see $\mathbf{B}_0$ can be non-diagonal in other circumstances. For example, suppose $D = 2$ and $\mu(\mathbf{x}) \approx (\eta_1 x_1 + \eta_2 x_2)$. Suppose for illustration $\eta_1 = 1$ and $\eta_2 = 2$. Then $\mu(\mathbf{x})$ is approximately constant when $x_1 + 2x_2$ is fixed. In Figure 5, there are two lines: $x_1 + 2x_2 = 0$ and $x_1 + 2x_2 = 1$, and three points: $\mathbf{x} = (0,0)$, $\mathbf{a}_1 = (1, -0.5)$ on line $x_1 + 2x_2 = 0$ and $\mathbf{a}_2 = (0.2, 0.4)$ on $x_1 + 2x_2 = 1$. Note that the line connecting $\mathbf{x}$ and $\mathbf{a}_2$ is perpendicular to both of the aforementioned lines. Therefore, $\mu(\mathbf{x})$ changes most quickly along the perpendicular line. If we are estimating $\mu(\mathbf{x})$ using centering points $\mathbf{a}_1$ and $\mathbf{a}_2$, by definition (18), we have $\widehat{\mu}(\mathbf{x}) = W_{\mathbf{a}_1,n}(\mathbf{x})\widetilde{\mu}_{J,\mathbf{a}_1}(\mathbf{x}) + W_{\mathbf{a}_2,n}(\mathbf{x})\widetilde{\mu}_{J,\mathbf{a}_2}(\mathbf{x})$.

Since $\widehat{\mu}(\mathbf{x})$ and $\widetilde{\mu}_{J,\mathbf{a}_1}(\mathbf{x})$ are estimating the same quantity, we expect $W_{\mathbf{a}_1,n}(\mathbf{x})$ to be big and $W_{\mathbf{a}_2,n}(\mathbf{x})$ to be small, even though $\mathbf{a}_2$ is closer to $\mathbf{x}$ than $\mathbf{a}_1$. Suppose $\mathbf{B}_0 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we need $(\mathbf{x}-\mathbf{a}_1)^T\mathbf{B}_0(\mathbf{x}-\mathbf{a}_1)$ to be smaller than $(\mathbf{x}-\mathbf{a}_2)^T\mathbf{B}_0(\mathbf{x}-\mathbf{a}_2)$, which is satisfied when $\mathbf{B}_0 = \begin{pmatrix} 1 & 0.94 \\ 0.94 & 1 \end{pmatrix}$.

Figure 5: Indication of choice of $\mathbf{B}_0$



Remark 4. Suppose $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ has the form $\sum_{i=1}^{n} r_{\mathbf{j},\mathbf{a},i}Y_i$,where $r_{\mathbf{j},\mathbf{a},1},\ldots,r_{\mathbf{j},\mathbf{a},n}$ do not depend on $Y_1,\ldots,Y_n$. The compound estimator and its derivatives are also linear functions of

$Y_1, \ldots, Y_n$,

$$\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) = \sum_{i=1}^{n} \left\{ \sum_{\mathbf{a}\in I_{0,n}} \sum_{\mathbf{k}:0\leq\mathbf{k}\leq\mathbf{j}} \binom{\mathbf{j}}{\mathbf{k}} \mathcal{D}_{\mathbf{k}} W_{\mathbf{a},n}(\mathbf{x}) \mathcal{D}_{\mathbf{j}-\mathbf{k}} \sum_{\mathbf{l}:0\leq|\mathbf{l}|\leq J} r_{\mathbf{l},\mathbf{a},i}(\mathbf{x}-\mathbf{a})^{\mathbf{l}} \right\} Y_i$$

for $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq J$. Moreover, if the error terms in model (15) follow a normal distribution, so do the compound estimator and its derivatives, which will be useful in constructing confidence intervals for $\mu(\mathbf{x})$ and its derivatives simultaneously.

2.3 Properties of compound estimator

In this section, we will show the consistency of the compound estimator $\widehat{\mu}(\mathbf{x})$ and its derivatives $\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x})$ under mild conditions.

Assume the pointwise estimator $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ satisfies the following conditions:

$$\sup_{\mathbf{a}\in I_{0,n}} |\widetilde{c}_{\mathbf{j},\mathbf{a}}| \leq C \tag{20}$$

and either

$$\sup_{\mathbf{a}\in I_{0,n}} \mathbb{E}[(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2] \leq C n^{-2\alpha_{|\mathbf{j}|}} \tag{21}$$

or

$$\sup_{\mathbf{a}\in I_{0,n}} |\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}| = O_p(n^{-\alpha_{|\mathbf{j}|}}) \tag{22}$$

for $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq J$ and some positive constants $C, \alpha_0, \ldots, \alpha_J$.

Assumption (20) may be achieved by truncating $|\widetilde{c}_{\mathbf{j},\mathbf{a}}|$ at C. More explicitly, define $\widetilde{c}_{\mathbf{j},\mathbf{a}} := \text{sign}(\widehat{c}_{\mathbf{j},\mathbf{a}}) \min\{|\widehat{c}_{\mathbf{j},\mathbf{a}}|, C\}$. To show that either (21) or (22) is also satisfied, we need to show $\mathbb{E}[(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2] \leq \mathbb{E}[(\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2]$ and $|\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}| \leq |\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}|$. Since $\mu(\mathbf{x})$ has continuous derivatives of order $(J+1)$, there exists a positive constant M such that $|\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})| \leq M$ for any $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq J$ and $\mathbf{x} \in [-1,1]^D$. Therefore, $|c_{\mathbf{j},\mathbf{a}}| \leq C$ for some constant $C \in (0,\infty)$. If $\widehat{c}_{\mathbf{j},\mathbf{a}} = 0$, it is obvious that $(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2 - (\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2 = (\widetilde{c}_{\mathbf{j},\mathbf{a}} - \widehat{c}_{\mathbf{j},\mathbf{a}})\left[(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}) + (\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})\right] \leq 0$. If $\widehat{c}_{\mathbf{j},\mathbf{a}} < 0$ and $|\widehat{c}_{\mathbf{j},\mathbf{a}}| \leq C$, then $\widetilde{c}_{\mathbf{j},\mathbf{a}} = \widehat{c}_{\mathbf{j},\mathbf{a}}$, $\widetilde{c}_{\mathbf{j},\mathbf{a}} - \widehat{c}_{\mathbf{j},\mathbf{a}} = 0$, so we have $(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2 - (\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2 \leq 0$. The discussion for the other three cases of $\widehat{c}_{\mathbf{j},\mathbf{a}} < 0$ and $|\widehat{c}_{\mathbf{j},\mathbf{a}}| > C$, $\widehat{c}_{\mathbf{j},\mathbf{a}} > 0$ and $|\widehat{c}_{\mathbf{j},\mathbf{a}}| \leq C$, $\widehat{c}_{\mathbf{j},\mathbf{a}} > 0$ and $|\widehat{c}_{\mathbf{j},\mathbf{a}}| > C$ are similar.

Finally, we obtain $(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2 \leq (\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2$, which is to say $|\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}| \leq |\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}}|$. Consequently, $\mathbb{E}[(\widetilde{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2] \leq \mathbb{E}[(\widehat{c}_{\mathbf{j},\mathbf{a}} - c_{\mathbf{j},\mathbf{a}})^2]$.

Theorem 2.1 *Consider model (15), assume the pointwise estimator $\widetilde{c}_{\mathbf{j},\mathbf{a}}$ satisfies (20) and either (21) or (22). Suppose there exists positive numbers $\delta, \gamma, w_0, w_1, \ldots, w_J$ and $\psi \in (0, \gamma/D)$ such that*

$$\psi D + \delta D + 2w_k\gamma < 2\alpha_k \quad \textit{for } k \in \{0, 1, \ldots, J\} \tag{23}$$

$$\textit{and} \quad 0 \leq \alpha_k - \alpha_{k+1} \leq \gamma \quad \textit{for } k \in \{0, 1, \ldots, J-1\}. \tag{24}$$

Then under assumption (21), there exist $\mathbf{B}_n$ and L_n such that, for $\mathbf{x} \in \liminf_{n\to\infty} I_n$ and $0 \leq |\mathbf{j}| \leq J$,

$$\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}) = O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\}+\gamma\max\{\max_{0\leq\mathbf{m}\leq\mathbf{j}}|\mathbf{m}|-w_{|\mathbf{j}-\mathbf{m}|},|\mathbf{j}|-(J+1)\}}). \tag{25}$$

Under assumption (22), there exist $\mathbf{B}_n$ and L_n such that, for $0 \leq |\mathbf{j}| \leq J$,

$$\sup_{\mathbf{x}\in I_n} |\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})| = O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\}+\gamma\max\{\max_{0\leq\mathbf{m}\leq\mathbf{j}}|\mathbf{m}|-w_{|\mathbf{j}-\mathbf{m}|},|\mathbf{j}|-(J+1)\}}). \tag{26}$$

Proof of Theorem 2.1. Let $\mathbf{B}_0$ be an arbitrary positive definite matrix and λ_0 the square root of its largest diagonal entry. Let L_0 be a positive constant. Choose $\mathbf{B}_n$ and L_n such that

$$\mathbf{B}_n := \mathbf{B}_0 n^{2\gamma+2\delta} \quad \text{and} \quad 1 \leq \frac{L_n}{2L_0 D\lambda_0 n^{\psi+\gamma+\delta}} \leq 3. \tag{27}$$

For $\mathbf{a} \in I_{0,n}$ and $\mathbf{x} \in \liminf_{n\to\infty} I_n$, let $I_{1,n}(\mathbf{x}) := \{\mathbf{a} \in I_{0,n} : (\mathbf{x}-\mathbf{a})^t\mathbf{B}_n(\mathbf{x}-\mathbf{a}) \leq n^{2\delta}\}$, $I_{2,n}(\mathbf{x}) := \{\mathbf{a} \in I_{0,n} : (\mathbf{x}-\mathbf{a})^t\mathbf{B}_n(\mathbf{x}-\mathbf{a}) \leq 1\}$, $E_{\mathbf{a}}(\mathbf{x}) := \widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) - \mu(\mathbf{x})$, $A_{\mathbf{a}}(\mathbf{x}) := \widetilde{\mu}_{J,\mathbf{a}}(\mathbf{x}) - \sum_{\mathbf{j}:0\leq|\mathbf{j}|\leq J} c_{\mathbf{j},\mathbf{a}}(\mathbf{x}-\mathbf{a})^{\mathbf{j}}$, and $T_{\mathbf{a}}(\mathbf{x}) := A_{\mathbf{a}}(\mathbf{x}) - E_{\mathbf{a}}(\mathbf{x}) = \mu(\mathbf{x}) - \sum_{\mathbf{j}:0\leq|\mathbf{j}|\leq J} c_{\mathbf{j},\mathbf{a}}(\mathbf{x}-\mathbf{a})^{\mathbf{j}}$.

Since $\mu(\mathbf{x})$ has continuous derivatives of order $(J+1)$ and $I_{0,n} \subset [-1,1]^D$, we may assume that C in (20) is large enough so $\sup_{\mathbf{a}\in I_{0,n}} |c_{\mathbf{j},\mathbf{a}}| \leq C$, for $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq (J+1)$, is satisfied. Therefore,

$$
\begin{aligned}
|\mathcal{D}_{\mathbf{j}}T_{\mathbf{a}}(\mathbf{x})| &= \Bigg| \sum_{0\le|\mathbf{p}|\le J-|\mathbf{j}|} \mathcal{D}_{\mathbf{p}+\mathbf{j}}\mu(\mathbf{a})\frac{(\mathbf{x}-\mathbf{a})^{\mathbf{p}}}{\mathbf{p}!} \qquad (28)\\
&\quad + \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} R_q(\mathbf{x})(\mathbf{x}-\mathbf{a})^{\mathbf{q}} - \sum_{0\le|\mathbf{p}|\le J-|\mathbf{j}|} c_{\mathbf{p}+\mathbf{j},\mathbf{a}}\frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!}(\mathbf{x}-\mathbf{a})^{\mathbf{p}} \Bigg|\\
&= \Bigg| \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} R_q(\mathbf{x})(\mathbf{x}-\mathbf{a})^{\mathbf{q}} \Bigg|\\
&\le \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} |R_q(\mathbf{x})||\mathbf{x}-\mathbf{a}|^{\mathbf{q}}\\
&\le C_0 \sum_{|\mathbf{q}|=J+1-|\mathbf{j}|} |\mathbf{x}-\mathbf{a}|^{\mathbf{q}} \qquad (29)
\end{aligned}
$$

where $R_q(\mathbf{x}) = \frac{|\mathbf{q}|}{\mathbf{q}!}\int_0^1 (1-t)^{|\mathbf{q}|-1}\mathcal{D}_{\mathbf{q}+\mathbf{j}}\mu(\mathbf{a}+\mathbf{t}(\mathbf{x}-\mathbf{a}))\mathbf{dt}$ and $C_0 = (J+1-|\mathbf{j}|)C$.

Define $\lambda(\mathbf{x},\mathbf{t}) := \sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{B}_n(\mathbf{x}-\mathbf{t})}$ for $\mathbf{x},\mathbf{t} \in \liminf_{n\to\infty} I_n$, (27) yields

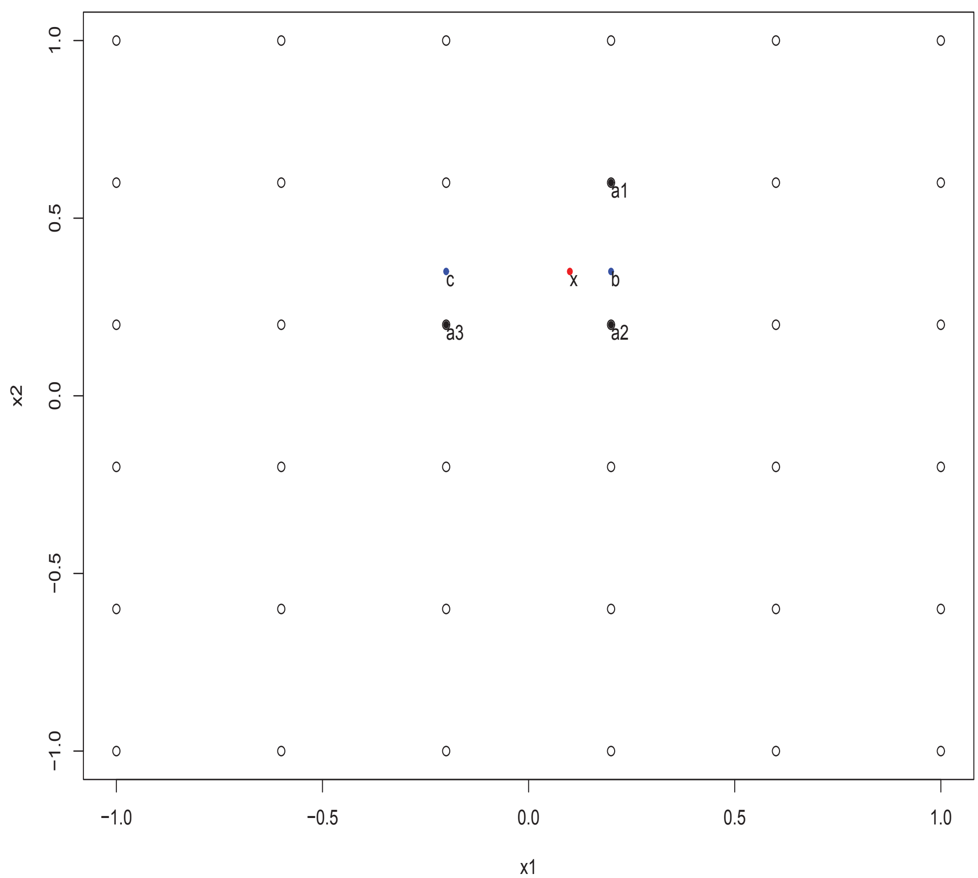
$$card\{I_{2,n}(\mathbf{x})\} \ge card\{\mathbf{a} \in I_{0,n} : \lambda(\mathbf{x},\mathbf{a}) \le 2L_0D\lambda_0 n^{\psi+\gamma+\delta}/L_n\} \ge C_1 n^{\psi D}$$

for some positive constant C_1. To show the second inequality holds, consider the case when $D = 2$ (Figure 6). The distance between $\mathbf{x}$ and $\mathbf{a}_1$ is less than the sum of the distance between $\mathbf{x}$ and $\mathbf{b}$, and the distance between $\mathbf{b}$ and $\mathbf{a}_1$. Therefore we have $\lambda(\mathbf{x},\mathbf{a}_1) \le \lambda(\mathbf{x},\mathbf{b}) + \lambda(\mathbf{b},\mathbf{a}_1) \le \lambda(\mathbf{c},\mathbf{b}) + \lambda(\mathbf{a}_1,\mathbf{a}_2) = \lambda(\mathbf{a}_2,\mathbf{a}_3) + \lambda(\mathbf{a}_1,\mathbf{a}_2)$. Let $\lambda(\mathbf{a}_1,\mathbf{a}_2) = v$ and $\lambda(\mathbf{a}_2,\mathbf{a}_3) = h$. Then $v = \sqrt{\begin{pmatrix}0 & \frac{2}{L_n}\end{pmatrix} B_n \begin{pmatrix}0\\ \frac{2}{L_n}\end{pmatrix}} = \frac{2}{L_n}\sqrt{(B_n)_{2,2}} = \frac{2n^{\gamma+\delta}}{L_n}(B_0)_{2,2} \le \frac{2n^{\gamma+\delta}}{L_n}\lambda_0$. Similarly, $h \le \frac{2n^{\gamma+\delta}}{L_n}\lambda_0$. Hence, $\lambda(\mathbf{x},\mathbf{a}_1) \le v + h \le \frac{4n^{\gamma+\delta}}{L_n}\lambda_0$. There are approximately 10^2 centering points such that $\lambda(\mathbf{x},\mathbf{a}) \le 5\frac{n^{\gamma+\delta}}{L_n}\lambda_0$. For generic D, there will be approximately $2^D K^D$ centering points $\mathbf{a}$ such that $\lambda(\mathbf{x},\mathbf{a}) \le K\frac{n^{\gamma+\delta}}{L_n}\lambda_0$. Let $K = 2L_0Dn^{\psi}$, the second inequality follows.

Thus,

$$
\begin{aligned}
&\sum_{\mathbf{c}\in I_{0,n}} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})] \ge \sum_{\mathbf{c}\in I_{2,n}(\mathbf{x})} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})]\\
&\ge \sum_{\mathbf{c}\in I_{2,n}(\mathbf{x})} \exp[-1] \ge C_1 n^{\psi D}\exp[-1]. \qquad (30)
\end{aligned}
$$

Figure 6: Indication of number of centering points



If (21) holds, we need to show that

$$\mathcal{D}_{\mathbf{j}}\left\{\sum_{\mathbf{a}\in I_{0,n}(\mathbf{x})} W_{\mathbf{a},n}(\mathbf{x})E_{\mathbf{a}}(\mathbf{x})\right\} = O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\}+\gamma\max\{\max_{0\le\mathbf{m}\le\mathbf{j}}|\mathbf{m}|-w_{|\mathbf{j}-\mathbf{m}|},|\mathbf{j}|-(J+1)\}}).$$

Consider the partition

$$\mathcal{D}_{\mathbf{j}}\sum_{\mathbf{a}\in I_{0,n}} W_{\mathbf{a},n}(\mathbf{x})E_{\mathbf{a}}(\mathbf{x}) = \mathcal{D}_{\mathbf{j}}\left\{\sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} W_{\mathbf{a},n}(\mathbf{x})E_{\mathbf{a}}(\mathbf{x})\right\} + \mathcal{D}_{\mathbf{j}}\left\{\sum_{\mathbf{a}\in I_{0,n}\cap\overline{I_{1,n}(\mathbf{x})}} W_{\mathbf{a},n}(\mathbf{x})E_{\mathbf{a}}(\mathbf{x})\right\}$$

for $\mathbf{j}$ with $0\le|\mathbf{j}|\le J$ and $\mathbf{x}\in\liminf_{n\to\infty} I_n$. Note that

$$\mathcal{D}_{\mathbf{j}}W_{\mathbf{a},n}(\mathbf{x}) = \frac{\sum_{\mathbf{z}_1\in I_{0,n}}\cdots\sum_{\mathbf{z}_{|\mathbf{j}|}\in I_{0,n}}\left\{S_{\mathbf{a}}(\mathbf{x})\prod_{k=1}^{|\mathbf{j}|}S_{\mathbf{z}_k}(\mathbf{x})\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}M_{k_l,l}\right\}}{(\sum_{\mathbf{c}\in I_{0,n}}\exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})])^{|\mathbf{j}|+1}} \tag{31}$$

for $\mathbf{j}=(j_1,\ldots,j_D)^T$ with $0\le|\mathbf{j}|\le J$, $\mathbf{a}\in I_{0,n}$, and $\mathbf{x}\in\liminf_{n\to\infty} I_n$, where $S_{\mathbf{z}}(\mathbf{x}) := \exp[-(\mathbf{x}-\mathbf{z})^T\mathbf{B}_n(\mathbf{x}-\mathbf{z})]$ and $M_{k_l,l} := 2\mathbf{b}_{l,n}^T(\mathbf{a}+\sum_{m=1}^{k_l-1}\mathbf{z}_m - k_l\mathbf{z}_{k_l})$ for $\mathbf{b}_{l,n}$ the l^{th} column of $\mathbf{B}_n$.

To show (31) holds, mathematical induction is employed. It is obvious that (31) holds when $|\mathbf{j}| = \mathbf{1}$. Suppose (31) holds with $r=|\mathbf{j}|$, next consider the case of $r+1$. Without loss of generality, we look at $\frac{\partial}{\partial x_1}\mathcal{D}_{\mathbf{j}}W_{\mathbf{a},n}(\mathbf{x})$. Since

$$\frac{\partial}{\partial x_1}S_{\mathbf{a}}(\mathbf{x}) = \mathbf{S}_{\mathbf{a}}(\mathbf{x})[-\mathbf{2b}^{\mathbf{T}}_{\mathbf{1},\mathbf{n}}(\mathbf{x}-\mathbf{a})],$$

$$\frac{\partial}{\partial x_1}S_{\mathbf{z}_k}(\mathbf{x}) = \mathbf{S}_{\mathbf{z_k}}(\mathbf{x})[-\mathbf{2b}^{\mathbf{T}}_{\mathbf{1},\mathbf{n}}(\mathbf{x}-\mathbf{z_k})],$$

$$\frac{\partial}{\partial x_1}\prod_{k=1}^{|\mathbf{j}|}S_{\mathbf{z}_k}(\mathbf{x}) = \prod_{\mathbf{k=1}}^{|\mathbf{j}|}\mathbf{S}_{\mathbf{z_k}}(\mathbf{x})\cdot\sum_{\mathbf{k=1}}^{|\mathbf{j}|}[-\mathbf{2b}^{\mathbf{T}}_{\mathbf{1},\mathbf{n}}(\mathbf{x}-\mathbf{z_k})],$$

$$\begin{aligned}&\frac{\partial}{\partial x_1}\left\{S_{\mathbf{a}}(\mathbf{x})\prod_{k=1}^{|\mathbf{j}|}S_{\mathbf{z}_k}(\mathbf{x})\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}M_{k_l,l}\right\}\\ =& \prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}M_{k_l,l}\left\{S_{\mathbf{a}}(\mathbf{x})\prod_{k=1}^{|\mathbf{j}|}S_{\mathbf{z}_k}(\mathbf{x})\cdot\left[-\mathbf{2b}^{\mathbf{T}}_{\mathbf{1},\mathbf{n}}\left(\mathbf{x}-\mathbf{a}+\sum_{\mathbf{k=1}}^{|\mathbf{j}|}(\mathbf{x}-\mathbf{z_k})\right)\right]\right\},\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial}{\partial x_1}(\sum_{\mathbf{c}\in I_{0,n}} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})])^{|\mathbf{j}|+1} \\
= \; & (|\mathbf{j}|+1)(\sum_{\mathbf{c}\in I_{0,n}} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})])^{|\mathbf{j}|} \\
& \left\{\sum_{\mathbf{c}\in I_{0,n}} \exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})][-2\mathbf{b}_{1,n}^T(\mathbf{x}-\mathbf{c})]\right\},
\end{aligned}$$

it follows that

$$\begin{aligned}
&\frac{\partial}{\partial x_1}\mathcal{D}_{\mathbf{j}}W_{\mathbf{a},n}(\mathbf{x}) = \\
&\frac{\sum_{\mathbf{z}_1\in I_{0,n}}\cdots\sum_{\mathbf{z}_{|\mathbf{j}|+1}\in I_{0,n}}\left\{S_{\mathbf{a}}(\mathbf{x})\prod_{k=1}^{|\mathbf{j}|+1}S_{\mathbf{z}_k}(\mathbf{x})\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}M_{k_l,l}\cdot M_{j_1+1,1}\right\}}{(\sum_{\mathbf{c}\in I_{0,n}}\exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})])^{|\mathbf{j}|+2}}
\end{aligned}$$

If $\mathbf{a}\in I_{0,n}\cap\overline{I_{1,n}(\mathbf{x})}=\{\mathbf{a}\in I_{0,n}:(\mathbf{x}-\mathbf{a})^t\mathbf{B}_n(\mathbf{x}-\mathbf{a})>n^{2\delta}\}$, for some positive constant C_2 we have

$$\begin{aligned}
&|\mathcal{D}_{\mathbf{j}}W_{\mathbf{a},n}(\mathbf{x})| \\
\leq\; & \frac{\sum_{\mathbf{z}_1\in I_{0,n}}\cdots\sum_{\mathbf{z}_{|\mathbf{j}|}\in I_{0,n}}\left\{S_{\mathbf{a}}(\mathbf{x})\prod_{k=1}^{|\mathbf{j}|}S_{\mathbf{z}_k}(\mathbf{x})\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}M_{k_l,l}|\right\}}{(\sum_{\mathbf{c}\in I_{0,n}}\exp[-(\mathbf{x}-\mathbf{c})^T\mathbf{B}_n(\mathbf{x}-\mathbf{c})])^{|\mathbf{j}|+1}} \\
\leq\; & \frac{card\{I_{0,n}\}^{|\mathbf{j}|}\exp[-n^{2\delta}]\cdot 1^{|\mathbf{j}|}}{(C_1 n^{\psi D}\exp[-1])^{|\mathbf{j}|+\mathbf{1}}} \\
& \frac{\cdot\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}n^{2(\gamma+\delta)}\left|2\mathbf{b}_{l,0}^T(\mathbf{a}+\sum_{m=1}^{k_l-1}\mathbf{z}_m-k_l\mathbf{z}_{k_l})\right|}{(C_1 n^{\psi D}\exp[-1])^{|\mathbf{j}|+\mathbf{1}}} \\
\leq\; & \frac{L_n^{D|\mathbf{j}|}\exp[-n^{2\delta}]n^{2(\gamma+\delta)|\mathbf{j}|}\prod_{l=1}^{D}\prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l}2k_l\left|2\mathbf{b}_{l,0}^T\right|\cdot\mathbf{1}}{(C_1\exp[-1])^{|\mathbf{j}|+\mathbf{1}}n^{\psi D(|\mathbf{j}|+\mathbf{1})}} \\
=\; & C_2 L_n^{D|\mathbf{j}|}\exp[-n^{2\delta}]n^{2(\gamma+\delta)|\mathbf{j}|}/n^{\psi D(|\mathbf{j}|+1)} \qquad (32)
\end{aligned}$$

where $\mathbf{b}_{l,0}$ is the l^{th} column of $\mathbf{B}_0$ and $\mathbf{1}$ is a column of 1's.

Recalling (20) and (28),

$$\begin{aligned}
|\mathcal{D}_{\mathbf{j}}E_{\mathbf{a}}(\mathbf{x})| &= |\mathcal{D}_{\mathbf{j}}A_{\mathbf{a}}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}T_{\mathbf{a}}(\mathbf{x})| \\
&\leq |\mathcal{D}_{\mathbf{j}}A_{\mathbf{a}}(\mathbf{x})| + |\mathcal{D}_{\mathbf{j}}T_{\mathbf{a}}(\mathbf{x})| \\
&\leq \left| \sum_{0 \leq |\mathbf{p}| \leq J-|\mathbf{j}|} (\widetilde{c}_{\mathbf{p}+\mathbf{j},\mathbf{a}} - c_{\mathbf{p}+\mathbf{j},\mathbf{a}}) \frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!} (\mathbf{x}-\mathbf{a})^{\mathbf{p}} \right| + C_0 \sum_{\mathbf{l}:|\mathbf{l}|=J+1-|\mathbf{j}|} |\mathbf{x}-\mathbf{a}|^{\mathbf{l}} \\
&\leq \sum_{0 \leq |\mathbf{p}| \leq J-|\mathbf{j}|} (C + |c_{\mathbf{p}+\mathbf{j},\mathbf{a}}|) \frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!} |\mathbf{x}-\mathbf{a}|^{\mathbf{p}} + C_0 \sum_{\mathbf{l}:|\mathbf{l}|=J+1-|\mathbf{j}|} |\mathbf{x}-\mathbf{a}|^{\mathbf{l}} \\
&\leq \sum_{0 \leq |\mathbf{p}| \leq J-|\mathbf{j}|} (C + |c_{\mathbf{p}+\mathbf{j},\mathbf{a}}|) \frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!} 2^{p_1} 2^{p_2} \cdots 2^{p_D} + C_0 \sum_{\mathbf{l}:|\mathbf{l}|=J+1-|\mathbf{j}|} 2^{l_1} 2^{l_2} \cdots 2^{l_D} \\
&= \sum_{0 \leq |\mathbf{p}| \leq J-|\mathbf{j}|} (C + |c_{\mathbf{p}+\mathbf{j},\mathbf{a}}|) \frac{(\mathbf{p}+\mathbf{j})!}{\mathbf{p}!} 2^{|\mathbf{p}|} + C_0 \sum_{\mathbf{l}:|\mathbf{l}|=J+1-|\mathbf{j}|} 2^{|\mathbf{l}|} \\
&= C_3 \qquad (33)
\end{aligned}$$

for some positive constant C_3. Since $card\{I_{0,n} \cap \overline{I_{1,n}(\mathbf{x})}\} \leq card\{I_{0,n}\} \leq L_n^D$,

$$\left| \sum_{\mathbf{a} \in I_{0,n} \cap \overline{I_{1,n}(\mathbf{x})}} \mathcal{D}_{\mathbf{l}} W_{\mathbf{a},n}(\mathbf{x}) \mathcal{D}_{\mathbf{j}-\mathbf{l}} E_{\mathbf{a}}(\mathbf{x}) \right| \leq C_2 C_3 L_n^{D(|\mathbf{l}|+1)} \exp[-n^{2\delta}] n^{2(\gamma+\delta)|\mathbf{l}|} / n^{\psi D(|\mathbf{l}|+1)}$$

for $\mathbf{j}$ and $\mathbf{l}$ with $0 \leq |\mathbf{j}| \leq J$ and $\mathbf{0} \leq \mathbf{l} \leq \mathbf{j}$. Since $\mathcal{D}_{\mathbf{j}} \left\{ \sum_{\mathbf{a} \in I_{0,n} \cap \overline{I_{1,n}(\mathbf{x})}} W_{\mathbf{a},n}(\mathbf{x}) E_{\mathbf{a}}(\mathbf{x}) \right\}$ consists of finitely many terms of $\sum_{\mathbf{a} \in I_{0,n} \cap \overline{I_{1,n}(\mathbf{x})}} \mathcal{D}_{\mathbf{l}} W_{\mathbf{a},n}(\mathbf{x}) \mathcal{D}_{\mathbf{j}-\mathbf{l}} E_{\mathbf{a}}(\mathbf{x})$, and $\exp[-n^{2\delta}]$ goes to 0 faster than any power of n, we conclude that

$$\mathcal{D}_{\mathbf{j}} \left\{ \sum_{\mathbf{a} \in I_{0,n} \cap \overline{I_{1,n}(\mathbf{x})}} W_{\mathbf{a},n}(\mathbf{x}) E_{\mathbf{a}}(\mathbf{x}) \right\} = O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\} + \gamma \max\{\max_{0 \leq \mathbf{m} \leq \mathbf{j}} |\mathbf{m}| - w_{|\mathbf{j}-\mathbf{m}|}, |\mathbf{j}| - (J+1)\}}). \qquad (34)$$

If $\mathbf{a} \in I_{0,n} \cap I_{1,n}(\mathbf{x}) = \{\mathbf{a} \in I_{0,n} : (\mathbf{x}-\mathbf{a})^t \mathbf{B}_n (\mathbf{x}-\mathbf{a}) \leq n^{2\delta}\}$, with $\max\{(\mathbf{z}_1 - \mathbf{x})^T \mathbf{B}_n (\mathbf{z}_1 - \mathbf{x}), \ldots, (\mathbf{z}_{|\mathbf{j}|} - \mathbf{x})^T \mathbf{B}_n (\mathbf{z}_{|\mathbf{j}|} - \mathbf{x})\} \geq n^{2\delta}$, the numerator of $\mathcal{D}_{\mathbf{j}} W_{\mathbf{a},n}(\mathbf{x})$ in (31) has at most L_n^D

summands and each summand is bounded above by

$$
\begin{aligned}
& \left| S_{\mathbf{a}}(\mathbf{x}) \prod_{k=1}^{|\mathbf{j}|} S_{\mathbf{z}_k}(\mathbf{x}) \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} M_{k_l,l} \right| \\
= & \left| S_{\mathbf{a}}(\mathbf{x}) \exp[-\sum_{k=1}^{|\mathbf{j}|} (\mathbf{x}-\mathbf{z}_k)^T \mathbf{B}_n (\mathbf{x}-\mathbf{z}_k)] \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} M_{k_l,l} \right| \\
\leq & \; S_{\mathbf{a}}(\mathbf{x}) \exp[-\max\{(\mathbf{z}_1-\mathbf{x})^T \mathbf{B}_n (\mathbf{z}_1-\mathbf{x}), \ldots, (\mathbf{z}_{|\mathbf{j}|}-\mathbf{x})^T \mathbf{B}_n (\mathbf{z}_{|\mathbf{j}|}-\mathbf{x})\}] \\
& \cdot n^{2(\gamma+\delta)|\mathbf{j}|} \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} 2k_l \left| 2\mathbf{b}_{l,0}^T \right| \cdot \mathbf{1} \\
\leq & \; 1 \cdot \exp[-n^{2\delta}] n^{2(\gamma+\delta)|\mathbf{j}|} \prod_{l=1}^{D} 4^{j_l} \frac{j_1+j_2+\cdots+j_l!}{j_1+j_2+\cdots+j_{l-1}!} \left| \mathbf{b}_{l,0}^T \right| \cdot \mathbf{1} \\
= & \; C_4 \exp[-n^{2\delta}] n^{2(\gamma+\delta)|\mathbf{j}|} \qquad (35)
\end{aligned}
$$

for some positive constant C_4, where $\mathbf{b}_{l,0}$ is the l^{th} column of $\mathbf{B}_0$ and $\mathbf{1}$ is a column of 1's.

If $\mathbf{a} \in I_{0,n} \cap I_{1,n}(\mathbf{x}) = \{\mathbf{a} \in I_{0,n} : (\mathbf{x}-\mathbf{a})^t \mathbf{B}_n (\mathbf{x}-\mathbf{a}) \leq n^{2\delta}\}$, and $\max\{(\mathbf{z}_1-\mathbf{x})^T \mathbf{B}_n (\mathbf{z}_1-\mathbf{x}), \ldots, (\mathbf{z}_{|\mathbf{j}|}-\mathbf{x})^T \mathbf{B}_n (\mathbf{z}_{|\mathbf{j}|}-\mathbf{x})\} < n^{2\delta}$, the numerator of $\mathcal{D}_{\mathbf{j}} W_{\mathbf{a},n}(\mathbf{x})$ in (31) has approximately $2^{D|\mathbf{j}|} \frac{(6L_0 D\lambda_0)^{D|\mathbf{j}|}}{\lambda_0^{D|\mathbf{j}|}} n^{(\psi+\gamma)D|\mathbf{j}|} = C_5 n^{(D\psi+D\delta)|\mathbf{j}|}$ summands, where C_5 is a positive constant. Each summand is bounded above by

$$
\begin{aligned}
& \left| S_{\mathbf{a}}(\mathbf{x}) \prod_{k=1}^{|\mathbf{j}|} S_{\mathbf{z}_k}(\mathbf{x}) \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} M_{k_l,l} \right| \\
\leq & \; 1 \cdot \prod_{k=1}^{|\mathbf{j}|} 1 \cdot \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} \left| 2\mathbf{b}_{l,n}^T [(\mathbf{a}-\mathbf{x}) + \sum_{m=1}^{k_l-1} (\mathbf{z}_m-\mathbf{x}) - k_l(\mathbf{z}_{k_l}-\mathbf{x})] \right| \\
\leq & \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} \left[\left| 2\mathbf{b}_{l,n}^T (\mathbf{a}-\mathbf{x}) \right| + \sum_{m=1}^{k_l-1} \left| 2\mathbf{b}_{l,n}^T (\mathbf{z}_m-\mathbf{x}) \right| + k_l \left| 2\mathbf{b}_{l,n}^T (\mathbf{z}_{k_l}-\mathbf{x}) \right| \right].
\end{aligned}
$$

Since $(\mathbf{z}-\mathbf{x})^T \mathbf{B}_n (\mathbf{z}-\mathbf{x}) \leq n^{2\delta}$ for all $\mathbf{z} \in I_{0,n}$, we have $(\mathbf{z}-\mathbf{x})^T \mathbf{B}_0 (\mathbf{z}-\mathbf{x}) \leq n^{-2\gamma}$. Therefore,

$|\mathbf{b}_{l,n}^T(\mathbf{z}-\mathbf{x})| = n^{2(\gamma+\delta)}\sqrt{(\mathbf{z}-\mathbf{x})^T\mathbf{b}_{l,0}\mathbf{b}_{l,0}^T(\mathbf{z}-\mathbf{x})} \le n^{2(\gamma+\delta)}C_6n^{-\gamma} = C_6n^{2\delta+\gamma}$. Hence,

$$\begin{aligned}
&\left| S_{\mathbf{a}}(\mathbf{x}) \prod_{k=1}^{|\mathbf{j}|} S_{\mathbf{z}_k}(\mathbf{x}) \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} M_{k_l,l} \right| \\
\le\;& \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} \left[2C_6n^{2\delta+\gamma} + 2(k_l-1)C_6n^{2\delta+\gamma} + 2k_lC_6n^{2\delta+\gamma}\right] \\
=\;& \prod_{l=1}^{D} \prod_{k_l=j_1+j_2+\cdots+j_{l-1}+1}^{j_1+j_2+\cdots+j_l} 4k_lC_6n^{2\delta+\gamma} \\
=\;& (4C_6)^{|\mathbf{j}|}\mathbf{j}!n^{(2\delta+\gamma)|\mathbf{j}|} \\
=\;& C_7n^{(2\delta+\gamma)|\mathbf{j}|} \qquad (36)
\end{aligned}$$

for some positive constant C_7.

Combining (30), (35) and (36), we have

$$\begin{aligned}
\sup_{\mathbf{a}\in I_{1,n}(\mathbf{x})} |\mathcal{D}_{\mathbf{j}}W_{\mathbf{a},n}(\mathbf{x})| \;\le\;& \left\{C_4L_n^{D|\mathbf{j}|}\exp[-2\delta]n^{2(\gamma+\delta)|\mathbf{j}|} + C_5n^{(D\psi+D\delta)|\mathbf{j}|}\cdot C_7n^{(2\delta+\gamma)|\mathbf{j}|}\right\} \qquad (37) \\
&\times(C_1n^{\psi D}\exp[-1])^{-(|\mathbf{j}|+1)} \\
=\;& O(n^{(D\psi+D\delta)|\mathbf{j}|+(2\delta+\gamma)|\mathbf{j}|-\psi D(|\mathbf{j}|+1)}) \\
=\;& O(n^{-\psi D+(D+2)\delta|\mathbf{j}|+\gamma|\mathbf{j}|})
\end{aligned}$$

for $\mathbf{x} \in \liminf_{n\to\infty} I_n$.

For any $\phi_n > 0$ and $\mathbf{l}$ with $0 \le |\mathbf{l}| \le J$, under assumption (21), we have

$$
\begin{aligned}
& \mathbb{P}\left[\sup_{\mathbf{a}\in I_{1,n}(\mathbf{x})} |\mathcal{D}_{\mathbf{l}} A_{\mathbf{a}}(\mathbf{x})| \ge \phi_n\right] \\
\le\ & \mathbb{P}\left[\bigcup_{\mathbf{a}\in I_{1,n}(\mathbf{x})} |\mathcal{D}_{\mathbf{l}} A_{\mathbf{a}}(\mathbf{x})| \ge \phi_n\right] \\
\le\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathbb{P}\left[|\mathcal{D}_{\mathbf{l}} A_{\mathbf{a}}(\mathbf{x})| \ge \phi_n\right] \qquad (38) \\
=\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathbb{P}\left[\left|\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} (\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}) \frac{(\mathbf{j}+\mathbf{l})!}{\mathbf{j}!}(\mathbf{x}-\mathbf{a})^{\mathbf{j}}\right| \ge \phi_n\right] \\
\le\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathbb{P}\left[\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} \frac{(\mathbf{j}+\mathbf{l})!}{\mathbf{j}!} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| \frac{n^{-\gamma|\mathbf{j}|}}{(2L_0 D)^{|\mathbf{j}|}} \ge \phi_n\right] \qquad (39) \\
\le\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathbb{P}\left[\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} C_8 |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{j}|} \ge \phi_n\right] \\
\le\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \phi_n^{-2} C_8^2\ \mathbb{E}\left[\left(\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{j}|}\right)^2\right] \qquad (40) \\
=\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \phi_n^{-2} C_8^2\ \mathbb{E}\left(\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{j}|}\right) \\
& \cdot \left(\sum_{\mathbf{k}:0\le|\mathbf{k}|\le J-|\mathbf{l}|} |\widetilde{c}_{\mathbf{k}+\mathbf{l},\mathbf{a}} - c_{\mathbf{k}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{k}|}\right) \\
\le\ & \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \phi_n^{-2} C_8^2 \sqrt{\mathbb{E}\left(\sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{j}|}\right)^2} \\
& \cdot \sqrt{\mathbb{E}\left(\sum_{\mathbf{k}:0\le|\mathbf{k}|\le J-|\mathbf{l}|} |\widetilde{c}_{\mathbf{k}+\mathbf{l},\mathbf{a}} - c_{\mathbf{k}+\mathbf{l},\mathbf{a}}| n^{-\gamma|\mathbf{k}|}\right)^2} \\
\le\ & \phi_n^{-2} C_9\ card(I_{1,n}(\mathbf{x})) \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} \sum_{\mathbf{k}:0\le|\mathbf{k}|\le J-|\mathbf{l}|} n^{-\alpha_{|\mathbf{j}+\mathbf{l}|}} n^{-\alpha_{|\mathbf{k}+\mathbf{l}|}} n^{-\gamma|\mathbf{j}|} n^{-\gamma|\mathbf{k}|} \\
<\ & \phi_n^{-2} C_9 O(n^{D(\psi+\delta)}) \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} \sum_{\mathbf{k}:0\le|\mathbf{k}|\le J-|\mathbf{l}|} n^{-\alpha_{|\mathbf{l}|}+|\mathbf{j}|\gamma} n^{-\alpha_{|\mathbf{l}|}+|\mathbf{k}|\gamma} n^{-\gamma|\mathbf{j}|} n^{-\gamma|\mathbf{k}|} \\
\le\ & \phi_n^{-2} O(n^{\psi D+\delta D-2\alpha_{|\mathbf{l}|}})
\end{aligned}
$$

where C_8 and C_9 are positive constants. Line (38) uses the Bonferroni inequality. Line (39) comes from the fact that $|x_k - a_k| \le \frac{n^{(\gamma+\delta)}}{L_n}\lambda_0 \le \frac{n^{-\gamma}}{2L_0 D}$ for $k = 1, \ldots, D$. Chebyshev's inequality is employed in line (40). And the last four lines use Cauchy-Schwarz inequality and assumption (24). Let $\phi_n = n^{-w_{|\mathbf{l}|}\gamma}$, assumption (23) yields

$$\sup_{\mathbf{a}\in I_{1,n}(\mathbf{x})} |\mathcal{D}_{\mathbf{l}} A_{\mathbf{a}}(\mathbf{x})| = O_p(n^{-w_{|\mathbf{l}|}\gamma}).$$

On the other hand, from (28) we obtain

$$\begin{aligned} \sup_{\mathbf{a}\in I_{1,n}(\mathbf{x})} |\mathcal{D}_{\mathbf{l}} T_{\mathbf{a}}(\mathbf{x})| &\leq C_0 \sum_{\mathbf{q}:|\mathbf{q}|=J+1-|\mathbf{l}|} \prod_{k=1}^{D} \frac{n^{-\gamma q_k}}{(2L_0 D)^{q_k}} \\ &= C_0 \sum_{\mathbf{q}:|\mathbf{q}|=J+1-|\mathbf{l}|} \frac{n^{-\gamma|\mathbf{q}|}}{(2L_0 D)^{|\mathbf{q}|}} = O(n^{-\gamma(J+1-|\mathbf{l}|)}). \end{aligned} \tag{41}$$

Therefore,

$$\begin{aligned} & \left| \sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathcal{D}_{\mathbf{l}} W_{\mathbf{a},n}(\mathbf{x}) \mathcal{D}_{\mathbf{j}-\mathbf{l}} E_{\mathbf{a}}(\mathbf{x}) \right| \\ \leq\; & O(n^{\psi D+\delta D}) \times O(n^{-\psi D+(D+2)\delta|\mathbf{l}|+\gamma|\mathbf{l}|}) \times \{O_p(n^{-w_{|\mathbf{j}-\mathbf{l}|}\gamma}) + O(n^{-\gamma(J+1-|\mathbf{j}|+|\mathbf{l}|)})\} \\ =\; & O_p(n^{\delta\{D+(D+2)|\mathbf{l}|\}+\gamma\max\{|\mathbf{l}|-w_{|\mathbf{j}-\mathbf{l}|},|\mathbf{j}|-(J+1)\}}) \\ =\; & O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\}+\gamma\max\{\max_{\mathbf{0}\leq\mathbf{m}\leq\mathbf{j}}|\mathbf{m}|-w_{|\mathbf{j}-\mathbf{m}|},|\mathbf{j}|-(J+1)\}}) \end{aligned}$$

for $\mathbf{j}$ and $\mathbf{l}$ with $0 \leq |\mathbf{j}| \leq J$ and $\mathbf{0} \leq \mathbf{l} \leq \mathbf{j}$. Since $\mathcal{D}_{\mathbf{j}}\left\{\sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} W_{\mathbf{a},n}(\mathbf{x}) E_{\mathbf{a}}(\mathbf{x})\right\}$ consists of finitely many terms of $\sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} \mathcal{D}_{\mathbf{l}} W_{\mathbf{a},n}(\mathbf{x}) \mathcal{D}_{\mathbf{j}-\mathbf{l}} E_{\mathbf{a}}(\mathbf{x})$, we conclude that

$$\mathcal{D}_{\mathbf{j}}\left\{\sum_{\mathbf{a}\in I_{1,n}(\mathbf{x})} W_{\mathbf{a},n}(\mathbf{x}) E_{\mathbf{a}}(\mathbf{x})\right\} = O_p(n^{\delta\{D+(D+2)|\mathbf{j}|\}+\gamma\max\{\max_{\mathbf{0}\leq\mathbf{m}\leq\mathbf{j}}|\mathbf{m}|-w_{|\mathbf{j}-\mathbf{m}|},|\mathbf{j}|-(J+1)\}}).$$

Together with (34), this yields result (25).

Under assumption (21), for $\mathbf{l}$ with $0 \le |\mathbf{l}| \le J$

$$\begin{aligned}
& \sup_{\mathbf{x}\in I_n} \sup_{\mathbf{a}\in I_{0,n}} |\mathcal{D}_{\mathbf{l}} A_{\mathbf{a}}(\mathbf{x})| \\
= \;& \sup_{\mathbf{x}\in I_n} \sup_{\mathbf{a}\in I_{0,n}} \left| \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} (\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}) \frac{(\mathbf{j}+\mathbf{l})!}{\mathbf{j}!} (\mathbf{x}-\mathbf{a})^{\mathbf{j}} \right| \\
\le \;& C_{10} \sup_{\mathbf{x}\in I_n} \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} \sup_{\mathbf{a}\in I_{0,n}} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| \, |\mathbf{x}-\mathbf{a}|^{\mathbf{j}} \\
\le \;& C_{11} \sup_{\mathbf{x}\in I_n} \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} \sup_{\mathbf{a}\in I_{0,n}} |\widetilde{c}_{\mathbf{j}+\mathbf{l},\mathbf{a}} - c_{\mathbf{j}+\mathbf{l},\mathbf{a}}| \, n^{-\gamma|\mathbf{j}|} \\
\le \;& C_{11} \sup_{\mathbf{x}\in I_n} \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} n^{-\alpha_{|\mathbf{j}+\mathbf{l}|}} n^{-\gamma|\mathbf{j}|} \\
\le \;& C_{11} \sum_{\mathbf{j}:0\le|\mathbf{j}|\le J-|\mathbf{l}|} n^{-\alpha_{|\mathbf{l}|}+\gamma|\mathbf{j}|} n^{-\gamma|\mathbf{j}|} \\
= \;& O_p(n^{-\alpha_{|\mathbf{l}|}}) \\
= \;& O_p(n^{-w_{|\mathbf{l}|}\gamma}),
\end{aligned}$$

where C_{10} and C_{11} are positive constants. Since (32), (33), (37), and (41) hold uniformly over $\mathbf{x} \in I_n$, result (26) is obtained.

Note that the convergence rate in Theorem 2.1 depends on $w_1, \ldots, w_J$ given $\alpha_1, \ldots, \alpha_J$. The following corollary shows that the compound estimator and its derivatives almost achieve the optimal convergence rate, assuming $\alpha_k = (J+1-k-\nu)/\{2(J+1)+D\}$ for $k \in \{0, 1, \ldots, J\}$ and $\nu \in (0, 1/(2J+2+D))$ an arbitrary constant.

Corollary 2.1 *Let $\nu \in (0, 1/(2J+2+D))$ be an arbitrary constant. If $\alpha_k = (J+1-k-\nu)/\{2(J+1)+D\}$ for $k \in \{0,1,\ldots,J\}$, then under assumption (21), there exist $\mathbf{B}_n$ and L_n such that*

$$\max_{\mathbf{j}:0\le|\mathbf{j}|\le J} \left(n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu} \{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\} \right) \xrightarrow{P} 0 \tag{42}$$

for $\mathbf{x} \in \liminf_{n\to\infty} I_n$. And under assumption (22), there exist $\mathbf{B}_n$ and L_n such that

$$\max_{\mathbf{j}:0\le|\mathbf{j}|\le J} \sup_{\mathbf{x}\in I_n} \left(n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu} \{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\} \right) \xrightarrow{P} 0. \tag{43}$$

Proof of Corollary 2.1. Let $\gamma := 1/\{2(J+1)+D\}$, $\delta := \{\nu/(2D)\}/\{2(J+1)+D\}$, $\psi := \{\nu/(2D)\}/\{2(J+1)+D\}$, and $w_k := J+1-k-2\nu$ for $k \in \{0,1,\ldots,J\}$. Then condition (23) in Theorem 2.1 is satisfied since

$$\psi D + \delta D + 2w_k\gamma - 2\alpha_k = \frac{2(J+1)-2k-3\nu-[2(J+1)-2k-2\nu]}{2(J+1)+D} = \frac{-\nu}{2(J+1)+D} < 0$$

for $k \in \{0,1,\ldots,J\}$, and condition (24) is satisfied since

$$\alpha_k - \alpha_{k+1} = 1/\{2(J+1)+D\} = \gamma$$

for $k \in \{0,1,\ldots,J-1\}$. Therefore, by Theorem 2.1, if (21) holds, since $\delta\{D+(D+2)|\mathbf{j}|\} + \gamma\max\{\max_{0\le\mathbf{m}\le\mathbf{j}}|\mathbf{m}| - w_{|\mathbf{j}-\mathbf{m}|}, |\mathbf{j}|-(J+1)\} = \delta\{D+(D+2)|\mathbf{j}|\} + \gamma(|\mathbf{j}|-(J+1)+2\nu)$, we have

$$\begin{aligned} n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu}\{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\} &= O_p\left(n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu+\delta\{D+(D+2)|\mathbf{j}|\}+\gamma(|\mathbf{j}|-(J+1)+2\nu)}\right) \\ &= O_p\left(n^{\frac{\frac{|\mathbf{j}|}{2}+\frac{1}{2}+\frac{|\mathbf{j}|}{D}-2J-D}{2(J+1)+D}\nu}\right) \\ &= O_p\left(n^{\frac{\frac{J}{2}+\frac{1}{2}+J-2J-D}{2(J+1)+D}\nu}\right) \\ &= O_p\left(n^{-\frac{(J+1)/2}{2(J+1)+D}\nu}\right) \end{aligned}$$

for $\mathbf{j}$ with $0 \le |\mathbf{j}| \le J$. Thus, $n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu}\{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\} = o_p(1)$. Since

$$\max_{\mathbf{j}:0\le|\mathbf{j}|\le J}\left(n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu}\{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\}\right)$$

has finitely many terms of $n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu}\{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\}$,

$$\max_{\mathbf{j}:0\le|\mathbf{j}|\le J}\left(n^{\frac{J+1-|\mathbf{j}|}{2(J+1)+D}-\nu}\{\mathcal{D}_{\mathbf{j}}\widehat{\mu}(\mathbf{x}) - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\}\right) \xrightarrow{P} 0.$$

Result (43) can be obtained similarly.

Stone (1980) showed that the optimal convergence rate for nonparametric estimators is

$\frac{J+1-|\mathbf{j}|}{2(J+1)+D}$ under certain assumptions. Corollary (2.1) states that the compound estimator and its derivatives obtain near optimal convergence rate. Meanwhile, the self-consistency property of compound estimation makes it perform better than spline smoothing or local regression in estimating $\mu(\mathbf{x})$ and its derivatives.

2.4 Filtration and extrapolation

In this section we propose an approach to filter some of the noise in the observed data, as well as to overcome the boundary issue in finite samples.

Suppose we observe data $(\mathbf{x}_i, Y_i)_{i=1}^n$ from model (15). Because of the noise in the observed data, better estimation in the mean response function and its derivatives would be expected if we had ideal data $(\mathbf{x}_i, \mu(\mathbf{x}_i))_{i=1}^n$. In practice, the ideal data are not available since $\mu(\mathbf{x})$ is unknown. However, we can still use better approximation of $\mu(\mathbf{x}_i)$ than Y_i .

An intuitive thought would be to apply any estimation method to approximate each $\mu(\mathbf{x}_i)$ and replace the original noisy data $(\mathbf{x}_i, Y_i))_{i=1}^n$ with the resulting approximation. One option is compound estimation because of the advatages over other major nonparametric estimation methods: near optimal convergence rate and self-consistency. Therefore, we perform compound estimation, setting J to be 0 in (16), to get synthetic data $(\mathbf{x}_i, \mu_0^*(\mathbf{x}_i))_{i=1}^n$. The subscript 0 in $\mu_0^*(\mathbf{x}_i)$ means the choice of J. Then we can apply compound estimation to the synthetic data with J set to a reasonable value.

A problem with the method above is that $\mu_0^*(\mathbf{x}_i)$ might perform poorly on the points near the boundaries. To fix this issue, we perform another compoud estimation with J set to 1, and define

$$\mu_{0,1}^*(\mathbf{x}) := (1 - \|\mathbf{x}\|_\infty)\, \mu_0^*(\mathbf{x}) + \|\mathbf{x}\|_\infty \mu_1^*(\mathbf{x}), \tag{44}$$

where $\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_D|)$, for $\mathbf{x} \in [-1, 1]^D$. Then apply compound estimation on synthetic data $\left(\mathbf{x}_i, \mu_{0,1}^*(\mathbf{x}_i)\right)_{i=1}^n$ with J set to a reasonable value. The idea of (44) is that $\mu_{0,1}^*(\mathbf{x})$ should be close to $\mu_1^*(\mathbf{x})$ when $\mathbf{x}$ is near a boundary point and close to $\mu_0^*(\mathbf{x})$ when $\mathbf{x}$ is not near a boundary point.

A possible issue with finite samples is the boundary effect since the pointwise estimators may not do well near boundary points (Loader, 1999). One way to overcome the issue is to extend data from $[-1,1]^D$ to $[-(1+\kappa_1),(1+\kappa_1)]\times[-(1+\kappa_2),(1+\kappa_2)]\times\cdots\times[-(1+\kappa_D),(1+\kappa_D)]$ where $\kappa_1,\ldots,\kappa_D$ are positive constants. Then the data points near boundaries in the original data set will become interior points in the extended data set.

To do so, let $n_i^* := \lfloor n(1+\kappa_i)\rfloor$, $n^* := \prod_{i=1}^D n_i^*$ and $x_{ij}^* := \left(-1+\frac{2j-1}{n_i^*}\right)(1+\kappa_i)$ for $i=1,2,\ldots,D$ and $j=1,2,\ldots,n_i^*$. Define

$$Z_i := \{1-\min(\|\mathbf{x}_i\|_\infty,1)\}\,\mu_0^*(\mathbf{x}_i^*)+\min(\|\mathbf{x}_i\|_\infty,1)\mu_1^*(\mathbf{x}_i^*)$$

for $1\le i\le n^*$, where $\mathbf{x}_i^*=(x_{1i}^*,x_{2i}^*,\ldots,x_{Di}^*)$. Then apply compound estimation to synthetic data $(\mathbf{x}_i^*,Z_i)_{i=1}^{n^*}$.

2.5 Numerical studies

In this section, we compare the performances of different nonparametric regression methods estimating up to the second order derivative when $\mathbf{x}=(x_1,x_2)\in[-1,1]\times[-1,1]$. Three test functions, $y=\cos(2\pi x_1)+\sin(2\pi x_2)$, $y=e^{x_1-x_2}$ and $y=(x_1-x_2)^4$ are used to generate 10 data sets of size 16^2 from model (15) with $\epsilon_i \overset{iid}{\sim} N(0,0.5^2)$ and $\mathbf{x}_i$'s equispaced on $[-1,1]\times[-1,1]$. Then we apply the following regression methods:

Method 1: Local regression. Choose the nearest neighbor fraction α from

$\{0.04,0.07,0.1,0.13\}$. Use tricube weights and local polynomial of degree 3 (to match the degree used in spline smooth). In R, the locfit function in locfit package is employed.

Method 2: Spline smoothing. Consider $\lambda\in\{10^{-8},2\times10^{-8},4\times10^{-8}\}$ and degrees of freedom $df=3$. The R function Tps from fields package is employed.

Method 3: Compound estimation without filtration and extrapolation. Consider $\mathbf{B}_n = \begin{bmatrix} b_n & 0\\ 0 & b_n\end{bmatrix}$ with $b_n\in\{10,20,30,40,50\}$ and $\alpha\in\{0.04,0.07,0.1,0.13\}$. For pointwise estimators, use local regression with tricube weights and spline smoothing with $\lambda\in\{10^{-8},2\times10^{-8},4\times10^{-8}\}$, based on local polynomial of degree 3, use $3^2\times3^2$ centering points.

Method 4: Compound estimation with filtration and extrapolation. Consider $\kappa_1=\kappa_2=0.2$, $\alpha_{FE}\in\{0.01,0.02,0.04\}$, $b_{FE}\in\{250,500,1000\}$ and $\lambda_{FE}\in\{10^{-9},2\times10^{-9},4\times10^{-9}\}$. Perform filtration and extrapolation process using bandwidth α_{FE} and tricube weights for the local regression pointwise estimators and smoothing parameter λ_{FE} for the spline smoothing. Also, b_{FE} denotes diagonal entry of the convolution matrix during filtration and extrapolation. The result is synthetic data $(x_i^*, Z_i)_{i=1}^{n^*}$. Then apply compound estimation to the synthetic data with tuning parameters α, λ and b_n selected by the ones producing the best results in compound estimation using local regression or spline smoothing pointwise estimators.

For evaluation purposes, we apply each method on a specified grid defined in $[-1,1] \times [-1,1]$ and calculate the average of

$$\frac{9}{4}\sum_{i=1}^{16^2}\left[\widehat{\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}_i)} - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x}_i)\right]^2 \approx \int_{-1}^{1}\int_{-1}^{1}\left[\widehat{\mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})} - \mathcal{D}_{\mathbf{j}}\mu(\mathbf{x})\right]^2 dx_1 dx_2$$

over the 10 simulated data sets for $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2$. Ideally, it has value zero. The regression methods which produce sum of squared errors close to zero are considered to be good ones.

Table 2: test function $\cos(2\pi x_1) + \sin(2\pi x_2)$

Strategy	y	$\frac{\partial y}{\partial x_1}$	$\frac{\partial y}{\partial x_2}$	$\frac{\partial^2 y}{\partial x_1{}^2}$	$\frac{\partial^2 y}{\partial x_2{}^2}$	$\frac{\partial^2 y}{\partial x_1 \partial x_2}$
$LR_{0.04}$	0.07	72.49	70.93	7545.48	13083.24	68.43
$LR_{0.07}$	0.01	0.67	1.04	1875.70	2243.38	1.57
$LR_{0.10}$	0.03	0.60	0.57	1811.08	2176.97	0.75
$LR_{0.13}$	0.09	2.88	1.51	2376.35	3933.30	0.64
$SS_{10^{-8}}$	0.003	0.12	0.10	683.32	365.56	13.58
$SS_{2\times10^{-8}}$	0.003	0.12	0.09	731.48	372.02	13.72
$SS_{4\times10^{-8}}$	0.003	0.13	0.08	828.93	398.86	7.36
$CELR_{10;0.04}$	0.39	0.21	1.63	288.93	169.80	2.39
$CELR_{10;0.07}$	0.53	4.81	11.86	453.45	142.92	2.39
$CELR_{10;0.1}$	0.51	5.82	13.15	471.81	6.73	1.16
$CELR_{10;0.13}$	0.40	3.95	12.51	352.08	3.60	4.05
$CELR_{20;0.04}$	0.15	1.16	0.71	628.49	965.92	6.22

$CELR_{20;0.07}$	0.16	1.61	2.34	229.72	339.08	0.54
$CELR_{20;0.1}$	0.13	1.33	1.99	249.53	14.05	1.05
$CELR_{20;0.13}$	0.10	0.74	2.04	174.61	8.12	3.93
$CELR_{30;0.04}$	0.07	3.99	1.00	872.19	803.69	9.56
$CELR_{30;0.07}$	0.06	0.30	0.52	157.06	261.55	1.01
$CELR_{30;0.1}$	0.05	0.04	0.22	141.45	11.43	0.61
$CELR_{30;0.13}$	0.03	0.05	0.34	68.13	8.13	2.25
$CELR_{40;0.04}$	0.03	6.91	2.03	955.24	222.18	10.86
$CELR_{40;0.07}$	0.03	0.12	0.28	29.21	58.16	1.75
$CELR_{40;0.1}$	0.02	0.13	0.09	30.98	2.39	0.44
$CELR_{40;0.13}$	0.01	0.34	0.14	21.11	15.09	1.43
$CELR_{50;0.04}$	0.01	9.75	3.87	980.41	139.66	9.46
$CELR_{50;0.07}$	0.01	0.11	0.28	49.55	13.61	1.41
$CELR_{50;0.1}$	0.01	0.27	0.16	220.83	23.73	0.35
$CELR_{50;0.13}$	0.003	0.54	0.15	295.68	41.65	1.06
$CELRFE_{40;0.1;250;0.01}$	0.11	1.41	2.42	294.31	389.39	0.53
$CELRFE_{40;0.1;250;0.02}$	0.02	1.27	4.21	56.36	115.57	1.00
$CELRFE_{40;0.1;250;0.04}$	0.03	2.51	2.70	38.05	262.47	0.22
$CELRFE_{40;0.1;500;0.01}$	0.13	1.82	6.59	348.75	363.69	0.65
$CELRFE_{40;0.1;500;0.02}$	0.03	1.65	7.52	45.43	122.46	0.94
$CELRFE_{40;0.1;500;0.04}$	0.05	2.98	5.36	28.77	269.32	1.52
$CELRFE_{40;0.1;1000;0.01}$	0.14	1.88	8.16	363.53	358.76	0.68
$CELRFE_{40;0.1;1000;0.02}$	0.04	1.73	8.56	42.97	123.54	1.11
$CELRFE_{40;0.1;1000;0.04}$	0.05	3.07	6.21	26.75	270.38	2.07
$CESS_{30;10^{-8}}$	0.43	2.15	1.28	1634.44	2127.33	14.14
$CESS_{30;2\times10^{-8}}$	0.43	2.20	1.29	1639.35	2120.32	11.64
$CESS_{30;4\times10^{-8}}$	0.43	2.28	1.32	1667.08	2109.13	9.50
$CESS_{40;10^{-8}}$	0.18	0.65	0.59	860.90	1122.36	20.97
$CESS_{40;2\times10^{-8}}$	0.18	0.59	0.59	868.47	1121.74	17.10

$CESS_{40;4\times10^{-8}}$	0.18	0.50	0.59	903.37	1122.93	13.63
$CESS_{50;10^{-8}}$	0.08	1.25	1.13	540.21	386.54	22.19
$CESS_{50;2\times10^{-8}}$	0.08	1.13	1.13	574.12	390.00	17.94
$CESS_{50;4\times10^{-8}}$	0.08	0.95	1.12	651.33	402.17	14.27
$CESSFE_{50;10^{-8};250;10^{-9}}$	0.55	115.11	202.61	25832.60	15629.09	590.01
$CESSFE_{50;10^{-8};250;2\times10^{-9}}$	0.55	115.08	202.62	25758.37	15618.79	585.63
$CESSFE_{50;10^{-8};250;4\times10^{-9}}$	0.55	114.78	202.53	25763.13	15594.51	585.46
$CESSFE_{50;10^{-8};500;10^{-9}}$	0.63	174.66	272.60	33393.60	15342.73	734.18
$CESSFE_{50;10^{-8};500;2\times10^{-9}}$	0.63	174.60	272.54	33482.29	15316.99	734.23
$CESSFE_{50;10^{-8};500;4\times10^{-9}}$	0.63	174.50	271.99	33405.39	15285.77	734.30
$CESSFE_{50;10^{-8};1000;10^{-9}}$	0.66	188.76	293.13	35521.62	15219.29	777.38
$CESSFE_{50;10^{-8};1000;2\times10^{-9}}$	0.66	188.70	293.05	35519.45	15209.45	777.45
$CESSFE_{50;10^{-8};1000;4\times10^{-9}}$	0.66	188.59	292.92	35516.59	15189.78	777.56

Table 3: test function $y = e^{x_1 - x_2}$

Strategy	y	$\frac{\partial y}{\partial x_1}$	$\frac{\partial y}{\partial x_2}$	$\frac{\partial^2 y}{\partial {x_1}^2}$	$\frac{\partial^2 y}{\partial {x_2}^2}$	$\frac{\partial^2 y}{\partial x_1 \partial x_2}$
$LR_{0.04}$	0.0024	7.64	7.63	1030.48	1034.42	26.75
$LR_{0.07}$	0.0006	0.05	0.07	95.17	97.27	1.25
$LR_{0.10}$	0.0003	0.02	0.03	6.22	5.13	0.65
$LR_{0.13}$	0.0002	0.01	0.02	3.46	2.38	0.36
$SS_{10^{-8}}$	0.0007	0.07	0.08	30.85	22.57	150.59
$SS_{2\times10^{-8}}$	0.0006	0.53	0.06	23.62	17.54	145.01
$SS_{4\times10^{-8}}$	0.0005	0.04	0.05	16.38	12.36	139.44
$CELR_{10;0.04}$	0.13	2.48	2.48	387.38	383.30	9.24
$CELR_{10;0.07}$	0.01	0.11	0.12	34.76	37.00	0.85
$CELR_{10;0.1}$	0.002	0.03	0.03	2.44	2.33	0.03
$CELR_{10;0.13}$	0.002	0.03	0.03	1.52	1.62	0.21

$CELR_{20;0.04}$	0.10	1.78	1.87	743.22	731.55	8.30
$CELR_{20;0.07}$	0.003	0.05	0.07	69.73	70.57	1.31
$CELR_{20;0.1}$	0.0004	0.02	0.02	4.96	4.78	0.64
$CELR_{20;0.13}$	0.0004	0.02	0.02	2.63	2.62	0.38
$CELR_{30;0.04}$	0.05	2.44	2.52	1030.10	1011.32	13.60
$CELR_{30;0.07}$	0.002	0.06	0.07	85.22	86.25	2.38
$CELR_{30;0.1}$	0.0004	0.03	0.03	6.89	5.90	1.07
$CELR_{30;0.13}$	0.0003	0.02	0.02	3.47	3.00	0.58
$CELR_{40;0.04}$	0.02	3.61	3.67	1329.78	1322.24	18.29
$CELR_{40;0.07}$	0.001	0.08	0.09	90.66	92.48	3.06
$CELR_{40;0.1}$	0.0003	0.03	0.03	7.62	6.24	1.27
$CELR_{40;0.13}$	0.0002	0.02	0.02	3.80	3.04	0.66
$CELR_{50;0.04}$	0.01	4.55	4.58	1621.09	1634.79	18.55
$CELR_{50;0.07}$	0.001	0.08	0.10	95.91	97.98	2.95
$CELR_{50;0.1}$	0.0004	0.03	0.03	7.67	6.36	1.21
$CELR_{50;0.13}$	0.0002	0.02	0.02	3.89	3.01	0.62
$CELRFE_{20;0.13;250;0.01}$	0.16	3.96	6.75	811.53	1064.21	29.94
$CELRFE_{20;0.13;250;0.02}$	0.02	0.30	0.32	23.82	27.32	1.97
$CELRFE_{20;0.13;250;0.04}$	0.02	0.30	0.32	21.28	22.68	3.25
$CELRFE_{20;0.13;500;0.01}$	0.15	3.31	5.00	721.74	945.74	24.62
$CELRFE_{20;0.13;500;0.02}$	0.03	0.60	0.62	34.54	38.36	3.06
$CELRFE_{20;0.13;500;0.04}$	0.03	0.58	0.59	30.89	32.64	4.28
$CELRFE_{20;0.13;1000;0.01}$	0.15	3.29	4.78	709.57	929.31	23.99
$CELRFE_{20;0.13;1000;0.02}$	0.04	0.70	0.71	37.99	41.93	3.39
$CELRFE_{20;0.13;1000;0.04}$	0.03	0.66	0.68	34.12	35.97	4.61
$CESS_{10;4\times10^{-8}}$	0.005	0.08	0.10	3.38	4.94	1.73
$CESS_{20;4\times10^{-8}}$	0.002	0.06	0.06	12.27	13.95	4.92
$CESS_{30;4\times10^{-8}}$	0.001	0.09	0.09	17.99	20.36	11.76
$CESS_{40;4\times10^{-8}}$	0.001	0.11	0.12	19.60	23.98	16.40

$CESS_{50;4\times10^{-8}}$	0.001	0.11	0.13	19.27	25.71	18.85
$CESSFE_{10;4\times10^{-8};250;10^{-9}}$	0.13	1.07	0.98	18.47	25.63	3.49
$CESSFE_{10;4\times10^{-8};250;2\times10^{-9}}$	0.13	1.07	0.98	18.57	25.53	3.48
$CESSFE_{10;4\times10^{-8};250;4\times10^{-9}}$	0.13	1.07	0.98	18.74	25.33	3.46
$CESSFE_{10;4\times10^{-8};500;10^{-9}}$	0.17	1.13	1.02	16.12	23.19	4.36
$CESSFE_{10;4\times10^{-8};500;2\times10^{-9}}$	0.17	1.13	1.02	16.22	23.10	4.35
$CESSFE_{10;4\times10^{-8};500;4\times10^{-9}}$	0.17	1.13	1.02	16.39	22.93	4.34
$CESSFE_{10;4\times10^{-8};1000;10^{-9}}$	0.17	1.16	1.04	15.97	22.82	4.60
$CESSFE_{10;4\times10^{-8};1000;2\times10^{-9}}$	0.17	1.16	1.04	16.07	22.74	4.59
$CESSFE_{10;4\times10^{-8};1000;4\times10^{-9}}$	0.17	1.16	1.04	16.22	22.57	4.57

Table 4: test function $y = (x_1 - x_2)^4$

Strategy	y	$\frac{\partial y}{\partial x_1}$	$\frac{\partial y}{\partial x_2}$	$\frac{\partial^2 y}{\partial {x_1}^2}$	$\frac{\partial^2 y}{\partial {x_2}^2}$	$\frac{\partial^2 y}{\partial x_1 \partial x_2}$
$LR_{0.04}$	0.01	44.08	44.22	13686.04	13781.59	718.03
$LR_{0.07}$	0.004	0.26	0.26	1134.85	1138.18	13.74
$LR_{0.10}$	0.002	0.03	0.03	47.17	30.22	27.47
$LR_{0.13}$	0.002	0.03	0.03	41.25	30.64	28.45
$SS_{10^{-8}}$	0.001	0.10	0.08	77.01	63.08	10905.14
$SS_{2\times10^{-8}}$	0.001	0.08	0.07	74.61	63.80	10815.04
$SS_{4\times10^{-8}}$	0.001	0.05	0.05	76.16	68.97	10667.97
$CELR_{10;0.04}$	0.79	24.82	24.54	6398.09	6443.77	79.95
$CELR_{10;0.07}$	0.19	2.92	2.86	409.83	447.95	10.84
$CELR_{10;0.1}$	0.27	2.41	2.32	20.94	25.57	4.87
$CELR_{10;0.13}$	0.28	2.32	2.26	22.20	24.04	5.50
$CELR_{20;0.04}$	0.44	15.70	15.69	12234.29	12452.06	100.11
$CELR_{20;0.07}$	0.03	0.89	0.90	681.50	727.72	5.55
$CELR_{20;0.1}$	0.03	0.42	0.42	16.30	10.56	1.47

$CELR_{20;0.13}$	0.03	0.38	0.38	12.64	8.21	1.26
$CELR_{30;0.04}$	0.27	22.96	23.05	15715.59	15988.89	225.09
$CELR_{30;0.07}$	0.02	0.58	0.62	859.63	892.92	8.52
$CELR_{30;0.1}$	0.01	0.14	0.16	16.71	6.66	1.65
$CELR_{30;0.13}$	0.01	0.12	0.13	11.02	4.19	0.75
$CELR_{40;0.04}$	0.14	33.45	33.52	17797.75	18031.66	324.10
$CELR_{40;0.07}$	0.01	0.61	0.65	966.29	988.91	13.44
$CELR_{40;0.1}$	0.004	0.08	0.10	25.88	14.30	2.56
$CELR_{40;0.13}$	0.004	0.07	0.07	20.51	12.85	1.13
$CELR_{50;0.04}$	0.06	39.97	40.01	19087.02	19253.67	360.27
$CELR_{50;0.07}$	0.01	0.61	0.64	1066.28	1082.73	14.08
$CELR_{50;0.1}$	0.002	0.06	0.07	46.80	32.53	4.09
$CELR_{50;0.13}$	0.001	0.05	0.05	42.66	33.46	2.37
$CELRFE_{30;0.13;250;0.01}$	0.36	53.59	53.16	15174.98	15183.93	597.28
$CELRFE_{30;0.13;250;0.02}$	0.06	3.06	3.77	1231.71	1345.74	89.56
$CELRFE_{30;0.13;250;0.04}$	0.06	7.11	7.05	1282.81	1287.04	239.07
$CELRFE_{30;0.13;500;0.01}$	0.26	38.34	37.99	13902.87	13937.15	513.52
$CELRFE_{30;0.13;500;0.02}$	0.10	2.98	3.65	1217.61	1332.93	90.45
$CELRFE_{30;0.13;500;0.04}$	0.08	6.75	6.69	1254.79	1258.94	236.38
$CELRFE_{30;0.13;1000;0.01}$	0.29	36.87	36.53	13741.36	13781.66	504.37
$CELRFE_{30;0.13;1000;0.02}$	0.11	3.02	3.68	1215.40	1331.17	90.98
$CELRFE_{30;0.13;1000;0.04}$	0.09	6.74	6.68	1250.81	1254.95	236.30
$CESS_{10;4\times10^{-8}}$	0.56	7.47	7.40	116.72	110.35	102.71
$CESS_{20;4\times10^{-8}}$	0.05	1.33	1.39	112.05	78.71	56.95
$CESS_{30;4\times10^{-8}}$	0.01	0.47	0.53	128.67	84.76	56.84
$CESS_{40;4\times10^{-8}}$	0.004	0.27	0.32	134.48	97.43	94.61
$CESS_{50;4\times10^{-8}}$	0.02	0.21	0.25	129.98	105.55	265.49
$CESSFE_{20;4\times10^{-8};250;10^{-9}}$	1.30	9.04	9.04	674.19	659.01	73.13
$CESSFE_{20;4\times10^{-8};250;2\times10^{-9}}$	1.30	9.04	9.04	675.12	659.68	73.09

$CESSFE_{20;4\times10^{-8};250;4\times10^{-9}}$	1.30	9.05	9.04	676.82	660.85	73.05
$CESSFE_{20;4\times10^{-8};500;10^{-9}}$	1.65	12.54	12.61	816.08	796.30	105.90
$CESSFE_{20;4\times10^{-8};500;2\times10^{-9}}$	1.65	12.54	12.61	817.23	797.16	105.87
$CESSFE_{20;4\times10^{-8};500;4\times10^{-9}}$	1.65	12.55	12.62	819.27	798.63	105.83
$CESSFE_{20;4\times10^{-8};1000;10^{-9}}$	1.70	13.15	13.24	838.64	818.83	112.13
$CESSFE_{20;4\times10^{-8};1000;2\times10^{-9}}$	1.70	13.15	13.24	839.80	819.71	112.11
$CESSFE_{20;4\times10^{-8};1000;4\times10^{-9}}$	1.70	13.16	13.25	841.86	821.21	112.08

The simulation results of test function $y = \cos(2\pi x_1) + \sin(2\pi x_2)$ are shown in Table 2. The rows LR_α pertain to local regression with nearest neighbor fraction α, rows SS_λ pertain to spline smoothing with smoothing parameter λ. Strategy rows $CELR_{b_n;\alpha}$ refer to compound estimation without filtration and extrapolation, where b_n is the diagonal entry of $\mathbf{B}_n$, using local regression pointwise estimators with nearest neighbor fraction α, while $CESS_\lambda$ refers to compound estimation without filtration and extrapolation using spline smoothing pointwise estimators with smoothing parameter λ. Rows $CELRFE_{b_n;\alpha;b_{FE};\alpha_{FE}}$ refer to compound estimation with filtration and extrapolation using local regression pointwise estimators, where b_{FE} is the diagonal element of $\mathbf{B}_n$ matrix and α_{FE} is the nearest neighbor fraction in filtration and extrapolation process. And rows $CESSFE_{b_n;\lambda;b_{FE};\lambda_{FE}}$ pertain to compound estimation with filtration and extrapolation using spline smoothing pointwise estimators, where λ_{FE} is the smoothing parameter in filtration and extrapolation process. Rows in Tables 3 and 4 are defined similarly for the other test functions.

The results show that for all three test functions, most of the strategies did very well in estimating the mean function and the first order derivatives, since they have small SSE values. On the other hand, compound estimations with filtration and extrapolation do not improve the estimations. For test function $y = \cos(2\pi x_1) + \sin(2\pi x_2)$, Table 2 shows that compound estimations without filtration and extrapolation using local regression pointwise estimators with $\alpha = 0.1$ and $b_n = 40$ did better in estimating the second order derivatives (SSE = 30.98, 2.39, 0.44). It is also shown that compound estimation using spline smoothing pointwise estimators, with filtration and extrapolation is the worst.

For test function $y = e^{x_1 - x_2}$ (Table 3), local regression with $\alpha = 0.13$, all compound estimations without filtration and extrapolation using local regression pointwise estimators with $\alpha = 0.13$ and $b_n \in \{10, 20, 30, 40, 50\}$ did very good jobs in estimating the mean function and its derivatives up to order 2 since they have sum of squared errors less than 4. Moreover, the compound estimation using local regression pointwise estimators with $\alpha = 0.13$ and $b_n = 10$ produces the smallest SSE when estimating $\frac{\partial^2 y}{\partial {x_1}^2}$ and $\frac{\partial^2 y}{\partial {x_2}^2}$ (1.52, 1.62). Even though its SSE for the mixed second partial derivative estimation (0.21) is larger than the smallest SSE (0.03) when $\alpha = 0.1$ and $b_n = 10$, both of them are very close to zero. Similar patterns apply with the mean function and first order derivative estimations.

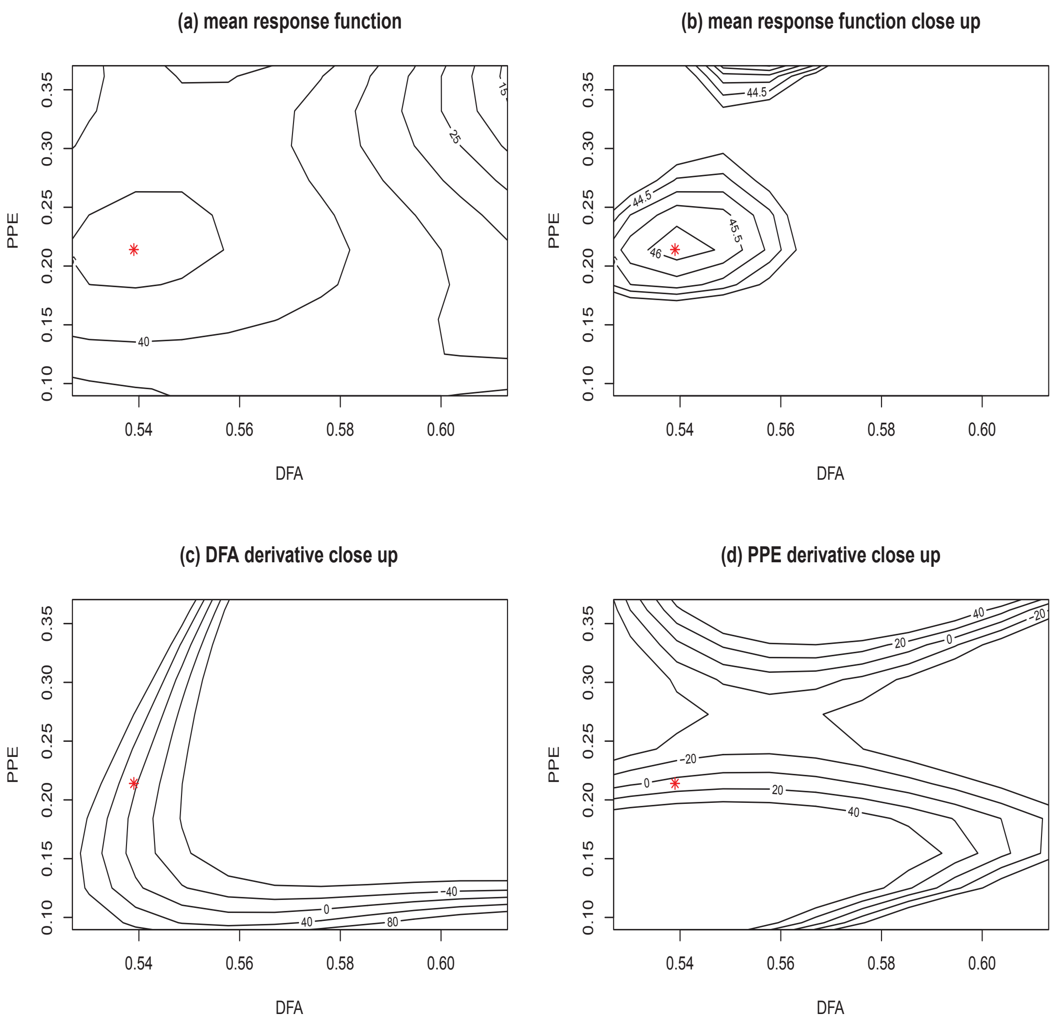
For the third test function $y = (x_1 - x_2)^4$ (Table 4), compound estimation without filtration and extrapolation using local regression pointwise estimators with $\alpha = 0.13$ and $b_n = 30$ dominates for the second order derivatives ($SSE = 11.02$, 4.19, 0.75). And its sum of squared errors for estimating the mean function and first order derivatives are around or below 0.10.

2.6 Parkinson's telemonitoring study

We revisit the Parkinson example and consider the first subject in the study. The response variable is total UPDRS score and the explanatory variables are DFA and PPE with $\mathcal{X} = [0.521, 0.622] \times [0.066, 0.391]$. Compound estimation using local regression pointwise estimators without filtration and extrapolation is applied on each subject to estimate the mean response function and its first order derivatives with respect to DFA and PPE. We considered several choices of tuning parameters, $b_n \in \{10, 20, 30, 40, 50, 100, 200\}$ and $\alpha \in \{0.04, 0.07, 0.1, 0.13, 0.3, 0.4, 0.5\}$. Since the range of DFA is much smaller than PPE, we also considered the convolution matrix with different diagonal entries. More specifically, let $\mathbf{B}_n = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}$, consider $(b_1, b_2) \in \{(50, 150), (50, 10), (100, 50), (150, 50), (200, 50), (200, 100)\}$. Detailed results are presented in Figure 7 for $\alpha = 0.5$, $(b_1, b_2) = (150, 50)$.

Panels a and b in Figure 7 present the estimated mean total UPDRS scores using com-

Figure 7: Parkinson's study: compound estimation



pound estimation on 12×12 points equispaced on $\mathcal{X} = [0.521, 0.622] \times [0.066, 0.391]$. Panel b shows that $(0.539, 0.214)$ is the interior maximizer of the mean response function, which is different from the maximizer from local regression, $(0.595, 0.172)$, in Figure 4. We would expect the estimated first order derivatives with respect to DFA and PPE to be zero at the optimal point. Recall that the estimated partial derivatives from local regression are not even close to zero at the optimal point. Panels c and d in Figure 7 show that the estimated partial derivatives are close to, but not exactly zero at the maximizer. Since we rounded the coordinates to three decimal places, the estimated partial derivatives may appear to be slightly different from zero at the maximizer.

3 Steinization in estimating multiple mean response functions and their derivatives at a fixed point

3.1 Local regression

Suppose we are interested in estimating the mean response functions $\mu(x)$ and $\nu(x)$ and their derivatives from the following model

$$Y_i = \mu(x_i) + \epsilon_i,$$

$$Z_i = \nu(x_i) + \delta_i$$

for $i \in \{1, \ldots, n\}$, where x_i belongs to a compact set $\mathcal{X} \subset \mathbb{R}$, $\mu(x)$ and $\nu(x)$ are two unknown, real-valued functions which have continous derivatives of order $(J+1)$, and $\begin{pmatrix} \epsilon_i \\ \delta_i \end{pmatrix} \overset{iid}{\sim} N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ for $\sigma^2 \in (0, \infty)$ and $-1 < \rho < 1$.

Before attempting compound estimation, we should check how local regression would be carried out. We would maximize the locally weighted log-likelihood

$$l = \sum_{i=1}^{n} \omega_i(x)\{-\log(2\pi) - \log\sigma^2 - \frac{1}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2(1-\rho^2)}\cdot$$

$$[(Y_i - \mu(x_i))^2 + (Z_i - \nu(x_i))^2 - 2\rho(Y_i - \mu(x_i))(Z_i - \nu(x_i))]\}$$

with respect to $a_0, \ldots, a_J, b_0, \ldots, b_J$ if ρ, σ^2 were known and $a_0, \ldots, a_J, b_0, \ldots, b_J, \rho, \sigma^2$ otherwise, where $a_0, a_1, \ldots, a_J, b_0, b_1, \ldots, b_J$ appear in Taylor expansions:

$$\mu(x_i) \approx a_0 + a_1(x_i - x) + \ldots + \frac{a_J}{J!}(x_i - x)^J,$$
$$\nu(x_i) \approx b_0 + b_1(x_i - x) + \ldots + \frac{b_J}{J!}(x_i - x)^J.$$

Recall that $\omega_i(x)$ is defined as $\omega_i(x) = W(\frac{x_i - x}{h})$ for some weight function W that is typically maximized at 0 and that approaches 0 as its argument becomes large in absolute value.

By taking partial derivatives of the locally weighted log-likelihood and setting them to

0, we have

$$
\left(\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \otimes W\right)
\begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \vdots \\ \widehat{a}_J \\ \widehat{b}_0 \\ \widehat{b}_1 \\ \vdots \\ \widehat{b}_J \end{pmatrix}
=
\begin{pmatrix}
\sum \omega_i(x)(Y_i - \rho Z_i) \\
\sum \omega_i(x)(Y_i - \rho Z_i)(x_i - x) \\
\vdots \\
\sum \omega_i(x)(Y_i - \rho Z_i)(x_i - x)^J \\
\sum \omega_i(x)(Z_i - \rho Y_i) \\
\sum \omega_i(x)(Z_i - \rho Y_i)(x_i - x) \\
\vdots \\
\sum \omega_i(x)(Z_i - \rho Y_i)(x_i - x)^J
\end{pmatrix}, \tag{45}
$$

where

$$
W = \begin{pmatrix}
\sum \frac{\omega_i(x)(x_i - x)^{0+0}}{0!} & \sum \frac{\omega_i(x)(x_i - x)^{1+0}}{1!} & \cdots & \sum \frac{\omega_i(x)(x_i - x)^{J+0}}{J!} \\
\sum \frac{\omega_i(x)(x_i - x)^{0+1}}{0!} & \sum \frac{\omega_i(x)(x_i - x)^{1+1}}{1!} & \cdots & \sum \frac{\omega_i(x)(x_i - x)^{J+1}}{J!} \\
\cdots & & & \\
\sum \frac{\omega_i(x)(x_i - x)^{0+J}}{0!} & \sum \frac{\omega_i(x)(x_i - x)^{1+J}}{1!} & \cdots & \sum \frac{\omega_i(x)(x_i - x)^{J+J}}{J!}
\end{pmatrix}
$$

and $\otimes$ denotes direct product. $\widehat{a}_0, \widehat{a}_1, \ldots, \widehat{a}_J$ in (45) are defined as $\widehat{a}_0 := \widehat{\mu(x)}, \widehat{a}_1 := \widehat{\mu'(x)}, \ldots, \widehat{a}_J = \widehat{\mu^{(J)}(x)}$ (Loader, 1999). $\widehat{b}_j$ with $j = 0, 1, \ldots, J$ are defined similarly. The solution turns out to be independent of ρ and σ^2 as shown in the following proposition, which is to say, we can estimate the mean functions and their derivatives using local regression "as if" the error terms are independent.

Proof of equation (45). After taking partial derivatives and setting them equal to 0, we can write the equations as following:

$$
\begin{aligned}
\frac{\partial l}{\partial a_0} &\propto \sum_{i=1}^{n} \omega_i(x)(Y_i - \rho Z_i) - a_0 \sum_{i=1}^{n} \omega_i(x) - \cdots - \frac{a_J}{J!} \sum_{i=1}^{n} \omega_i(x)(x_i - x)^J \\
&+ \rho b_0 \sum_{i=1}^{n} \omega_i(x) + \rho b_1 \sum_{i=1}^{n} \omega_i(x)(x_i - x) + \cdots + \rho \frac{b_J}{J!} \sum_{i=1}^{n} \omega_i(x)(x_i - x)^J = 0 \\
\frac{\partial l}{\partial a_1} &\propto \sum_{i=1}^{n} \omega_i(x)(Y_i - \rho Z_i)(x_i - x) - a_0 \sum_{i=1}^{n} \omega_i(x)(x_i - x) - a_1 \sum_{i=1}^{n} \omega_i(x)(x_i - x)^2 \\
&- \cdots - \frac{a_J}{J!} \sum_{i=1}^{n} \omega_i(x)(x_i - x)^{J+1} + \rho b_0 \sum_{i=1}^{n} \omega_i(x)(x_i - x) + \rho b_1 \sum_{i=1}^{n} \omega_i(x)(x_i - x)^2 \\
&+ \cdots + \rho \frac{b_J}{J!} \sum_{i=1}^{n} \omega_i(x)(x_i - x)^{J+1} = 0 \\
\cdots
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l}{\partial a_J} &\propto \sum_{i=1}^{n}\omega_i(x)(Y_i-\rho Z_i)(x_i-x)^J - a_0\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J - a_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} \\
&- \cdots - \frac{a_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+J} + \rho b_0\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J + \rho b_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} \\
&+ \cdots + \rho\frac{b_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+J} = 0 \\
\frac{\partial l}{\partial b_0} &\propto \sum_{i=1}^{n}\omega_i(x)(Z_i-\rho Y_i) + \rho a_0\sum_{i=1}^{n}\omega_i(x) + \cdots + \rho\frac{a_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J \\
&- b_0\sum_{i=1}^{n}\omega_i(x) - b_1\sum_{i=1}^{n}\omega_i(x)(x_i-x) - \cdots - \frac{b_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J = 0 \\
\frac{\partial l}{\partial b_1} &\propto \sum_{i=1}^{n}\omega_i(x)(Z_i-\rho Y_i)(x_i-x) + \rho a_0\sum_{i=1}^{n}\omega_i(x)(x_i-x) + \rho a_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^2 \\
&+ \cdots + \rho\frac{a_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} - b_0\sum_{i=1}^{n}\omega_i(x)(x_i-x) - b_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^2 \\
&- \cdots - \frac{b_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} = 0 \\
&\cdots \\
\frac{\partial l}{\partial b_J} &\propto \sum_{i=1}^{n}\omega_i(x)(Z_i-\rho Y_i)(x_i-x)^J + \rho a_0\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J + \rho a_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} \\
&+ \cdots + \rho\frac{a_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+J} - b_0\sum_{i=1}^{n}\omega_i(x)(x_i-x)^J - b_1\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+1} \\
&- \cdots - \frac{b_J}{J!}\sum_{i=1}^{n}\omega_i(x)(x_i-x)^{J+J} = 0
\end{aligned}
$$

And equation (45) is the matrix form of the equations above.

Proposition 3.1 *The solutions to equation (45) are*

$$\begin{pmatrix} \widehat{a}_0 \\ \widehat{a}_1 \\ \vdots \\ \widehat{a}_J \\ \widehat{b}_0 \\ \widehat{b}_1 \\ \vdots \\ \widehat{b}_J \end{pmatrix} = \begin{pmatrix} W^{-1} \begin{pmatrix} \sum \omega_i(x) Y_i \\ \sum \omega_i(x)(x_i - x) Y_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Y_i \end{pmatrix} \\ W^{-1} \begin{pmatrix} \sum \omega_i(x) Z_i \\ \sum \omega_i(x)(x_i - x) Z_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Z_i \end{pmatrix} \end{pmatrix}. \tag{46}$$

Proof of Proposition 3.1.

$$
\begin{pmatrix} W & -\rho W \\ -\rho W & W \end{pmatrix}
\begin{pmatrix}
W^{-1}\begin{pmatrix} \sum \omega_i(x) Y_i \\ \sum \omega_i(x)(x_i - x) Y_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Y_i \end{pmatrix} \\
W^{-1}\begin{pmatrix} \sum \omega_i(x) Z_i \\ \sum \omega_i(x)(x_i - x) Z_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Z_i \end{pmatrix}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
WW^{-1}\begin{pmatrix} \sum \omega_i(x) Y_i \\ \sum \omega_i(x)(x_i - x) Y_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Y_i \end{pmatrix} - \rho WW^{-1}\begin{pmatrix} \sum \omega_i(x) Z_i \\ \sum \omega_i(x)(x_i - x) Z_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Z_i \end{pmatrix} \\
-\rho WW^{-1}\begin{pmatrix} \sum \omega_i(x) Y_i \\ \sum \omega_i(x)(x_i - x) Y_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Y_i \end{pmatrix} + WW^{-1}\begin{pmatrix} \sum \omega_i(x) Z_i \\ \sum \omega_i(x)(x_i - x) Z_i \\ \vdots \\ \sum \omega_i(x)(x_i - x)^J Z_i \end{pmatrix}
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sum \omega_i(x)(Y_i - \rho Z_i) \\
\sum \omega_i(x)(Y_i - \rho Z_i)(x_i - x) \\
\vdots \\
\sum \omega_i(x)(Y_i - \rho Z_i)(x_i - x)^J \\
\sum \omega_i(x)(Z_i - \rho Y_i) \\
\sum \omega_i(x)(Z_i - \rho Y_i)(x_i - x) \\
\vdots \\
\sum \omega_i(x)(Z_i - \rho Y_i)(x_i - x)^J
\end{pmatrix}.
$$

Next, we consider a more general case with an arbitrary number of outcomes p ($p \geq 1$).

Suppose we have

$$\begin{aligned} Y_{1i} &= \mu_1(x_i) + \epsilon_{1i} \\ Y_{2i} &= \mu_2(x_i) + \epsilon_{2i} \\ & \cdots \\ Y_{pi} &= \mu_p(x_i) + \epsilon_{pi} \end{aligned} \tag{47}$$

for $i \in \{1, \ldots, n\}$, where x_i belongs to a compact set $\mathcal{X} \subset \mathbb{R}$, $\mu_1(x), \mu_2(x), \ldots, \mu_p(x)$ are unknown, real-valued functions, and $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma)$ where

$\Sigma = \begin{pmatrix} {\sigma_1}^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1p}\sigma_1\sigma_p \\ \rho_{12}\sigma_1\sigma_2 & {\sigma_2}^2 & \cdots & \rho_{2p}\sigma_2\sigma_p \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p}\sigma_1\sigma_p & \rho_{2p}\sigma_2\sigma_p & \cdots & {\sigma_p}^2 \end{pmatrix}$ for ${\sigma_j}^2 \in (0, \infty)$ with $j = 1, 2, \ldots, p$ and the correlations between -1 and 1.

Proposition 3.2 *Similar to Proposition 3.1, the local regression estimators of $\mu_j(x)$, for $j = 1, 2, \ldots, p$, do not depend on the ${\sigma_j}^2$'s or the correlations.*

When there are more than two outcomes, it is possible to apply James-Stein shrinkage (Stein 1956; James and Stein 1961) on the estimators of $\mu_j(x)$ $j = 1, 2, \ldots, p$ to get better estimation with smaller mean squared error(MSE).

3.2 James-Stein shrinkage

Hereafter we allow general linear estimator for $\mu_1(x), \mu_2(x), \ldots, \mu_p(x)$ and their derivatives. That is, $\widehat{\mu}_j^{(k)}(x) = \sum_{i=1}^n l_{j,i}^{(k)}(x) Y_{ji}$ for some functions $l_{j,i}^{(k)}(x)$ with $j = 1, 2, \ldots, p$, $k = 0, 1, \ldots, J$ and $i = 1, 2, \ldots, n$. This includes but is not limited to local regression. Where needed for theoretical results, additional assumptions may be made regarding the nature of these linear estimators.

To start, we consider the simplest case where $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma = \sigma^2 I\right)$ with known $\sigma^2 \in (0, \infty)$. Let $\widehat{a_j}^{(k)}$ be the local regression estimator of $\frac{d^k}{dx^k}\mu_j(x)$, $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$. Define the James-Stein estimator as $\widetilde{a_j}^{(k)} := \widehat{a_j}^{(k)} - \frac{b(p,n,k)}{\sum_{j=1}^p (\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\widehat{a_j}^{(k)}$, where $b(p, n, k)$ is a function of p, n and k, $\sigma_{(k)}{}^2(x) = var(\widehat{a_j}^{(k)})$. The question now is how to find $b(p, n, k)$ such that the James-Stein estimator has smaller MSE than the local regression estimator does.

Theorem 3.1 *Let $\widehat{a_j}^{(k)}$ be the linear estimator of $\frac{d^k}{dx^k}\mu_j(x)$, $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$, $\widetilde{a_j}^{(k)} := \widehat{a_j}^{(k)} - \frac{b(p,n,k)}{\sum_{j=1}^p (\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\widehat{a_j}^{(k)}$ be the James-Stein estimator. Let $b_j^{(k)} = E[\widehat{a_j}^{(k)}] - \mu_j^{(k)}(x)$, $\sigma_{(k)}{}^2(x) = var(\widehat{a_j}^{(k)})$. Then*

$$MSE(\widetilde{a_j}^{(k)}) \leq MSE(\widehat{a_j}^{(k)}) \tag{48}$$

if $b(p, n, k) = p-2+\frac{\sum_{j=1}^p b_j^{(k)} E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^p [\widehat{a_j}^{(k)}]^2}}{\sigma_{(k)}{}^2(x) E\frac{1}{\sum_{j=1}^p [\widehat{a_j}^{(k)}]^2}}$. In (48), "=" holds when $-\sigma_{(k)}^2(x) E\frac{1}{\sum_{j=1}^p [\widehat{a_j}^{(k)}]^2} = \sum_{j=1}^p b_j^{(k)} E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^p [\widehat{a_j}^{(k)}]^2}$.

Proof of Theorem 3.1. MSE of James-Stein estimator is

$$\sum_{j=1}^{p} E[\widetilde{a_j}^{(k)} - \mu_j^{(k)}(x)]^2 = \sum_{j=1}^{p} E[\widehat{a_j}^{(k)} - \mu_j^{(k)}(x) - \frac{b(p,n,k)}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\widehat{a_j}^{(k)}]^2$$

$$= \sum_{j=1}^{p} E[\widehat{a_j}^{(k)} - \mu_j^{(k)}(x)]^2 + \sum_{j=1}^{p} E[\frac{b(p,n,k)}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\widehat{a_j}^{(k)}]^2$$

$$-2\sum_{j=1}^{p} E\{[\widehat{a_j}^{(k)} - \mu_j^{(k)}(x)]\frac{b(p,n,k)}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\widehat{a_j}^{(k)}\}$$

$$= \sum_{j=1}^{p} E[\widehat{a_j}^{(k)} - \mu_j^{(k)}(x)]^2 + [b(p,n,k)]^2 \sum_{j=1}^{p} E[\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}]^2$$

$$-2b(p,n,k)\sum_{j=1}^{p} E\{[\widehat{a_j}^{(k)} - E\widehat{a_j}^{(k)}]\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}\}$$

$$-2b(p,n,k)\sum_{j=1}^{p} b_j^{(k)} E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2/\sigma_{(k)}{}^2(x)}$$

$$= \sum_{j=1}^{p} E[\widehat{a_j}^{(k)} - \mu_j^{(k)}(x)]^2 + [b(p,n,k)]^2 \sigma_{(k)}{}^4(x) E\frac{1}{\sum_{j=1}^{p}[\widehat{a_j}^{(k)}]^2}$$

$$-2b(p,n,k)(p-2)\sigma_{(k)}{}^4(x)E\frac{1}{\sum_{j=1}^{p}[\widehat{a_j}^{(k)}]^2} - 2b(p,n,k)\sigma_{(k)}{}^2(x)\sum_{j=1}^{p} b_j^{(k)} E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}(\widehat{a_j}^{(k)})^2}$$

which achieves its minimum when $b(p,n,k) = p - 2 + \frac{\sum_{j=1}^{p} b_j^{(k)} E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}[\widehat{a_j}^{(k)}]^2}}{\sigma_{(k)}{}^2(x)E\frac{1}{\sum_{j=1}^{p}[\widehat{a_j}^{(k)}]^2}}$. The last equation follows from Lemma 3.1 in Falaki. With this choice of $b(p,n,k)$, it is obvious that $MSE(\widetilde{a_j}^{(k)}) \le MSE(\widehat{a_j}^{(k)})$ by simple calculas.

Remark 1. If the bias $b_j^{(k)}$ is equal to 0, then $b(p,n,k) = p - 2$, which agrees with the results in James and Stein (1961).

Remark 2. The condition under which the equality holds is equivalent to $\sum_{j=1}^{p} \mu_j^{(k)}(x)$ $E\frac{\widehat{a_j}^{(k)}}{\sum_{j=1}^{p}[\widehat{a_j}^{(k)}]^2} = 1$.

Verification of Remark 2.

$$
\begin{aligned}
& -\sigma^2_{(k)}(x) E \frac{1}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} \\
= & \sum_{j=1}^p b_j^{(k)} E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} \\
= & \sum_{j=1}^p [E\widehat{a}_j^{(k)} - \mu_j^{(k)}(x)] E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} \\
= & \sum_{j=1}^p E\widehat{a}_j^{(k)} E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} - \sum_{j=1}^p \mu_j^{(k)}(x) E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} \\
= & 1 - \sigma^2_{(k)}(x) E \frac{1}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} - \sum_{j=1}^p \mu_j^{(k)}(x) E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2}, \qquad (49)
\end{aligned}
$$

which can be simplified to $1 = \sum_{j=1}^p \mu_j^{(k)}(x) E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2}$.

Equation (49) can be verified as following:

$$
\begin{aligned}
\sigma^2_{(k)}(x) E \frac{1}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2} &= \sum_{j=1}^p E\{[\widehat{a}_j^{(k)} - E\widehat{a}_j^{(k)}] \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p (\widehat{a}_j^{(k)})^2}\} \\
&= \sum_{j=1}^p E \frac{(\widehat{a}_j^{(k)})^2}{\sum_{j=1}^p (\widehat{a}_j^{(k)})^2} - \sum_{j=1}^p E\widehat{a}_j^{(k)} E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p (\widehat{a}_j^{(k)})^2} \\
&= 1 - \sum_{j=1}^p E\widehat{a}_j^{(k)} E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p (\widehat{a}_j^{(k)})^2}.
\end{aligned}
$$

Remark 3. $b(p, n, k)$ involves unknown parameters such as $b_j^{(k)}$ and $\sigma_{(k)}{}^2(x)$, which need to be estimated. However, using estimates does not guarantee the inequlity (3) holds. Also, it is difficult to compute $E \frac{\widehat{a}_j^{(k)}}{\sum_{j=1}^p [\widehat{a}_j^{(k)}]^2}$ in $b(p, n, k)$ since it is a fraction of a normal random variable and a non-central chi-square random variable which are dependent.

3.2.1 James-Stein estimator and empirical Bayes

Efron and Morris (1973) showed how to get James-Stein estimators using empirical Bayes approach (Robbins 1956). In this section, we will present James-Stein estimators of $\mu_j(x)$, $j = 1, 2, \ldots, p$ in (90), and show that the James-Stein estimators improve the local regression estimators by reducing risk. Assume the error terms have distribution $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim}$

$N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} {\sigma_1}^2 & 0 & \cdots & 0 \\ 0 & {\sigma_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\sigma_p}^2 \end{pmatrix}\right)$, ${\sigma_j}^2 > 0, j = 1, 2, \ldots, p$, where Σ is known. We have the following theorem.

Theorem 3.2 *Let $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_{j,i}^{(k)}(x) Y_{ji}$ be the local regression estimator of $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$. Assume bias $b_j^{(k)}(x) = E[\widehat{\mu}_j^{(k)}(x)] - \mu_j^{(k)}(x)$ is known for all $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$. Define $\widetilde{\mu}_j^{(k)}(x) := (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x)) - \frac{p-2}{\sum_{s=1}^p [\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2 / ||l_j^{(k)}(x)||^2 \sigma_s^2} (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))$, where $||l_j^{(k)}(x)||^2 = \sum_{i=1}^n l_{j,i}^{(k)}(x)^2$. If the loss function is $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^p (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T (||l_j^{(k)}(x)||^2 \mathbf{\Sigma})^{-1} (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, then*

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)),$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$, and $\mathbf{b}^{(k)}(x) = (b_1^{(k)}(x), b_2^{(k)}(x), \ldots, b_p^{(k)}(x))^T$.

Proof of Theorem 3.2. Since $\mathbf{Y}_i = (Y_{1i}, Y_{2i}, \ldots, Y_{pi})^T \overset{indep.}{\sim} N(\mu(x_i) = (\mu_1(x_i), \mu_2(x_i), \ldots, \mu_p(x_i))^T, \mathbf{\Sigma})$, for $i = 1, 2, \ldots, n$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}^{(k)}(x) + \mathbf{b}^{(k)}(x) = (\mu_1^{(k)}(x) + b_1^{(k)}(x), \mu_2^{(k)}(x) + b_2^{(k)}(x), \ldots, \mu_p^{(k)}(x) + b_p^{(k)}(x))^T$ and variance $||r(x)||^2\mathbf{\Sigma}$. Therefore, $(\widehat{\mu}_1^{(k)}(x) - b_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x) - b_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x) - b_p^{(k)}(x))^T$ is normally distributed with mean $\boldsymbol{\mu}^{(k)}(x)$ and variance

$||l_j^{(k)}(x)||^2\boldsymbol{\Sigma}$. Therefore, we have

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^{p} E \frac{\left[\widetilde{\mu}_j^{(k)}(x) - \mu_j^{(k)(x)}\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} \\
= & \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E \left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right. \\
& \left. - \frac{p-2}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2} \cdot (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))\right]^2 \\
= & \sum_{j=1}^{p} E \frac{\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} \\
& +(p-2)^2 \sum_{j=1}^{p} E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2/||l_j^{(k)}(x)||^2\sigma_j^2}{\left\{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2\right\}^2} \\
& -2(p-2) \sum_{j=1}^{p} E \left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right] \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))/||l_j^{(k)}(x)||^2\sigma_j^2}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2} \\
= & \; p + (p-2)^2 E \frac{1}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2} \\
& -2(p-2)^2 E \frac{1}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2} \qquad (50) \\
= & \; p - (p-2)^2 E \frac{1}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_j^{(k)}(x)||^2\sigma_s^2}.
\end{aligned}
$$

Line (50) comes from the fact that if $X \sim N(\mu, \sigma^2)$, then $E((X-\mu)g(X)) = \sigma^2 E(g'(X))$. Meawhile, note that $EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^{p} E \frac{\left[\widehat{\mu}_j^{(k)} - b_j^{(k)}(x) - \mu_j^{(k)}\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} = p$. It is obvious that $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))$.

The James-Stein type estimator presented above assumes variance σ_j^2 and bias $b_j^{(k)}(x)$ are known for $j = 1, 2, \ldots, p$ and $k = 0, 1, 2, \ldots, J$. However, variance and bias are usually unknown in practice. Next, we consider the scenario when the variance σ_j^2 is unknown and bias $b_j^{(k)}(x)$ is known. Since the variance is unknown, we need to seek a consistent estimator of it.

Proposition 3.3 *Suppose the error terms in (90)* $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\right.$

$\left.\mathbf{\Sigma} = \begin{pmatrix} {\sigma_1}^2 & 0 & \cdots & 0 \\ 0 & {\sigma_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & {\sigma_p}^2 \end{pmatrix}\right)$, ${\sigma_j}^2 > 0, j = 1, 2, \ldots, p$. *Assume that the linear estimator is local regression, kernel smoothing, or asymptotically equivalent to kernel smoothing. Assume* $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ *uniformly. Then* $\widetilde{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^n [Y_{ji} - \widehat{\mu}_j(x_i)]^2$, $\overline{\sigma}_j^2 := \frac{1}{n-2\gamma_1+\gamma_2}\sum_{i=1}^n [Y_{ji} - \widehat{\mu}_j(x_i)]^2$, *where* $\gamma_1 = \sum_{i=1}^n l_i(x_i)$ *and* $\gamma_2 = \sum_{i=1}^n ||l(x_i)||^2 = \sum_{i=1}^n \sum_{j=1}^n l_j(x_i)^2$, *and* $\widehat{\sigma}_j^2 = \frac{1}{m}\sum_{i:|x_i-x|>2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2$, *where* $m = \sum_{i:|x_i-x|>2h}[1 - 2l_i(x_i) + ||l(x_i)||^2]$, *are consistent estimators of* σ_j^2.

Proof of Proposition 3.3. We have

$$\begin{aligned} \widetilde{\sigma}_j^2 &= \frac{1}{n}\sum_{i=1}^n [Y_{ji} - \widehat{\mu}_j(x_i)]^2 = \frac{1}{n}\sum_{i=1}^n [\mu_j(x_i) + \epsilon_{ji} - \widehat{\mu}_j(x_i)]^2 \\ &= \frac{1}{n}\sum_{i=1}^n \epsilon_{ji}^2 + \frac{1}{n}\sum_{i=1}^n [\mu_j(x_i) - \widehat{\mu}_j(x_i)]^2 + \frac{2}{n}\sum_{i=1}^n \epsilon_{ji}[\mu_j(x_i) - \widehat{\mu}_j(x_i)] \end{aligned} \quad (51)$$

As $n \to \infty$, we have $\frac{1}{n}\sum_{i=1}^n \epsilon_{ji}^2 \xrightarrow{P} E\epsilon_{j1}^2 = \sigma_j^2$ by weak law of large numbers. Since $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ uniformly, the last two terms in line (51) converge to 0 in probability. Hence $\widetilde{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$.

Next, consider $\overline{\sigma}_j^2 := \frac{1}{n-2\gamma_1+\gamma_2}\sum_{i=1}^n [Y_{ji} - \widehat{\mu}_j(x_i)]^2 = \frac{n}{n-2\gamma_1+\gamma_2}\widetilde{\sigma}_j^2$. If we can show $\frac{n}{n-2\gamma_1+\gamma_2} \to 1$ as $n \to \infty$, then $\overline{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$ by Slutsky's theorem. Letting W denote the weight function in the asymptotic kernel representation, we have

$$\gamma_1 = \sum_{i=1}^n l_i(x_i) = \sum_{i=1}^n \frac{\omega_i(x_i)}{\sum_{j=1}^n \omega_j(x_i)} \approx \sum_{i=1}^n \frac{W\left(\frac{x_i - x_i}{h}\right)}{nh} = W(0) \cdot \frac{1}{h} = W(0)n^{1/(2J+2+d)},$$

$$\gamma_2 = \sum_{i=1}^n ||l(x_i)||^2 \approx \sum_{i=1}\sum_{j=1}\left[\frac{W\left(\frac{x_j - x_i}{h}\right)}{nh}\right]^2 \le \sum_{i=1}\sum_{j=1}\frac{[W(0)]^2}{(nh)^2} = [W(0)]^2 n^{2/(2J+2+d)},$$

$$\frac{2\gamma_1}{n} \approx 2W(0)n^{1/(2J+2+d)-1} \to 0,$$

and

$$0 \le \frac{\gamma_2}{n} \le [W(0)]^2 n^{2/(2J+2+d)-1} \to 0.$$

Therefore, we have $\frac{\gamma_2}{n} \to 0$ and $\frac{n}{n-2\gamma_1+\gamma_2} = \frac{1}{1-\frac{2\gamma_1}{n}+\frac{\gamma_2}{n}} \to 1$. Thus $\overline{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$.

Last, consider

$$\begin{aligned}
\widehat{\sigma}_j^2 &= \frac{1}{m} \sum_{i:|x_i-x|>2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 \\
&= \frac{1}{m} \left\{ \sum_{i=1}^{n} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 - \sum_{i:|x_i-x|\le 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 \right\} \\
&= \frac{1}{m} \left\{ (n - 2\gamma_1 + \gamma_2)\overline{\sigma}_j^2 - \sum_{i:|x_i-x|\le 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 \right\} \\
&= \frac{n - 2\gamma_1 + \gamma_2}{m}\overline{\sigma}_j^2 - \frac{1}{m} \sum_{i:|x_i-x|\le 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2.
\end{aligned}$$

Since $m = n - 2\gamma_1 + \gamma_2 - \sum_{i:|x_i-x|\le 2h}[1 - 2l_i(x_i) + ||l(x_i)||^2]$, and we have

$$\begin{aligned}
\sum_{i:|x_i-x|\le 2h} [1 - 2l_i(x_i) + ||l(x_i)||^2] &\approx 2n \cdot 2h - 2 \cdot 4nh \cdot \frac{W(0)}{nh} + \sum_{i:|x_i-x|\le 2h} ||l(x_i)||^2 \\
&\le 4nh - 8W(0) + 4nh \cdot n \cdot \frac{[W(0)]^2}{(nh)^2} \\
&= 4nh - 8W(0) + \frac{2[W(0)]^2}{h},
\end{aligned}$$

$\frac{m}{n-2\gamma_1+\gamma_2} = 1 - \frac{n}{n-2\gamma_1+\gamma_2} \cdot \frac{1}{n} \sum_{i:|x_i-x|\le 2h}[1 - 2l_i(x_i) + ||l(x_i)||^2]$, the last term on the right hand side converges to 0 as $n \to \infty$. Therefore, $\frac{n-2\gamma_1+\gamma_2}{m} \to 1$ and $\frac{n-2\gamma_1+\gamma_2}{m}\overline{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$.

On the other hand, let $n_x = card\{i : |x_i - x| \leq 2h\}$. We have

$$\frac{1}{m} \sum_{i:|x_i - x| \leq 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 = \frac{n_x}{m} \cdot \frac{1}{n_x} \sum_{i:|x_i - x| \leq 2h} [\mu_j(x_i) + \epsilon_{ji} - \widehat{\mu}_j(x_i)]^2$$

$$= \frac{n_x}{m} \left\{ \frac{1}{n_x} \sum_{i:|x_i - x| \leq 2h} \epsilon_{ji}^2 + \frac{1}{n_x} \sum_{i:|x_i - x| \leq 2h} [\mu_j(x_i) - \widehat{\mu}_j(x_i)]^2 \right.$$
$$\left. + \frac{2}{n_x} \sum_{i:|x_i - x| \leq 2h} \epsilon_{ji} [\mu_j(x_i) - \widehat{\mu}_j(x_i)] \right\}.$$

Similar to the proof above,

$$\frac{1}{n_x} \sum_{i:|x_i - x| \leq 2h} \epsilon_{ji}^2 + \frac{1}{n_x} \sum_{i:|x_i - x| \leq 2h} [\mu_j(x_i) - \widehat{\mu}_j(x_i)]^2 + \frac{2}{n_x} \sum_{i:|x_i - x| \leq 2h} \epsilon_{ji} [\mu_j(x_i) - \widehat{\mu}_j(x_i)] \xrightarrow{P} \sigma_j^2.$$

Since $\frac{n_x}{m} = \frac{n_x}{n} \cdot \frac{n}{m} \approx \frac{4nh}{n} \cdot \frac{n}{m} \to 0 \cdot 1 = 0$, $\frac{1}{m} \sum_{i:|x_i - x| \leq 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2 \xrightarrow{P} 0$. Therefore, $\widehat{\sigma}_j^2 \xrightarrow{P} \sigma_j^2$.

Then we substitute the variance σ_j^2, $j = 1, 2, \ldots, p$, in James-Stein estimator $\widetilde{\mu}_j^{(k)}(x)$ in Theorem 3.2 with the consistent estimator $\widehat{\sigma}_j^2$ which is independent of $\widehat{\mu}_j^{(k)}$. And the following result is obtained.

Theorem 3.3 *Let $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_{j,i}^{(k)}(x) Y_{ji}$ be the linear estimator of $\mu_j^{(k)}(x) = \frac{d^k}{dx^k} \mu_j(x)$. Suppose bias $b_j^{(k)}(x) = E[\widehat{\mu}_j^{(k)}(x)] - \mu_j^{(k)}(x)$ is known for all $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$. Assume $\sigma_{(1)}^2 \leq \sigma_j^2 \leq \sigma_{(p)}^2$ and $b_j(x) \leq B$. Let $\widehat{\sigma}_j^2$ be the consistent estimator of σ_j^2 defined in Proposition 3.3. Define $\widetilde{\mu}_j^{(k)}(x) := (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x)) - \frac{c||l_j^{(k)}(x)||^2 \widehat{\sigma}_j^2}{\sum_{s=1}^p [\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2} (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))$, where $0 < c < \frac{2p\sigma_{(1)}^2 - 4(\sigma_{(p)}^2 + \frac{1}{m} \sum_{i:|x_i - x| > 2h} B^2)}{F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}}$, $F = \frac{1}{m^2} \sum_{i:|x_i - x| > 2h} 3[1 - 2l_i(x_i) + ||l(x_i)||^2]^2 + \frac{1}{m^2} \sum_{r \neq s} [1 - 2l_r(x_r) + ||l(x_r)||^2][1 - 2l_s(x_s) + ||l(x_s)||^2] + 2[-l_r(x_s) - l_s(x_r) + \sum_{q=1}^n l_q(x_r) l_q(x_s)]^2$, $G = \frac{1}{m^2} \sum_{i:|x_i - x| > 2h} 6B^2 [1 - 2l_i(x_i) + ||l(x_i)||^2] + \frac{1}{m^2} \sum_{r \neq s} B^2 [1 - 2l_r(x_r) + ||l(x_r)||^2] + 4B^2 \left| -l_r(x_s) - l_s(x_r) + \sum_{q=1}^n l_q(x_r) l_q(x_s) \right| + B^2 [1 - 2l_s(x_s) + ||l(x_s)||^2]$, $H = \frac{1}{m^2} \sum_{i:|x_i - x| > 2h} B^4 + \frac{1}{m^2} \sum_{r \neq s} B^2$ and $||l_j^{(k)}(x)||^2 = \sum_{i=1}^n l_{j,i}^{(k)}(x)^2$. If the loss function is $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^p (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T (||l_j^{(k)}(x)||^2 \boldsymbol{\Sigma})^{-1} (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, then*

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)),$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$, *and* $\mathbf{b}^{(k)}(x) = (b_1^{(k)}(x), b_2^{(k)}(x), \ldots, b_p^{(k)}(x))^T$.

Proof of Theorem 3.3 Since $(\widehat{\mu}_1^{(k)}(x) - b_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x) - b_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x) - b_p^{(k)}(x))^T$ follows a normal distribution with mean $(\mu_1^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$ and variance $||l_j^{(k)}(x)||^2\boldsymbol{\Sigma}$, we have

$$
\begin{aligned}
&EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^{p} E\frac{\left[\widetilde{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} \\
= \;& \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right. \\
&\left. - \frac{c||l_j^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2}(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))\right]^2 \\
= \;& \sum_{j=1}^{p} E\frac{\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right]}{||l_j^{(k)}(x)||^2\sigma_j^2} \\
& + c^2 \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E\frac{||l_j^{(k)}(x)||^4\widehat{\sigma}_j^4(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^{p}\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& - 2c \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right] \frac{||l_j^{(k)}(x)||^2\widehat{\sigma}_j^2(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2} \\
= \;& p + c^2 \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E\widehat{\sigma}_j^4 \cdot E\frac{||l_j^{(k)}(x)||^4(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^{p}\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& - 2c \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2\sigma_j^2} E\widehat{\sigma}_j^2 \cdot E\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right] \frac{||l_j^{(k)}(x)||^2(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2} \\
= \;& p + c^2 \sum_{j=1}^{p} \frac{||l_j^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 \cdot E\frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^{p}\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& - 2c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 \cdot E\frac{1}{\sum_{s=1}^{p}\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
& + 4c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 \cdot E\frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^{p}\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2}. \qquad (52)
\end{aligned}
$$

Note that

$$
\begin{aligned}
E\widehat{\sigma}_j^2 = \frac{1}{m} \sum_{i:|x_i - x| > 2h} E\left[Y_{ji} - \widehat{\mu}_j(x_i)\right]^2 \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} E\left[Y_{ji}^2 + \widehat{\mu}_j(x_i)^2 - 2Y_{ji}\widehat{\mu}_j(x_i)\right]^2 \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 - 2E\left[Y_{ji} \sum_{q=1}^{n} l_q(x_i) Y_{jq}\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2E\left[Y_{ji} \cdot l_i(x_i) Y_{ji} + Y_{ji} \sum_{q \neq i} l_q(x_i) Y_{jq}\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2\left[l_i(x_i) EY_{ji}^2 + EY_{ji} \cdot \sum_{q \neq i} l_q(x_i) EY_{jq}\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2\left[l_i(x_i)\left(\sigma_j^2 + \mu_j(x_i)^2\right) + \mu_j(x_i) \sum_{q \neq i} l_q(x_i) \mu_j(x_q)\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2\left[l_i(x_i)\sigma_j^2 + \mu_j(x_i) \sum_{q=1}^{n} l_q(x_i) \mu_j(x_q)\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2\left[l_i(x_i)\sigma_j^2 + \mu_j(x_i) E\widehat{\mu}_j(x_i)\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \sigma_j^2 + \mu_j(x_i)^2 + ||l(x_i)||^2\sigma_j^2 + [\mu_j(x_i) + b_j(x_i)]^2 \\
& -2\left[l_i(x_i)\sigma_j^2 + \mu_j(x_i)\left(\mu_j(x_i) + b_j(x_i)\right)\right] \\
= \quad & \frac{1}{m} \sum_{i:|x_i - x| > 2h} \left[1 - 2l_i(x_i) + ||l(x_i)||^2\right]\sigma_j^2 + b_j(x_i)^2 \\
= \quad & \sigma_j^2 + \frac{1}{m} \sum_{i:|x_i - x| > 2h} b_j(x_i)^2.
\end{aligned}
$$

Since $0 \le b_j(x_i)^2 \le B^2$, $\sigma_j^2 \le E\widehat{\sigma}_j^2 \le \sigma_j^2 + \frac{1}{m}\sum_{i:|x_i-x|>2h} B^2$. And

$$
\begin{aligned}
E\widehat{\sigma}_j^4 &= E\left\{\frac{1}{m}\sum_{i:|x_i-x|>2h}[Y_{ji}-\widehat{\mu}_j(x_i)]^2\right\}^2 \\
&= \frac{1}{m^2}E\left\{\sum_{i:|x_i-x|>2h}[Y_{ji}-\widehat{\mu}_j(x_i)]^4 + \sum_{r\neq s}[Y_{jr}-\widehat{\mu}_j(x_r)]^2[Y_{js}-\widehat{\mu}_j(x_s)]^2\right\}. \quad (53)
\end{aligned}
$$

To find out the expectation of $[Y_{ji}-\widehat{\mu}_j(x_i)]^4$, observe that $Y_{ji}-\widehat{\mu}_j(x_i) = Y_{ji}-\sum_{q=1}^n l_q(x_i)Y_{jq}$ $= \mu_j(x_i)+\epsilon_{ji}-\sum_{q=1}^n l_q(x_i)\left[\mu_j(x_q)+\epsilon_{jq}\right] = \epsilon_{ji}-\sum_{q=1}^n l_q(x_i)\epsilon_{jq} - b_j(x_i) = [1-l_i(x_i)]\,\epsilon_{ji} - \sum_{q\neq i} l_q(x_i)\epsilon_{jq} - b_j(x_i) \sim N\left(-b_j(x_i), [1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2\right)$. Then we have $\frac{[Y_{ji}-\widehat{\mu}_j(x_i)]^2}{[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2} \sim \chi_1^2\left(\frac{b_j(x_i)^2}{[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2}\right)$. Therefore

$$
E\frac{[Y_{ji}-\widehat{\mu}_j(x_i)]^4}{[1-2l_i(x_i)+||l(x_i)||^2]^2\sigma_j^4} = \frac{b_j(x_i)^4}{[1-2l_i(x_i)+||l(x_i)||^2]^2\sigma_j^4} + 6\frac{b_j(x_i)^2}{[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2} + 3.
$$

And

$$
E\left[Y_{ji}-\widehat{\mu}_j(x_i)\right]^4 = b_j(x_i)^4 + 6b_j(x_i)^2[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2 + 3[1-2l_i(x_i)+||l(x_i)||^2]^2\sigma_j^4. \quad (54)
$$

The second expectation in line 53 can be written as

$$
\begin{aligned}
& E\left[Y_{jr}-\widehat{\mu}_j(x_r)\right]^2\left[Y_{js}-\widehat{\mu}_j(x_s)\right]^2 \\
= \; & E\left[\epsilon_{jr}-\sum_{q=1}^{n} l_q(x_r)\epsilon_{jq}-b_j(x_r)\right]^2\left[\epsilon_{js}-\sum_{q=1}^{n} l_q(x_s)\epsilon_{jq}-b_j(x_s)\right]^2 \\
= \; & E\left[\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2-2b_j(x_r)\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)+b_j(x_r)^2\right]\cdot \\
& \left[\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2-2b_j(x_s)\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)+b_j(x_s)^2\right] \\
= \; & E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2-2b_j(x_s)E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right) \\
& +b_j(x_s)^2E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2-2b_j(x_r)E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2 \\
& +4b_j(x_r)b_j(x_s)E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)-2b_j(x_r)b_j(x_s)^2E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right) \\
& +b_j(x_r)^2E\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2-2b_j(x_s)b_j(x_r)^2E\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)+b_j(x_r)^2b_j(x_s)^2 \\
= \; & E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2-0+b_j(x_s)^2\sum_{q=1}^{n} a_q^2\sigma_j^2 \\
& -0+4b_j(x_r)b_j(x_s)\sum_{q=1}^{n} a_qd_q\sigma_j^2+0 \\
& +b_j(x_r)^2\sum_{q=1}^{n} d_q^2\sigma_j^2-0+b_j(x_r)^2b_j(x_s)^2 \\
= \; & E\left(\sum_{q=1}^{n} a_q\epsilon_{jq}\right)^2\left(\sum_{q=1}^{n} d_q\epsilon_{jq}\right)^2+b_j(x_s)^2[1-2l_r(x_r)+||l(x_r)||^2]\sigma_j^2 \\
& +4b_j(x_r)b_j(x_s)[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^{n} l_q(x_r)l_q(x_s)]\sigma_j^2 \\
& +b_j(x_r)^2[1-2l_s(x_s)+||l(x_s)||^2]\sigma_j^2+b_j(x_r)^2b_j(x_s)^2, \qquad (55)
\end{aligned}
$$

where $a_q = 1 - l_r(x_r)$ if $q = r$ and $-l_q(x_r)$ otherwise, d_q is defined similarly. Now consider $E\left(\sum_{q=1}^n a_q\epsilon_{jq}\right)^2 \left(\sum_{q=1}^n d_q\epsilon_{jq}\right)^2$, we have

$$\begin{aligned}
& E\left(\sum_{q=1}^n a_q\epsilon_{jq}\right)^2 \left(\sum_{q=1}^n d_q\epsilon_{jq}\right)^2 \\
= \; & E\left(\sum_{q=1}^n a_q^2\epsilon_{jq}^2 + \sum_{i\neq k} a_i a_k \epsilon_{ji}\epsilon_{jk}\right)\left(\sum_{q=1}^n d_q^2\epsilon_{jq}^2 + \sum_{i\neq k} d_i d_k \epsilon_{ji}\epsilon_{jk}\right) \\
= \; & E\left(\sum_{q=1}^n a_q^2\epsilon_{jq}^2\right)\left(\sum_{q=1}^n d_q^2\epsilon_{jq}^2\right) + E\left(\sum_{q=1}^n a_q^2\epsilon_{jq}^2\right)\left(\sum_{i\neq k} d_i d_k \epsilon_{ji}\epsilon_{jk}\right) \\
& + E\left(\sum_{i\neq k} a_i a_k \epsilon_{ji}\epsilon_{jk}\right)\left(\sum_{q=1}^n d_q^2\epsilon_{jq}^2\right) + E\left(\sum_{i\neq k} a_i a_k \epsilon_{ji}\epsilon_{jk}\right)\left(\sum_{i\neq k} d_i d_k \epsilon_{ji}\epsilon_{jk}\right) \\
= \; & 3\sigma_j^4 \sum_{q=1}^n a_q^2 d_q^2 + \sigma_j^4 \sum_{i\neq k} a_i^2 d_k^2 + 0 + 0 + 2\sigma_j^4 \sum_{i\neq k} a_i a_k d_i d_k \\
= \; & 3\sigma_j^4 \sum_{q=1}^n a_q^2 d_q^2 + \sigma_j^4 \left(\sum_{i=1}^n\sum_{k=1}^n a_i^2 d_k^2 - \sum_{q=1}^n a_q^2 d_q^2\right) + 2\sigma_j^4 \left(\sum_{i=1}^n\sum_{k=1}^n a_i a_k d_i d_k - \sum_{q=1}^n a_q^2 d_q^2\right) \\
= \; & \sigma_j^4 \sum_{i=1}^n a_i^2 \sum_{k=1}^n d_k^2 + 2\sigma_j^4 \left(\sum_{i=1}^n a_i d_i \sum_{k=1}^n a_k d_k\right) \\
= \; & \sigma_j^4[1 - 2l_r(x_r) + ||l(x_r)||^2][1 - 2l_s(x_s) + ||l(x_s)||^2] + 2\sigma_j^4[-l_r(x_s) - l_s(x_r) \\
& + \sum_{q=1}^n l_q(x_r) l_q(x_s)]^2. \qquad (56)
\end{aligned}$$

Combine (53), (54), (55) and (56), the following result is obtained assuming $|b_j(x_i)| \leq B$ for all $j = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$.

$$
\begin{aligned}
& E\widehat{\sigma}_j^4 = E\left\{\frac{1}{m}\sum_{i:|x_i-x|>2h}[Y_{ji}-\widehat{\mu}_j(x_i)]^2\right\}^2 \\
= \; & \frac{1}{m^2}\sum_{i:|x_i-x|>2h} b_j(x_i)^4 + 6b_j(x_i)^2[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2 \\
& +\frac{1}{m^2}\sum_{i:|x_i-x|>2h} 3[1-2l_i(x_i)+||l(x_i)||^2]^2\sigma_j^4 \\
& +\frac{1}{m^2}\sum_{r\neq s}\sigma_j^4[1-2l_r(x_r)+||l(x_r)||^2][1-2l_s(x_s)+||l(x_s)||^2] \\
& +\frac{1}{m^2}\sum_{r\neq s}2\sigma_j^4[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)]^2 \\
& +\frac{1}{m^2}\sum_{r\neq s}b_j(x_s)^2[1-2l_r(x_r)+||l(x_r)||^2]\sigma_j^2 \\
& +b_j(x_r)^2[1-2l_s(x_s)+||l(x_s)||^2]\sigma_j^2+b_j(x_r)^2b_j(x_s)^2 \\
& +\frac{1}{m^2}\sum_{r\neq s}4b_j(x_r)b_j(x_s)[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)]\sigma_j^2 \\
\leq \; & \frac{1}{m^2}\sum_{i:|x_i-x|>2h} B^4+6B^2[1-2l_i(x_i)+||l(x_i)||^2]\sigma_j^2+3[1-2l_i(x_i)+||l(x_i)||^2]^2\sigma_j^4 \\
& +\frac{1}{m^2}\sum_{r\neq s}\sigma_j^4[1-2l_r(x_r)+||l(x_r)||^2][1-2l_s(x_s)+||l(x_s)||^2] \\
& +\frac{1}{m^2}\sum_{r\neq s}2\sigma_j^4[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)]^2 \\
& +\frac{1}{m^2}\sum_{r\neq s}B^2[1-2l_r(x_r)+||l(x_r)||^2]\sigma_j^2+4B^2\left|-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)\right|\sigma_j^2 \\
& +\frac{1}{m^2}\sum_{r\neq s}B^2[1-2l_s(x_s)+||l(x_s)||^2]\sigma_j^2+B^2 \\
= \; & F\sigma_j^4+G\sigma_j^2+H, \qquad (57)
\end{aligned}
$$

where $F = \frac{1}{m^2}\sum_{i:|x_i-x|>2h}3[1-2l_i(x_i)+||l(x_i)||^2]^2+\frac{1}{m^2}\sum_{r\neq s}[1-2l_r(x_r)+||l(x_r)||^2][1-2l_s(x_s)+||l(x_s)||^2]+2[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)]^2$,
$G = \frac{1}{m^2}\sum_{i:|x_i-x|>2h}6B^2[1-2l_i(x_i)+||l(x_i)||^2]+\frac{1}{m^2}\sum_{r\neq s}B^2[1-2l_r(x_r)+||l(x_r)||^2]+4B^2\left|-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)\right|+B^2[1-2l_s(x_s)+||l(x_s)||^2]$,

and $H = \frac{1}{m^2}\sum_{i:|x_i-x|>2h} B^4 + \frac{1}{m^2}\sum_{r\neq s} B^2$.

By substituting $E\widehat{\sigma}_j^2$ and $E\widehat{\sigma}_j^4$ in (52), we obtain an upper bound of $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))$ under the assumption that σ_j^2 is bounded between $\sigma_{(1)}^2$ and $\sigma_{(p)}^2$.

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
= \; & p + c^2 \sum_{j=1}^{p} \frac{||l_j^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 \cdot E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& -2c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 \cdot E \frac{1}{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
& +4c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 \cdot E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
\leq \; & p + c^2 \sum_{j=1}^{p} \frac{||l_j^{(k)}(x)||^2}{\sigma_j^2} (F\sigma_j^4 + G\sigma_j^2 + H) E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& -2c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 \sigma_j^2 \frac{1}{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
& +4c \sum_{j=1}^{p} ||l_j^{(k)}(x)||^2 (\sigma_j^2 + \frac{1}{m} \sum_{i:|x_i-x|>2h} B^2) E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
\leq \; & p + c^2 ||l_j^{(k)}(x)||^2 \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}\right) \sum_{s=1}^{p} E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2} \\
& -2c ||l_j^{(k)}(x)||^2 \sigma_{(1)}^2 \sum_{j=1}^{p} E \frac{1}{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
& +4c ||l_j^{(k)}(x)||^2 (\sigma_{(p)}^2 + \frac{1}{m} \sum_{i:|x_i-x|>2h} B^2) \sum_{j=1}^{p} E \frac{(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))^2}{\left\{\sum_{s=1}^p \left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2\right\}^2}
\end{aligned}
$$

$$
\begin{aligned}
&= p + c^2||l_j^{(k)}(x)||^2\left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}\right) E\frac{1}{\sum_{s=1}^p\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
&\quad -2pc||l_j^{(k)}(x)||^2\sigma_{(1)}^2 E\frac{1}{\sum_{s=1}^p\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
&\quad +4c||l_j^{(k)}(x)||^2(\sigma_{(p)}^2 + \frac{1}{m}\sum_{i:|x_i-x|>2h} B^2) E\frac{1}{\sum_{s=1}^p\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2} \\
&= p + \left[c^2||l_j^{(k)}(x)||^2\left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}\right) - 2pc||l_j^{(k)}(x)||^2\sigma_{(1)}^2\right. \\
&\quad \left. +4c||l_j^{(k)}(x)||^2(\sigma_{(p)}^2 + \frac{1}{m}\sum_{i:|x_i-x|>2h} B^2)\right] E\frac{1}{\sum_{s=1}^p\left[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)\right]^2}.
\end{aligned}
$$

Let $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - p < 0$, we need c to satisfy

$$
c^2||l_j^{(k)}(x)||^2\left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}\right) - 2c||l_j^{(k)}(x)||^2\left[p\sigma_{(1)}^2 - 2c(\sigma_{(p)}^2 + \frac{1}{m}\sum_{i:|x_i-x|>2h} B^2)\right] < 0,
$$

since c is required to be positive, the inequality above is equivalent to

$$
c < \frac{2p\sigma_{(1)}^2 - 4(\sigma_{(p)}^2 + \frac{1}{m}\sum_{i:|x_i-x|>2h} B^2)}{F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}}}.
$$

Theorem 3.4 *Let $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_{j,i}^{(k)}(x)Y_{ji}$ be the linear estimator of $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$. Assume $p \geq 4$, $\sigma_{(1)}^2 \leq \sigma_j^2 \leq \sigma_{(p)}^2$ and bias $b_j^{(k)}(x) = E[\widehat{\mu}_j^{(k)}(x)] - \mu_j^{(k)}(x) \leq B$ for all $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$. Let $\widehat{\sigma}_j^2$ be the consistent estimator of σ_j^2 defined in Proposition*

3.3. Define $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l_j^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$*, where c is a positive constant satisfying*

$$\left[\frac{c^2L_1(x)}{\sigma_{(1)}^2L_0(x)}\left(F\sigma_{(p)}^2+G+\frac{H}{\sigma_{(1)}^2}\right)+\frac{4cL_1(x)}{\sigma_{(1)}^2L_0(x)}\left(\sigma_{(p)}^2+\frac{n_x}{m}B^2\right)\right]\frac{\sigma_{(p)}^2L_1(x)}{p(M_0^2-2M_1B)}$$
$$\cdot\left(1-e^{-\frac{p(M_1+B)^2}{2\sigma_{(1)}^2L_0(x)}}\right)-\frac{2pcL_0(x)\sigma_{(1)}^2}{\sigma_{(p)}^2L_1(x)}\frac{1}{p-2}\cdot\frac{2\sigma_{(1)}^2L_0(x)}{p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\sigma_{(p)}^2L_1(x)}}\right)$$
$$+\frac{2pcB}{\sigma_{(1)}^2L_0(x)}\left(1+\frac{n_xB^2}{m\sigma_{(1)}^2}\right)\left(\sigma_{(p)}\sqrt{\frac{2L_1(x)}{\pi}}+M_1\right)\frac{2\sigma_{(p)}^2L_1(x)}{(p-1)(M_0^2-2M_1B)}$$
$$\left(1-e^{-\frac{(p-1)(M_1+B)^2}{2\sigma_{(1)}^2L_0(x)}}\right)<0,$$

$F = \frac{1}{m^2}\sum_{i:|x_i-x|>2h}3[1-2l_i(x_i)+||l(x_i)||^2]^2+\frac{1}{m^2}\sum_{r\neq s}[1-2l_r(x_r)+||l(x_r)||^2][1-2l_s(x_s)+||l(x_s)||^2]+2[-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)]^2$, $G = \frac{1}{m^2}\sum_{i:|x_i-x|>2h}6B^2[1-2l_i(x_i)+||l(x_i)||^2]+\frac{1}{m^2}\sum_{r\neq s}B^2[1-2l_r(x_r)+||l(x_r)||^2]+4B^2\left|-l_r(x_s)-l_s(x_r)+\sum_{q=1}^n l_q(x_r)l_q(x_s)\right|+B^2[1-2l_s(x_s)+||l(x_s)||^2]$, $H = \frac{1}{m^2}\sum_{i:|x_i-x|>2h}B^4+\frac{1}{m^2}\sum_{r\neq s}B^2$. *Suppose* $L_0(x)\leq ||l_j^{(k)}(x)||^2\leq L_1(x)$ *for* $j=1,2,\ldots,p$*. If the loss function is* $L(\mathbf{Z}^{(k)},\boldsymbol{\mu}^{(k)}(x))=\sum_{j=1}^p(\mathbf{Z}^{(k)}-\boldsymbol{\mu}^{(k)}(x))^T(||l_j^{(k)}(x)||^2\boldsymbol{\Sigma})^{-1}(\mathbf{Z}^{(k)}-\boldsymbol{\mu}^{(k)}(x))$*, then*

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)),$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x)=(\widetilde{\mu}_1^{(k)}(x),\widetilde{\mu}_2^{(k)}(x),\ldots,\widetilde{\mu}_p^{(k)}(x))^T$, $\boldsymbol{\mu}^{(k)}(x)=(\mu_1^{(k)}(x),\mu_2^{(k)}(x),\ldots,\mu_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x)=(\widehat{\mu}_1^{(k)}(x),\widehat{\mu}_2^{(k)}(x),\ldots,\widehat{\mu}_p^{(k)}(x))^T$*, and* $\mathbf{b}^{(k)}(x)=(b_1^{(k)}(x),b_2^{(k)}(x),\ldots,b_p^{(k)}(x))^T$*.*

Proof of Theorem 3.4 We have

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^p E\frac{\left[\widetilde{\mu}_j^{(k)}(x)-\mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2}$$
$$= \sum_{j=1}^p E\frac{\left[\widetilde{\mu}_j^{(k)}(x)-\widehat{\mu}_j^{(k)}(x)+\widehat{\mu}_j^{(k)}(x)-\mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2}.$$

Therefore,

$$\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
= & \sum_{j=1}^{p} E \frac{\left[\widetilde{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2 \sigma_j^2} - \sum_{j=1}^{p} E \frac{\left[\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2 \sigma_j^2} \\
= & \sum_{j=1}^{p} \frac{E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)^2 + 2E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right)}{||l_j^{(k)}(x)||^2 \sigma_j^2}. \quad (58)
\end{aligned}$$

Consider the first term in (58), we have

$$\begin{aligned}
& \sum_{j=1}^{p} \frac{E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)^2}{||l_j^{(k)}(x)||^2 \sigma_j^2} = \sum_{j=1}^{p} \frac{1}{||l_j^{(k)}(x)||^2 \sigma_j^2} E \frac{c^2 ||l_j^{(k)}(x)||^4 \widehat{\sigma}_j^4 \widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
= & \sum_{j=1}^{p} \frac{c^2 ||l_j^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
\leq & \sum_{j=1}^{p} \frac{c^2 ||l_j^{(k)}(x)||^2}{\sigma_j^2} \left(F\sigma_j^4 + G\sigma_j^2 + H\right) E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} \quad (59) \\
= & \sum_{j=1}^{p} c^2 ||l_j^{(k)}(x)||^2 \left(F\sigma_j^2 + G + \frac{H}{\sigma_j^2}\right) E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
\leq & c^2 L_1(x) \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2}\right) \sum_{j=1}^{p} E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
= & c^2 L_1(x) \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2}\right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2}. \quad (60)
\end{aligned}$$

Line (59) is obtained from line (57).

Next, consider the second term in (58). Let $n_x = card\{i : |x_i - x| > 2h\}$, recall that $\sigma_j^2 \le E\widehat{\sigma}_j^2 \le \sigma_j^2 + \frac{n_x}{m}B^2$. We have

$$
\begin{aligned}
& \sum_{j=1}^{p} \frac{2E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right)}{||l_j^{(k)}(x)||^2\sigma_j^2} \\
= & \sum_{j=1}^{p} \frac{2}{||l_j^{(k)}(x)||^2\sigma_j^2} E\left(-\frac{c||l_j^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)\right)\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right) \\
= & -\sum_{j=1}^{p} \frac{2c}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right) \\
= & -\sum_{j=1}^{p} \frac{2c}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x) - b_j^{(k)}(x) + b_j^{(k)}(x)\right) \\
= & -\sum_{j=1}^{p} \frac{2c}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x) - b_j^{(k)}(x)\right) \\
& -\sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} \\
= & -\sum_{j=1}^{p} \frac{2c}{\sigma_j^2} E\widehat{\sigma}_j^2 \cdot ||l_j^{(k)}(x)||^2\sigma_j^2 E\frac{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2 - 2\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
& -\sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} \qquad (61) \\
= & -\sum_{j=1}^{p} 2c||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E\frac{1}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} + \sum_{j=1}^{p} 4c||l_j^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
& -\sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} \\
\le & -\sum_{j=1}^{p} 2c||l_j^{(k)}(x)||^2\sigma_j^2 E\frac{1}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} \\
& +\sum_{j=1}^{p} 4c||l_j^{(k)}(x)||^2\left(\sigma_j^2 + \frac{n_x}{m}B^2\right) E\frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2\right]^2} \\
& +\left|\sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq -2pcL_0(x)\sigma_{(1)}^2 E\frac{1}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} + 4cL_1(x)\left(\sigma_{(p)}^2 + \frac{n_x}{m}B^2\right)E\frac{1}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} \\
&\quad + \sum_{j=1}^p \frac{2c\left|b_j^{(k)}(x)\right|}{\sigma_j^2} E\widehat{\sigma}_j^2 \left| E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}\right| \\
&\leq -2pcL_0(x)\sigma_{(1)}^2 E\frac{1}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} + 4cL_1(x)\left(\sigma_{(p)}^2 + \frac{n_x}{m}B^2\right)E\frac{1}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} \\
&\quad + \sum_{j=1}^p \frac{2cB}{\sigma_j^2}\left(\sigma_j^2 + \frac{n_x}{m}B^2\right)E\frac{\left|\widehat{\mu}_j^{(k)}(x)\right|}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}. \qquad (62)
\end{aligned}
$$

Line (61) follows the fact that if $X \sim N(\mu, \sigma^2)$, then $E((X-\mu)g(X)) = \sigma^2 E(g'(X))$. Now consider

$$
\begin{aligned}
&\sum_{j=1}^p \frac{2cB}{\sigma_j^2}\left(\sigma_j^2 + \frac{n_x}{m}B^2\right)E\frac{\left|\widehat{\mu}_j^{(k)}(x)\right|}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} \leq \sum_{j=1}^p \frac{2cB}{\sigma_j^2}\left(\sigma_j^2 + \frac{n_x}{m}B^2\right)E\frac{\left|\widehat{\mu}_j^{(k)}(x)\right|}{\sum_{s\neq j} \widehat{\mu}_s^{(k)}(x)^2} \\
&= \sum_{j=1}^p \frac{2cB}{\sigma_j^2}\left(\sigma_j^2 + \frac{n_x}{m}B^2\right)E\left|\widehat{\mu}_j^{(k)}(x)\right| E\frac{1}{\sum_{s\neq j} \widehat{\mu}_s^{(k)}(x)^2} \\
&= \sum_{j=1}^p 2cB\left(1 + \frac{n_x B^2}{m\sigma_j^2}\right)\left(\sigma_j ||l_j^{(k)}(x)||\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{\left(\mu_j^{(k)}(x) + b_j^{(k)}(x)\right)^2}{2\sigma_j^2||l_j^{(k)}(x)||^2}\right\}\right. \\
&\quad \left. + \mu_j^{(k)}(x)\left[1 - 2\Phi(-\mu_j^{(k)}(x)/\sigma_j||l_j^{(k)}(x)||)\right]\right)E\frac{1}{\sum_{s\neq j} \widehat{\mu}_s^{(k)}(x)^2} \\
&\leq \sum_{j=1}^p 2cB\left(1 + \frac{n_x B^2}{m\sigma_j^2}\right)\left(\sigma_j ||l_j^{(k)}(x)||\sqrt{\frac{2}{\pi}} + \left|\mu_j^{(k)}(x)\right|\right)E\frac{1}{\sum_{s\neq j} \widehat{\mu}_s^{(k)}(x)^2}. \qquad (63)
\end{aligned}
$$

Combine (60), (62) and (63), we obtain an upper bound of $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))$

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
\leq \; & c^2 L_1(x) \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2} \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} \\
& -2pcL_0(x)\sigma_{(1)}^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} + 4cL_1(x) \left(\sigma_{(p)}^2 + \frac{n_x}{m} B^2 \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} \\
& + \sum_{j=1}^{p} 2cB \left(1 + \frac{n_x B^2}{m\sigma_j^2} \right) \left(\sigma_j ||l_j^{(k)}(x)|| \sqrt{\frac{2}{\pi}} + \left| \mu_j^{(k)}(x) \right| \right) E \frac{1}{\sum_{s \neq j} \widehat{\mu}_s^{(k)}(x)^2} \\
\leq \; & c^2 L_1(x) \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2} \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2 \cdot \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& -2pcL_0(x)\sigma_{(1)}^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2 \cdot \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& +4cL_1(x) \left(\sigma_{(p)}^2 + \frac{n_x}{m} B^2 \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2 \cdot \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& + \sum_{j=1}^{p} 2cB \left(1 + \frac{n_x B^2}{m\sigma_{(1)}^2} \right) \left(\sigma_{(p)} \sqrt{\frac{2L_1(x)}{\pi}} + M_1 \right) E \frac{1}{\sum_{s \neq j} \widehat{\mu}_s^{(k)}(x)^2} \\
\leq \; & \frac{c^2 L_1(x)}{\sigma_{(1)}^2 L_0(x)} \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2} \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& - \frac{2pcL_0(x)\sigma_{(1)}^2}{\sigma_{(p)}^2 L_1(x)} E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& + \frac{4cL_1(x)}{\sigma_{(1)}^2 L_0(x)} \left(\sigma_{(p)}^2 + \frac{n_x}{m} B^2 \right) E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
& + \sum_{j=1}^{p} \frac{2cB}{\sigma_{(1)}^2 L_0(x)} \left(1 + \frac{n_x B^2}{m\sigma_{(1)}^2} \right) \left(\sigma_{(p)} \sqrt{\frac{2L_1(x)}{\pi}} + M_1 \right) E \frac{1}{\sum_{s \neq j} \widehat{\mu}_s^{(k)}(x)^2 / \sigma_s^2 ||l_s^{(k)}(x)||^2} \\
\leq \; & \frac{c^2 L_1(x)}{\sigma_{(1)}^2 L_0(x)} \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2} \right) E \frac{1}{p - 2 + 2K} - \frac{2pcL_0(x)\sigma_{(1)}^2}{\sigma_{(p)}^2 L_1(x)} E \frac{1}{p - 2 + 2K} \\
& + \frac{4cL_1(x)}{\sigma_{(1)}^2 L_0(x)} \left(\sigma_{(p)}^2 + \frac{n_x}{m} B^2 \right) E \frac{1}{p - 2 + 2K} \\
& + \sum_{j=1}^{p} \frac{2cB}{\sigma_{(1)}^2 L_0(x)} \left(1 + \frac{n_x B^2}{m\sigma_{(1)}^2} \right) \left(\sigma_{(p)} \sqrt{\frac{2L_1(x)}{\pi}} + M_1 \right) E \frac{1}{p - 2 + 2K_j},
\end{aligned}
$$

where K and K_j follow Poisson distributions with parameters $\lambda = \sum_{s=1}^{p} \frac{\left[\mu_s^{(k)}(x)+b_s^{(k)}(x)\right]^2}{2\sigma_s^2||l_s^{(k)}(x)||^2}$ and $\lambda_j = \sum_{s\neq j} \frac{\left[\mu_s^{(k)}(x)+b_s^{(k)}(x)\right]^2}{2\sigma_s^2||l_s^{(k)}(x)||^2}$, respectively. Since $\frac{p(M_0^2-2M_1B)}{2\sigma_{(p)}^2 L_1(x)} \leq \lambda \leq \frac{p(M_1+B)^2}{2\sigma_{(1)}^2 L_0(x)}$,

$$
\begin{aligned}
E\frac{1}{p-2+2K} &= \sum_{k=0}^{\infty} \frac{1}{p-2+2k} \cdot \frac{e^{-\lambda}\lambda^k}{k!} \\
&\leq \sum_{k=0}^{\infty} \frac{1}{2+2k} \cdot \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \frac{1}{2}\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k+1)!} \\
&= \frac{1}{2}\sum_{j=1}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \cdot \frac{1}{\lambda} \\
&= \frac{1}{2\lambda}\left(1-e^{-\lambda}\right) \\
&\leq \frac{\sigma_{(p)}^2 L_1(x)}{p(M_0^2-2M_1B)} \left(1-e^{-\frac{p(M_1+B)^2}{2\sigma_{(1)}^2 L_0(x)}}\right).
\end{aligned}
$$

On the other hand, a lower bound of $E\frac{1}{p-2+2K}$ is

$$
\begin{aligned}
E\frac{1}{p-2+2K} &\geq E\frac{1}{p-2+(p-2)K} \\
&= \frac{1}{p-2}E\frac{1}{1+K} \\
&= \frac{1}{p-2}\sum_{k=0}^{\infty} \frac{1}{k+1} \cdot \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \frac{1}{p-2}\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k+1)!} \\
&= \frac{1}{p-2}\sum_{j=1}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \cdot \frac{1}{\lambda} \\
&= \frac{1}{p-2} \cdot \frac{1}{\lambda}\left(1-e^{-\lambda}\right) \\
&\geq \frac{1}{p-2} \cdot \frac{2\sigma_{(1)}^2 L_0(x)}{p(M_1+B)^2} \left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\sigma_{(p)}^2 L_1(x)}}\right).
\end{aligned}
$$

Similarly, an upper bound of $E\frac{1}{p-2+2K_j}$ is $\frac{2\sigma^2_{(p)}L_1(x)}{(p-1)(M_0^2-2M_1B)}\left(1-e^{-\frac{(p-1)(M_1+B)^2}{2\sigma^2_{(1)}L_0(x)}}\right)$. To have $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x))-EL(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x))<0$, we need c to satisfy that

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x))-EL(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)) \\
\leq\ & \frac{c^2L_1(x)}{\sigma^2_{(1)}L_0(x)}\left(F\sigma^2_{(p)}+G+\frac{H}{\sigma^2_{(1)}}\right)E\frac{1}{p-2+2K}-\frac{2pcL_0(x)\sigma^2_{(1)}}{\sigma^2_{(p)}L_1(x)}E\frac{1}{p-2+2K} \\
& +\frac{4cL_1(x)}{\sigma^2_{(1)}L_0(x)}\left(\sigma^2_{(p)}+\frac{n_x}{m}B^2\right)E\frac{1}{p-2+2K} \\
& +\sum_{j=1}^{p}\frac{2cB}{\sigma^2_{(1)}L_0(x)}\left(1+\frac{n_xB^2}{m\sigma^2_{(1)}}\right)\left(\sigma_{(p)}\sqrt{\frac{2L_1(x)}{\pi}}+M_1\right)E\frac{1}{p-2+2K_j} \\
\leq\ & \left[\frac{c^2L_1(x)}{\sigma^2_{(1)}L_0(x)}\left(F\sigma^2_{(p)}+G+\frac{H}{\sigma^2_{(1)}}\right)+\frac{4cL_1(x)}{\sigma^2_{(1)}L_0(x)}\left(\sigma^2_{(p)}+\frac{n_x}{m}B^2\right)\right]\frac{\sigma^2_{(p)}L_1(x)}{p(M_0^2-2M_1B)} \\
& \cdot\left(1-e^{-\frac{p(M_1+B)^2}{2\sigma^2_{(1)}L_0(x)}}\right)-\frac{2pcL_0(x)\sigma^2_{(1)}}{\sigma^2_{(p)}L_1(x)}\frac{1}{p-2}\cdot\frac{2\sigma^2_{(1)}L_0(x)}{p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\sigma^2_{(p)}L_1(x)}}\right) \\
& +\frac{2pcB}{\sigma^2_{(1)}L_0(x)}\left(1+\frac{n_xB^2}{m\sigma^2_{(1)}}\right)\left(\sigma_{(p)}\sqrt{\frac{2L_1(x)}{\pi}}+M_1\right)\frac{2\sigma^2_{(p)}L_1(x)}{(p-1)(M_0^2-2M_1B)} \\
& \cdot\left(1-e^{-\frac{(p-1)(M_1+B)^2}{2\sigma^2_{(1)}L_0(x)}}\right) \\
<\ & 0.
\end{aligned}
$$

So far we have assumed independent error terms ϵ_{ji} where $j=1,2,\dots,p$ and $i=1,2,\dots,n$. Now we relax the assumption to any symmetric, positive definite variance-covariance matrix and the following result is obtained. Before stating the theorem, we introduce some notation. Put $\widetilde{\boldsymbol{\mu}}^{(k)}(x)=(\widetilde{\mu}_1^{(k)}(x),\widetilde{\mu}_2^{(k)}(x),\dots,\widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x)=(\widehat{\mu}_1^{(k)}(x),\widehat{\mu}_2^{(k)}(x),\dots,\widehat{\mu}_p^{(k)}(x))^T$, $\boldsymbol{\mu}^{(k)}(x)=(\mu_1^{(k)}(x),\mu_2^{(k)}(x),\dots,\mu_p^{(k)}(x))^T$ and $\mathbf{b}^{(k)}(x)=(b_1^{(k)}(x),b_2^{(k)}(x),\dots,b_p^{(k)}(x))^T$. Let $\boldsymbol{\Sigma}_0:=var(\widehat{\boldsymbol{\mu}}^{(k)}(x))$, where $(\boldsymbol{\Sigma}_0)_{ij}=\sum_{s=1}^n l_{i,s}^{(k)}(x)l_{j,s}^{(k)}(x)\,(\boldsymbol{\Sigma})_{ij}$.

Theorem 3.5 *Assume the error terms have distribution* $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{\Sigma}\right)$, *where* $\mathbf{\Sigma}$ *is a known, symmetric and positive definite matrix. Let* $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_{j,i}^{(k)}(x) Y_{ji}$ *be the linear estimator of* $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$*. Suppose* $p \geq 4$, $M_0 \leq \left|\mu_j^{(k)}(x)\right| \leq M_1$, *and that the unknown bias* $b_j^{(k)}(x) = E[\widehat{\mu}_j^{(k)}(x)] - \mu_j^{(k)}(x) \in [-B, B]$ *for all* $k = 0, 1, \ldots, J$ *and* $j = 1, 2, \ldots, p$*. Define* $\widetilde{\boldsymbol{\mu}}^{(k)}(x) := \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)}\right] \widehat{\boldsymbol{\mu}}^{(k)}(x)$, *where* c *satisfies*

$$
\begin{aligned}
& \frac{2p\sqrt{p}B}{p-3}\left(\sqrt{\frac{2}{\pi}} \cdot \textit{largest eigenvalue of } \mathbf{\Sigma_0}^{-1}\right. \\
& \left. + \sqrt{p}(M_1 + B) \cdot \textit{largest eigenvalue of } \mathbf{\Sigma_0}^{-1}\right) \\
- \quad & \frac{4}{\textit{largest eigenvalue of } \mathbf{\Sigma_0}^{-1} \cdot p(M_1+B)^2}\left(1 - e^{-\frac{p(M_0^2 - 2M_1 B)}{2 \cdot \textit{largest eigenvalue of } \mathbf{\Sigma_0}}}\right) \\
+ \quad & c\frac{\textit{largest eigenvalue of } \mathbf{\Sigma_0}}{p(M_0^2 - 2M_1 B)}\left(1 - e^{-\frac{1}{2} \cdot \textit{largest eigenvalue of } \mathbf{\Sigma_0}^{-1} \cdot p(M_1+B)^2}\right) \\
< \quad & 0.
\end{aligned}
$$

If the loss function is $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T \mathbf{\Sigma_0}^{-1} (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, *then*

$$
EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)).
$$

Proof of Theorem 3.5. We have

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = E(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T \mathbf{\Sigma_0}^{-1} (\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)) \\
= \quad & E\left[\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} \widehat{\boldsymbol{\mu}}^{(k)}(x)\right]^T \mathbf{\Sigma_0}^{-1} \\
& \cdot \left[\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} \widehat{\boldsymbol{\mu}}^{(k)}(x)\right] \\
= \quad & E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right)^T \mathbf{\Sigma_0}^{-1} \left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\
& -2cE\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} \widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\
& +c^2 E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma_0}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)}.
\end{aligned} \tag{64}
$$

Since $\mathbf{\Sigma}_0$ is symmetric and positive definite, there exists a symmetric, positive definite matrix $\mathbf{A}$ such that $\mathbf{\Sigma}_0 = \mathbf{A}\mathbf{A}$. Note that $\widehat{\boldsymbol{\mu}}^{(k)}(x) \sim N\left(\boldsymbol{\xi}^{(k)} = \boldsymbol{\mu}^{(k)}(x) + \mathbf{b}^{(k)}(x), \mathbf{\Sigma}_0\right)$, we have $\mathbf{Z} = \mathbf{A}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x) \sim N\left(\boldsymbol{\psi}^{(k)} = \mathbf{A}^{-1}\boldsymbol{\xi}^{(k)}, \mathbf{A}^{-1}\mathbf{\Sigma}_0\mathbf{A} = \mathbf{I}\right)$. First, notice that

$$E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)} = E\frac{1}{\mathbf{Z}^T\mathbf{Z}} = E\frac{1}{\sum_{s=1}^p Z_s^2} = E\frac{1}{p-2+2K}, \tag{65}$$

where K has a Poisson distribution with parameter $\lambda = \frac{1}{2}\boldsymbol{\psi}^{(k)\prime}\boldsymbol{\psi}^{(k)}$. Second, we have

$$\begin{aligned}
& E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\
= \; & E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)} + \mathbf{b}^{(k)}(x)\right) \\
= \; & E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\
& + E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\mathbf{b}^{(k)}(x).
\end{aligned} \tag{66}$$

The first term in (66) can be written as

$$\begin{aligned}
& E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T{\mathbf{\Sigma}_0}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\
= \; & E\frac{1}{\mathbf{Z}^T\mathbf{Z}}(\mathbf{A}\mathbf{Z})^T{\mathbf{\Sigma}_0}^{-1}(\mathbf{A}\mathbf{Z} - \mathbf{A}\boldsymbol{\psi}^{(k)}) \\
= \; & E\frac{1}{\mathbf{Z}^T\mathbf{Z}}\mathbf{Z}^T\mathbf{A}^T{\mathbf{\Sigma}_0}^{-1}\mathbf{A}(\mathbf{Z} - \boldsymbol{\psi}^{(k)}) \\
= \; & E\frac{1}{\mathbf{Z}^T\mathbf{Z}}\mathbf{Z}^T(\mathbf{Z} - \boldsymbol{\psi}^{(k)}) \\
= \; & \sum_{j=1}^{p} E\frac{1}{\sum_{s=1}^p Z_s^2}(Z_j - \psi_j^{(k)})Z_j \\
= \; & (p-2)E\frac{1}{\sum_{s=1}^p Z_s^2} \\
= \; & (p-2)E\frac{1}{p-2+2K},
\end{aligned} \tag{67}$$

where K has a Poisson distribution with parameter $\lambda = \frac{1}{2}\boldsymbol{\psi}^{(k)\prime}\boldsymbol{\psi}^{(k)}$. Consider the second term in (66), we have

$$\begin{aligned} E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x) &= E\frac{1}{\sum_{s=1}^p Z_s^2}\mathbf{Z}^T\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x) \\ &= \sum_{j=1}^p E\frac{1}{\sum_{s=1}^p Z_s^2}g_j Z_j, \end{aligned} \tag{68}$$

where g_j is the j-th element of $\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x)$.

Combing (64), (65), (66), (67) and (68), we have

$$\begin{aligned} & EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\ = \; & E(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T\boldsymbol{\Sigma}_0{}^{-1}(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)) \\ & -E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\ = \; & -2cE\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\ & +c^2E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)} \\ = \; & -2c(p-2)E\frac{1}{p-2+2K} - 2c\sum_{j=1}^p E\frac{1}{\sum_{s=1}^p Z_s^2}g_j Z_j + c^2E\frac{1}{p-2+2K}. \end{aligned} \tag{69}$$

We need the equation above to be less than 0 to have

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)).$$

Since the calculations of the expectations are difficult , upper and lower bounds will be obtained.

Consider

$$\begin{aligned}
-2c\sum_{j=1}^{p}E\frac{1}{\sum_{s=1}^{p}Z_s^2}g_jZ_j &\leq \left|-2c\sum_{j=1}^{p}E\frac{1}{\sum_{s=1}^{p}Z_s^2}g_jZ_j\right| \\
&\leq 2c\sum_{j=1}^{p}E\frac{|Z_j|}{\sum_{s=1}^{p}Z_s^2}|g_j| \\
&\leq 2c\sum_{j=1}^{p}E\frac{|Z_j|}{\sum_{s\neq j}Z_s^2}|g_j| \\
&= 2c\sum_{j=1}^{p}E|Z_j|E\frac{1}{\sum_{s\neq j}Z_s^2}|g_j|. \qquad (70)
\end{aligned}$$

We have $E|Z_j| \leq \sqrt{\frac{2}{\pi}} + |\psi_j^{(k)}|$ from line (63). Furthermore, an upper bound of $|\psi_j^{(k)}|$ is

$$\begin{aligned}
|\psi_j^{(k)}| &\leq ||\boldsymbol{\psi}^{(k)}||_2 = ||\mathbf{A}^{-1}\boldsymbol{\xi}^{(k)}||_2 \leq ||\mathbf{A}^{-1}||_2||\boldsymbol{\xi}^{(k)}||_2 \\
&\leq \sqrt{\text{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}} \cdot \sqrt{\sum_{j=1}^{p}\left[\mu_j^{(k)}(x) + b_j^{(k)}(x)\right]^2} \\
&\leq \sqrt{\text{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}} \cdot \sqrt{p(M_1+B)^2}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
|g_j| &\leq ||\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x)||_2 \leq ||\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}||_2||\mathbf{b}^{(k)}(x)||_2 \\
&\leq \sqrt{\text{largest eigenvalue of } \left(\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}\right)^T\mathbf{A}^T\boldsymbol{\Sigma}_0{}^{-1}} \cdot \sqrt{\sum_{j=1}^{p}b_j^{(k)}(x)^2} \\
&\leq \sqrt{\text{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}} \cdot \sqrt{pB^2}.
\end{aligned}$$

An upper bound of $E\frac{1}{\sum_{s\neq j}Z_s^2}$ is

$$E\frac{1}{\sum_{s\neq j}Z_s^2} = E\frac{1}{p-3+2K_j} \leq \frac{1}{p-3}.$$

Therefore, we can obtain an upper bound of $-2c\sum_{j=1}^{p} E\frac{1}{\sum_{s=1}^{p} Z_s^2} g_j Z_j$ as following

$$\begin{aligned}
& -2c\sum_{j=1}^{p} E\frac{1}{\sum_{s=1}^{p} Z_s^2} g_j Z_j \\
\leq\ & 2c\sum_{j=1}^{p}\left(\sqrt{\frac{2}{\pi}} + \sqrt{\text{largest eigenvalue of } \boldsymbol{\Sigma_0}^{-1}}\cdot\sqrt{p(M_1+B)^2}\right) \\
& \cdot\sqrt{\text{largest eigenvalue of } \boldsymbol{\Sigma_0}^{-1}}\cdot\sqrt{pB^2}\cdot\frac{1}{p-3} \\
=\ & \frac{2cp\sqrt{p}B}{p-3}\left(\sqrt{\frac{2}{\pi}\cdot\text{largest eigenvalue of } \boldsymbol{\Sigma_0}^{-1}}\right. \qquad (71) \\
& \left. +\sqrt{p}(M_1+B)\cdot\text{largest eigenvalue of } \boldsymbol{\Sigma_0}^{-1}\right). \qquad (72)
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
E\frac{1}{p-2+2K} \geq E\frac{1}{p-2+(p-2)K} &= \frac{1}{p-2}E\frac{1}{K+1} \\
&= \frac{1}{p-2}\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{(k+1)k!} \\
&= \frac{1}{p-2}\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\cdot\frac{1}{\lambda} \\
&= \frac{1}{p-2}\cdot\frac{1}{\lambda}\left(1-e^{-\lambda}\right).
\end{aligned}$$

And

$$E\frac{1}{p-2+2K} \leq E\frac{1}{2+2K} = \frac{1}{2}E\frac{1}{K+1} = \frac{1}{2}\cdot\frac{1}{\lambda}\left(1-e^{-\lambda}\right).$$

Since

$$\lambda = \frac{1}{2}\boldsymbol{\psi}^{(k)\prime}\boldsymbol{\psi}^{(k)} = \frac{1}{2}||\boldsymbol{\psi}^{(k)}||_2^2 \leq \frac{1}{2}\cdot\text{largest eigenvalue of } \boldsymbol{\Sigma_0}^{-1}\cdot p(M_1+B)^2,$$

and $\boldsymbol{\xi}^{(k)} = \mathbf{A}\boldsymbol{\psi}^{(k)}$, we have $||\boldsymbol{\xi}^{(k)}||_2^2 \leq ||\mathbf{A}||_2^2||\boldsymbol{\psi}^{(k)}||_2^2$. Therefore

$$\lambda = \frac{1}{2}||\boldsymbol{\psi}^{(k)}||_2^2 \geq \frac{1}{2}\cdot\frac{||\boldsymbol{\xi}^{(k)}||_2^2}{||\mathbf{A}||_2^2} \geq \frac{1}{2}\cdot\frac{p(M_0^2-2M_1B)}{\text{largest eigenvalue of } \boldsymbol{\Sigma_0}}.$$

Hence,

$$E\frac{1}{p-2+2K} \geq \frac{1}{p-2}\cdot\frac{2}{\text{largest eigenvalue of } {\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\cdot\text{largest eigenvalue of }\mathbf{\Sigma}_0}}\right). \tag{73}$$

And an upper bound is

$$E\frac{1}{p-2+2K} \leq \frac{\text{largest eigenvalue of }\mathbf{\Sigma}_0}{p(M_0^2-2M_1B)}\left(1-e^{-\frac{1}{2}\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\right). \tag{74}$$

Finally, an upper bound of equation (69) can be obtained by (72), (73) and (74)

$$\begin{aligned}
& -2c(p-2)E\frac{1}{p-2+2K} - 2c\sum_{j=1}^{p}E\frac{1}{\sum_{s=1}^{p}Z_s^2}g_jZ_j + c^2E\frac{1}{p-2+2K} \\
\leq\; & \frac{2cp\sqrt{p}B}{p-3}\left(\sqrt{\frac{2}{\pi}}\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\right. \\
& \left.+\sqrt{p}(M_1+B)\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\right) \\
& -\frac{4c}{\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\cdot\text{largest eigenvalue of }\mathbf{\Sigma}_0}}\right) \\
& +c^2\frac{\text{largest eigenvalue of }\mathbf{\Sigma}_0}{p(M_0^2-2M_1B)}\left(1-e^{-\frac{1}{2}\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\right).
\end{aligned}$$

To make $EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))$, set the right hand side of the inequality above to be less than 0 and we have

$$\begin{aligned}
& \frac{2p\sqrt{p}B}{p-3}\left(\sqrt{\frac{2}{\pi}}\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\right. \\
& \left.+\sqrt{p}(M_1+B)\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\right) \\
- \;& \frac{4}{\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\cdot\text{largest eigenvalue of }\mathbf{\Sigma}_0}}\right) \\
+ \;& c\frac{\text{largest eigenvalue of }\mathbf{\Sigma}_0}{p(M_0^2-2M_1B)}\left(1-e^{-\frac{1}{2}\cdot\text{largest eigenvalue of }{\mathbf{\Sigma}_0}^{-1}\cdot p(M_1+B)^2}\right) \\
< \;& 0.
\end{aligned}$$

Last, we consider the most general situation when both $\mathbf{\Sigma}$ and the bias $b_j^{(k)}$ are unknown. Before stating the main result, we present a consistent estimator of $\mathbf{\Sigma}$.

Proposition 3.4 *Suppose the error terms* $\begin{pmatrix} \epsilon_{1k} \\ \epsilon_{2k} \\ \vdots \\ \epsilon_{pk} \end{pmatrix}$ *follow a multivariate normal distribution with mean* $\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ *and variance* $\mathbf{\Sigma} = \begin{pmatrix} {\sigma_1}^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & {\sigma_2}^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & {\sigma_p}^2 \end{pmatrix}$, *with* $\sigma_{ij}, {\sigma_j}^2 > 0, i, j = 1, 2, \ldots, p$ *and* $k = 1, 2, \ldots, n$. *Assume that the linear estimator is local regression, kernel smoothing, or asymptotically equivalent to kernel smoothing. Assume* $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ *uniformly. Then*

$$\widehat{\mathbf{\Sigma}} = \frac{1}{m} \sum_{k:|x_k - x| > 2h} \left(\widehat{\mathbf{\Sigma}}\right)_{ij},$$

where $\left(\widehat{\mathbf{\Sigma}}\right)_{ij} = [Y_{ik} - \widehat{\mu}_i(x_k)]\,[Y_{jk} - \widehat{\mu}_j(x_k)]$ *is the* (i, j)*-th element of* $\widehat{\mathbf{\Sigma}}$, *and* $m = \sum_{k:|x_k - x| > 2h} [1 - 2l_k(x_k) + ||l(x_k)||^2]$, *is a consistent estimator of* $\mathbf{\Sigma}$.

Proof of Proposition 3.4. In Proposition 3.3, we showed that

$$\frac{1}{m} \sum_{k:|x_k - x| > 2h} [Y_{jk} - \widehat{\mu}_j(x_k)]^2$$

is a consistent estimator of σ_j^2 for $j = 1, 2, \ldots, p$.

To show that $\frac{1}{m} \sum_{k:|x_k - x| > 2h} [Y_{ik} - \widehat{\mu}_i(x_k)]\,[Y_{jk} - \widehat{\mu}_j(x_k)]$ consistently estimates σ_{ij}, similar to the proof of Proposition 3.3, first we consider

$$\begin{aligned}
& \frac{1}{n} \sum_{k=1}^{n} [Y_{ik} - \widehat{\mu}_i(x_k)]\,[Y_{jk} - \widehat{\mu}_j(x_k)] \\
= \quad & \frac{1}{n} \sum_{k=1}^{n} [\mu_i(x_k) - \widehat{\mu}_i(x_k) + \epsilon_{ik}]\,[\mu_j(x_k) - \widehat{\mu}_j(x_k) + \epsilon_{jk}] \\
= \quad & \frac{1}{n} \sum_{k=1}^{n} \{[\mu_i(x_k) - \widehat{\mu}_i(x_k)]\,[\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{jk}\,[\mu_i(x_k) - \widehat{\mu}_i(x_k)] \\
& + \epsilon_{ik}\,[\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{ik}\epsilon_{jk}\} \\
\xrightarrow{P} \quad & \sigma_{ij},
\end{aligned}$$

since the first three summands converge in probability to 0 under the assumpation that $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ uniformly, and $\frac{1}{n}\sum_{k=1}^n \epsilon_{ik}\epsilon_{jk} \xrightarrow{P} E\epsilon_{i1}\epsilon_{j1} = \sigma_{ij}$ by weak law of large numbers.

Next, we have

$$\begin{aligned}
& \frac{1}{n-2\gamma_1+\gamma_2}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] \\
= \quad & \frac{n}{n-2\gamma_1+\gamma_2} \cdot \frac{1}{n}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right],
\end{aligned}$$

where $\gamma_1 = \sum_{k=1}^n l_k(x_k)$ and $\gamma_2 = \sum_{k=1}^n ||l(x_k)||^2$. It is shown that $\frac{n}{n-2\gamma_1+\gamma_2} \to 1$ in the proof of Proposition 3.3. Therefore, $\frac{1}{n-2\gamma_1+\gamma_2}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] \xrightarrow{P} \sigma_{ij}$ by Slutsky's theorem.

Finally, consider

$$\begin{aligned}
& \frac{1}{m}\sum_{k:|x_k-x|>2h} \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] \\
= \quad & \frac{1}{m}\left\{\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] - \sum_{k:|x_k-x|\le 2h} \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right]\right\} \\
= \quad & \frac{1}{m}\left\{(n-2\gamma_1+\gamma_2)\cdot\frac{1}{n-2\gamma_1+\gamma_2}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right]\right. \\
& \left. - \sum_{k:|x_k-x|\le 2h} \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right]\right\} \\
= \quad & \frac{n-2\gamma_1+\gamma_2}{m}\cdot\frac{1}{n-2\gamma_1+\gamma_2}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] \\
& -\frac{1}{m}\sum_{k:|x_k-x|\le 2h} \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right].
\end{aligned}$$

Since $\frac{n-2\gamma_1+\gamma_2}{m} \to 1$ from Proposition 3.3,

$$\frac{n-2\gamma_1+\gamma_2}{m}\cdot\frac{1}{n-2\gamma_1+\gamma_2}\sum_{k=1}^n \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right] \xrightarrow{P} \sigma_{ij}.$$

Put $n_x = card\{k : |x_k - x| \le 2h\}$, we have

$$\begin{aligned}
& \frac{1}{m} \sum_{k:|x_k-x|\le 2h} [Y_{ik} - \widehat{\mu}_i(x_k)] [Y_{jk} - \widehat{\mu}_j(x_k)] \\
= & \frac{n_x}{m} \cdot \frac{1}{n_x} \sum_{k:|x_k-x|\le 2h} [Y_{ik} - \widehat{\mu}_i(x_k)] [Y_{jk} - \widehat{\mu}_j(x_k)] \\
= & \frac{n_x}{m} \cdot \frac{1}{n_x} \sum_{k:|x_k-x|\le 2h} [\mu_i(x_k) - \widehat{\mu}_i(x_k) + \epsilon_{ik}] [\mu_j(x_k) - \widehat{\mu}_j(x_k) + \epsilon_{jk}] \\
= & \frac{n_x}{m} \cdot \frac{1}{n_x} \sum_{k:|x_k-x|\le 2h} \{[\mu_i(x_k) - \widehat{\mu}_i(x_k)] [\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{jk} [\mu_i(x_k) - \widehat{\mu}_i(x_k)] \\
& + \epsilon_{ik} [\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{ik}\epsilon_{jk}\} .
\end{aligned}$$

Similarly, $\frac{1}{n_x} \sum_{k:|x_k-x|\le 2h} \{[\mu_i(x_k) - \widehat{\mu}_i(x_k)] [\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{jk} [\mu_i(x_k) - \widehat{\mu}_i(x_k)]$ $+ \epsilon_{ik} [\mu_j(x_k) - \widehat{\mu}_j(x_k)] + \epsilon_{ik}\epsilon_{jk}\} \xrightarrow{P} \sigma_{ij}$ by weak law of large numbers and the assumpation that $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ uniformly. Since $\frac{n_x}{m} \to 0$, $\frac{1}{m} \sum_{k:|x_k-x|\le 2h} [Y_{ik} - \widehat{\mu}_i(x_k)] [Y_{jk} - \widehat{\mu}_j(x_k)]$ $\xrightarrow{P} 0$ by Slutsky's theorem. Therefore, $\frac{1}{m} \sum_{k:|x_k-x|>2h} [Y_{ik} - \widehat{\mu}_i(x_k)] \cdot [Y_{jk} - \widehat{\mu}_j(x_k)] \xrightarrow{P} \sigma_{ij}$.

We end this chapter with the following theorem where both $\mathbf{\Sigma}$ and bias $b_j^{(k)}$ are unknown. Put $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$, $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$ and $\mathbf{b}^{(k)}(x) = (b_1^{(k)}(x), b_2^{(k)}(x), \ldots, b_p^{(k)}(x))^T$. Let $\mathbf{\Sigma}_0 := var(\widehat{\boldsymbol{\mu}}^{(k)}(x))$, where $(\mathbf{\Sigma}_0)_{ij} = \sum_{s=1}^n l_{i,s}^{(k)}(x) l_{j,s}^{(k)}(x) (\mathbf{\Sigma})_{ij}$.

Theorem 3.6 *Assume the error terms have distribution* $\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{\Sigma} \right)$, *where* $\mathbf{\Sigma}$ *is unknown, symmetric and positive definite. Let* $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_i^{(k)}(x) Y_{ji}$ *be the linear estimator of* $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$. *Suppose* $p \ge 4$, $M_0 \le \left|\mu_j^{(k)}(x)\right| \le M_1$, *and that the unknown bias* $b_j^{(k)}(x) = E[\widehat{\mu}_j^{(k)}(x)] - \mu_j^{(k)}(x) \in [-B, B]$ *for all* $k = 0, 1, \ldots, J$ *and* $j = 1, 2, \ldots, p$. *Let* $\mathbf{S}$ *be a consistent estimator of* $\mathbf{\Sigma}$ *defined in Proposition 3.4 and define*

$\mathbf{S_0} = ||l^{(k)}(x)||^2\mathbf{S}$. *Put* $\widetilde{\boldsymbol{\mu}}^{(k)}(x) := \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\mathbf{S_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\right]\widehat{\boldsymbol{\mu}}^{(k)}(x)$, *where* c *satisfies*

$$\begin{aligned} & \frac{2p\sqrt{p}B}{p-3}\left(\sqrt{\frac{2}{\pi}\cdot \textit{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}}\right. \\ & \left. +\sqrt{p}(M_1+B)\cdot \textit{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}\right) \\ - & \frac{4}{\textit{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}\cdot p(M_1+B)^2}\left(1-e^{-\frac{p(M_0^2-2M_1B)}{2\cdot \textit{largest eigenvalue of } \boldsymbol{\Sigma}_0}}\right) \\ + & c\frac{\textit{largest eigenvalue of } \boldsymbol{\Sigma}_0}{p(M_0^2-2M_1B)}\left(1-e^{-\frac{1}{2}\cdot \textit{largest eigenvalue of } \boldsymbol{\Sigma}_0{}^{-1}\cdot p(M_1+B)^2}\right) \\ < & 0. \end{aligned}$$

If the loss function is $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T\boldsymbol{\Sigma}_0{}^{-1}(\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, *then*

$$EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) < EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)).$$

Proof of Theorem 3.6. Since $\widehat{\boldsymbol{\mu}}^{(k)}(x) \sim N\left(\boldsymbol{\xi}^{(k)} = \boldsymbol{\mu}^{(k)}(x) + \mathbf{b}^{(k)}(x), \boldsymbol{\Sigma}_0\right)$, we have

$$\begin{aligned} & EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\ = & E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)} + \mathbf{b}^{(k)}(x)\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)} + \mathbf{b}^{(k)}(x)\right) \\ = & E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \qquad (75) \\ & +2\mathbf{b}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) + \mathbf{b}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x) \\ = & p + \mathbf{b}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x). \qquad (76) \end{aligned}$$

On the other hand, we have

$$\begin{aligned} & EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)\right) \\ = & E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)} + \mathbf{b}^{(k)}(x)\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)} + \mathbf{b}^{(k)}(x)\right) \\ = & E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) + 2\mathbf{b}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\ & +\mathbf{b}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x). \end{aligned}$$

(77)

Put $\mathbf{Y_i} = (Y_{1i}, Y_{2i}, \ldots, Y_{pi})^T$, for $i = 1, 2, \ldots, n$. From (90), we have $\mathbf{Y_i} \overset{indep.}{\sim} N\left(\boldsymbol{\mu}(x_i) = \begin{pmatrix} \mu_1(x_i) \\ \mu_2(x_i) \\ \vdots \\ \mu_p(x_i) \end{pmatrix}, \boldsymbol{\Sigma}\right)$. Since $\boldsymbol{\Sigma}$ is positive definite and symmetric, there exists a symmetric, positive definite matirx $\mathbf{A}$ such that $\boldsymbol{\Sigma} = \mathbf{AA}$. Let $\mathbf{Z}_i = \mathbf{A}^{-1}\mathbf{Y}_i$, which has a multivariate normal distribution with mean $\boldsymbol{\mu}^*(x_i) = \mathbf{A}^{-1}\boldsymbol{\mu}(x_i)$ and variance $\mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1} = \mathbf{I}$. Therefore,

$$\begin{aligned}
& E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\
= & \int \left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \frac{1}{(2\pi)^{pn/2}} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \\
& \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{Y_i} - \boldsymbol{\mu}(x_i)\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{Y_i} - \boldsymbol{\mu}(x_i)\right)\right\} d\mathbf{Y_1}\, d\mathbf{Y_2} \cdots d\mathbf{Y_n} \\
= & \int \left(\widetilde{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\widetilde{\boldsymbol{\mu}}^{(k)}(x) - 2\boldsymbol{\xi}^{(k)T}\boldsymbol{\Sigma}_0{}^{-1}\widetilde{\boldsymbol{\mu}}^{(k)}(x) + \boldsymbol{\xi}^{(k)T}\boldsymbol{\Sigma}_0{}^{-1}\boldsymbol{\xi}^{(k)}\right) \frac{1}{(2\pi)^{pn/2}} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \\
& \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{AZ_i} - \mathbf{A}\boldsymbol{\mu}^*(x_i)\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{AZ_i} - \mathbf{A}\boldsymbol{\mu}^*(x_i)\right)\right\} |\mathbf{A}|^n d\mathbf{Z_1}\, d\mathbf{Z_2} \cdots d\mathbf{Z_n} \\
= & \int \left(\widetilde{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\widetilde{\boldsymbol{\mu}}^{(k)}(x) - 2\boldsymbol{\xi}^{(k)T}\boldsymbol{\Sigma}_0{}^{-1}\widetilde{\boldsymbol{\mu}}^{(k)}(x) + \boldsymbol{\xi}^{(k)T}\boldsymbol{\Sigma}_0{}^{-1}\boldsymbol{\xi}^{(k)}\right) \frac{1}{(2\pi)^{pn/2}} \\
& \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)^T \left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)\right\} d\mathbf{Z_1}\, d\mathbf{Z_2} \cdots d\mathbf{Z_n}. \qquad (78)
\end{aligned}$$

We assume $\widehat{\mu}_j^{(k)}(x)$ has the form $\sum_{i=1}^n l_i^{(k)}(x)Y_{ji}$, hence $\boldsymbol{\Sigma}_0 = var(\widehat{\boldsymbol{\mu}}^{(k)}(x)) = ||l^{(k)}(x)||^2\boldsymbol{\Sigma}$. Let $\widehat{\boldsymbol{\mu}}^{(k)^*}(x) = \begin{pmatrix} \widehat{\mu}_1^{(k)^*}(x) \\ \widehat{\mu}_2^{(k)^*}(x) \\ \vdots \\ \widehat{\mu}_p^{(k)^*}(x) \end{pmatrix} = \sum_{i=1}^n l_i^{(k)}(x)\mathbf{Z}_i = \sum_{i=1}^n l_i^{(k)}(x)\mathbf{A}^{-1}\mathbf{Y}_i = \mathbf{A}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)$, which follows a multivariate normal distribution $N\left(\boldsymbol{\xi}^{(k)^*} = \mathbf{A}^{-1}\boldsymbol{\xi}^{(k)}, \mathbf{A}^{-1}\boldsymbol{\Sigma}_0\mathbf{A}^{-1} = ||l^{(k)}(x)||^2\mathbf{I}\right)$. Since the estimator of $\boldsymbol{\Sigma}$ can be written as

$$
\begin{aligned}
\mathbf{S} &= \frac{1}{m} \sum_{i:|x_i - x| > 2h} \begin{pmatrix} Y_{1i} - \widehat{\mu}_1(x_i) \\ Y_{2i} - \widehat{\mu}_2(x_i) \\ \vdots \\ Y_{pi} - \widehat{\mu}_p(x_i) \end{pmatrix} \Big(Y_{1i} - \widehat{\mu}_1(x_i) \quad Y_{2i} - \widehat{\mu}_2(x_i) \quad \cdots \quad Y_{pi} - \widehat{\mu}_p(x_i) \Big) \\
&= \frac{1}{m} \sum_{i:|x_i - x| > 2h} \Big(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}^{(k)}(x_i) \Big) \Big(\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}^{(k)}(x_i) \Big)^T,
\end{aligned}
$$

where $m = \sum_{i:|x_i - x| > 2h} [1 - 2l_i(x_i) + ||l(x_i)||^2]$, we have

$$
\begin{aligned}
\mathbf{S}^* &:= \frac{1}{m} \sum_{i:|x_i - x| > 2h} (\mathbf{Z}_i - \widehat{\boldsymbol{\mu}}^*(x_i)) (\mathbf{Z}_i - \widehat{\boldsymbol{\mu}}^*(x_i))^T \\
&= \frac{1}{m} \sum_{i:|x_i - x| > 2h} \big(\mathbf{A}^{-1}\mathbf{Y}_i - \mathbf{A}^{-1}\widehat{\boldsymbol{\mu}}(x_i) \big) \big(\mathbf{A}^{-1}\mathbf{Y}_i - \mathbf{A}^{-1}\widehat{\boldsymbol{\mu}}(x_i) \big)^T \\
&= \mathbf{A}^{-1} \frac{1}{m} \sum_{i:|x_i - x| > 2h} (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}(x_i)) (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}(x_i))^T \mathbf{A}^{-1} \\
&= \mathbf{A}^{-1}\mathbf{S}\mathbf{A}^{-1}.
\end{aligned}
$$

Define $\mathbf{S}_0^* = ||l^{(k)}(x)||^2 \mathbf{S}^*$, then we have $\mathbf{S}_0^* = ||l^{(k)}(x)||^2 \mathbf{A}^{-1}\mathbf{S}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{S}_0\mathbf{A}^{-1}$. Put $\widetilde{\boldsymbol{\mu}}^{(k)^*}(x) := \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \right] \widehat{\boldsymbol{\mu}}^{(k)^*}(x)$. Consider (78) and the following results are obtained

$$
\begin{aligned}
\widetilde{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1} \widetilde{\boldsymbol{\mu}}^{(k)}(x) &= \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{S_0}{}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} \right]^2 \widehat{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x) \\
&= \left[1 - \frac{c}{\Big(\mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \Big)^T (\mathbf{A}\mathbf{S}_0^*\mathbf{A})^{-1} \mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \right]^2 \Big(\mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \Big)^T \boldsymbol{\Sigma}_0{}^{-1} \Big(\mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \Big) \\
&= \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{A}\mathbf{A}^{-1}\mathbf{S}_0^{*-1}\mathbf{A}^{-1}\mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \right]^2 \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{A} ||l^{(k)}(x)||^{-2} \boldsymbol{\Sigma}^{-1} \mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
&= \left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \right]^2 ||l^{(k)}(x)||^{-2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
&= ||l^{(k)}(x)||^{-2} \widetilde{\boldsymbol{\mu}}^{(k)^*}(x)^T \widetilde{\boldsymbol{\mu}}^{(k)^*}(x),
\end{aligned}
$$

similarly

$$
\begin{aligned}
& \boldsymbol{\xi}^{(k)^T}\boldsymbol{\Sigma}_0{}^{-1}\widetilde{\boldsymbol{\mu}}^{(k)}(x) = \left[1-\frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\mathbf{S_0}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\right]\boldsymbol{\xi}^{(k)^T}\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x) \\
= & \left[1-\frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\mathbf{S}_0^{*-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\right]\left(\mathbf{A}\boldsymbol{\xi}^{(k)^*}\right)^T||l^{(k)}(x)||^{-2}\boldsymbol{\Sigma}^{-1}\mathbf{A}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= & ||l^{(k)}(x)||^{-2}\left[1-\frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\mathbf{S}_0^{*-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\right]\boldsymbol{\xi}^{(k)^{*T}}\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= & ||l^{(k)}(x)||^{-2}\boldsymbol{\xi}^{(k)^{*T}}\widetilde{\boldsymbol{\mu}}^{(k)^*}(x),
\end{aligned}
$$

and

$$
\boldsymbol{\xi}^{(k)^T}\boldsymbol{\Sigma}_0{}^{-1}\boldsymbol{\xi}^{(k)} = \left(\mathbf{A}\boldsymbol{\xi}^{(k)^*}\right)^T||l^{(k)}(x)||^{-2}\boldsymbol{\Sigma}^{-1}\mathbf{A}\boldsymbol{\xi}^{(k)^*} = ||l^{(k)}(x)||^{-2}\boldsymbol{\xi}^{(k)^{*T}}\boldsymbol{\xi}^{(k)^*}.
$$

Therefore, equation (78) can be written as

$$
\begin{aligned}
& E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x)-\boldsymbol{\xi}^{(k)}\right)^T\boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x)-\boldsymbol{\xi}^{(k)}\right) \\
= & \int ||l^{(k)}(x)||^{-2}\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)^T\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-2\boldsymbol{\xi}^{(k)^{*T}}\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)+\boldsymbol{\xi}^{(k)^{*T}}\boldsymbol{\xi}^{(k)^*}\right) \\
& \cdot\frac{1}{(2\pi)^{pn/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{Z_i}-\boldsymbol{\mu}^*(x_i)\right)^T\left(\mathbf{Z_i}-\boldsymbol{\mu}^*(x_i)\right)\right\}d\mathbf{Z_1}\,d\mathbf{Z_2}\cdots d\mathbf{Z_n} \\
= & ||l^{(k)}(x)||^{-2}E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right). \qquad (79)
\end{aligned}
$$

Put $\boldsymbol{\xi}^{(k)^*} := \mathbf{A}^{-1}\boldsymbol{\xi}^{(k)}$, $\mathbf{b}^{(k)^*}(x) := \mathbf{A}^{-1}\mathbf{b}^{(k)}(x)$. And we obtain the following result

$$
\begin{aligned}
& \mathbf{b}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1} E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\
= & \int \mathbf{b}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \frac{1}{(2\pi)^{pn/2}} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \\
& \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \left(\mathbf{Y_i} - \boldsymbol{\mu}(x_i)\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{Y_i} - \boldsymbol{\mu}(x_i)\right)\right\} d\mathbf{Y_1}\, d\mathbf{Y_2} \cdots d\mathbf{Y_n} \\
= & \int \left(\mathbf{A}\mathbf{b}^{(k)^*}(x)\right)^T ||l^{(k)}(x)||^{-2} \boldsymbol{\Sigma}^{-1}\mathbf{A}\left(\left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\right]\widehat{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right) \\
& \cdot \frac{1}{(2\pi)^{pn/2}} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)^T\left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)\right\} d\mathbf{Z_1}\, d\mathbf{Z_2} \cdots d\mathbf{Z_n} \\
= & ||l^{(k)}(x)||^{-2} \int \mathbf{b}^{(k)^*}(x)^T \left(\left[1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\right]\widehat{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right) \\
& \cdot \frac{1}{(2\pi)^{pn/2}} \cdot \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)^T\left(\mathbf{Z_i} - \boldsymbol{\mu}^*(x_i)\right)\right\} d\mathbf{Z_1}\, d\mathbf{Z_2} \cdots d\mathbf{Z_n} \\
= & ||l^{(k)}(x)||^{-2}\mathbf{b}^{(k)^*}(x)^T E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right). \qquad (80)
\end{aligned}
$$

By substituting (79) and (80) in (77), we get

$$
\begin{aligned}
& E\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\xi}^{(k)}\right) \\
= & ||l^{(k)}(x)||^{-2} E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right)^T\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right) + 2||l^{(k)}(x)||^{-2}\mathbf{b}^{(k)^*}(x)^T \\
& \cdot E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right) + \mathbf{b}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x). \qquad (81)
\end{aligned}
$$

Before presenting any further results, we propose a property of $\mathbf{S}_0^*$ that will be used in future proof. Let $t = card\{i : |x_i - x| > 2h\}$. Define

$$
L = \begin{pmatrix}
1 - l_1(x_1) & -l_2(x_1) & \cdots & -l_n(x_1) \\
-l_1(x_2) & 1 - l_2(x_2) & \cdots & -l_n(x_2) \\
\vdots & \vdots & & \vdots \\
-l_1(x_t) & -l_2(x_t) & \cdots & 1 - l_n(x_t)
\end{pmatrix}
= \begin{pmatrix}
1 - l_1(x_1) & L_{12} \\
L_{21} & L_{22}
\end{pmatrix}.
$$

Then $L^TL = \begin{pmatrix} [1-l_1(x_1)]^2 + L_{12}L_{12}^T & [1-l_1(x_1)]L_{21}^T + L_{12}L_{22}^T \\ [1-l_1(x_1)]L_{21} + L_{22}L_{12}^T & L_{21}L_{21}^T + L_{22}L_{22}^T \end{pmatrix} = \begin{pmatrix} \tilde{L}_{11} & \tilde{L}_{12} \\ \tilde{L}_{21} & \tilde{L}_{22} \end{pmatrix}$. If $p < t$ and $\tilde{L}_{11} - \tilde{L}_{12}\tilde{L}_{22}^{-1}\tilde{L}_{21} \neq 0$, $\mathbf{S}_0^*$ is full rank with probability 1.

Similar to Dykstra (1970), let $\mathbf{W}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\mu}}^*(x_i)$ and $\hat{U} = \left(\hat{U}\right)_{ij}$, where the ij-th element of $\hat{U}$ is $\left(\hat{U}\right)_{ij} = Z_{ij} - \widehat{\mu}_i^*(x_j)$. We have $\mathbf{S}_0^* = \frac{1}{m}||l^{(k)}(x)||^2\hat{U}\hat{U}^T$. Therefore $\text{rank}(\mathbf{S}_0^*) = \text{rank}(\hat{U})$. We can write $\hat{U} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \cdots & \mathbf{W}_t \end{pmatrix}$. Since $\hat{U}$ has a multivariate normal distribution with variance $LL^T \otimes I$, where $\otimes$ denotes direct product and I is a $p \times p$ indentity matrix. The conditional distribution of $\mathbf{W}_1|\mathbf{W}_2, \ldots, \mathbf{W}_t$ is multivariate normal with variance $\tilde{L}_{11} \otimes I - \left(\tilde{L}_{12} \otimes I\right)\left(\tilde{L}_{22} \otimes I\right)^{-1}\left(\tilde{L}_{21} \otimes I\right) = \tilde{L}_{11} \otimes I - \left(\tilde{L}_{12}\tilde{L}_{22}^{-1} \otimes I\right)\left(\tilde{L}_{21} \otimes I\right) = \left(\tilde{L}_{11} - \tilde{L}_{12}\tilde{L}_{22}^{-1}\tilde{L}_{21}\right) \otimes I$. By the assumption that $\tilde{L}_{11} - \tilde{L}_{12}\tilde{L}_{22}^{-1}\tilde{L}_{21} \neq 0$, the variance matrix $\left(\tilde{L}_{11} - \tilde{L}_{12}\tilde{L}_{22}^{-1}\tilde{L}_{21}\right) \otimes I$ is full rank. Let $S\{\mathbf{v}_i, i = 1, 2, \ldots, t-1\}$ be the space spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{t-1}$, for any set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{t-1}\}$ in R^p. Hence,

$$\begin{aligned}
& P\left(\text{rank}\left(\mathbf{S}_0^*\right) < p\right) = P\left(\text{rank}(\hat{U}) < p\right) \\
\leq\ & \sum_{i=1}^{t} P\left(\mathbf{W}_i \in S\left\{\mathbf{W}_1, \ldots, \mathbf{W}_{i-1}, \ldots, \mathbf{W}_{i+1}, \ldots, \mathbf{W}_t\right\}\right) \\
\leq\ & t \cdot P\left(\mathbf{W}_1 \in S\left\{\mathbf{W}_2, \ldots, \mathbf{W}_t\right\}\right) \\
=\ & t \cdot E\left(P\left[\mathbf{W}_1 \in S\left\{\mathbf{W}_2, \ldots, \mathbf{W}_t\right\}|\mathbf{W}_2, \ldots, \mathbf{W}_t\right]\right) \\
=\ & t \cdot \int P\left(\mathbf{W}_1 \in S\left\{\mathbf{w}_2, \ldots, \mathbf{w}_t\right\}|\mathbf{W}_2 = \mathbf{w}_2, \ldots, \mathbf{W}_t = \mathbf{w}_t\right) dF\left(\mathbf{w}_2, \ldots, \mathbf{w}_t\right) \\
=\ & t \cdot \int 0 dF\left(\mathbf{w}_2, \ldots, \mathbf{w}_t\right) \qquad (82) \\
=\ & 0.
\end{aligned}$$

Equation (82) holds by the fact that $\left(\tilde{L}_{11} - \tilde{L}_{12}\tilde{L}_{22}^{-1}\tilde{L}_{21}\right) \otimes I$ is nonsingular with probability 1. Therefore, $\mathbf{S}_0^*$ is full rank. And $\mathbf{S}_0^*$ can be decomposed as $\mathbf{S}_0^* = UDU^T$ where U is a orthogonal matrix, and D is a diagonal matrix whose entries are the eigenvalues of $\mathbf{S}_0^*$.

Now consider the first term in (81), we have

$$
\begin{aligned}
& ||l^{(k)}(x)||^{-2} E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right) \\
= \quad & ||l^{(k)}(x)||^{-2} E\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}-\frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T \\
& \cdot\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}-\frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right) \\
= \quad & ||l^{(k)}(x)||^{-2} E\left[\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)-2\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\right. \\
& \left.\cdot \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)+E \frac{c^2}{\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \mathbf{S}_0^{*-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right] \\
= \quad & ||l^{(k)}(x)||^{-2}\left[p||l^{(k)}(x)||^2+c^2 E \frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\left(\mathbf{U D U}^T\right)^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right. \\
& \left.-\frac{2 c}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\left(\mathbf{U D U}^T\right)^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right] \\
= \quad & p-2 c||l^{(k)}(x)||^{-2} \frac{1}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\left(\mathbf{U D U}^T\right)^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
& +c^2||l^{(k)}(x)||^{-2} E \frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\left(\mathbf{U D U}^T\right)^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= \quad & p-2 c||l^{(k)}(x)||^{-2} \frac{1}{\left(\mathbf{U}^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T \mathbf{D}^{-1}\left(\mathbf{U}^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
& +c^2||l^{(k)}(x)||^{-2} E \frac{1}{\left(\left(\mathbf{U}^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T \mathbf{D}^{-1}\left(\mathbf{U}^{-1} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x). \qquad (83)
\end{aligned}
$$

Let $\mathbf{V} = \mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)$ and $\boldsymbol{\xi}^{(k)^{**}} = \mathbf{U}^{-1}\boldsymbol{\xi}^{(k)^*}$. Let λ_k, $k = 1, 2, \ldots, p$ be the eigenvalues of $\mathbf{S}_0^*$, and $\Lambda_0 \leq \lambda_k \leq \Lambda_1$. We have

$$\begin{aligned}
& -2c||l^{(k)}(x)||^{-2}\frac{1}{\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T\mathbf{D}^{-1}\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= & -2c||l^{(k)}(x)||^{-2}\frac{\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\left(\mathbf{U}^{-1}\right)^T\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)}{\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T\mathbf{D}^{-1}\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)} \\
= & -2c||l^{(k)}(x)||^{-2}\frac{1}{\mathbf{V}^T\mathbf{D}^{-1}\mathbf{V}}\left[\mathbf{U}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)\right]^T\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right) \\
= & -2c||l^{(k)}(x)||^{-2}\frac{1}{\mathbf{V}^T\mathbf{D}^{-1}\mathbf{V}}\left(\mathbf{V}-\boldsymbol{\xi}^{(k)^{**}}\right)^T\mathbf{V} \\
= & -2c||l^{(k)}(x)||^{-2}\sum_{j=1}^{p}\frac{\left(V_j-\xi_j^{(k)^{**}}\right)V_j}{\sum_{k=1}^{p}\lambda_k^{-1}V_k^2} \\
= & -2c||l^{(k)}(x)||^{-2}\sum_{j=1}^{p}||l^{(k)}(x)||^2 E\frac{\sum_{k=1}^{p}\lambda_k^{-1}V_k^2-2\lambda_j^{-1}V_j^2}{\left[\sum_{k=1}^{p}\lambda_k^{-1}V_k^2\right]^2} && (84) \\
= & -2c\sum_{j=1}^{p}E\frac{1}{\sum_{k=1}^{p}\lambda_k^{-1}V_k^2}+4cE\frac{\sum_{j=1}^{p}\lambda_j^{-1}V_j^2}{\left[\sum_{k=1}^{p}\lambda_k^{-1}V_k^2\right]^2} \\
= & -2c(p-2)E\frac{1}{\sum_{k=1}^{p}\lambda_k^{-1}V_k^2} \\
\leq & -2c(p-2)E\frac{1}{\Lambda_0^{-1}\sum_{k=1}^{p}V_k^2} \\
= & -2c\Lambda_0(p-2)E\frac{1}{\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)} \\
= & -2c\Lambda_0(p-2)E\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T\widehat{\boldsymbol{\mu}}^{(k)^*}(x)/||l^{(k)}(x)||^2}\cdot\frac{1}{||l^{(k)}(x)||^2} \\
= & -\frac{2c\Lambda_0(p-2)}{||l^{(k)}(x)||^2}E\frac{1}{p-2+2K}, && (85)
\end{aligned}$$

where $K \sim$ Poisson $\left(\lambda = \frac{\boldsymbol{\xi}^{(k)^{*T}}\boldsymbol{\xi}^{(k)^*}}{2||l^{(k)}(x)||^2}\right)$. Line (84) comes from the fact that if $Z \sim N\left(\mu, \sigma^2\right)$, then $E\left[(Z-\mu)g(Z)\right] = \sigma^2 Eg'(Z)$. In Theorem 3.5, we showed that the upper and lower bounds for λ are

$$\begin{aligned}
\lambda &= \frac{||\boldsymbol{\xi}^{(k)^*}||_2^2}{2||l^{(k)}(x)||^2} = \frac{1}{2||l^{(k)}(x)||^2}||\mathbf{A}^{-1}\boldsymbol{\xi}^{(k)}||_2^2 \\
&\leq \frac{1}{2||l^{(k)}(x)||^2} \cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1} \cdot \sum_{j=1}^{p} \left[\mu_j^{(k)}(x) + b_j^{(k)}(x)\right]^2 \\
&\leq \frac{1}{2||l^{(k)}(x)||^2} \cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1} \cdot p(M_1 + B)^2,
\end{aligned}$$

and

$$\lambda = \frac{||\boldsymbol{\xi}^{(k)^*}||_2^2}{2||l^{(k)}(x)||^2} \geq \frac{1}{2||l^{(k)}(x)||^2} \cdot \frac{||\boldsymbol{\xi}^{(k)}||_2^2}{||\mathbf{A}||_2^2} \geq \frac{1}{2||l^{(k)}(x)||^2} \cdot \frac{p(M_0^2 - 2M_1 B)}{\text{largest eigenvalue of } \boldsymbol{\Sigma}}.$$

As shown in Theorem 3.5, an upper bound for (85) is

$$\begin{aligned}
&-2c||l^{(k)}(x)||^{-2} \frac{1}{\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T \mathbf{D}^{-1} \left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)} \left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x) - \boldsymbol{\xi}^{(k)^*}\right)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
\leq\; & -\frac{2c\Lambda_0(p-2)}{||l^{(k)}(x)||^2} E \frac{1}{p-2+2K} \\
\leq\; & -\frac{2c\Lambda_0(p-2)}{||l^{(k)}(x)||^2} \cdot \frac{1}{p-2} \cdot \frac{1}{\lambda}\left(1 - e^{-\lambda}\right) \\
\leq\; & -\frac{2c\Lambda_0(p-2)}{||l^{(k)}(x)||^2} \cdot \frac{1}{p-2} \cdot \frac{2||l^{(k)}(x)||^2}{\text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1} \cdot p(M_1+B)^2} \\
& \left(1 - e^{-\frac{1}{2||l^{(k)}(x)||^2} \cdot \frac{p(M_0^2 - 2M_1B)}{\text{largest eigenvalue of } \boldsymbol{\Sigma}}}\right) \\
=\; & -\frac{4c\Lambda_0}{\text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1} \cdot p(M_1+B)^2} \left(1 - e^{-\frac{1}{2||l^{(k)}(x)||^2} \cdot \frac{p(M_0^2 - 2M_1B)}{\text{largest eigenvalue of } \boldsymbol{\Sigma}}}\right). \quad (86)
\end{aligned}$$

Consider the last term in (83), we have

$$
\begin{aligned}
& c^2||l^{(k)}(x)||^{-2} E \frac{1}{\left(\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^T \mathbf{D}^{-1}\left(\mathbf{U}^{-1}\widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= \;& c^2||l^{(k)}(x)||^{-2} E \frac{1}{(\mathbf{V}^T\mathbf{D}^{-1}\mathbf{V})^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= \;& c^2||l^{(k)}(x)||^{-2} E \frac{1}{\left(\sum_{k=1}^p \lambda_k^{-1} V_k^2\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
\leq \;& c^2||l^{(k)}(x)||^{-2} E \frac{1}{\Lambda_1^{-2}\left(\sum_{k=1}^p V_k^2\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= \;& c^2||l^{(k)}(x)||^{-2} E \frac{1}{\Lambda_1^{-2}\left(\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x) \\
= \;& c^2||l^{(k)}(x)||^{-2} \Lambda_1^2 E \frac{1}{\widehat{\boldsymbol{\mu}}^{(k)^*}(x)^T \widehat{\boldsymbol{\mu}}^{(k)^*}(x)} \\
= \;& c^2||l^{(k)}(x)||^{-4} \Lambda_1^2 E \frac{1}{p-2+2K} \\
\leq \;& c^2||l^{(k)}(x)||^{-4} \Lambda_1^2 \cdot \frac{1}{2} \cdot \frac{1}{\lambda}\left(1-e^{-\lambda}\right) \qquad (87) \\
\leq \;& c^2||l^{(k)}(x)||^{-4} \Lambda_1^2 \cdot \frac{1}{2} \cdot \frac{2||l^{(k)}(x)||^2 \cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}}{p(M_0^2-2M_1B)} \\
& \left(1-e^{-\frac{1}{2||l^{(k)}(x)||^2}\cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1}\cdot p(M_1+B)^2}\right) \\
= \;& c^2||l^{(k)}(x)||^{-2} \Lambda_1^2 \frac{\text{largest eigenvalue of } \boldsymbol{\Sigma}}{p(M_0^2-2M_1B)}\left(1-e^{-\frac{p(M_1+B)^2}{2||l^{(k)}(x)||^2}\cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1}}\right) \qquad (88)
\end{aligned}
$$

Line (87) is obtained in Theorem 3.5. By combining (86) and (88), we can get an upper bound of (83)

$$
\begin{aligned}
& ||l^{(k)}(x)||^{-2} E\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right)^T\left(\widetilde{\boldsymbol{\mu}}^{(k)^*}(x)-\boldsymbol{\xi}^{(k)^*}\right) \\
\leq \;& p - \frac{4c\Lambda_0^{-1}}{\text{largest eigenvalue of } \boldsymbol{\Sigma}\cdot p(M_1+B)^2}\left(1-e^{-\frac{1}{2||l^{(k)}(x)||^2}\cdot\frac{p(M_0^2-2M_1B)}{\text{largest eigenvalue of } \boldsymbol{\Sigma}}}\right) \\
& + c^2||l^{(k)}(x)||^{-2}\Lambda_1^2 \cdot \frac{\text{largest eigenvalue of } \boldsymbol{\Sigma}}{p(M_0^2-2M_1B)} \\
& \cdot\left(1-e^{-\frac{1}{2||l^{(k)}(x)||^2}\cdot \text{largest eigenvalue of } \boldsymbol{\Sigma}^{-1}\cdot p(M_1+B)^2}\right). \qquad (89)
\end{aligned}
$$

4 James-Stein type estimators over an interval

Suppose we have the general model

$$
\begin{aligned}
Y_{1i} &= \mu_1(x_i) + \epsilon_{1i} \\
Y_{2i} &= \mu_2(x_i) + \epsilon_{2i} \\
&\cdots \\
Y_{pi} &= \mu_p(x_i) + \epsilon_{pi}
\end{aligned}
\tag{90}
$$

for $i \in \{1, \ldots, n\}$, where x_i belongs to a compact set $\mathcal{X} \subset \mathbb{R}$, and
$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{pi} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{\Sigma}\right)$ with $\mathbf{\Sigma}$ semi-positive definite.

Let $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_{j,i}^{(k)}(x) Y_{ji}$ be the linear estimator of $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$, where $k = 0, 1, \ldots, J$ and $j = 1, 2, \ldots, p$. Let $\widetilde{\mu}_j^{(k)}(x)$ be the James-Stein type estimator of $\mu_j^{(k)}(x)$. In this chapter, we will study the Steinized estimators in reducing the integrated risk over $\mathcal{X}$. Define loss function $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^p (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T (||l_j^{(k)}(x)||^2 \mathbf{\Sigma})^{-1} (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, where $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$ and
$\boldsymbol{Z}^{(k)}(x) = (Z_1^{(k)}(x), Z_2^{(k)}(x), \ldots, Z_p^{(k)}(x))^T$. And we will look at the integral of risk

$$
\int_{\mathcal{X}} \mathrm{E} L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) dx.
$$

4.1 Error terms independent

In section 4.1, we assume the error terms ϵ_{1i}, ϵ_{2i}, ..., ϵ_{pi} are independent for $i = 1, 2, \ldots, n$. Suppose $\mathbf{\Sigma}$ has form $\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix}$, where $\sigma_j^2 > 0$. Define bias $b_j^{(k)}(x) = E\left[\widehat{\mu}_j^{(k)}(x)\right] - \mu_j^{(k)}(x)$. Different scenarios will be studied depending on whether bias $b_j^{(k)}(x)$ or variance

σ_j^2 is known.

4.1.1 Bias and variance known

If both bias and variance are known, we have the following theorem.

Theorem 4.1 *Assume bias* $b_j^{(k)}(x)$ *and variance* σ_j^2 *are known for all* $k = 1, 2, \ldots, J$ *and* $j = 1, 2, \ldots, p$*. Define* $\widetilde{\mu}_j^{(k)}(x) := (\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x)) - \frac{c}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_s^{(k)}(x)||^2\sigma_s^2}(\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x))$*, where c is a constant satisfying* $0 < c < 2(p-2)$*, and* $||l_j^{(k)}(x)||^2 = \sum_{i=1}^{n} l_{j,i}^{(k)}(x)^2$*. Then*

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$*,* $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$*, and* $\mathbf{b}^{(k)}(x) = (b_1^{(k)}(x), b_2^{(k)}(x), \ldots, b_p^{(k)}(x))^T$*.*

Proof of Theorem 4.1. From the proof of Theorem 3.2, the following results are obtained:

$$\begin{aligned} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) &= \sum_{j=1}^{p} E\frac{\left[\widetilde{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} \\ &= p + \left[c^2 - 2c(p-2)\right] E\frac{1}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_s^{(k)}(x)||^2\sigma_s^2}, \end{aligned}$$

and

$$EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = \sum_{j=1}^{p} E\frac{\left[\widehat{\mu}_j^{(k)}(x) - b_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l_j^{(k)}(x)||^2\sigma_j^2} = p.$$

Therefore,

$$\begin{aligned} &\int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx - \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx \\ = \ &\left[c^2 - 2c(p-2)\right] \int_{\mathcal{X}} E\frac{1}{\sum_{s=1}^{p}[\widehat{\mu}_s^{(k)}(x) - b_s^{(k)}(x)]^2/||l_s^{(k)}(x)||^2\sigma_s^2}dx. \end{aligned} \tag{91}$$

Since the integration in (91) is positive, we need $c^2 - 2c(p-2) < 0$, i.e. $0 < c < 2(p-2)$, such that $\int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx$.

4.1.2 Bias unknown variance known

In Section 4.1.2, we use the following notations: $A := \int_{\mathcal{X}} \mathrm{E}\frac{1}{p-2+2K(x)}dx$ where $K(x)$ has a Poisson distribution with parameter $\lambda = \sum_{s=1}^{p} \frac{\left[\mu_s^{(k)}(x)+b_s^{(k)}(x)\right]^2}{2\sigma_s^2||l_s^{(k)}(x)||^2}$,
$B := \int_{\mathcal{X}} \mathrm{E}\sum_{j=1}^{p} b_j^{(k)}(x)\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}dx$, and $D := -2(p-2)A - 2B$.

Assume bias $b_j^{(k)}(x)$ is unknown and that variance σ_j^2 is known for all $k = 1, 2, \ldots, J$ and $j = 1, 2, \ldots, p$. Define Stein estimator $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l_j^{(k)}(x)||^2\sigma_j^2}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$. We consider theoretically whether and how the positive constant c may be chosen such that $\int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx$. Numerically, we will closely examine different examples of mean response functions for $p = 3, 4$.

In the rest of Section 4.1.2, let $\widehat{\mu}_j^{(k)}(x)$ be the local regression estimator of $\mu_j^{(k)}(x)$, assume the variance-covariance matrix $\boldsymbol{\Sigma}$ is an identity matrix, and assume tricube weight function with the same nearest neighbor parameter is applied on all p mean response functions in local regression estimation. Therefore, since $\sigma_j^2 = 1$ and $l_{j,i}^{(k)}(x) = l_i^{(k)}(x)$ for all j, we can write $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^{n} l_i^{(k)}(x)Y_{ji}$ and $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$.

Theorem 4.2 *Assume bias $b_j^{(k)}(x)$ is unknown and variance σ_j^2 is known, which is 1 for all $k = 1, 2, \ldots, J$ and $j = 1, 2, \ldots, p$. Define $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$, where c is a constant satisfying $0 < c < -\frac{D}{A}$, and $||l^{(k)}(x)||^2 = \sum_{i=1}^{n} l_i^{(k)}(x)^2$. Then*

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$.

Proof of Theorem 4.2. Similar to the proof of Theorem 3.4, we have

$$
\begin{aligned}
EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)) &= \sum_{j=1}^{p} E\frac{\left[\widetilde{\mu}_j^{(k)}(x)-\mu_j^{(k)}(x)\right]^2}{||l^{(k)}(x)||^2\sigma_j^2} \\
&= \sum_{j=1}^{p} E\frac{\left[\widetilde{\mu}_j^{(k)}(x)-\widehat{\mu}_j^{(k)}(x)+\widehat{\mu}_j^{(k)}(x)-\mu_j^{(k)}(x)\right]^2}{||l^{(k)}(x)||^2} \\
&= EL\left(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)\right)+\sum_{j=1}^{p}\frac{E\left(\widetilde{\mu}_j^{(k)}(x)-\widehat{\mu}_j^{(k)}(x)\right)^2}{||l^{(k)}(x)||^2} \\
&\quad +\sum_{j=1}^{p}\frac{2E\left(\widetilde{\mu}_j^{(k)}(x)-\widehat{\mu}_j^{(k)}(x)\right)\left(\widehat{\mu}_j^{(k)}(x)-\mu_j^{(k)}(x)\right)}{||l^{(k)}(x)||^2} \\
&= EL\left(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)\right)+c^2||l^{(k)}(x)||^2 E\frac{1}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}-\sum_{j=1}^{p}2c||l^{(k)}(x)||^2 \\
&\quad \cdot E\frac{1}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}+4c||l^{(k)}(x)||^2 E\frac{1}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}-\sum_{j=1}^{p}2cb_j^{(k)}(x)E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2} \\
&= EL\left(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x)\right)+c^2E\frac{1}{p-2+2K}-2(p-2)cE\frac{1}{p-2+2K} \\
&\quad -2c\sum_{j=1}^{p}b_j^{(k)}(x)E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2},
\end{aligned}
$$

where $K(x)\sim \text{Poisson}\left(\lambda=\frac{\sum_{s=1}^{p}\left[\mu_s^{(k)}(x)+b_s^{(k)}(x)\right]^2}{2||l^{(k)}(x)||^2}\right)$.

Therefore,

$$
\begin{aligned}
&\int_{\mathcal{X}}\mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x))dx-\int_{\mathcal{X}}\mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x),\boldsymbol{\mu}^{(k)}(x))dx \\
=\;& \int_{\mathcal{X}}\left[c^2E\frac{1}{p-2+2K(x)}-2(p-2)cE\frac{1}{p-2+2K(x)}\right. \\
&\left.-2c\sum_{j=1}^{p}b_j^{(k)}(x)E\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\right]dx \\
=\;& Ac^2+Dc. \qquad (92)
\end{aligned}
$$

Note that since both A and c are positive, (92) is positive if D is positive. Therefore, when D is positive, no Steinization should be applied. If D is negative, James-Stein type estimators with $0<c<-\frac{D}{A}$ will improve local regression estimators in reducing risk. Also, note that (92) achieves its minimum at $c = -\frac{D}{2A}$.

In the following two examples, we use different mean response functions to calculate A, B and D. Data sets of size 100 are generated from model (90) with equispaced x_i on $\mathcal{X} = [-1, 1]$. Simpson's rule and Monte Carlo algorithm are applied in the calculations of A and B, respectively. In estimating $\mu_j(x)$, $\mu_j^{(1)}(x)$ and $\mu_j^{(2)}(x)$, J is set to be 0, 1 and 2 respectively. Let α be the nearest neighbor parameter in local regression, consider $\alpha \in \{0.15, 0.10, 0.05\}$ for $J = 0$, $\alpha \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$ for $J = 1$ and $\alpha \in \{0.65, 0.45, 0.25, 0.15, 0.05\}$ for $J = 2$.

Example 1. Consider $p = 3$. Results are shown in Table (5) - Table (14) for various mean response functions and the following patterns can be observed.

1. Steinization will reduce risk in all cases since D is always negative.

2. As the bandwidth α decreases, the value of A increases. We have $\lambda = \sum_{s=1}^{p} \frac{\left[\mu_s^{(k)}(x)+b_s^{(k)}(x)\right]^2}{2||l^{(k)}(x)||^2}$. Note that as α decreases, the numerator decreases because the bias $b_j^{(k)}$ goes to 0. On the other hand, the denominator gets larger since $var\left(\widehat{\mu}_j^{(k)}\right) = ||l^{(k)}(x)||^2\sigma_j^2$ increases as α decreases. Therefore, λ becomes smaller when α is smaller. Since $A = \int_{-1}^{1} \mathrm{E}\frac{1}{p-2+2K(x)}dx$ and $\mathrm{E}\frac{1}{p-2+2K(x)} = \sum_{k=0}^{\infty} \frac{1}{p-2+2k(x)} \cdot \frac{\lambda^k e^{-\lambda}}{k!}$, a smaller λ will result in a larger A.

3. As α decreases, B goes close to 0 since the bias $b_j^{(k)}$ goes to 0 and $B := \int_{\mathcal{X}} \mathrm{E}\sum_{j=1}^{p} b_j^{(k)}(x)\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}dx$. No patterns were observed about the sign of B. Also note that the optimal point $-\frac{D}{2A}$ becomes closer to 1 as α gets smaller, which agrees with $c = p - 2 = 1$ in James and Stein(1961) when there is no bias. Since $-\frac{D}{2A} = -\frac{-2(p-2)A-2B}{2A} = 1 + \frac{B}{A}$, as B gets close to 0, $-\frac{D}{2A}$ goes to 1.

4. Define naive Steinization $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{(p-2)||l^{(k)}(x)||^2}{\sum_{s=1}^{p}\widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$ as if there is no bias. It will reduce risk when $-\frac{D}{2A} > \frac{1}{2}$.

5. Define integrated mean squared error

$$MSE(\boldsymbol{Z}^{(k)}(x)) = \int_{\mathcal{X}} \mathrm{E}(\mathbf{Z}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T(||l^{(k)}(x)||^2\boldsymbol{\Sigma})^{-1}(\mathbf{Z}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\boldsymbol{Z}^{(k)}(x)$ is an estimator of $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$. We observe that integrated MSE of local regression estimator converges to 6 as α decreases, which can be explained by following

$$\begin{aligned}
& \mathrm{E}(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T (||l^{(k)}(x)||^2 \boldsymbol{\Sigma})^{-1} (\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)) \\
= \; & ||l^{(k)}(x)||^{-2} \sum_{j=1}^{p} \mathrm{E}\left[\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2 \\
= \; & ||l^{(k)}(x)||^{-2} \sum_{j=1}^{p} \left[\text{bias } \widehat{\mu}_j^{(k)}(x)\right]^2 + \text{var } \widehat{\mu}_j^{(k)}(x) \\
= \; & ||l^{(k)}(x)||^{-2} \sum_{j=1}^{p} \left\{\left[\text{bias } \widehat{\mu}_j^{(k)}(x)\right]^2 + ||l^{(k)}(x)||^2\right\} \\
= \; & ||l^{(k)}(x)||^{-2} \sum_{j=1}^{p} \left[\text{bias } \widehat{\mu}_j^{(k)}(x)\right]^2 + p.
\end{aligned}$$

As α gets smaller, the bias converges to 0. Therefore, $\mathrm{E}(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T$ $(||l^{(k)}(x)||^2 \boldsymbol{\Sigma})^{-1} (\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))$ becomes very close to p if α is small. In this example, since $p = 3$ and $\mathcal{X} = [-1, 1]$, the integrated MSE of local regression estimator goes to 6 as α decreases. Moreover, it is observed that Steinization reduces risk by up to 33%.

Example 2. Consider $p = 4$. Results are shown in Table (15) - Table (30) for various mean response functions. We can observe similar patterns as in Example 1, as well as the following

1. Steinization reduces risk except for a few cases when α is large.

2. The value of A can be 0 for mean response functions which take large values on $\mathcal{X} = [-1, 1]$. Since $\lambda = \sum_{s=1}^{p} \frac{\left[\mu_s^{(k)}(x) + b_s^{(k)}(x)\right]^2}{2||l^{(k)}(x)||^2}$, if the absolute value of $\mu_s^{(k)}(x)$ is large for some s or k, then λ becomes big and A will be very small even 0.

3. Since $p = 4$, the optimal point converges to $p - 2 = 2$, and the integrated MSE of local regression estimator goes to 8 as α gets smaller.

4. Steinization reduces risk by up to 50%.

Table 5: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05705	-0.01116	-0.09177	0.80431	6.19804	6.16332	6.16113
	0.10	0.09368	-0.00462	-0.17811	0.95067	6.03662	5.95219	5.95196
	0.05	0.20583	-0.00101	-0.40964	0.99511	6.00191	5.79810	5.79809
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.04917	0	-0.09835	1	6	5.95083	5.95083
	0.35	0.11163	0	-0.22325	1	6	5.88837	5.88837
	0.25	0.37004	0	-0.74009	1	6	5.62996	5.62996
	0.15	1.29302	0	-2.58603	1	6	4.70699	4.70699
	0.05	1.97350	0	-3.94700	1	6	4.02650	4.02650
$\widehat{\mu}_j^{(2)}(x)$	0.65	2	0	-4	1	6	4	4
	0.45	2	0	-4	1	6	4	4
	0.25	2	0	-4	1	6	4	4
	0.15	2	0	-4	1	6	4	4
	0.05	2	0	-4	1	6	4	4

Table 6: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.23641	-0.01389	-0.44504	0.94123	6.10333	5.89471	5.89389
	0.10	0.36531	-0.00612	-0.71839	0.98325	6.02200	5.66892	5.66882
	0.05	0.67268	-0.00137	-1.34262	0.99797	6.00132	5.33138	5.33138
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.74599	-0.21158	-1.06882	0.71638	7.84738	7.52455	7.46454
	0.35	1.09073	-0.11500	-1.95145	0.89456	6.51356	5.65283	5.64071
	0.25	1.49799	-0.04231	-2.91136	0.97175	6.08447	4.67111	4.66991
	0.15	1.85443	-0.00590	-3.69707	0.99682	6.00466	4.16202	4.16200
	0.05	1.99512	-5.72e-05	-3.99012	0.99997	6.00001	4.00500	4.00500
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.38093	-0.29993	-2.16202	0.78281	7.98982	7.20874	7.14360
	0.45	1.82245	-0.05364	-3.53762	0.97057	6.14485	4.42968	4.42810
	0.25	1.98619	-0.00171	-3.96897	0.99914	6.00166	4.01888	4.01888
	0.15	1.99889	-6.75e-05	-3.99764	0.99997	6.00003	4.00128	4.00128
	0.05	2.00000	-7.09e-07	-3.99999	1.00000	6	4.00000	4.00000

Table 7: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15254	-0.02556	-0.25395	0.83243	6.02814	5.92672	5.92244
	0.10	0.24497	-0.00933	-0.47130	0.96193	6.00321	5.77689	5.77654
	0.05	0.55204	-0.00151	-1.10106	0.99727	6.00015	5.45113	5.45112
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.30359	-0.08107	-0.44505	0.73298	12.46979	12.3283	12.3067
	0.35	0.49313	0.06086	-1.10799	1.12342	8.35948	7.74463	7.73712
	0.25	0.98375	0.04327	-2.05404	1.04398	6.46865	5.39837	5.39646
	0.15	1.68393	-0.00178	-3.36430	0.99894	6.02598	4.34561	4.34561
	0.05	1.98880	-7.28e-05	-3.97746	0.99996	6.00003	4.01137	4.01137
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.32017	1.42028	-3.48089	5.43600	66.26759	63.10687	56.80653
	0.45	1.40192	-0.22455	-2.35473	0.83982	9.81032	8.85751	8.82154
	0.25	1.88696	-0.01641	-3.7411	0.99130	6.01318	4.15904	4.15889
	0.15	1.98959	-0.00013	-3.97892	0.99994	6.00004	4.01071	4.01071
	0.05	1.99998	2.99e-06	-3.99996	1.00000	6	4.00002	4.00002

Table 8: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17848	-0.00949	-0.33798	0.94683	6.04931	5.88981	5.88930
	0.10	0.28643	-0.00404	-0.56479	0.98591	6.00964	5.73128	5.73122
	0.05	0.57865	-0.00087	-1.15556	0.99849	6.00053	5.42362	5.42362
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.60857	-0.08159	-1.05396	0.86594	6.22796	5.78256	5.77162
	0.35	1.02705	-0.03930	-1.97551	0.96174	6.05147	5.10301	5.10151
	0.25	1.51727	-0.01178	-3.01097	0.99224	6.00690	4.51319	4.51310
	0.15	1.87948	-0.00125	-3.75647	0.99934	6.00031	4.12332	4.12332
	0.05	1.99629	-5.90e-05	-3.99246	0.99997	6.00000	4.00383	4.00383
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.68256	-0.01025	-3.34461	0.99391	6.01737	4.35531	4.35525
	0.45	1.94667	-0.00107	-3.89121	0.99945	6.00108	4.05654	4.05654
	0.25	1.99733	-7.91e-05	-3.99450	0.99996	6.00001	4.00284	4.00284
	0.15	1.99981	1.61e-06	-3.99962	1.00000	6.00000	4.00019	4.00019
	0.05	2.00000	-9.09e-07	4.00000	1.00000	6	4.00000	4.00000

Table 9: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05480	-0.01044	-0.08872	0.80947	6.19250	6.15858	6.15659
	0.10	0.08999	-0.00433	-0.17133	0.95189	6.03583	5.95450	5.95429
	0.05	0.20076	-0.00095	-0.39962	0.99526	6.00188	5.80302	5.80302
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05381	-0.00653	-0.09457	0.87870	6.08346	6.04271	6.04192
	0.35	0.12253	-0.00433	-0.23640	0.96466	6.02008	5.90621	5.90606
	0.25	0.40679	-0.00224	-0.80910	0.99449	6.00287	5.60056	5.60055
	0.15	1.33356	-0.00040	-2.66632	0.99970	6.00014	4.66738	4.66738
	0.05	1.97541	-1.81e-05	-3.95078	0.99999	6.00000	4.02463	4.02463
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.91179	-0.01091	-3.80176	0.99429	6.01737	4.12741	4.12735
	0.45	1.98463	-0.00120	-3.96687	0.99940	6.00108	4.01884	4.01884
	0.25	1.99918	-7.28e-05	-3.99821	0.99996	6.00001	4.00098	4.00098
	0.15	1.99994	-6.81e-06	-3.99987	1.00000	6.00000	4.00007	4.00007
	0.05	2	-6.52e-07	-4.00000	1.00000	6	4.00000	4.00000

Table 10: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17130	-0.00698	-0.32865	0.95927	6.01324	5.85589	5.85560
	0.10	0.27984	-0.00289	-0.55390	0.98967	6.00245	5.72839	5.72836
	0.05	0.62278	-0.00057	-1.24441	0.99908	6.00013	5.37850	5.37850
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.87099	-0.00281	-1.73636	0.99678	6.03581	5.17044	5.17043
	0.35	1.33243	-0.00122	-2.66242	0.99909	6.00782	4.67783	4.67783
	0.25	1.72546	-0.00029	-3.45035	0.99983	6.00101	4.27612	4.27612
	0.15	1.93997	7.14e-06	-3.87996	1.00000	6.00004	4.06006	4.06006
	0.05	1.99822	-4.58e-06	-3.99643	1.00000	6.00000	4.00179	4.00179
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.93820	0.00020	-3.87679	1.00010	6.00885	4.07026	4.07026
	0.45	1.99051	-3.21e-05	-3.98095	0.99998	6.00046	4.01002	4.01002
	0.25	1.99953	7.90e-07	-3.99906	1	6.00000	4.00047	4.00047
	0.15	1.99996	-2.80e-06	-3.99992	1.00000	6	4.00004	4.00004
	0.05	2	-5.83e-08	-4	1	6	4	4

Table 11: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.22849	-0.01149	-0.43399	0.94971	6.17416	5.96866	5.96808
	0.10	0.28164	-0.00475	-0.55379	0.98314	6.03217	5.76002	5.75994
	0.05	0.40132	-0.00103	-0.80059	0.99744	6.00168	5.60241	5.60241
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05527	0.00168	-0.11391	1.03043	6.01585	5.95721	5.95716
	0.35	0.12637	0.00097	-0.25467	1.00765	6.00393	5.87563	5.87562
	0.25	0.42139	0.00044	-0.84366	1.00104	6.00057	5.57830	5.57830
	0.15	1.35295	5.08e-05	-2.70600	1.00004	6.00003	4.64698	4.64698
	0.05	1.97637	4.13e-06	-3.95275	1.00000	6	4.02362	4.02362
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.99049	-0.00356	-3.97386	0.99821	6.00595	4.02257	4.02257
	0.45	1.99787	-0.00030	-3.99514	0.99985	6.00026	4.00299	4.00299
	0.25	1.99988	-1.41e-07	-3.99975	1.00000	6.00000	4.00013	4.00013
	0.15	1.9999	-1.49e-06	-3.99998	1.00000	6	4.00001	4.00001
	0.05	2	-1.72e-07	-4	1.00000	6	4	4

Table 12: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.19192	-0.00977	-0.36429	0.94909	6.02848	5.85610	5.85560
	0.10	0.31422	-0.00419	-0.62005	0.98666	6.00555	5.69971	5.69966
	0.05	0.68331	-0.00086	-1.36491	0.99875	6.00030	5.31870	5.31870
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.92082	-0.08181	-1.67802	0.91115	6.18030	5.42310	5.41583
	0.35	1.32154	-0.03370	-2.57568	0.97450	6.03921	4.78507	4.78421
	0.25	1.69870	-0.00850	-3.38041	0.99500	6.00504	4.32333	4.32329
	0.15	1.93039	-0.00077	-3.85924	0.99960	6.00022	4.07136	4.07136
	0.05	1.99791	-2.58e-05	-3.99577	0.99999	6.00000	4.00214	4.00214
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.70492	7.01e-05	-3.40997	1.00004	6.00885	4.30379	4.30379
	0.45	1.95240	4.91e-05	-3.90489	1.00003	6.00046	4.04797	4.04797
	0.25	1.99763	1.67e-05	-3.99530	1.00001	6.00000	4.00234	4.00234
	0.15	1.99984	-8.00e-06	-3.99966	1.00000	6	4.00018	4.00018
	0.05	2.00000	-4.96e-07	-4.00000	1.00000	6	4.00000	4.00000

Table 13: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.93160	-0.01852	-1.82617	0.98013	6.08499	5.19043	5.19006
	0.10	1.07737	-0.00798	-2.13878	0.99259	6.01833	4.95692	4.95687
	0.05	1.34226	-0.00172	-2.68108	0.99872	6.00111	4.66230	4.66230
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.82768	-0.21905	-1.21726	0.73534	7.77976	7.39018	7.33221
	0.35	1.18799	-0.11702	-2.14193	0.90149	6.49741	5.54347	5.53194
	0.25	1.57607	-0.04134	-3.06945	0.97377	6.08217	4.58878	4.58770
	0.15	1.88221	-0.00559	-3.75324	0.99703	6.00455	4.13352	4.13351
	0.05	1.99609	-6.33e-05	-3.99206	0.99997	6.00001	4.00404	4.00404
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.43133	-0.30156	-2.25954	0.78931	7.97840	7.15019	7.08666
	0.45	1.83409	-0.05403	-3.56011	0.97054	6.14403	4.41801	4.41641
	0.25	1.98688	-0.00168	-3.97040	0.99915	6.00165	4.01813	4.01813
	0.15	1.99894	-5.58e-05	-3.99776	0.99997	6.00003	4.00120	4.00120
	0.05	2.00000	1.05e-05	-4.00002	1.00001	6	3.99998	3.99998

Table 14: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.16061	-0.00740	-0.30642	0.95394	6.02999	5.88418	5.88383
	0.10	0.25994	-0.00307	-0.51373	0.98819	6.00574	5.75195	5.75191
	0.05	0.54379	-0.00065	-1.08627	0.99880	6.00031	5.45783	5.45783
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.61426	-0.01745	-1.19362	0.97159	6.09931	5.51995	5.51946
	0.35	1.06723	-0.01119	-2.11209	0.98952	6.02401	4.97915	4.97903
	0.25	1.56319	-0.00387	-3.11864	0.99752	6.00344	4.44799	4.44798
	0.15	1.89612	-0.00047	-3.79130	0.99975	6.00017	4.10499	4.10499
	0.05	1.99684	-4.31e-05	-3.99360	0.99998	6.00000	4.00324	4.00324
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.90286	-0.01425	-3.77722	0.99251	6.02332	4.14896	4.14885
	0.45	1.98253	-0.00154	-3.96197	0.99922	6.00133	4.02190	4.02189
	0.25	1.99906	-5.23e-05	-3.99801	0.99997	6.00001	4.00106	4.00106
	0.15	1.99993	-5.49e-06	-3.99985	1.00000	6.00000	4.00008	4.00008
	0.05	2	9.91e-07	-4.00000	1	6	4.00000	4.00000

Table 15: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$, $\mu_4(x) = 20x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01272	-0.01155	-0.02777	1.09177	10.83857	10.82351	10.82341
	0.10	0.01985	-0.00476	-0.06987	1.76023	8.52492	8.47490	8.46343
	0.05	0.04036	-0.00104	-0.15935	1.97432	8.02736	7.90836	7.87005
$\widehat{\mu}_j^{(1)}(x)$	0.45	0	0	0	-	8	8	-
	0.35	1.09e-05	0	-4.36e-05	2	8	7.99996	7.99996
	0.25	0.02079	0	-0.08316	2	8	7.91684	7.91684
	0.15	0.10218	0	-0.40871	2	8	7.59129	7.59129
	0.05	0.87005	0	-3.48020	2	8	4.51980	4.51980
$\widehat{\mu}_j^{(2)}(x)$	0.65	1	0	-4	2	8	4	4
	0.45	1	0	-4	2	8	4	4
	0.25	1	0	-4	2	8	4	4
	0.15	1	0	-4	2	8	4	4
	0.05	1	0	-4	2	8	4	4

Table 16: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$, $\mu_4(x) = 2x + 100$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0	-7.29e-05	0.00015	-	8.22445	8.22459	-
	0.10	0	-9.81e-05	0.00020	-	8.04151	8.04170	-
	0.05	0	-4.87e-05	9.73e-05	-	8.00216	8.00226	-
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.04228	0	-0.16911	2	8	7.83089	7.83089
	0.35	0.09243	0	-0.36973	2	8	7.63027	7.63027
	0.25	0.25673	0	-1.02692	2	8	6.97308	6.97308
	0.15	0.69473	0	-2.77894	2	8	5.22106	5.22106
	0.05	0.98873	0	-3.95493	2	8	4.04507	4.04507
$\widehat{\mu}_j^{(2)}(x)$	0.65	1	0	-4	2	8	4	4
	0.45	1	0	-4	2	8	4	4
	0.25	1	0	-4	2	8	4	4
	0.15	1	0	-4	2	8	4	4
	0.05	1	0	-4	2	8	4	4

Table 17: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$, $\mu_4(x) = \exp(-x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.06159	-0.00949	-0.22737	1.84588	8.12420	7.95842	7.91434
	0.10	0.09403	-0.00417	-0.36778	1.95563	8.02609	7.75234	7.66647
	0.05	0.19433	-0.00097	-0.77540	1.99503	8.00155	7.42048	7.22807
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.33184	-0.16774	-0.99190	1.49453	9.93085	9.27442	9.18963
	0.35	0.51482	-0.09321	-1.87286	1.81895	8.53363	6.84718	6.83031
	0.25	0.73710	-0.03478	-2.87884	1.95281	8.08734	5.27807	5.27643
	0.15	0.92600	-0.00490	-3.69423	1.99471	8.00480	4.32036	4.32034
	0.05	0.99756	-9.97e-05	-3.99005	1.9999	8.00001	4.01016	4.01016
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.73502	-0.23724	-2.46560	1.67723	10.00719	8.17105	7.93950
	0.45	0.92731	-0.03942	-3.63040	1.95749	8.14592	4.61310	4.59268
	0.25	0.99451	-0.00139	-3.97526	1.99860	8.00167	4.02933	4.02919
	0.15	0.99956	-0.00024	-3.99776	1.99976	8.00003	4.00277	4.00275
	0.05	1.00000	0	-4	2	8	4	4

Table 18: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin 5x$, $\mu_4(x) = \cos 10x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.10972	-0.08352	-0.27182	1.23874	8.32433	8.16223	8.15597
	0.10	0.15897	-0.03187	-0.57212	1.79950	8.03376	7.62060	7.51899
	0.05	0.30777	-0.00522	-1.22061	1.98302	8.00142	7.08857	6.79117
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.15227	-1.06666	1.52425		95.50266		
	0.35	0.17871	-0.63082	0.54682		31.9365		
	0.25	0.27218	-0.14999	-0.78876	1.44895	12.82696	12.33817	12.25553
	0.15	0.61776	-0.01554	-2.43996	1.97485	8.39507	5.98618	5.98579
	0.05	0.98080	-0.00043	-3.92233	1.99956	8.00042	4.07894	4.07894
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.15577	-0.36898	0.11488		1002.147		
	0.45	0.19014	-0.24917	-0.26222	0.68954	131.7425	131.7355	131.6521
	0.25	0.71460	-0.22014	-2.41812	1.69194	9.79097	7.98321	7.74531
	0.15	0.94787	-0.00726	-3.77696	1.99234	8.00972	4.25682	4.24722
	0.05	0.99984	0	-3.99936	2	8	4.0006	4.00064

Table 19: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin(5x)$, $\mu_4(x) = \cos(2x) + 100$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0	0.00014	-0.00029	-	8.05312	8.05283	-
	0.10	0	6.93e-05	-0.00014	-	8.00768	8.00754	-
	0.05	0	8.45e-05	-0.00017	-	8.00037	8.00020	-
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.17095	-0.09256	-0.49868	1.45856	14.49459	14.18103	14.13091
	0.35	0.28273	0.02817	-1.18725	2.09964	10.36221	9.11862	9.11581
	0.25	0.52826	0.02500	-2.16302	2.04733	8.46913	6.25611	6.25492
	0.15	0.85555	-0.00132	-3.41956	1.99846	8.02603	4.60911	4.60910
	0.05	0.99500	-0.00020	-3.97962	1.9998	8.00003	4.02081	4.02081
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.20940	1.23828	-3.31416	7.91347	68.71267	63.02405	55.59942
	0.45	0.74241	-0.16954	-2.63056	1.77164	11.83789	9.69179	9.50769
	0.25	0.95632	-0.01186	-3.80156	1.98760	8.01342	4.24238	4.23543
	0.15	0.99602	-0.00015	-3.98378	1.99985	8.00005	4.01667	4.01657
	0.05	0.99999	-1.66E-06	-3.99996	2.00000	8	4.00004	4.00004

Table 20: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 20x$, $\mu_4(x) = \exp(x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.00860	-0.01137	-0.01166	0.67826	10.6944	10.69133	10.69044
	0.10	0.01379	-0.00469	-0.04578	1.66014	8.49850	8.46651	8.46050
	0.05	0.03043	-0.00111	-0.11950	1.96357	8.02599	7.93692	7.90867
$\widehat{\mu}_j^{(1)}(x)$	0.45	0	-0.00049	0.00097		8.08346		
	0.35	0.00020	-0.00034	-0.00012	0.29303	8.02008	8.02064	8.02006
	0.25	0.02198	-0.00020	-0.08753	1.99086	8.00287	7.91574	7.91574
	0.15	0.10802	-5.45e-05	-0.43196	1.99950	8.00014	7.56829	7.56829
	0.05	0.87641	-2.10e-05	-3.50558	1.99998	8.00000	4.49446	4.49446
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.96669	-0.00723	-3.8523	1.99252	8.01737	4.18395	4.17948
	0.45	0.99423	-0.00050	-3.97592	1.99950	8.00108	4.02636	4.02616
	0.25	0.99969	-4.17E-05	-3.99868	1.99996	8.00001	4.00139	4.00142
	0.15	0.99998	1.29E-05	-3.99995	2.00001	8.00000	4.00006	4.00003
	0.05	1.00000	0	-4	2	8	4	4

Table 21: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 20x$, $\mu_4(x) = \sin(5x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01279	-0.01250	-0.02615	1.02221	10.69504	10.68168	10.68167
	0.10	0.01995	-0.00517	-0.06947	1.74106	8.49639	8.44687	8.43592
	0.05	0.04062	-0.00113	-0.16023	1.97230	8.02585	7.90625	7.86785
$\widehat{\mu}_j^{(1)}(x)$	0.45	0	-0.00103	0.00206		14.43398		
	0.35	0.00013	0.00160	-0.00372	14.43885	10.35167	10.34475	10.32485
	0.25	0.02148	0.00072	-0.08738	2.03348	8.46765	8.37883	8.37881
	0.15	0.10552	-0.00055	-0.42098	1.99480	8.02594	7.60605	7.60605
	0.05	0.87371	-0.00014	-3.49456	1.99984	8.00003	4.50574	4.50574
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.23588	1.25942	-3.46236	7.33924	68.25874	62.38290	55.55319
	0.45	0.75630	-0.17933	-2.66654	1.76289	11.80986	9.64302	9.45946
	0.25	0.95742	-0.01294	-3.8038	1.98649	8.01317	4.24185	4.23508
	0.15	0.99610	-8.64E-05	-3.98423	1.99991	8.00004	4.01604	4.01599
	0.05	0.99999	-1.01E-06	-3.99996	2.00000	8	4.00008	4.00004

Table 22: $\mu_1(x) = x$, $\mu_2(x) = 20x$, $\mu_3(x) = x^4$, $\mu_4(x) = \exp(x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
	0.15	0.01306	-0.01134	-0.02957	1.13168	10.72862	10.71212	10.71189
$\widehat{\mu}_j(x)$	0.10	0.020417	-0.00469	-0.07228	1.77006	8.50720	8.45534	8.44323
	0.05	0.04157	-0.00118	-0.16393	1.97157	8.02659	7.90424	7.86499
	0.45	0	-0.00272	0.00545		9.70289		
	0.35	0.00025	-0.00238	0.00379		8.48217		
$\widehat{\mu}_j^{(1)}(x)$	0.25	0.02213	-0.00169	-0.08513	1.92370	8.08044	7.99869	7.99856
	0.15	0.10862	-0.00080	-0.43288	1.99265	8.00449	7.57321	7.57320
	0.05	0.87687	-8.01e-05	-3.50733	1.99991	8.00001	4.49284	4.49284
	0.65	0.82739	-0.25234	-2.80488	1.69502	9.98982	7.77538	7.61266
	0.45	0.94571	-0.04070	-3.70144	1.95696	8.14485	4.53751	4.52306
$\widehat{\mu}_j^{(2)}(x)$	0.25	0.99550	-0.00145	-3.9791	1.99854	8.00166	4.02554	4.02546
	0.15	0.99963	-0.00020	-3.99812	1.9998	8.00003	4.00231	4.00231
	0.05	1.00000	-3.03E-06	-3.99999	2.00000	8	4.00002	4.00001

4.1.3 Bias and variance unknown

In this section, we assume both bias $b_j^{(k)}(x)$ and variance σ_j^2 are unknown for all $k = 1, 2, \ldots, J$ and $j = 1, 2, \ldots, p$. Let $\widehat{\mu}_j^{(k)}(x)$ be the local regression estimator of $\mu_j^{(k)}(x)$. Assume tricube weight function with the same nearest neighbor parameter is applied on all p mean response functions in local regression estimation. Then we have $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^n l_i^{(k)}(x)Y_{ji}$. Recall that $\widehat{\sigma}_j^2 = \frac{1}{m}\sum_{i:|x_i-x|>2h}[Y_{ji} - \widehat{\mu}_j(x_i)]^2$, where $m = \sum_{i:|x_i-x|>2h}[1 - 2l_i(x_i) + ||l(x_i)||^2]$, is a consistent estimator of σ_j^2 which is independent of $\widehat{\mu}_j^{(k)}(x)$. Define Stein estimator $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$, where the positive constant c is chosen such that $\int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx$. Put $A := \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \sum_{j=1}^p \frac{\mathrm{E}\widehat{\sigma}_j^4}{\sigma_j^2} \cdot \mathrm{E}\frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2\right]^2}dx$, $B := \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \sum_{j=1}^p \mathrm{E}\widehat{\sigma}_j^2 \mathrm{E}\frac{1}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}dx$, $D := \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \sum_{j=1}^p \mathrm{E}\widehat{\sigma}_j^2 \mathrm{E}\frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2\right]^2}dx$, $E := \int_{\mathcal{X}} b_j^{(k)}(x) \cdot \frac{\mathrm{E}\widehat{\sigma}_j^2}{\sigma_j^2} \cdot \mathrm{E}\frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}dx$, and $F := B + D + E$. The following theorem is obtained.

Theorem 4.3 *Assume bias $b_j^{(k)}(x)$ and variance σ_j^2 are unknown, for all $k = 1, 2, \ldots, J$ and $j = 1, 2, \ldots, p$. Define $\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x)$, where c is a constant satisfying $0 < c < -\frac{F}{A}$, and $||l^{(k)}(x)||^2 = \sum_{i=1}^n l_i^{(k)}(x)^2$. Then*

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$.

Table 23: $\mu_1(x) = x$, $\mu_2(x) = 20x$, $\mu_3(x) = \cos x$, $\mu_4(x) = \sin(5x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01282	-0.01271	-0.02585	1.00821	10.67274	10.65971	10.65971
	0.10	0.01999	-0.00525	-0.06944	1.73739	8.49231	8.44285	8.43198
	0.05	0.04072	-0.00114	-0.16061	1.97196	8.02564	7.90575	7.86728
$\widehat{\mu}_j^{(1)}(x)$	0.45	0	-0.00132	0.00264		14.45394		
	0.35	0.00016	0.00156	-0.00377	11.59558	10.35555	10.34867	10.33371
	0.25	0.02168	0.00071	-0.08814	2.03279	8.46809	8.37852	8.37850
	0.15	0.10648	-0.00055	-0.42480	1.99483	8.02595	7.60225	7.60225
	0.05	0.87473	-3.63e-05	-3.49885	1.99996	8.00003	4.50125	4.50125
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.23339	1.25084	-3.43524	7.35944	68.26165	62.42897	55.62093
	0.45	0.75438	-0.17329	-2.67094	1.77029	11.81006	9.62718	9.44589
	0.25	0.95730	-0.01248	-3.80424	1.98696	8.01318	4.24054	4.23374
	0.15	0.99609	-0.00014	-3.98408	1.99986	8.00004	4.01628	4.01624
	0.05	0.99999	-3.25E-06	-3.99995	2.00000	8	4.00004	4.00005

Table 24: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$, $\mu_4(x) = 5x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
	0.15	0.05116	-0.01107	-0.18250	1.78368	8.26837	8.13702	8.10560
$\widehat{\mu}_j(x)$	0.10	0.07923	-0.00464	-0.30763	1.94144	8.05252	7.82411	7.75389
	0.05	0.15416	-0.00103	-0.61457	1.99332	8.00291	7.54250	7.39039
	0.45	0.05130	-0.04434	-0.11653	1.13576	9.84738	9.81952	9.78121
	0.35	0.11066	-0.03264	-0.37736	1.70504	8.51356	8.20148	8.19185
$\widehat{\mu}_j^{(1)}(x)$	0.25	0.29618	-0.01793	-1.14884	1.93945	8.08447	6.97150	6.97041
	0.15	0.72370	-0.00383	-2.88714	1.99470	8.00466	5.12519	5.12517
	0.05	0.98989	-7.45e-05	-3.9594	1.99993	8.00001	4.04076	4.04076
	0.65	0.75728	-0.23802	-2.55308	1.68569	9.98982	8.04920	7.83797
	0.45	0.93246	-0.04096	-3.64792	1.95607	8.14485	4.59561	4.57705
$\widehat{\mu}_j^{(2)}(x)$	0.25	0.99482	-0.00090	-3.97748	1.99910	8.00166	4.02616	4.02598
	0.15	0.99958	-0.00018	-3.99796	1.99982	8.00003	4.00239	4.00243
	0.05	1.00000	0	-4	2	8	4	4

Table 25: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$, $\mu_4(x) = \sin(5x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
	0.15	0.12981	-0.02793	-0.46339	1.78483	8.12483	7.79126	7.71130
$\widehat{\mu}_j(x)$	0.10	0.18831	-0.01097	-0.73131	1.94174	8.02399	7.48099	7.31398
	0.05	0.33409	-0.00208	-1.33220	1.99378	8.00140	7.00330	6.67335
	0.45	0.18507	-0.16125	-0.41779	1.12871	16.28136	16.18608	16.04558
	0.35	0.29074	-0.03758	-1.08779	1.87073	10.86522	9.85260	9.84774
$\widehat{\mu}_j^{(1)}(x)$	0.25	0.51318	-0.01350	-2.02573	1.97370	8.55212	6.55337	6.55302
	0.15	0.83910	-0.00645	-3.34352	1.99232	8.03060	4.69997	4.69992
	0.05	0.99431	-7.74e-05	-3.97707	1.99992	8.00003	4.02312	4.02312
	0.65	0.21026	0.96361	-2.76826	6.58295	70.24857	65.63379	61.13692
	0.45	0.69759	-0.21213	-2.3661	1.69591	11.9547	10.17584	9.94835
$\widehat{\mu}_j^{(2)}(x)$	0.25	0.95256	-0.01365	-3.78294	1.98567	8.01483	4.26703	4.25899
	0.15	0.99569	-0.00014	-3.98248	1.99986	8.00007	4.01797	4.01787
	0.05	1.00000	-2.55E-06	-4.00000	2.00000	8	4.00006	4.00001

Table 26: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$, $\mu_4(x) = \cos(10x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.13518	-0.07713	-0.38645	1.42941	8.39952	8.14825	8.12333
	0.10	0.19005	-0.03034	-0.69955	1.84039	8.05254	7.54305	7.40882
	0.05	0.33339	-0.00532	-1.32291	1.98404	8.00259	7.01307	6.69024
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.25691	-1.79865	2.56966		90.88025		
	0.35	0.26415	-0.90399	0.75136		30.09058		
	0.25	0.35306	-0.21856	-0.97510	1.38095	12.44277	11.90479	11.76949
	0.15	0.65859	-0.02659	-2.58119	1.95962	8.37375	5.84575	5.84468
	0.05	0.98310	-0.00040	-3.93162	1.99960	8.00040	4.06958	4.06958
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.46299	-3.53395	5.21594		943.8696		
	0.45	0.25498	-0.37225	-0.27542	0.54008	128.077	128.6165	128.0026
	0.25	0.74152	-0.21504	-2.536	1.71000	9.77945	7.83189	7.61117
	0.15	0.95112	-0.00718	-3.79012	1.99245	8.0097	4.24262	4.23389
	0.05	0.99985	4.39E-05	-3.99949	2.00004	8	4.00042	4.00042

Table 27: $\mu_1(x) = x^4$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = \sin(5x)$, $\mu_4(x) = \cos x$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.08583	-0.02148	-0.30037	1.74974	8.10710	7.89256	7.84432
	0.10	0.12875	-0.00857	-0.49786	1.93343	8.02046	7.65135	7.53917
	0.05	0.25213	-0.00169	-1.00513	1.99331	8.00121	7.24820	6.99944
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.19724	-0.14955	-0.48985	1.24178	16.15683	15.96608	15.85269
	0.35	0.30383	-0.02560	-1.16413	1.91574	10.83772	9.72480	9.72264
	0.25	0.53032	-0.00847	-2.10432	1.98402	8.54852	6.46115	6.46102
	0.15	0.84842	-0.00621	-3.38127	1.99268	8.03044	4.66160	4.66156
	0.05	0.99466	-7.47e-05	-3.97849	1.99993	8.00003	4.02169	4.02169
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.21793	0.99234	-2.8564	6.55348	70.25147	65.49443	60.89179
	0.45	0.70532	-0.21703	-2.38722	1.69230	11.95491	10.16069	9.93497
	0.25	0.95309	-0.01393	-3.7845	1.98538	8.01483	4.26587	4.25799
	0.15	0.99573	-0.00028	-3.98236	1.99972	8.00007	4.01835	4.01827
	0.05	0.99999	1.09E-06	-3.99996	2.00000	8	4.00004	4.00004

Table 28: $\mu_1(x) = 5x$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin(5x)$, $\mu_4(x) = \cos(10x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.03899	-0.04049	-0.07496	0.96136	8.48684	8.45086	8.45080
	0.10	0.05948	-0.01718	-0.20356	1.71114	8.06385	7.91977	7.88969
	0.05	0.12009	-0.00349	-0.47336	1.97090	8.00299	7.64972	7.53652
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.03902	-0.40472	0.65336		95.48682		
	0.35	0.06818	-0.33664	0.40055		31.93258		
	0.25	0.15145	-0.10601	-0.39376	1.29999	12.82639	12.64466	12.57045
	0.15	0.49503	-0.01591	-1.94831	1.96787	8.39504	6.47854	6.47803
	0.05	0.97295	-0.00145	-3.88891	1.99851	8.00042	4.11442	4.11442
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.15595	-0.36466	0.10552		1002.141		
	0.45	0.19018	-0.24945	-0.26182	0.68835	131.7422	131.7356	131.6521
	0.25	0.71464	-0.21930	-2.41996	1.69313	9.79097	7.97965	7.74231
	0.15	0.94787	-0.00692	-3.77764	1.99270	8.00972	4.25546	4.24587
	0.05	0.99984	-7.73E-06	-3.99935	1.99999	8	4.00064	4.00067

Table 29: $\mu_1(x) = 5x$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin 5x$, $\mu_4(x) = \cos(2x) + 100$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0	0.00012	-0.00024	-	8.21563	8.21539	-
	0.10	0	5.78e-05	-0.00012	-	8.03777	8.03765	-
	0.05	0	7.29e-05	-0.00015	-	8.00194	8.00179	-
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.04090	-0.02417	-0.11526	1.40896	14.47874	14.41183	14.39755
	0.35	0.08556	0.01273	-0.36772	2.14880	10.35828	9.96510	9.96320
	0.25	0.23425	0.01011	-0.95721	2.04317	8.46856	7.49112	7.49068
	0.15	0.66547	-0.00142	-2.65903	1.99787	8.02600	5.36981	5.36981
	0.05	0.98700	-0.00012	-3.94775	1.99987	8.00003	4.05253	4.05253
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.20964	1.24217	-3.3229	7.92525	68.70673	63.00195	55.53932
	0.45	0.74309	-0.17639	-2.61958	1.76263	11.83763	9.71613	9.52896
	0.25	0.95636	-0.01211	-3.80122	1.98734	8.01342	4.24320	4.23627
	0.15	0.99602	-8.26E-05	-3.98392	1.99992	8.00005	4.01635	4.0163
	0.05	0.99999	2.52E-06	-3.99997	2.00000	8	4.00002	4.00003

Table 30: $\mu_1(x) = \exp(x)$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin(5x)$, $\mu_4(x) = \cos(10x)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.07477	-0.06041	-0.17826	1.19208	8.34267	8.23918	8.23642
	0.10	0.10929	-0.02351	-0.39014	1.78485	8.03742	7.75657	7.68925
	0.05	0.21687	-0.00407	-0.85932	1.98122	8.00162	7.35917	7.15037
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.14414	-1.04867	1.52078		95.57028		
	0.35	0.17285	-0.62441	0.55745		31.95265		
	0.25	0.26708	-0.15352	-0.76127	1.42519	12.82926	12.37502	12.28678
	0.15	0.61139	-0.01672	-2.41213	1.97265	8.39518	6.01649	6.01603
	0.05	0.98044	-0.00138	-3.91899	1.99859	8.00042	4.08419	4.08419
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.15421	-0.37784	0.13884		1002.159		
	0.45	0.18977	-0.25346	-0.25216	0.66438	131.7433	131.7459	131.6595
	0.25	0.71442	-0.22015	-2.41738	1.69185	9.79098	7.98402	7.74606
	0.15	0.94785	-0.00707	-3.77726	1.99254	8.00972	4.25616	4.24655
	0.05	0.99984	4.54E-05	-3.99945	2.00005	8	4.00044	4.00046

Proof of Theorem 4.3. From the proof of Theorem 3.4, we have

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
= & \sum_{j=1}^{p} E \frac{\left[\widetilde{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l^{(k)}(x)||^2 \sigma_j^2} - \sum_{j=1}^{p} E \frac{\left[\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right]^2}{||l^{(k)}(x)||^2 \sigma_j^2} \\
= & \sum_{j=1}^{p} \frac{E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)^2 + 2E\left(\widetilde{\mu}_j^{(k)}(x) - \widehat{\mu}_j^{(k)}(x)\right)\left(\widehat{\mu}_j^{(k)}(x) - \mu_j^{(k)}(x)\right)}{||l^{(k)}(x)||^2 \sigma_j^2} \qquad (93) \\
= & \sum_{j=1}^{p} \frac{c^2 ||l^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} - \sum_{j=1}^{p} 2c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} \\
& + \sum_{j=1}^{p} 4c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} - \sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx - \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx \\
= & \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{c^2||l^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& - \int_{\mathcal{X}} \sum_{j=1}^{p} 2c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
& + \int_{\mathcal{X}} \sum_{j=1}^{p} 4c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& - \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
= & Ac^2 + Bc + Dc + Ec \\
= & Ac^2 + Fc. \qquad (94)
\end{aligned}
$$

Since A and c are positive, if F in (94) is negative, James-Stein estimators with $0 < c < -\frac{F}{A}$ will improve local regression estimators in reducing risk. Moreover, (94) achieves its minimum at $c = -\frac{F}{2A}$.

In the following two examples, different mean response functions will be closely examined and numerical calculations of A, B, D, E and F are carried out. Data sets of size 100 are generated from model (90) with equispaced x_i on $\mathcal{X} = [-1, 1]$. Monte Carlo algorithm is applied in the calculations of A, B, D, E and F. In estimating $\mu_j(x)$, $\mu_j^{(1)}(x)$ and $\mu_j^{(2)}(x)$, J is set to be 0, 1 and 2 respectively. Let α be the nearest neighbor parameter in local regression, consider $\alpha \in \{0.15, 0.10, 0.05\}$ for $J = 0$, $\alpha \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$ for $J = 1$ and $\alpha \in \{0.45, 0.25, 0.15, 0.05\}$ for $J = 2$.

Example 3. Consider $p = 3$. Simulation results are shown in Table (31) - Table (40) for various mean response functions. And we can observe the following:

1. Steinization reduced risk in all ten examples of mean response function.

2. The optimal point $-\frac{F}{2A}$ does not converge to 1 as in Example 1. It actually gets close

to 0.96 as α becomes small. If there is no bias, we have

$$
\begin{aligned}
& \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{c^2||l^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& - \int_{\mathcal{X}} \sum_{j=1}^{p} 2c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
& + \int_{\mathcal{X}} \sum_{j=1}^{p} 4c||l^{(k)}(x)||^2 E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& - \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{2cb_j^{(k)}(x)}{\sigma_j^2} E\widehat{\sigma}_j^2 E \frac{\widehat{\mu}_j^{(k)}(x)}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
= & \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{c^2||l^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& - \int_{\mathcal{X}} \sum_{j=1}^{p} 2c||l^{(k)}(x)||^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
& + \int_{\mathcal{X}} \sum_{j=1}^{p} 4c||l^{(k)}(x)||^2 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
= & \int_{\mathcal{X}} \sum_{j=1}^{p} \frac{c^2||l^{(k)}(x)||^2}{\sigma_j^2} E\widehat{\sigma}_j^4 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& -2(p-2)c \int_{\mathcal{X}} ||l^{(k)}(x)||^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \\
\approx & \ 1.04c^2 \int_{\mathcal{X}} \sum_{j=1}^{p} ||l^{(k)}(x)||^2 E \frac{\widehat{\mu}_j^{(k)}(x)^2}{\left[\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2\right]^2} dx \\
& -2(p-2)c \int_{\mathcal{X}} ||l^{(k)}(x)||^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx \qquad (95) \\
= & \left[1.04c^2 - 2(p-2)c\right] \int_{\mathcal{X}} ||l^{(k)}(x)||^2 E \frac{1}{\sum_{s=1}^{p} \widehat{\mu}_s^{(k)}(x)^2} dx,
\end{aligned}
$$

which is minimized at $c = -\frac{p-2}{1.04} \approx 0.96$. The number 1.04 in line (95) comes from numerical calculations. When α is small, bias is close to zero and therefore the optimal point is at $c \approx 0.96$.

3. Steinization reduces risk by up to 33%.

Table 31: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05962	0.17201	0.05738	-0.01125	-0.09201	0.77158	6.19804	6.16566	6.16255
	0.10	0.09684	0.28140	0.09380	-0.00472	-0.17816	0.91984	6.03662	5.95531	5.95468
	0.05	0.21459	0.61760	0.20587	-0.00094	-0.40983	0.95493	6.00191	5.80666	5.80623
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.13432	0.18729	0.07640	0	-0.06899	0.25681	6	6.06533	5.99114
	0.35	0.15820	0.36184	0.12897	0	-0.20781	0.65681	6	5.95039	5.93175
	0.25	0.41210	1.13566	0.38229	0	-0.74215	0.90045	6	5.66995	5.66587
	0.15	1.34589	3.89554	1.29876	0	-2.59605	0.96444	6	4.74984	4.74814
	0.05	2.05705	5.92158	1.97378	0	-3.94804	0.95964	6	4.10901	4.10566
$\widehat{\mu}_j^{(2)}(x)$	0.45	3.65807	7.58906	2.49811	0	-5.18566	0.70880	6	4.47241	4.16221
	0.25	2.17196	6.13541	2.04358	0	-4.09650	0.94304	6	4.07545	4.06841
	0.15	2.07586	6.0254	2.00974	0	-4.01184	0.96631	6	4.06402	4.06166
	0.05	2.08471	6.0011	2.00044	0	-4.00046	0.95948	6	4.08425	4.08082

Table 32: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.24484	0.71095	0.23680	-0.01405	-0.44659	0.91201	6.10333	5.90158	5.89969
	0.10	0.37681	1.09677	0.36551	-0.00550	-0.72050	0.95605	6.02200	5.67831	5.67758
	0.05	0.70096	2.01830	0.67271	-0.00150	-1.34276	0.95780	6.00132	5.35952	5.35827
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.97219	2.41649	0.80379	-0.22795	-1.16190	0.59757	7.84738	7.65766	7.50022
	0.35	1.21904	3.3663	1.12165	-0.11055	-2.02490	0.83053	6.51356	5.70770	5.67269
	0.25	1.58996	4.53805	1.51161	-0.04250	-2.94463	0.92601	6.08447	4.72980	4.72109
	0.15	1.92235	5.57583	1.85828	-0.00616	-3.70624	0.96399	6.00466	4.22077	4.21828
	0.05	2.07900	5.98611	1.99518	5.55e-06	-3.99150	0.95996	6.00001	4.08751	4.08417
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.26408	5.83936	1.94155	-0.05859	-3.79533	0.83816	6.14485	4.61360	4.55430
	0.25	2.10515	6.01263	2.00332	-0.00174	-4.00852	0.95207	6.00166	4.09829	4.09346
	0.15	2.06832	6.00991	2.00265	-0.00020	-4.00883	0.96910	6.00003	4.05952	4.05754
	0.05	2.08418	6.00075	2.00006	-7.79e-05	-4.00111	0.95987	6	4.08307	4.07972

Table 33: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = \sin(5x)$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15761	0.45785	0.15260	-0.02563	-0.25404	0.80592	6.02814	5.93171	5.92577
	0.10	0.25245	0.73501	0.24499	-0.00933	-0.47140	0.93366	6.00321	5.78426	5.78315
	0.05	0.57522	1.65613	0.55204	-0.00148	-1.10114	0.95715	6.00015	5.47423	5.47317
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.45155	1.00728	0.34968	-0.07758	-0.46070	0.51013	12.46979	12.46064	12.35228
	0.35	0.57172	1.52789	0.51004	0.05767	-1.13099	0.98910	8.35948	7.80022	7.80015
	0.25	1.03901	2.97112	0.99149	0.04508	-2.06644	0.99443	6.46865	5.44122	5.44119
	0.15	1.73948	5.05446	1.68411	-0.00317	-3.36615	0.96758	6.02598	4.39931	4.39748
	0.05	2.07236	5.96648	1.98882	4.03e-05	-3.97777	0.95972	6.00003	4.09461	4.09125
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.97804	4.68588	1.56373	-0.28732	-2.54222	0.64261	9.81032	9.24614	8.99349
	0.25	1.98651	5.69654	1.90243	-0.01684	-3.74968	0.94379	6.01318	4.25001	4.24373
	0.15	2.05472	5.97192	1.99003	-0.00218	-3.97938	0.96835	6.00004	4.07539	4.07333
	0.05	2.08400	6.00000	1.99999	-2.43e-05	-3.99998	0.95969	6	4.08402	4.08063

Table 34: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.18473	0.53607	0.17870	-0.01010	-0.33715	0.91255	6.04931	5.89689	5.89547
	0.10	0.29534	0.85959	0.28652	-0.00386	-0.56537	0.95716	6.00964	5.73960	5.73906
	0.05	0.60302	1.73605	0.57866	-0.00097	-1.15551	0.95811	6.00053	5.44804	5.44698
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.75209	1.92766	0.63880	-0.08475	-1.13061	0.75164	6.22796	5.84944	5.80305
	0.35	1.12269	3.13614	1.04446	-0.02925	-2.03592	0.90672	6.05147	5.13823	5.12846
	0.25	1.59377	4.57606	1.52442	-0.01252	-3.02943	0.95039	6.00690	4.57125	4.56732
	0.15	1.94320	5.64445	1.88157	-0.00113	-3.76036	0.96757	6.00031	4.18314	4.18110
	0.05	2.08023	5.98917	1.99636	-9.17e-06	-3.99289	0.95973	6.00000	4.08734	4.08396
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.36963	6.12080	2.04285	-0.00149	-4.06723	0.85820	6.00108	4.30348	4.25584
	0.25	2.09713	6.02219	2.00278	8.65e-05	-4.03342	0.96165	6.00001	4.06372	4.06064
	0.15	2.06857	6.00575	2.00091	8.98e-05	-4.00803	0.96879	6.00000	4.06055	4.05853
	0.05	2.08407	6.00030	2.00009	4.13e-06	-4.00025	0.95972	6	4.08381	4.08043

Table 35: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05727	0.16520	0.05503	-0.01063	-0.08903	0.77730	6.19250	6.16074	6.15790
	0.10	0.09291	0.27033	0.09010	-0.00440	-0.17148	0.92277	6.03583	5.95727	5.95672
	0.05	0.20927	0.60240	0.20079	-0.00092	-0.39978	0.95520	6.00188	5.81136	5.81094
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.15393	0.20270	0.08629	-0.00653	-0.04720	0.15332	6.08346	6.19020	6.07985
	0.35	0.17635	0.39596	0.14275	-0.00401	-0.21290	0.60361	6.02008	5.98353	5.95582
	0.25	0.45129	1.24758	0.42247	-0.00224	-0.80080	0.88724	6.00287	5.65336	5.64762
	0.15	1.38838	4.01728	1.33868	-0.00044	-2.67895	0.96478	6.00014	4.70957	4.70784
	0.05	2.05904	5.92729	1.97573	8.35e-07	-3.95165	0.95959	6.00000	4.10739	4.10403
$\widehat{\mu}_j^{(2)}(x)$	0.45	3.60768	7.44312	2.46928	-0.00102	-5.00709	0.69395	6.00108	4.60166	4.26374
	0.25	2.16884	6.12760	2.04150	-6.93e-05	-4.08904	0.94268	6.00001	4.07981	4.07268
	0.15	2.08054	6.02451	2.00846	9.29e-05	-4.01539	0.96499	6.00000	4.06515	4.06260
	0.05	2.08412	6.00108	2.00033	1.31e-06	-4.00083	0.95984	6	4.08329	4.07993

Table 36: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17690	0.51407	0.17134	-0.00694	-0.32888	0.92958	6.01324	5.86125	5.86038
	0.10	0.28836	0.83960	0.27986	-0.00293	-0.55390	0.96043	6.00245	5.73691	5.73645
	0.05	0.64894	1.86836	0.62278	-0.00043	-1.24474	0.95906	6.00013	5.40433	5.40324
$\widehat{\mu}_j^{(1)}(x)$	0.45	1.00347	2.65940	0.88576	-0.00311	-1.76954	0.88171	6.03581	5.26973	5.25569
	0.35	1.42247	4.01843	1.33826	-0.00142	-2.68098	0.94236	6.00782	4.74931	4.74459
	0.25	1.79996	5.18420	1.72796	-0.00046	-3.45564	0.95992	6.00101	4.34533	4.34244
	0.15	2.00338	5.82156	1.94037	-0.00011	-3.88143	0.96872	6.00004	4.12199	4.12003
	0.05	2.08220	5.99473	1.99822	-6.67e-05	-3.99643	0.95967	6.00000	4.08576	4.08238
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.29130	6.07627	2.02512	-0.00041	-4.05125	0.88405	6.00046	4.24051	4.20971
	0.25	2.08594	6.00766	2.02512	-0.00020	-3.91445	0.93829	6.00000	4.17150	4.16356
	0.15	2.06524	6.00160	2.00046	-7.47e-05	-4.00122	0.96871	6	4.06402	4.06200
	0.05	2.08403	6.00007	2.00001	4.07e-06	-4.00012	0.95971	6	4.08392	4.08053

Table 37: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.23809	0.68851	0.22920	-0.01137	-0.43745	0.91865	6.17416	5.97480	5.97326
	0.10	0.29098	0.84591	0.28196	-0.00470	-0.55458	0.95296	6.03217	5.76857	5.76792
	0.05	0.41840	1.20417	0.40140	-0.00105	-0.80063	0.95678	6.00168	5.61944	5.61866
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.15854	0.20534	0.08875	0.00203	-0.05975	0.18844	6.01585	6.11463	6.01022
	0.35	0.18212	0.40601	0.14742	0.00096	-0.22426	0.61567	6.00393	5.96180	5.93490
	0.25	0.47084	1.28971	0.43636	0.00057	-0.83513	0.88685	6.00057	5.63628	5.63025
	0.15	1.40620	4.07401	1.35468	-8.34e-05	-2.72912	0.97039	6.00003	4.67711	4.67587
	0.05	2.05947	5.93006	1.97658	-7.17e-05	-3.95367	0.95987	6	4.10581	4.10249
$\widehat{\mu}_j^{(2)}(x)$	0.45	3.53573	7.39884	2.45276	0.00024	-4.98714	0.70525	6.00026	4.54885	4.24167
	0.25	2.1601	6.11897	2.03903	0.00028	-4.08238	0.94495	6.00000	4.07772	4.07118
	0.15	2.07862	6.02231	2.00755	-7.87e-05	-4.01428	0.96561	6	4.06435	4.06189
	0.05	2.08452	6.00097	2.00008	6.33e-07	-4.00161	0.95984	6	4.08291	4.07956

Table 38: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.19824	0.57612	0.19198	-0.00961	-0.36510	0.92085	6.02848	5.86162	5.86038
	0.10	0.32384	0.94283	0.31424	-0.00421	-0.62028	0.95770	6.00555	5.70911	5.70853
	0.05	0.71202	2.04999	0.68333	-0.00094	-1.3648	0.95840	6.00030	5.34752	5.34629
$\widehat{\mu}_j^{(1)}(x)$	0.45	1.05545	2.84434	0.94997	-0.08533	-1.71814	0.81393	6.18030	5.51762	5.48108
	0.35	1.42602	4.00317	1.33537	-0.03469	-2.59549	0.91005	6.03921	4.86974	4.85820
	0.25	1.77758	5.11139	1.70319	-0.00886	-3.39229	0.95419	6.00504	4.39032	4.38659
	0.15	1.99474	5.79471	1.93159	-0.00023	-3.86257	0.96819	6.00022	4.13238	4.13036
	0.05	2.08183	5.99390	1.99793	3.76e-05	-3.99616	0.95978	6.00000	4.08567	4.08230
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.28526	6.01814	2.00550	0.00011	-4.01450	0.87835	6.00046	4.27122	4.23740
	0.25	2.09009	6.01055	2.00217	-0.00016	-4.01211	0.95979	6.00000	4.07800	4.07461
	0.15	2.06669	6.00316	2.00102	-8.66e-05	-4.00206	0.96823	6	4.06462	4.06254
	0.05	2.08402	6.00017	2.00009	8.45e-07	-3.99998	0.95968	6	4.08405	4.08066

Table 39: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.96548	2.80069	0.93306	-0.01845	-1.83222	0.94887	6.08499	5.21825	5.21573
	0.10	1.11052	3.23420	1.07779	-0.00833	-2.14059	0.96378	6.01833	4.98826	4.98680
	0.05	1.39875	4.02722	1.34236	-0.00188	-2.68126	0.95845	6.00111	4.71861	4.71620
$\widehat{\mu}_j^{(1)}(x)$	0.45	1.01739	2.62530	0.87105	-0.23727	-1.29184	0.63488	7.77976	7.50531	7.36968
	0.35	1.31011	3.64075	1.21218	-0.12176	-2.18927	0.83553	6.49741	5.61825	5.58281
	0.25	1.66293	4.76457	1.58717	-0.04109	-3.09828	0.93157	6.08217	4.64682	4.63903
	0.15	1.94964	5.65704	1.88552	-0.00523	-3.76153	0.96467	6.00455	4.19266	4.19023
	0.05	2.08005	5.98891	1.99611	-9.83e-05	-3.99319	0.95988	6.00001	4.08687	4.08352
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.26862	5.78986	1.93715	-0.05747	-3.7162	0.81905	6.14403	4.69644	4.62216
	0.25	2.09094	6.00401	2.00007	-0.00174	-4.00426	0.95753	6.00165	4.08833	4.08456
	0.15	2.07071	6.00770	2.00214	9.29e-06	-4.00689	0.96752	6.00003	4.06385	4.06166
	0.05	2.08424	6.00063	2.00002	4.40e-05	-4.00129	0.95989	6	4.08295	4.07960

Table 40: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$

	α	A	B	D	E	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.16608	0.48217	0.16072	-0.00745	-0.30656	0.92293	6.02999	5.88951	5.88852
	0.10	0.26797	0.77996	0.25999	-0.00281	-0.51431	0.95965	6.00574	5.75940	5.75896
	0.05	0.56666	1.63141	0.54380	-0.00059	-1.08644	0.95865	6.00031	5.48052	5.47955
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.73325	1.91517	0.63489	-0.01672	-1.25734	0.85737	6.09931	5.57523	5.56031
	0.35	1.15506	3.23907	1.07967	-0.01106	-2.13737	0.92522	6.02401	5.04170	5.03524
	0.25	1.63576	4.70526	1.56805	-0.00382	-3.13068	0.95695	6.00344	4.50851	4.50548
	0.15	1.95802	5.69204	1.89708	-0.00053	-3.79469	0.96901	6.00017	4.16350	4.16162
	0.05	2.08078	5.99071	1.9969	-1.66e-05	-3.99378	0.95968	6.00000	4.08700	4.08362
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.35239	6.15091	2.05197	-0.00112	-4.09168	0.86969	6.00133	4.26204	4.22209
	0.25	2.09279	6.01642	2.00545	9.99e-06	-4.01107	0.95830	6.00001	4.08174	4.07810
	0.15	2.06578	6.00364	2.00101	-8.99e-05	-4.00308	0.96890	6.00000	4.06270	4.06070
	0.05	2.08402	6.00018	2.00004	-1.38e-05	-4.00016	0.95972	6	4.08386	4.08048

4.2 Error terms dependent

In the last section, we looked at theoretical calculations and simulation results assuming error terms in (90) are independent. Next we will relax the independence assumption and let the variance-covariance matrix $\mathbf{\Sigma}$ be any arbitrary, symmetric and positive definite matrix. Define the bias vector $\mathbf{b}^{(k)}(x) = E(\widehat{\boldsymbol{\mu}}^{(k)}(x)) - \boldsymbol{\mu}^{(k)}(x)$, where $\boldsymbol{\mu}^{(k)}(x) = (\mu_1^{(k)}(x), \mu_2^{(k)}(x), \ldots, \mu_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$ with $\widehat{\mu}_j^{(k)}(x) := \sum_{i=1}^{n} l_{j,i}^{(k)}(x) Y_{ji}$ the linear estimator of $\mu_j^{(k)}(x) = \frac{d^k}{dx^k}\mu_j(x)$. Three different scenarios will be considered: both $\mathbf{b}^{(k)}(x)$ and $\mathbf{\Sigma}$ matrix known, $\mathbf{b}^{(k)}(x)$ unknown $\mathbf{\Sigma}$ known, and both $\mathbf{b}^{(k)}(x)$ and $\mathbf{\Sigma}$ unknown.

4.2.1 Bias and variance matrix known

First we consider the simplest situation when both $\mathbf{b}^{(k)}(x)$ and $\mathbf{\Sigma}$ matrix are known. The following theorem is obtained. Recall that the loss function is $L(\mathbf{Z}^{(k)}, \boldsymbol{\mu}^{(k)}(x)) = (\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))^T \mathbf{\Sigma}_0{}^{-1}(\mathbf{Z}^{(k)} - \boldsymbol{\mu}^{(k)}(x))$, where $\mathbf{Z}^{(k)} = \left(Z_1^{(k)}, Z_1^{(k)}, \ldots, Z_p^{(k)}\right)^T$.

Theorem 4.4 *Assume bias $\mathbf{b}^{(k)}(x)$ and the variance-covariance matrix of the error terms $\mathbf{\Sigma}$ are known for all $k = 0, 1, 2, \ldots, J$. Let $\mathbf{\Sigma}_0 := var(\widehat{\boldsymbol{\mu}}^{(k)}(x))$, where $(\mathbf{\Sigma}_0)_{ij} = \sum_{s=1}^{n} l_{i,s}^{(k)}(x) l_{j,s}^{(k)}(x) \, (\mathbf{\Sigma})_{ij}$. Define*

$$\widetilde{\boldsymbol{\mu}}^{(k)}(x) := \left[1 - \frac{c}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)^T \mathbf{\Sigma}_0{}^{-1} \left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)}\right] \left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right),$$

where c is a positive constant satisfying $0 < c < 2(p-2)$. Then

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$, and $\mathbf{b}^{(k)}(x) = (b_1^{(k)}(x), b_2^{(k)}(x), \ldots, b_p^{(k)}(x))^T$.

Proof of Theorem 4.4. Similar to the proof of Theorem 3.5, we have

$$
\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = E(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T \boldsymbol{\Sigma}_0{}^{-1}(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)) \\
= \; & E\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right) \\
& -2cE\frac{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)} \\
& +c^2E\frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)^T \boldsymbol{\Sigma}_0{}^{-1}\left(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x)\right)} \\
= \; & \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) + \left[c^2 - 2c(p-2)\right]\mathrm{E}\frac{1}{\sum_{s=1}^{p}\left(V_s - \widetilde{b}_s\right)^2},
\end{aligned}
$$

where $\mathbf{V} := (V_1, V_2, \ldots, V_p)^T = \mathbf{A}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)$, $\widetilde{\mathbf{b}} := (\widetilde{b}_1, \widetilde{b}_2, \ldots, \widetilde{b}_p)^T = \mathbf{A}^{-1}\mathbf{b}$ with $\boldsymbol{\Sigma}_0 = \mathbf{AA}$.

Therefore

$$
\begin{aligned}
& \int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx - \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx \\
= \; & \left[c^2 - 2c(p-2)\right]\int_{\mathcal{X}} \mathrm{E}\frac{1}{\sum_{s=1}^{p}\left(V_s - \widetilde{b}_s\right)^2}dx.
\end{aligned}
$$

Since the integral in the equality above is positive, we need $\left[c^2 - 2c(p-2)\right]$, i.e. $0 < c < 2(p-2)$, such that $\int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x) - \mathbf{b}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx$.

4.2.2 Bias unknown variance matrix known

In this section, theoretical calculations and simulation study will be performed assuming unknown bias $\mathbf{b}^{(k)}(x)$ and known variance matrix $\boldsymbol{\Sigma}$. Let $A = \int \mathrm{E}\frac{1}{p-2+2K}dx$ where K has a Poisson distribution with parameter $\lambda = \frac{1}{2}\left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right]^T \boldsymbol{\Sigma}_0{}^{-1}\left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right]$, $B = \int \mathrm{E}\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x)dx$, and $D = -2(p-2)A - 2B$. And we have the following theorem

Theorem 4.5 *Assume bias $\mathbf{b}^{(k)}(x)$ is unknown and variance-covariance matrix $\boldsymbol{\Sigma}$ is known, for $k = 0, 1, 2, \ldots, J$. Define $\widetilde{\boldsymbol{\mu}}^{(k)}(x) := \left(1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\right)\widehat{\boldsymbol{\mu}}^{(k)}(x)$, where c is a con-*

stant satisfying $0 < c < -\frac{D}{A}$ *if* $D < 0$, *or* $-\frac{D}{A} < c < 0$ *if* $D > 0$. *Then*

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$.

Proof of Theorem 4.5. From the proof of Theorem 3.5, we obtain the following

$$\begin{aligned}
& \int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx - \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx \\
= \; & c^2 \int_{\mathcal{X}} \mathrm{E}\frac{1}{p-2+2K}dx - 2c(p-2)\int_{\mathcal{X}} \mathrm{E}\frac{1}{p-2+2K}dx \\
& -2c\int_{\mathcal{X}} \mathrm{E}\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}_0{}^{-1}\mathbf{b}^{(k)}(x)dx \\
= \; & Ac^2 - 2Ac(p-2) - 2Bc \\
= \; & Ac^2 + Dc, \qquad (96)
\end{aligned}$$

where K follows a Poisson distribution with parameter $\lambda = \frac{1}{2}\left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right]^T \boldsymbol{\Sigma}_0{}^{-1}\left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right]$.

Note that if the bias is very small, the quantity B will be close to zero and D is close to $-2(p-2)$. Therefore the optimum point in (96) is approximately at $c = p-2$, which agrees with the result in James and Stein (1961).

In the following example, we consider $p = 3$. Quantities A, B and D are calculated based on different mean response functions. Data sets of size 100 are generated from model (90) with equispaced x_i on $\mathcal{X} = [-1, 1]$. Simpson's rule and Monte Carlo algorithm are applied in the calculations of A and B, respectively. In estimating $\mu_j(x)$, $\mu_j^{(1)}(x)$ and $\mu_j^{(2)}(x)$, J is set to be 0, 1 and 2 respectively. Let α be the nearest neighbor parameter in local regression, consider $\alpha \in \{0.15, 0.10, 0.05\}$ for $J = 0$, $\alpha \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$ for $J = 1$ and $\alpha \in \{0.65, 0.45, 0.25, 0.15, 0.05\}$ for $J = 2$.

Example. Let $\widehat{\mu}_j^{(k)}(x) = \sum_{i=1}^n l_i^{(k)}(x) Y_{ji}$ be the local regression estimator of $\mu_j^{(k)}$, for $j = 1, 2, 3$ and $k = 0, 1, 2$. Consider $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$, where $\rho \in \{0.2, 0.4, 0.8\}$. Simulation results are shown in Table (41) - Table (70). And we can observe the following patterns.

1. In this example, Steinization always does a better job in reducing risk than local regression.

2. As the bandwidth α decreases, the value of A increases. Since the same weight functions, i.e., $l_i^{(k)}(x)$, are applied for all $j = 1, 2, \ldots, p$, we have $\mathbf{\Sigma}_0 = ||l^{(k)}(x)||^2 \mathbf{\Sigma}$. Therefore, the Poisson mean λ can be written as $\lambda = \frac{||l^{(k)}(x)||^{-2}}{2} \left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x) \right]^T \mathbf{\Sigma}^{-1} \left[\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x) \right]$. On the other hand, $\mathrm{E}\widehat{\boldsymbol{\mu}}^{(k)}(x) = \boldsymbol{\mu}^{(k)}(x) + \mathbf{b}^{(k)}(x)$. As α decreases, $\mathbf{b}^{(k)}(x)$ gets closer to 0. Also note that $||l^{(k)}(x)||^2$ increases as α decreases because $var\left(\widehat{\mu}_j^{(k)}\right) = ||l^{(k)}(x)||^2 \sigma_j^2$, where σ_j^2 is the (j, j)-th element of $\mathbf{\Sigma}$, becomes larger with a smaller α. Hence, λ is small when α is small. Since $A = \mathrm{E}\frac{1}{p-2+2K}$ where K has a Poisson distribution with parameter λ, we conclude that A increases as α decreases.

3. As α decreases, B goes close to 0 since the bias $\mathbf{b}^{(k)}(x)$ goes to 0.

4. Define naive Steinization $\widetilde{\mu}_j^{(k)}(x) := \left(1 - \frac{p-2}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \mathbf{\Sigma}_0{}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)}\right) \widehat{\boldsymbol{\mu}}^{(k)}(x)$ as if there is no bias. It will reduce risk when $-\frac{D}{2A} > \frac{1}{2}$.

5. At fixed α, J, and with the same mean response functions, as the correlation ρ approches 1, James-Stein estimators do not perform as well as the estimators with ρ closer to 0, in reducing mean squared errors. An intuitive explanation would be, consider the extreme case where $\rho = 1$, the mean response functions are perfectly linearly related. Therefore, any response function can be written as a linear combination of the rest, and we are only estimating one mean response function, in which case $p = 1$. And we already know that steinization works when $p \geq 3$. So Steinization does not work as well when ρ is closer to 1.

Table 41: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05449	-0.01097	-0.08704	0.79869	6.19804	6.16550	6.16329
	0.10	0.08918	-0.00456	-0.16925	0.94892	6.03662	5.95655	5.95632
	0.05	0.19999	-0.00100	-0.39797	0.99501	6.00191	5.80393	5.80393
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05683	0	-0.11365	1	6	5.94317	5.94317
	0.35	0.13007	0	-0.26014	1	6	5.86993	5.86993
	0.25	0.43356	0	-0.86715	1	6	5.56642	5.56642
	0.15	1.36537	0	-2.73074	1	6	4.63463	4.63463
	0.05	1.97694	0	-3.95389	1	6	4.02306	4.02306
$\widehat{\mu}_j^{(2)}(x)$	0.65	2	0	-4	1	6	4	4
	0.45	2	0	-4	1	6	4	4
	0.25	2	0	-4	1	6	4	4
	0.15	2	0	-4	1	6	4	4
	0.05	2	0	-4	1	6	4	4

Table 42: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ $(\rho = 0.2)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.26106	-0.01509	-0.49193	0.94219	6.10333	5.87234	5.87148
	0.10	0.39444	-0.00633	-0.77622	0.98396	6.02200	5.64073	5.64062
	0.05	0.70183	-0.00138	-1.40089	0.99803	6.00132	5.30228	5.30227
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.75320	-0.19385	-1.11869	0.74262	7.84738	7.48189	7.43200
	0.35	1.12325	-0.10623	-2.03402	0.90542	6.51356	5.60278	5.59273
	0.25	1.53944	-0.03800	-3.00287	0.97531	6.08447	4.62104	4.62010
	0.15	1.87280	-0.00529	-3.73502	0.99718	6.00466	4.14244	4.14243
	0.05	1.99581	-3.83e-05	-3.99154	0.99998	6.00001	4.00427	4.00427
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.48419	0.28834	-2.39171	0.80573	7.98982	7.08231	7.02629
	0.45	1.84825	-0.05208	-3.59234	0.97182	6.14485	4.40076	4.39929
	0.25	1.98769	-0.00101	-3.97337	0.99949	6.00166	4.01598	4.01598
	0.15	1.99899	0.00011	-3.99820	1.00005	6.00003	4.00082	4.00082
	0.05	2	0	-4	1.00000	6	4	4

Table 43: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ $(\rho = 0.2)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.14592	-0.02602	-0.23979	0.82167	6.02814	5.93434	5.92968
	0.10	0.23369	-0.00844	-0.45051	0.96388	6.00321	5.78846	5.78808
	0.05	0.52398	-0.00158	-1.04480	0.99698	6.00015	5.47924	5.47923
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.28794	-0.08480	-0.40627	0.70549	12.46979	12.35145	12.32648
	0.35	0.46664	0.06010	-1.05349	1.12880	8.35948	7.77264	7.76490
	0.25	0.93495	0.04650	-1.96288	1.04973	6.46865	5.44072	5.43840
	0.15	1.65949	-0.00128	-3.31640	0.99923	6.02598	4.36906	4.36906
	0.05	1.98782	0.00012	-3.97589	1.00006	6.00003	4.01196	4.01196
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.30588	1.45005	-3.51185	5.74063	66.26759	63.06162	56.18750
	0.45	1.38247	-0.24426	-2.27640	0.823313	9.81032	8.91638	8.87322
	0.25	1.87908	-0.01864	-3.72089	0.99008	6.01318	4.17137	4.17118
	0.15	1.98882	-0.00036	-3.97691	0.99982	6.00004	4.01195	4.01195
	0.05	1.99997	1.60e-05	-3.99998	1.00001	6	3.99999	3.99999

Table 44: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15566	-0.00965	-0.29202	0.93804	6.04931	5.91291	5.91232
	0.10	0.25195	-0.00411	-0.49570	0.98371	6.00964	5.76576	5.76569
	0.05	0.53305	-0.00088	-1.06434	0.99835	6.00053	5.46927	5.46927
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.62853	-0.08502	-1.08703	0.86474	6.22796	5.76946	5.75796
	0.35	1.05813	-0.03966	-2.03695	0.96252	6.05147	5.07265	5.07117
	0.25	1.54639	-0.01146	-3.06986	0.99259	6.00690	4.48343	4.48335
	0.15	1.88967	-0.00120	-3.77694	0.99936	6.00031	4.11305	4.11305
	0.05	1.99663	-1.12e-05	-3.99324	0.99999	6.00000	4.00339	4.00339
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.70357	-0.00945	-3.38824	0.99445	6.01737	4.33271	4.33265
	0.45	1.95055	-0.00086	-3.89937	0.99956	6.00108	4.05225	4.05225
	0.25	1.99753	-5.50e-05	-3.99494	0.99997	6.00001	4.00259	4.00259
	0.15	1.99982	4.23e-05	-3.99973	1.00002	6.00000	4.00009	4.00009
	0.05	2.00000	5.86e-07	-4	1	6	4.00000	4.00000

Table 45: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ $(\rho = 0.2)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05275	-0.01023	-0.08504	0.80612	6.19250	6.16035	6.15834
	0.10	0.08651	-0.00426	-0.16449	0.95070	6.03583	5.95784	5.95763
	0.05	0.19500	-0.00094	-0.38813	0.99518	6.00188	5.80875	5.80875
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05861	-0.00528	-0.10665	0.90986	6.08346	6.03542	6.03494
	0.35	0.13418	-0.00335	-0.26165	0.97503	6.02008	5.89260	5.89252
	0.25	0.44620	-0.00168	-0.88905	0.99625	6.00287	5.56002	5.56001
	0.15	1.37675	-0.00028	-2.75294	0.99980	6.00014	4.62395	4.62395
	0.05	1.97744	-0.00010	-3.95468	0.99995	6.00000	4.02276	4.02276
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.90578	-0.01159	-3.78838	0.99392	6.01737	4.13477	4.13470
	0.45	1.98355	-0.00147	-3.96415	0.99926	6.00108	4.02048	4.02048
	0.25	1.99912	6.71e-05	-3.99838	1.00003	6.00001	4.00076	4.00076
	0.15	1.99994	9.77e-05	-4.00007	1.00005	6.00000	3.99987	3.99987
	0.05	2	-2.02e-06	-4.00000	1.00000	6	4.00000	4.00000

Table 46: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17803	-0.00678	-0.34248	0.96189	6.01324	5.84787	5.84764
	0.10	0.29071	-0.00288	-0.57566	0.99008	6.00245	5.71721	5.71718
	0.05	0.63785	-0.00053	-1.27464	0.99917	6.00013	5.36341	5.36341
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.93667	-0.00612	-1.86111	0.99347	6.03581	5.11138	5.11134
	0.35	1.37931	-0.00320	-2.75222	0.99768	6.00782	4.63491	4.63491
	0.25	1.74632	-0.00093	-3.49076	0.99946	6.00101	4.25656	4.25656
	0.15	1.94464	-2.21e-06	-3.88928	1.00000	6.00004	4.05540	4.05540
	0.05	1.99836	2.18e-05	-3.99676	1.00001	6.00000	4.00160	4.00160
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.93389	0.00014	-3.86806	1.00007	6.00885	4.07468	4.07468
	0.45	1.98983	-0.00022	-3.97923	0.99989	6.00046	4.01106	4.01106
	0.25	1.99949	2.14e-05	-3.99903	1.00001	6.00000	4.00047	4.00047
	0.15	1.99996	6.42e-05	-4.00005	1.00003	6	3.99991	3.99991
	0.05	2	5.09e-07	-4.00000	1	6	4.00000	4.00000

Table 47: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.23731	-0.01147	-0.45167	0.95165	6.17416	5.95995	5.95938
	0.10	0.29212	-0.00477	-0.57471	0.98367	6.03217	5.74958	5.74951
	0.05	0.41581	-0.00104	-0.82955	0.99750	6.00168	5.58794	5.58794
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05929	-0.00052	-0.11753	0.99116	6.01585	5.95760	5.95760
	0.35	0.13606	-0.00031	-0.27150	0.99771	6.00393	5.86849	5.86849
	0.25	0.45334	-0.00060	-0.90547	0.99867	6.00057	5.54844	5.54844
	0.15	1.38512	-0.00019	-2.76986	0.99987	6.00003	4.61528	4.61528
	0.05	1.97784	2.49e-05	-3.95573	1.00001	6	4.02211	4.02211
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.98981	-0.00395	-3.97171	0.99801	6.00595	4.02404	4.02404
	0.45	1.99772	-0.00018	-3.99507	0.99991	6.00026	4.00290	4.00290
	0.25	1.99987	1.08e-05	-3.99975	1.00001	6.00000	4.00011	4.00011
	0.15	1.99999	3.66e-05	-4.00005	1.00002	6	3.99994	3.99994
	0.05	2	8.54e-07	-4.00000	1	6	4.00000	4.00000

Table 48: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ $(\rho = 0.2)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.19730	-0.01069	-0.37321	0.94582	6.02848	5.85256	5.85198
	0.10	0.32218	-0.00468	-0.63500	0.98547	6.00555	5.69273	5.69266
	0.05	0.69227	-0.00084	-1.38285	0.99878	6.00030	5.30972	5.30972
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.84549	-0.08883	-1.51333	0.89494	6.18030	5.51246	5.50313
	0.35	1.25120	-0.03881	-2.42479	0.96898	6.03921	4.86562	4.86442
	0.25	1.65746	-0.00981	-3.29530	0.99408	6.00504	4.36720	4.36714
	0.15	1.91939	-0.00088	-3.83703	0.99954	6.00022	4.08258	4.08258
	0.05	1.99757	1.55e-05	-3.99517	1.00001	6.00000	4.00240	4.00240
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.65223	3.80e-05	-3.30454	1.00002	6.00885	4.35654	4.35654
	0.45	1.94323	0.00016	-3.88679	1.00009	6.00046	4.05690	4.05690
	0.25	1.99725	1.85e-06	-3.99450	1.00000	6.00000	4.00275	4.00275
	0.15	1.99981	5.42e-07	-3.99961	1	6	4.00019	4.00019
	0.05	2.00000	9.81e-07	-4.00000	1	6	4.00000	4.00000

Table 49: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.92779	-0.01837	-1.81884	0.98020	6.08499	5.19394	5.19358
	0.10	1.07665	-0.00772	-2.13785	0.99283	6.01833	4.95713	4.95708
	0.05	1.34639	-0.00163	-2.68951	0.99879	6.00111	4.65799	4.65799
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.84799	-0.21041	-1.27516	0.75187	7.77976	7.35259	7.30038
	0.35	1.21107	-0.11227	-2.19758	0.90729	6.49741	5.51089	5.50048
	0.25	1.59457	-0.03946	-3.11022	0.97525	6.08217	4.56652	4.56554
	0.15	1.88891	-0.00503	-3.76776	0.99734	6.00455	4.12570	4.12568
	0.05	1.99633	-5.82e-05	-3.99254	0.99997	6.00001	4.00379	4.00379
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.46865	-0.29999	-2.33730	0.79573	7.97840	7.10974	7.04846
	0.45	1.84240	-0.05420	-3.5764	0.97058	6.14403	4.41003	4.40844
	0.25	1.98723	-0.00182	-3.97082	0.99908	6.00165	4.01806	4.01806
	0.15	1.99896	0.00011	-3.99813	1.00006	6.00003	4.00085	4.00085
	0.05	2.00000	-9.17e-06	-3.99998	1.00000	6	4.00002	4.00002

Table 50: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.2$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.14297	-0.00649	-0.27295	0.95460	6.02999	5.90000	5.89971
	0.10	0.23272	-0.00270	-0.46004	0.98840	6.00574	5.77842	5.77839
	0.05	0.51032	-0.00059	-1.01946	0.99884	6.00031	5.49118	5.49118
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.78472	-0.02668	-1.51608	0.96600	6.09931	5.36795	5.36704
	0.35	1.22648	-0.01466	-2.42364	0.98805	6.02401	4.82685	4.82667
	0.25	1.65339	-0.00460	-3.29759	0.99722	6.00344	4.35924	4.35922
	0.15	1.91940	-0.00028	-3.83824	0.99985	6.00017	4.08132	4.08132
	0.05	1.99757	-5.43e-05	-3.99502	0.99997	6.00000	4.00255	4.00255
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.89316	-0.01672	-3.75288	0.99117	6.02332	4.16360	4.16345
	0.45	1.98045	-0.00167	-3.95757	0.99916	6.00133	4.02422	4.02422
	0.25	1.99894	1.94e-06	-3.99787	1.00000	6.00001	4.00107	4.00107
	0.15	1.99992	2.58e-05	-3.99990	1.00001	6.00000	4.00003	4.00003
	0.05	2	-1.19e-06	-4.00000	1.00000	6	4.00000	4.00000

Table 51: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04652	-0.01096	-0.07113	0.76449	6.19804	6.17343	6.17085
	0.10	0.07597	-0.00451	-0.14294	0.94069	6.03662	5.96966	5.96939
	0.05	0.17376	-0.00098	-0.34555	0.99435	6.00191	5.83012	5.83011
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05632	0	-0.11265	1	6	5.94368	5.94368
	0.35	0.12885	0	-0.25770	1	6	5.87115	5.87115
	0.25	0.42947	0	-0.85894	1	6	5.57053	5.57053
	0.15	1.36109	0	-2.72218	1	6	4.63891	4.63891
	0.05	1.97675	0	-3.95349	1	6	4.02325	4.02325
$\widehat{\mu}_j^{(2)}(x)$	0.65	2	0	-4	1	6	4	4
	0.45	2	0	-4	1	6	4	4
	0.25	2	0	-4	1	6	4	4
	0.15	2	0	-4	1	6	4	4
	0.05	2	0	-4	1	6	4	4

Table 52: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ $(\rho = 0.4)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.25626	-0.01558	-0.48135	0.93920	6.10333	5.87824	5.87729
	0.10	0.38238	-0.00691	-0.75093	0.98192	6.02200	5.65344	5.65332
	0.05	0.67313	-0.00135	-1.34356	0.99799	6.00132	5.33089	5.33089
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.70619	-0.17988	-1.05261	0.74528	7.84738	7.50096	7.45514
	0.35	1.09749	-0.10167	-1.99164	0.90736	6.51356	5.61940	5.60999
	0.25	1.53422	-0.03787	-2.99271	0.97532	6.08447	4.62598	4.62505
	0.15	1.87319	-0.00545	-3.73548	0.99709	6.00466	4.14238	4.14236
	0.05	1.99583	-7.02e-05	-3.99152	0.99996	6.00001	4.00432	4.00432
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.52615	-0.30945	-2.43342	0.79724	7.98982	7.08256	7.01982
	0.45	1.85166	-0.05856	-3.58621	0.96838	6.14485	4.41030	4.40845
	0.25	1.98727	-0.00203	-3.97049	0.99898	6.00166	4.01844	4.01844
	0.15	1.99894	-5.36e-05	-3.99777	0.99997	6.00003	4.00120	4.00120
	0.05	2.00000	3.65e-05	-4.00007	1.00002	6	3.99993	3.99993

Table 53: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12990	-0.02638	-0.20702	0.79688	6.02814	5.95101	5.94565
	0.10	0.20686	-0.00851	-0.39671	0.95887	6.00321	5.81337	5.81302
	0.05	0.45929	-0.00164	-0.91530	0.99644	6.00015	5.54413	5.54413
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.25028	-0.09221	-0.31614	0.63156	12.46979	12.40393	12.36996
	0.35	0.40836	0.05855	-0.93383	1.14339	8.35948	7.83402	7.82562
	0.25	0.82927	0.05814	-1.77481	1.07011	6.46865	5.52311	5.51903
	0.15	1.59723	-0.00078	-3.19292	0.99951	6.02598	4.43030	4.43030
	0.05	1.98513	-9.45e-05	-3.97007	0.99995	6.00003	4.01509	4.01509
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.26491	1.51636	-3.56253	6.72403	66.26759	62.96997	54.29030
	0.45	1.32571	-0.28223	-2.08697	0.78711	9.81032	9.04906	8.98898
	0.25	1.85513	-0.02137	-3.66753	0.98848	6.01318	4.20078	4.20053
	0.15	1.98647	-0.00053	-3.97188	0.99974	6.00004	4.01463	4.01463
	0.05	1.99997	-1.68e-05	-3.99990	0.99999	6	4.00007	4.00007

Table 54: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12254	-0.00968	-0.22572	0.92099	6.04931	5.94613	5.94536
	0.10	0.19844	-0.00407	-0.38874	0.97949	6.00964	5.81933	5.81925
	0.05	0.44372	-0.00090	-0.88565	0.99798	6.00053	5.55860	5.55860
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.60335	-0.08846	-1.02979	0.85339	6.22796	5.80152	5.78855
	0.35	1.02871	-0.04218	-1.97307	0.95900	6.05147	5.10711	5.10538
	0.25	1.52646	-0.01214	-3.02864	0.99205	6.00690	4.50472	4.50463
	0.15	1.88402	-0.00127	-3.76552	0.99933	6.00031	4.11882	4.11882
	0.05	1.99645	-7.38e-05	-3.99276	0.99996	6.00000	4.00369	4.00369
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.68382	-0.00983	3.34798	0.99416	6.01737	4.35322	4.35316
	0.45	1.94694	-0.00108	-3.89172	0.99945	6.00108	4.05630	4.05630
	0.25	1.99734	-8.03e-05	-3.99452	0.99996	6.00001	4.00283	4.00283
	0.15	1.99981	-6.91e-05	-3.99948	0.99997	6.00000	4.00033	4.00033
	0.05	2.00000	-1.61e-07	-4.00000	1.00000	6	4.00000	4.00000

Table 55: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04549	-0.01008	-0.07081	0.77836	6.19250	6.16717	6.16494
	0.10	0.07454	-0.00420	-0.14068	0.94362	6.03583	5.96969	5.96946
	0.05	0.17076	-0.00092	-0.33969	0.99461	6.00188	5.83296	5.83295
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05488	-0.00348	-0.10280	0.93652	6.08346	6.03555	6.03533
	0.35	0.12517	-0.00253	-0.24528	0.97978	6.02008	5.89997	5.89992
	0.25	0.41627	-0.00155	-0.82943	0.99627	6.00287	5.58970	5.58970
	0.15	1.34579	-0.00011	-2.69135	0.99992	6.00014	4.65457	4.65457
	0.05	1.97601	2.90e-05	-3.95209	1.00002	6.00000	4.02393	4.02393
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.88708	-0.01452	-3.74510	0.99230	6.01737	4.15935	4.15923
	0.45	1.98013	-0.00153	-3.95721	0.99923	6.00108	4.02400	4.02400
	0.25	1.99894	-9.11e-05	-3.99769	0.99995	6.00001	4.00126	4.00126
	0.15	1.99992	8.04e-05	-4.00001	1.00004	6.00000	3.99992	3.99992
	0.05	2	3.79e-06	-4.00001	1.00000	6	3.99999	3.99999

Table 56: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ $(\rho = 0.4)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17449	-0.00651	-0.33597	0.96271	6.01324	5.85176	5.85152
	0.10	0.28294	-0.00275	-0.56039	0.99030	6.00245	5.72500	5.72497
	0.05	0.60599	-0.00046	-1.21106	0.99925	6.00013	5.39505	5.39505
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.92449	-0.00848	-1.83201	0.99083	6.03581	5.12829	5.12821
	0.35	1.35902	-0.00378	-2.71048	0.99722	6.00782	4.65636	4.65635
	0.25	1.73268	-0.00104	-3.46328	0.99940	6.00101	4.27041	4.27041
	0.15	1.94099	-6.09e-05	-3.88185	0.99997	6.00004	4.05918	4.05918
	0.05	1.99825	-2.87e-05	-3.99643	0.99999	6.00000	4.00181	4.00181
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.92042	-0.00173	-3.83739	0.99910	6.00885	4.09188	4.09188
	0.45	1.98771	0.00044	-3.97630	1.00022	6.00046	4.01187	4.01187
	0.25	1.99939	1.26e-05	-3.99880	1.00001	6.00000	4.00059	4.00059
	0.15	1.99995	-3.53e-05	-3.99984	0.99998	6	4.00012	4.00012
	0.05	2	3.77e-06	-4.00001	1.00000	6	3.99999	3.99999

Table 57: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.22888	-0.01158	-0.43461	0.94943	6.17416	5.96843	5.96785
	0.10	0.28211	-0.00476	-0.55469	0.98312	6.03217	5.75958	5.75950
	0.05	0.40182	-0.00106	-0.80154	0.99737	6.00168	5.60196	5.60196
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05460	-0.00281	-0.10358	0.94853	6.01585	5.96687	5.96672
	0.35	0.12455	-0.00172	-0.24566	0.98622	6.00393	5.88282	5.88280
	0.25	0.41438	-0.00100	-0.82675	0.99758	6.00057	5.58819	5.58819
	0.15	1.34413	-5.16e-05	-2.68815	0.99996	6.00003	4.65600	4.65600
	0.05	1.97595	-7.87e-05	-3.95174	0.99996	6	4.02421	4.02421
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.98768	0.00483	-3.96570	0.99757	6.00595	4.02793	4.02792
	0.45	1.99724	-6.82e-05	-3.99434	0.99997	6.00026	4.00316	4.00316
	0.25	1.99984	-9.50e-06	-3.99966	1.00000	6.00000	4.00018	4.00018
	0.15	1.99999	-2.82e-05	-3.99992	0.99999	6	4.00007	4.00007
	0.05	2	-2.59e-07	-4.00000	1.00000	6	4.00000	4.00000

Table 58: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ $(\rho = 0.4)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.18762	-0.01221	-0.35081	0.93490	6.02848	5.86529	5.86449
	0.10	0.30339	-0.00535	-0.59607	0.98237	6.00555	5.71286	5.71277
	0.05	0.64471	-0.00107	-1.28728	0.99834	6.00030	5.35773	5.35773
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.71843	-0.10107	-1.23472	0.85932	6.18030	5.66401	5.64979
	0.35	1.12670	-0.04409	-2.16522	0.96087	6.03921	5.00069	4.99896
	0.25	1.58012	-0.01234	-3.13557	0.99219	6.00504	4.44959	4.44949
	0.15	1.89794	-0.00116	-3.79355	0.99939	6.00022	4.10461	4.10460
	0.05	1.99689	-9.09e-05	-3.99360	0.99995	6.00000	4.00329	4.00329
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.56202	0.00336	-3.13074	1.00215	6.00885	4.44012	4.44012
	0.45	1.92659	0.00014	-3.85290	0.99993	6.00046	4.07416	4.07416
	0.25	1.99643	5.94e-05	-3.99298	1.00003	6.00000	4.00346	4.00346
	0.15	1.99975	2.38e-05	-3.99954	1.00001	6	4.00021	4.00021
	0.05	2.00000	-1.23e-06	-4.00000	1.00000	6	4.00000	4.00000

4.2.3 Bias and variance matrix unknown

At the end of Chapter 4, we consider the situation where both bias $\mathbf{b}^{(\mathbf{k})}(\mathbf{x})$ and variance-covariance matrix $\mathbf{\Sigma}$ are unknown. Proposition 3.4 showed that a consistent estimator of $\mathbf{\Sigma}$ is $\widehat{\mathbf{\Sigma}} = \frac{1}{m}\sum_{k:|x_k-x|>2h}\left(\widehat{\mathbf{\Sigma}}\right)_{ij}$, where the ij-th element of $\widehat{\mathbf{\Sigma}}$ is

$$\left(\widehat{\mathbf{\Sigma}}\right)_{ij} = \left[Y_{ik} - \widehat{\mu}_i(x_k)\right]\left[Y_{jk} - \widehat{\mu}_j(x_k)\right],$$

$m = \sum_{k:|x_k-x|>2h}[1-2l_k(x_k)+||l(x_k)||^2]$, assuming $\widehat{\mu}_j(x) - \mu_j(x) \xrightarrow{P} 0$ uniformly. Put $A := \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \mathrm{E}\frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right)^2}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\mathbf{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)dx$, $B := \int_{\mathcal{X}} \mathrm{E}\frac{||l^{(k)}(x)||^2}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}dx$, $D := \int_{\mathcal{X}} \mathrm{E}\frac{\mathbf{b}^{(k)}(x)^T\mathbf{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}dx$, and $F := -2(p-2)B - 2D$. We obtain the following result.

Theorem 4.6 *Assume that both bias $\mathbf{b}^{(k)}(x)$ and variance-covariance matrix $\mathbf{\Sigma}$ are unknown, for $k = 0, 1, 2, \ldots, J$. Let $\widehat{\mathbf{\Sigma}}$ be the consistent estimator of $\mathbf{\Sigma}$ defined in Proposition 3.4. Define $\widetilde{\boldsymbol{\mu}}^{(k)}(x) := \left(1 - \frac{c}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T||l^{(k)}(x)||^{-2}\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}\right)\widehat{\boldsymbol{\mu}}^{(k)}(x)$, where c is a constant satisfying $Ac^2 + Fc < 0$. Then*

$$\int_{\mathcal{X}} EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx < \int_{\mathcal{X}} EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx,$$

where $\widetilde{\boldsymbol{\mu}}^{(k)}(x) = (\widetilde{\mu}_1^{(k)}(x), \widetilde{\mu}_2^{(k)}(x), \ldots, \widetilde{\mu}_p^{(k)}(x))^T$, $\widehat{\boldsymbol{\mu}}^{(k)}(x) = (\widehat{\mu}_1^{(k)}(x), \widehat{\mu}_2^{(k)}(x), \ldots, \widehat{\mu}_p^{(k)}(x))^T$.

Proof of Theorem 4.6. As shown in Theorem 3.6, we have

$$\begin{aligned}
& \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
= \; & c^2||l^{(k)}(x)||^2\mathrm{E}\frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)\right)^2}\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\mathbf{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x) \\
& -2c(p-2)||l^{(k)}(x)||^2\mathrm{E}\frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)} - 2c\mathrm{E}\frac{\mathbf{b}^{(k)}(x)^T\mathbf{\Sigma}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T\widehat{\mathbf{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}^{(k)}(x)}.
\end{aligned}$$

Table 59: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.89328	-0.01780	-1.75096	0.98007	6.08499	5.22732	5.22696
	0.10	1.04383	-0.00786	-2.07193	0.99247	6.01833	4.99023	4.99017
	0.05	1.31637	-0.00157	-2.62959	0.99881	6.00111	4.68789	4.68788
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.81558	-0.21315	-1.20487	0.73865	7.77976	7.39048	7.33477
	0.35	1.17522	-0.11095	-2.12855	0.90559	6.49741	5.54408	5.53361
	0.25	1.56866	-0.04122	-3.05489	0.97372	6.08217	4.59595	4.59486
	0.15	1.88006	-0.00586	-3.74839	0.99688	6.00455	4.13621	4.13619
	0.05	1.99601	-6.01e-05	-3.99189	0.99997	6.00001	4.00412	4.00412
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.44334	-0.33109	-2.22450	0.77061	7.97840	7.19723	7.12128
	0.45	1.82876	-0.06245	-3.53263	0.96585	6.14403	4.44016	4.43803
	0.25	1.98568	-0.00174	-3.96789	0.99912	6.00165	4.01944	4.01944
	0.15	1.99882	-0.00012	-3.9974	0.99994	6.00003	4.00145	4.00145
	0.05	2.00000	-3.62e-05	-3.99992	0.99998	6	4.00008	4.00008

Table 60: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.4$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.11521	-0.00592	-0.21859	0.94865	6.02999	5.92661	5.92630
	0.10	0.18640	-0.00245	-0.36791	0.98688	6.00574	5.82423	5.82420
	0.05	0.42926	-0.00055	-0.85742	0.99873	6.00031	5.57215	5.57215
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.89773	-0.03789	-1.71969	0.95780	6.09931	5.27735	5.27575
	0.35	1.31526	-0.01932	-2.59188	0.98531	6.02401	4.74739	4.74710
	0.25	1.69690	-0.00580	-3.38219	0.99658	6.00344	4.31814	4.31813
	0.15	1.92963	-0.00049	-3.85829	0.99975	6.00017	4.07151	4.07151
	0.05	1.99787	3.42e-05	-3.99581	1.00002	6.00000	4.00206	4.00206
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.86965	-0.02114	-3.69702	0.98869	6.02332	4.19595	4.19571
	0.45	1.97568	-0.00218	-3.94699	0.99889	6.00133	4.03002	4.03002
	0.25	1.99867	-5.34e-05	-3.99722	0.99997	6.00001	4.00145	4.00145
	0.15	1.99990	-8.25e-05	-3.99964	0.99996	6.00000	4.00026	4.00026
	0.05	2	-4.42e-07	-4.00000	1.00000	6	4.00000	4.00000

Table 61: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01739	-0.01070	-0.01338	0.38486	6.19804	6.20204	6.19546
	0.10	0.02880	-0.00443	-0.04873	0.84606	6.03662	6.01669	6.01601
	0.05	0.06779	-0.00097	-0.13364	0.98571	6.00191	5.93606	5.93604
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.02830	0	-0.05661	1	6	5.97170	5.97170
	0.35	0.06307	0	-0.12614	1	6	5.93693	5.93693
	0.25	0.19618	0	-0.39236	1	6	5.80382	5.80382
	0.15	0.97356	0	-1.94712	1	6	5.02644	5.02644
	0.05	1.95475	0	-3.90951	1	6	4.04525	4.04525
$\widehat{\mu}_j^{(2)}(x)$	0.65	2	0	-4	1	6	4	4
	0.45	2	0	-4	1	6	4	4
	0.25	2	0	-4	1	6	4	4
	0.15	2	0	-4	1	6	4	4
	0.05	2	0	-4	1	6	4	4

Table 62: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ $(\rho = 0.8)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15313	-0.02283	-0.26060	0.85093	6.10333	5.99586	5.99246
	0.10	0.22998	-0.00967	-0.44063	0.95796	6.02200	5.81135	5.81095
	0.05	0.42464	-0.00192	-0.84545	0.99549	6.00132	5.58051	5.58050
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.37074	-0.13967	-0.46214	0.62327	7.84738	7.75598	7.70336
	0.35	0.73963	-0.10334	-1.27257	0.86028	6.51356	5.98061	5.96617
	0.25	1.27235	-0.04923	-2.44623	0.96131	6.08447	4.91059	4.90868
	0.15	1.77610	-0.00896	-3.53428	0.99496	6.00466	4.24648	4.24643
	0.05	1.99210	-9.36e-05	-3.98401	0.99995	6.00001	4.00810	4.00810
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.37757	-0.58801	-1.57912	0.57315	7.98982	7.78827	7.53728
	0.45	1.74789	-0.12503	-3.24573	0.92847	6.14485	4.64701	4.63807
	0.25	1.97396	-0.00455	-3.93883	0.99769	6.00166	4.03679	4.03678
	0.15	1.99774	-0.00019	-3.99511	0.99991	6.00003	4.00266	4.00266
	0.05	2.00000	-5.47e-05	-3.99988	0.99997	6	4.00012	4.00012

Table 63: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.06305	-0.02720	-0.07169	0.56853	6.02814	6.01950	6.00776
	0.10	0.09913	-0.01011	-0.17805	0.89806	6.00321	5.92430	5.92326
	0.05	0.22026	-0.00184	-0.43684	0.99165	6.00015	5.78357	5.78356
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.11063	-0.11542	0.00959	-0.04333	12.46979	12.59001	12.46958
	0.35	0.19841	0.04705	-0.49091	1.23712	8.35948	8.06698	8.05583
	0.25	0.42140	0.09412	-1.03103	1.22334	6.46865	5.85902	5.83800
	0.15	1.18231	0.00278	-2.37019	1.00235	6.02598	4.83810	4.83810
	0.05	1.96023	-0.00013	-3.92019	0.99993	6.00003	4.04006	4.04006
$\widehat{\mu}_j^{(2)}(x)$	0.65	0.12997	1.67021	-3.60035	13.85096	66.26759	62.79721	41.33346
	0.45	1.06088	-0.55743	-1.00690	0.47456	9.81032	9.86430	9.57140
	0.25	1.65076	-0.05125	-3.19902	0.96895	6.01318	4.46491	4.46332
	0.15	1.96416	-0.00121	-3.92590	0.99939	6.00004	4.03830	4.03830
	0.05	1.99991	3.76e-05	-3.99990	1.00002	6	4.00001	4.00001

Table 64: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04319	-0.00938	-0.06763	0.78292	6.04931	6.02486	6.02283
	0.10	0.06701	-0.00394	-0.12614	0.94120	6.00964	5.95051	5.95027
	0.05	0.15121	-0.00089	-0.30063	0.99409	6.00053	5.85111	5.85110
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.36201	-0.10748	-0.50907	0.70311	6.22796	6.08090	6.04899
	0.35	0.69549	-0.05657	-1.27783	0.91866	6.05147	5.46912	5.46452
	0.25	1.21954	-0.02005	-2.39899	0.98356	6.00690	4.82745	4.82713
	0.15	1.76870	-0.00248	-3.53245	0.99860	6.00031	4.23656	4.23656
	0.05	1.99243	-9.36e-05	-3.98467	0.99995	6.00000	4.00776	4.00776
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.37906	-0.01903	-2.72005	0.98620	6.01737	4.67638	4.67612
	0.45	1.88458	-0.00271	-3.76373	0.99856	6.00108	4.12192	4.12192
	0.25	1.9941	-9.87e-05	-3.98800	0.99995	6.00001	4.00611	4.00611
	0.15	1.99958	6.87e-06	-3.99917	1.00000	6.00000	4.00041	4.00041
	0.05	2.00000	4.69e-06	-4.00001	1.00000	6	3.99999	3.99999

Table 65: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ $(\rho = 0.8)$

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01748	-0.01001	-0.01493	0.42710	6.19250	6.19505	6.18931
	0.10	0.02885	-0.00413	-0.04943	0.85673	6.03583	6.01525	6.01466
	0.05	0.06837	-0.00091	-0.13492	0.98672	6.00188	5.93533	5.93532
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.02483	0.00146	-0.05258	1.05876	6.08346	6.05572	6.05563
	0.35	0.05522	0.00027	-0.11097	1.00483	6.02008	5.96433	5.96433
	0.25	0.16945	-0.00024	-0.33841	0.99858	6.00287	5.83391	5.83391
	0.15	0.88894	-0.00011	-1.77767	0.99988	6.00014	5.11141	5.11141
	0.05	1.94816	-2.37e-05	-3.89627	0.99999	6.00000	4.05189	4.05189
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.72377	-0.03178	-3.38399	0.98157	6.01737	4.35715	4.35657
	0.45	1.94792	-0.00404	-3.88777	0.99793	6.00108	4.06123	4.06122
	0.25	1.99717	-0.00014	-3.99405	0.99993	6.00001	4.00313	4.00313
	0.15	1.99979	8.10e-05	-3.99975	1.00004	6.00000	4.00005	4.00005
	0.05	2.00000	-1.16e-06	-4.00000	1.00000	6	4.00000	4.00000

Table 66: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12795	-0.01067	-0.23456	0.91658	6.01324	5.90663	5.90574
	0.10	0.20213	-0.00455	-0.39515	0.97747	6.00245	5.80943	5.80932
	0.05	0.40609	-0.00092	-0.81035	0.99774	6.00013	5.59587	5.59587
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.62927	-0.00714	-1.24427	0.98866	6.03581	5.42081	5.42073
	0.35	1.01945	-0.00772	-2.02347	0.99243	6.00782	5.00381	5.00375
	0.25	1.50038	-0.00265	-2.99546	0.99823	6.00101	4.50593	4.50593
	0.15	1.87475	-0.00023	-3.74905	0.99988	6.00004	4.12575	4.12575
	0.05	1.99614	-0.00010	-3.99208	0.99995	6.00000	4.00406	4.00406
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.79712	0.00077	-3.59578	1.00043	6.00885	4.21019	4.21019
	0.45	1.96742	0.00012	-3.93507	1.00006	6.00046	4.03281	4.03281
	0.25	1.99837	4.77e-05	-3.99683	1.00002	6.00000	4.00154	4.00154
	0.15	1.99989	7.14e-05	-3.99992	1.00004	6	3.99997	3.99997
	0.05	2	4.04e-06	-4.00001	1.00000	6	3.99999	3.99999

Table 67: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15364	-0.01158	-0.28412	0.92464	6.17416	6.04368	6.04280
	0.10	0.18972	-0.00478	-0.36988	0.97478	6.03217	5.85201	5.85189
	0.05	0.27452	-0.00104	-0.54696	0.99620	6.00168	5.72924	5.72923
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.02383	-0.00650	-0.03466	0.72717	6.01585	6.00502	6.00324
	0.35	0.05282	-0.00420	-0.09723	0.92041	6.00393	5.95952	5.95919
	0.25	0.16084	-0.00221	-0.31725	0.98625	6.00057	5.84415	5.84412
	0.15	0.86103	-0.00061	-1.72085	0.99930	6.00003	5.14021	5.14021
	0.05	1.94606	5.96e-05	-3.89224	1.00003	6	4.05382	4.05382
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.96738	-0.01224	-3.91028	0.99378	6.00595	4.06305	4.06297
	0.45	1.99264	-0.00111	-3.98306	0.99945	6.00026	4.00983	4.00983
	0.25	1.99957	-8.38e-05	-3.99897	0.99996	6.00000	4.00060	4.00060
	0.15	1.99997	-3.23e-05	-3.99987	0.99998	6	4.00010	4.00010
	0.05	2	2.05e-07	-4	1	6	4	4

Table 68: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12429	-0.01900	-0.21059	0.84715	6.02848	5.94218	5.93928
	0.10	0.19337	-0.00842	-0.36988	0.95643	6.00555	5.82903	5.82866
	0.05	0.38784	-0.00174	-0.77219	0.99550	6.00030	5.61595	5.61594
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.28011	-0.12505	-0.31011	0.55355	6.18030	6.15030	6.09447
	0.35	0.57461	-0.06580	-1.01763	0.88549	6.03921	5.59619	5.58865
	0.25	1.11093	-0.02416	-2.17354	0.97825	6.00504	4.94243	4.94190
	0.15	1.73022	-0.00289	-3.45467	0.99833	6.00022	4.27577	4.27576
	0.05	1.99112	-7.73e-05	-3.98209	0.99996	6.00000	4.00903	4.00903
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.01372	0.00987	-2.04719	1.00974	6.00885	4.97538	4.97529
	0.45	1.79499	0.00090	-3.59177	1.0005	6.00046	4.20368	4.20368
	0.25	1.98963	3.97e-05	-3.97934	1.00002	6.00000	4.01029	4.01029
	0.15	1.99927	-6.51e-05	-3.99840	0.99997	6	4.00087	4.00087
	0.05	2.00000	6.21e-06	-4.00001	1.00000	6	3.99999	3.99999

Table 69: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
	0.15	0.65430	-0.01761	-1.27339	0.97309	6.08499	5.46590	5.46543
$\widehat{\mu}_j(x)$	0.10	0.79775	-0.00808	-1.57935	0.98988	6.01833	5.23673	5.23665
	0.05	1.07595	-0.00179	-2.14832	0.99834	6.00111	4.92874	4.92874
	0.45	0.53462	-0.29378	-0.48168	0.45049	7.77976	7.83270	7.67127
	0.35	0.82912	-0.16744	-1.32338	0.79806	6.49741	6.00315	5.96934
$\widehat{\mu}_j^{(1)}(x)$	0.25	1.26714	-0.06822	-2.39784	0.94616	6.08217	4.95147	4.94780
	0.15	1.75332	-0.01152	-3.48358	0.99343	6.00455	4.27428	4.27421
	0.05	1.99096	-7.96e-05	-3.98176	0.99996	6.00001	4.00921	4.00921
	0.65	1.05790	-0.59219	-0.93142	0.44022	7.97840	8.10487	7.77338
	0.45	1.64766	-0.13580	-3.02373	0.91758	6.14403	4.76796	4.75677
$\widehat{\mu}_j^{(2)}(x)$	0.25	1.96620	-0.00566	-3.92108	0.99712	6.00165	4.04677	4.04676
	0.15	1.99714	-6.40e-05	-3.99416	0.99997	6.00003	4.00301	4.00301
	0.05	1.99999	-1.91e-05	-3.99995	0.99999	6	4.00005	4.00005

Table 70: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.8$)

	α	A	B	D	$-D/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04273	-0.00485	-0.07577	0.88661	6.02999	5.99695	5.99640
	0.10	0.06629	-0.00204	-0.12850	0.96926	6.00574	5.94353	5.94347
	0.05	0.14863	-0.00050	-0.29626	0.99662	6.00031	5.85268	5.85268
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.91888	-0.08398	-1.6698	0.90860	6.09931	5.34839	5.34072
	0.35	1.28976	-0.04209	-2.49535	0.96737	6.02401	4.81842	4.81704
	0.25	1.65398	-0.01368	-3.28060	0.99173	6.00344	4.37681	4.37670
	0.15	1.91182	-0.00151	-3.82061	0.99921	6.00017	4.09137	4.09137
	0.05	1.99724	-7.82e-05	-3.99432	0.99996	6.00000	4.00292	4.00292
$\widehat{\mu}_j^{(2)}(x)$	0.65	1.68259	-0.05351	-3.25815	0.96820	6.02332	4.44775	4.44605
	0.45	1.93424	-0.00633	-3.85583	0.99673	6.00133	4.07975	4.07973
	0.25	1.99628	-0.00014	-3.99227	0.99993	6.00001	4.00402	4.00402
	0.15	1.99973	-5.16e-05	-3.99935	0.99997	6.00000	4.00038	4.00038
	0.05	2.00000	1.90e-06	-4.00000	1.00000	6	4.00000	4.00000

Therefore, the difference of integrated MSE is

$$
\begin{aligned}
& \int_{\mathcal{X}} \mathrm{E}L(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx - \int_{\mathcal{X}} \mathrm{E}L(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x))dx \\
= \quad & c^2 \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \mathrm{E} \frac{1}{\left(\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)\right)^2} \widehat{\boldsymbol{\mu}}^{(k)}(x)^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x) dx \\
& -2c(p-2) \int_{\mathcal{X}} ||l^{(k)}(x)||^2 \mathrm{E} \frac{1}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} dx - 2c \int_{\mathcal{X}} \mathrm{E} \frac{\mathbf{b}^{(k)}(x)^T \boldsymbol{\Sigma}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)}{\widehat{\boldsymbol{\mu}}^{(k)}(x)^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{\mu}}^{(k)}(x)} dx \\
= \quad & Ac^2 + Fc < 0.
\end{aligned}
$$

Next we consider an example where $p = 3$. Quantities A, B and D are calculated based on different mean response functions. Data sets of size 100 are generated from model (90) with equispaced x_i on $\mathcal{X} = [-1, 1]$. Monte Carlo algorithm is used to calculate A, B and D. In estimating $\mu_j(x)$, $\mu_j^{(1)}(x)$ and $\mu_j^{(2)}(x)$, J is set to be 0, 1 and 2 respectively. Let α be the nearest neighbor parameter in local regression, consider $\alpha \in \{0.15, 0.10, 0.05\}$ for $J = 0$, $\alpha \in \{0.45, 0.35, 0.25, 0.15, 0.05\}$ for $J = 1$ and $\alpha \in \{0.65, 0.45, 0.25, 0.15, 0.05\}$ for $J = 2$.

Example. Let $\widehat{\mu}_j^{(k)}(x) = \sum_{i=1}^n l_i^{(k)}(x) Y_{ji}$ be the local regression estimator of $\mu_j^{(k)}$, for $j = 1, 2, 3$ and $k = 0, 1, 2$. Consider $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$, where $\rho \in \{0.2, 0.4, 0.8\}$. Simulation results are shown in Tables (71) - (**??**). Since James-Stein estimators with $c \in [-\frac{D}{A}, 0]$ or $c \in [0, -\frac{D}{A}]$ produce smaller integred MSE than local regression estimators, we look at the optimum point, $-\frac{D}{2A}$, which minimizes the integrated MSE. For most cases, the optimum point is around $p - 2 = 1$. In some cases when α is big, the optimum point gets far away from 1.

Table 71: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05518	0.05389	-0.01090	-0.08598	0.77914	6.19804	6.16724	6.16454
	0.10	0.08644	0.08558	-0.00443	-0.16228	0.93874	6.03662	5.96078	5.96045
	0.05	0.18711	0.18403	-0.00096	-0.36614	0.97842	6.00191	5.82288	5.82279
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08529	0.06925	0	-0.13849	0.81190	6	5.94680	5.94378
	0.35	0.15186	0.14054	0	-0.28108	0.92546	6	5.87078	5.86994
	0.25	0.42035	0.40908	0	-0.81815	0.97317	6	5.60220	5.60190
	0.15	1.32004	1.30471	0	-2.60943	0.98839	6	4.71061	4.71043
	0.05	1.83148	1.83144	0	-3.66287	0.99998	6	4.16861	4.16861
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.30899	2.22915	0	-4.45831	0.96543	6	3.85068	3.84792
	0.25	2.02002	1.91781	0	-3.83561	0.94940	6	4.18441	4.17924
	0.15	1.87712	1.87755	0	-3.75510	1.00023	6	4.12202	4.12202
	0.05	1.66717	1.65313	0	-3.30626	0.99158	6	4.36091	4.36080

Table 72: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.25225	0.24282	-0.01443	-0.45678	0.90539	6.10333	5.89881	5.89655
	0.10	0.38173	0.38262	-0.00636	-0.75253	0.98568	6.02200	5.65120	5.65112
	0.05	0.66170	0.65924	-0.00136	-1.31576	0.99422	6.00132	5.34726	5.34724
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.73445	0.72725	-0.19641	-1.06168	0.72277	7.84738	7.52015	7.46371
	0.35	1.10484	1.11454	-0.10482	-2.01943	0.91390	6.51356	5.59897	5.59078
	0.25	1.44622	1.42204	-0.03688	-2.77033	0.95778	6.08447	4.76036	4.75778
	0.15	1.70473	1.69523	-0.00421	-3.38204	0.99196	6.00466	4.32735	4.32724
	0.05	1.83623	1.83187	8.32e-05	-3.66391	0.99767	6.00001	4.17233	4.17232
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.97666	1.91807	-0.05843	-3.71928	0.94080	6.14485	4.40222	4.39530
	0.25	1.95340	1.92143	-0.00201	-3.83883	0.98260	6.00166	4.11623	4.11564
	0.15	1.98950	1.95320	-8.69e-05	-3.90622	0.98171	6.00003	4.08331	4.08264
	0.05	1.66167	1.66521	3.85e-06	-3.33043	1.00214	6	4.33123	4.33123

Table 73: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.14264	0.14153	-0.02561	-0.23184	0.81263	6.02814	5.93895	5.93394
	0.10	0.22447	0.22712	-0.00823	-0.43779	0.97518	6.00321	5.78989	5.78975
	0.05	0.49383	0.49418	-0.00149	-0.98539	0.99771	6.00015	5.50859	5.50858
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.31790	0.31218	-0.08944	-0.44547	0.70065	12.46979	12.34222	12.31373
	0.35	0.50846	0.45396	0.06804	-1.04400	1.02664	8.35948	7.82394	7.82358
	0.25	0.93947	0.90902	0.04396	-1.90596	1.01438	6.46865	5.50216	5.50197
	0.15	1.56969	1.54003	-0.00076	-3.07855	0.98062	6.02598	4.51712	4.51653
	0.05	1.82266	1.81533	3.93e-05	-3.63073	0.99600	6.00003	4.19196	4.19193
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.44906	1.42720	-0.27904	-2.29632	0.79235	9.81032	8.96305	8.90057
	0.25	1.78747	1.70531	-0.01780	-3.37503	0.94408	6.01318	4.42562	4.42003
	0.15	1.84043	1.80561	2.57e-05	-3.61127	0.98109	6.00004	4.22920	4.22855
	0.05	1.59717	1.60699	-7.65e-07	-3.21397	1.00614	6	4.38320	4.38314

Table 74: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.14913	0.14745	-0.00943	-0.27604	0.92550	6.04931	5.92240	5.92157
	0.10	0.25629	0.24606	-0.00411	-0.48390	0.94406	6.00964	5.78202	5.78122
	0.05	0.51645	0.50628	-0.00087	-1.01082	0.97862	6.00053	5.50616	5.50592
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.59997	0.59390	-0.08619	-1.01542	0.84623	6.22796	5.81250	5.79832
	0.35	1.05590	1.01514	-0.04025	-1.94978	0.92328	6.05147	5.15758	5.15137
	0.25	1.50725	1.49699	-0.01206	-2.96986	0.98519	6.00690	4.54430	4.54397
	0.15	1.75169	1.73335	-0.00103	-3.46465	0.98895	6.00031	4.28735	4.28714
	0.05	1.79212	1.80997	2.14e-05	-3.61998	1.00997	6.00000	4.17215	4.17197
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.94190	1.93652	-0.00128	-3.87048	0.99657	6.00108	4.07249	4.07247
	0.25	1.94894	1.93494	6.43e-05	-3.87000	0.99285	6.00001	4.07895	4.07885
	0.15	1.81974	1.80638	2.19e-05	-3.61279	0.99267	6.00000	4.20694	4.20685
	0.05	1.60274	1.62726	-9.23e-07	-3.25452	1.01530	6	4.34822	4.34785

Table 75: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ $(\rho = 0.2)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.05274	0.05223	-0.01004	-0.08438	0.79988	6.19250	6.16086	6.15875
	0.10	0.08135	0.08024	-0.00413	-0.15221	0.93556	6.03583	5.96500	5.96463
	0.05	0.18603	0.18351	-0.00088	-0.36526	0.98172	6.00188	5.82265	5.82259
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08758	0.07131	-0.00594	-0.13075	0.74648	6.08346	6.04029	6.03467
	0.35	0.15244	0.14167	-0.00386	-0.27562	0.90402	6.02008	5.89690	5.89550
	0.25	0.44520	0.42363	-0.00177	-0.84372	0.94758	6.00287	5.60434	5.60312
	0.15	1.29588	1.28552	-0.00018	-2.57069	0.99187	6.00014	4.72533	4.72525
	0.05	1.85649	1.86767	-1.96e-05	-3.73529	1.00601	6.00000	4.12120	4.12113
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.25312	2.09849	-0.00089	-4.19520	0.93098	6.00108	4.05900	4.04826
	0.25	1.84022	1.81281	-9.37e-05	-3.62544	0.98506	6.00001	4.21479	4.21438
	0.15	1.85070	1.82599	2.54e-05	-3.65204	0.98666	6.00000	4.19866	4.19833
	0.05	1.70472	1.70561	1.54e-07	-3.41121	1.00052	6	4.29350	4.29350

Table 76: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17787	0.17279	-0.00656	-0.33246	0.93456	6.01324	5.85864	5.85788
	0.10	0.28467	0.28612	-0.00260	-0.56704	0.99596	6.00245	5.72008	5.72007
	0.05	0.61333	0.61684	-0.00044	-1.23280	1.00500	6.00013	5.38066	5.38064
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.96811	0.93818	-0.00554	-1.86528	0.96336	6.03581	5.13864	5.13734
	0.35	1.29325	1.27506	-0.00170	-2.54665	0.98460	6.00782	4.75441	4.75410
	0.25	1.63951	1.63720	-0.00101	-3.27238	0.99797	6.00101	4.36814	4.36813
	0.15	1.86441	1.86092	4.82e-05	-3.72193	0.99815	6.00004	4.14253	4.14252
	0.05	1.84781	1.84066	6.45e-06	-3.68134	0.99614	6.00000	4.16647	4.16645
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.95483	1.91994	7.26e-05	-3.84002	0.98219	6.00046	4.11528	4.11466
	0.25	1.81785	1.79713	-5.93e-05	-3.59413	0.98857	6.00000	4.22372	4.22348
	0.15	1.78718	1.78274	1.60e-05	-3.56551	0.99752	6	4.22167	4.22166
	0.05	1.67331	1.67516	1.78e-07	-3.35032	1.00111	6	4.32299	4.32299

Table 77: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.23251	0.22819	-0.01131	-0.43374	0.93273	6.17416	5.97293	5.97188
	0.10	0.28883	0.27842	-0.00461	-0.54762	0.94799	6.03217	5.77338	5.77260
	0.05	0.37451	0.37823	-0.00102	-0.75442	1.00720	6.00168	5.62177	5.62175
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08742	0.07184	-0.00096	-0.14175	0.81071	6.01585	5.96152	5.95839
	0.35	0.16076	0.14526	-0.00055	-0.28941	0.90014	6.00393	5.87528	5.87368
	0.25	0.43361	0.40728	-0.00029	-0.81397	0.93860	6.00057	5.62021	5.61857
	0.15	1.29714	1.27908	-8.43e-05	-2.55798	0.98601	6.00003	4.73919	4.73893
	0.05	1.84062	1.84680	-3.61e-05	-3.69354	1.00334	6	4.14708	4.14706
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.19972	2.05911	0.00015	-4.11851	0.93615	6.00026	4.08147	4.07250
	0.25	1.93424	1.87848	-5.20e-05	-3.75686	0.97115	6.00000	4.17738	4.17577
	0.15	1.93644	1.95341	1.52e-05	-3.90686	1.00878	6	4.02958	4.02943
	0.05	1.59193	1.63405	-3.16e-08	-3.26810	1.02646	6	4.32383	4.32271

Table 78: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.19680	0.19530	-0.01079	-0.36902	0.93756	6.02848	5.85626	5.85549
	0.10	0.31115	0.30854	-0.00457	-0.60795	0.97692	6.00555	5.70876	5.70859
	0.05	0.67377	0.68353	-0.00093	-1.36520	1.01311	6.00030	5.31411	5.31400
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.79398	0.82529	-0.09138	-1.46780	0.92433	6.18030	5.50648	5.50193
	0.35	1.22147	1.21018	-0.03713	-2.34611	0.96037	6.03921	4.91456	4.91264
	0.25	1.60519	1.60884	-0.00965	-3.19837	0.99626	6.00504	4.41186	4.41183
	0.15	1.76975	1.76831	-0.00088	-3.53486	0.99869	6.00022	4.23511	4.23511
	0.05	1.82694	1.82927	-1.12e-05	-3.65851	1.00127	6.00000	4.16843	4.16843
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.96732	1.90984	0.00011	-3.81990	0.97084	6.00046	4.14788	4.14621
	0.25	1.98445	1.95321	-5.79e-05	-3.90630	0.98423	6.00000	4.07816	4.07766
	0.15	1.99483	1.96314	-3.19e-05	-3.92621	0.98410	6	4.06862	4.06811
	0.05	1.66752	1.68081	2.37e-07	-3.36163	1.00797	6	4.30589	4.30579

Table 79: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ $(\rho = 0.2)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.93860	0.90797	-0.01725	-1.78142	0.94898	6.08499	5.24217	5.23973
	0.10	1.03464	1.03089	-0.00759	-2.04660	0.98904	6.01833	4.99359	4.99346
	0.05	1.28373	1.27595	-0.00165	-2.54859	0.99265	6.00111	4.73625	4.73618
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.81688	0.80612	-0.21157	-1.18908	0.72782	7.77976	7.40756	7.34704
	0.35	1.18191	1.11343	-0.11832	-1.99020	0.84194	6.49741	5.68912	5.65959
	0.25	1.51299	1.46104	-0.04078	-2.84051	0.93871	6.08217	4.75465	4.74896
	0.15	1.83330	1.76418	-0.00482	-3.51872	0.95967	6.00455	4.31913	4.31615
	0.05	1.84012	1.84083	-7.62e-05	-3.68150	1.00034	6.00001	4.15863	4.15863
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.89534	1.89055	-0.05605	-3.66900	0.96790	6.14403	4.37036	4.36841
	0.25	1.89857	1.80940	-0.00182	-3.61518	0.95208	6.00165	4.28504	4.28068
	0.15	1.88114	1.85597	7.62e-05	-3.71209	0.98666	6.00003	4.16907	4.16874
	0.05	1.66009	1.65265	3.44e-06	-3.30530	0.99552	6	4.35479	4.35475

Table 80: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.2$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.13830	0.13884	-0.00630	-0.26509	0.95839	6.02999	5.90320	5.90296
	0.10	0.21958	0.22635	-0.00270	-0.44729	1.01853	6.00574	5.77803	5.77795
	0.05	0.47702	0.48652	-0.00058	-0.97188	1.01869	6.00031	5.50545	5.50529
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.80495	0.77141	-0.02695	-1.48893	0.92486	6.09931	5.41532	5.41078
	0.35	1.25092	1.21447	-0.01433	-2.40028	0.95941	6.02401	4.87464	4.87258
	0.25	1.58079	1.53595	-0.00425	-3.06339	0.96894	6.00344	4.52083	4.51931
	0.15	1.82433	1.82499	-0.00054	-3.64890	1.00007	6.00017	4.17559	4.17559
	0.05	1.82560	1.84340	-5.58e-05	-3.68670	1.00972	6.00000	4.13891	4.13874
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.98932	1.89121	-0.00212	-3.77819	0.94962	6.00133	4.21247	4.20742
	0.25	1.99131	1.88757	-5.36e-05	-3.77503	0.94788	6.00001	4.21629	4.21088
	0.15	1.81984	1.76791	-1.61e-05	-3.53579	0.97145	6.00000	4.28406	4.28257
	0.05	1.64055	1.65116	1.65e-07	-3.30233	1.00647	6	4.33823	4.33816

Table 81: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04961	0.04704	-0.01072	-0.07263	0.73207	6.19804	6.17502	6.17145
	0.10	0.07458	0.07383	-0.00441	-0.13884	0.93086	6.03662	5.97236	5.97200
	0.05	0.16450	0.16192	-0.00094	-0.32196	0.97858	6.00191	5.84445	5.84438
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08667	0.06946	0	-0.13891	0.80138	6	5.94776	5.94434
	0.35	0.15308	0.13673	0	-0.27346	0.89320	6	5.87962	5.87787
	0.25	0.41963	0.40737	0	-0.81475	0.97079	6	5.60489	5.60453
	0.15	1.28917	1.26604	0	-2.53209	0.98206	6	4.75709	4.75667
	0.05	1.82402	1.82234	0	-3.64467	0.99908	6	4.17935	4.17935
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.17183	2.11105	0	-4.22210	0.97201	6	3.94973	3.94803
	0.25	1.97579	1.98134	0	-3.96267	1.00281	6	4.01312	4.01310
	0.15	1.88333	1.88204	0	-3.76407	0.99932	6	4.11925	4.11925
	0.05	1.63888	1.63469	0	-3.26938	0.99745	6	4.36949	4.36948

Table 82: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.24818	0.23333	-0.01581	-0.43505	0.87647	6.10333	5.91647	5.91268
	0.10	0.35856	0.35881	-0.00683	-0.70395	0.98163	6.02200	5.67661	5.67648
	0.05	0.63566	0.63463	-0.00135	-1.26655	0.99625	6.00132	5.37043	5.37042
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.66589	0.67962	-0.18423	-0.99076	0.74393	7.84738	7.52251	7.47885
	0.35	1.08794	0.99879	-0.10039	-1.79679	0.82578	6.51356	5.80470	5.77168
	0.25	1.45478	1.45008	-0.03951	-2.82114	0.96961	6.08447	4.71811	4.71676
	0.15	1.77408	1.73732	-0.00499	-3.46466	0.97646	6.00466	4.31409	4.31310
	0.05	1.82127	1.81137	-5.68e-05	-3.62263	0.99454	6.00001	4.19864	4.19858
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.92151	1.73386	-0.06204	-3.34365	0.87007	6.14485	4.72271	4.69027
	0.25	1.94476	1.92101	-0.00179	-3.83843	0.98687	6.00166	4.10799	4.10765
	0.15	1.87035	1.85456	-7.52e-05	-3.70896	0.99151	6.00003	4.16142	4.16129
	0.05	1.58497	1.64005	-4.50e-07	-3.28011	1.03475	6	4.30487	4.30295

Table 83: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\hat{\mu}_j(x)$	0.15	0.12588	0.12462	-0.02601	-0.19722	0.78336	6.02814	5.95680	5.95089
	0.10	0.20080	0.20173	-0.00964	-0.38420	0.95665	6.00321	5.81982	5.81944
	0.05	0.44684	0.43894	-0.00157	-0.87474	0.97882	6.00015	5.57225	5.57205
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.26666	0.25953	-0.09587	-0.32732	0.61374	12.46979	12.40913	12.36934
	0.35	0.40635	0.39189	0.06273	-0.90924	1.11879	8.35948	7.85660	7.85086
	0.25	0.78211	0.75768	0.06303	-1.64141	1.04934	6.46865	5.60936	5.60745
	0.15	1.58349	1.48770	0.00082	-2.97706	0.94003	6.02598	4.63241	4.62672
	0.05	1.83666	1.82471	-0.00016	-3.64909	0.99340	6.00003	4.18760	4.18752
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.64907	1.32185	-0.30304	-2.03762	0.61781	9.81032	9.42177	9.19089
	0.25	1.74567	1.72140	-0.02234	-3.39811	0.97330	6.01318	4.36073	4.35949
	0.15	1.83016	1.80157	-0.00031	-3.60252	0.98421	6.00004	4.22769	4.22723
	0.05	1.60829	1.60616	-4.57e-06	-3.21231	0.99867	6	4.39598	4.39598

Table 84: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ $(\rho = 0.4)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.11934	0.11406	-0.00948	-0.20917	0.87637	6.04931	5.95947	5.95765
	0.10	0.19266	0.18462	-0.00399	-0.36125	0.93755	6.00964	5.84104	5.84029
	0.05	0.43659	0.42395	-0.00088	-0.84614	0.96903	6.00053	5.59098	5.59056
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.59996	0.58315	-0.08966	-0.98697	0.82253	6.22796	5.84095	5.82205
	0.35	0.98584	0.92712	-0.04108	-1.77208	0.89877	6.05147	5.26522	5.25512
	0.25	1.34590	1.31952	-0.01183	-2.61537	0.97161	6.00690	4.73743	4.73634
	0.15	1.75993	1.75691	-0.00125	-3.51131	0.99757	6.00031	4.24894	4.24893
	0.05	1.79001	1.80287	2.25e-05	-3.60578	1.00720	6.00000	4.18423	4.18414
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.93512	1.88268	-0.00070	-3.76396	0.97254	6.00108	4.17224	4.17078
	0.25	1.95982	1.93944	-6.65e-05	-3.87874	0.98956	6.00001	4.08109	4.08088
	0.15	1.78573	1.75616	-9.52e-06	-3.51231	0.98344	6.00000	4.27343	4.27294
	0.05	1.68999	1.69225	-9.05e-07	-3.38449	1.00133	6	4.30550	4.30550

Table 85: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ $(\rho = 0.4)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04702	0.04716	-0.00999	-0.07434	0.79052	6.19250	6.16518	6.16311
	0.10	0.07214	0.07030	-0.00408	-0.13244	0.91796	6.03583	5.97553	5.97505
	0.05	0.16160	0.15349	-0.00088	-0.30522	0.94434	6.00188	5.85827	5.85777
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08240	0.06824	-0.00378	-0.12894	0.78234	6.08346	6.03693	6.03303
	0.35	0.14875	0.13540	-0.00253	-0.26573	0.89321	6.02008	5.90310	5.90140
	0.25	0.41417	0.39841	-0.00151	-0.79379	0.95829	6.00287	5.62325	5.62253
	0.15	1.27183	1.25543	-0.00030	-2.51025	0.98686	6.00014	4.76172	4.76150
	0.05	1.78397	1.80128	2.78e-05	-3.60262	1.00972	6.00000	4.18135	4.18118
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.32946	1.96215	-0.00088	-3.92255	0.84194	6.00108	4.40799	4.34980
	0.25	2.02957	1.82235	-8.59e-05	-3.64453	0.89786	6.00001	4.38505	4.36388
	0.15	1.87072	1.81470	2.24e-05	-3.62945	0.97007	6.00000	4.24127	4.23959
	0.05	1.66309	1.64085	-3.31e-07	-3.28170	0.98663	6	4.38139	4.38110

Table 86: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.17055	0.16366	-0.00648	-0.31435	0.92158	6.01324	5.86944	5.86839
	0.10	0.25947	0.25336	-0.00272	-0.50127	0.96594	6.00245	5.76065	5.76035
	0.05	0.55251	0.55251	-0.00049	-1.10402	0.99911	6.00013	5.44861	5.44861
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.91618	0.90119	-0.00607	-1.79024	0.97702	6.03581	5.16174	5.16126
	0.35	1.23423	1.22174	-0.00503	-2.43342	0.98581	6.00782	4.80863	4.80838
	0.25	1.64008	1.61821	-0.00046	-3.23550	0.98639	6.00101	4.40558	4.40528
	0.15	1.79849	1.81451	-0.00011	-3.62879	1.00884	6.00004	4.16975	4.16961
	0.05	1.86132	1.85140	-2.96e-05	-3.70273	0.99465	6.00000	4.15859	4.15854
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.99714	1.91572	-0.00029	-3.83086	0.95909	6.00046	4.16674	4.16340
	0.25	1.82672	1.80456	-2.24e-05	-3.60908	0.98786	6.00000	4.21764	4.21737
	0.15	1.83938	1.79614	1.79e-05	-3.59231	0.97650	6	4.24707	4.24606
	0.05	1.69659	1.68367	1.79e-05	-3.36737	0.99239	6	4.32922	4.32912

Table 87: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.22285	0.21750	-0.01140	-0.41219	0.92480	6.17416	5.98482	5.98356
	0.10	0.26158	0.25245	-0.00467	-0.49557	0.94727	6.03217	5.79818	5.79745
	0.05	0.34339	0.34627	-0.00099	-0.69056	1.00550	6.00168	5.65451	5.65450
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.08410	0.06726	-0.00325	-0.12800	0.76101	6.01585	5.97194	5.96714
	0.35	0.15082	0.13334	-0.00197	-0.26274	0.87103	6.00393	5.89201	5.88950
	0.25	0.42391	0.39801	-0.00100	-0.79403	0.93657	6.00057	5.63044	5.62874
	0.15	1.26031	1.24660	-0.00013	-2.49292	0.98901	6.00003	4.76741	4.76726
	0.05	1.83350	1.82186	-3.56e-05	-3.64364	0.99363	6	4.18986	4.18979
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.16248	2.08931	-6.93e-05	-4.17848	0.96613	6.00026	3.98426	3.98178
	0.25	1.92013	1.91130	-4.89e-05	-3.82251	0.99538	6.00000	4.09762	4.09758
	0.15	1.84408	1.84392	-1.61e-05	-3.68781	0.99991	6	4.15626	4.15626
	0.05	1.68369	1.68921	-2.91e-07	-3.37842	1.00328	6	4.30527	4.30526

Table 88: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.18682	0.18111	-0.01204	-0.33814	0.90499	6.02848	5.87716	5.87547
	0.10	0.29023	0.28311	-0.00515	-0.55591	0.95770	6.00555	5.73987	5.73935
	0.05	0.61079	0.60017	-0.00097	-1.19841	0.98102	6.00030	5.41269	5.41247
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.68977	0.67233	-0.10429	-1.13608	0.82352	6.18030	5.73399	5.71251
	0.35	1.11269	1.07559	-0.04649	-2.05820	0.92487	6.03921	5.09370	5.08742
	0.25	1.50544	1.50406	-0.01249	-2.98313	0.99079	6.00504	4.52734	4.52722
	0.15	1.72658	1.71512	-0.00112	-3.42798	0.99271	6.00022	4.29881	4.29872
	0.05	1.77958	1.80601	-7.33e-05	-3.61187	1.01481	6.00000	4.16771	4.16732
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.96788	1.92858	-0.00051	-3.85612	0.97976	6.00046	4.11222	4.11142
	0.25	1.98007	1.96076	-6.08e-05	-3.92140	0.99022	6.00000	4.05867	4.05848
	0.15	1.84349	1.83231	-3.42e-06	-3.66461	0.99393	6	4.17888	4.17881
	0.05	1.69004	1.70002	-2.76e-07	-3.40005	1.00591	6	4.28999	4.28994

Table 89: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.85115	0.80395	-0.01765	-1.57261	0.92382	6.08499	5.36353	5.35859
	0.10	1.02702	0.99477	-0.00741	-1.97473	0.96139	6.01833	5.07062	5.06909
	0.05	1.31229	1.27589	-0.00152	-2.54873	0.97110	6.00111	4.76467	4.76357
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.77677	0.75320	-0.21665	-1.07311	0.69074	7.77976	7.48343	7.40914
	0.35	1.13714	1.08879	-0.11119	-1.95519	0.85970	6.49741	5.67936	5.65697
	0.25	1.49207	1.42617	-0.04094	-2.77045	0.92839	6.08217	4.80378	4.79613
	0.15	1.75828	1.71646	-0.00561	-3.42169	0.97302	6.00455	4.34114	4.33986
	0.05	1.80841	1.80984	-9.26e-05	-3.61949	1.00074	6.00001	4.18892	4.18892
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.85343	1.85012	-0.06582	-3.56860	0.96270	6.14403	4.42886	4.42628
	0.25	1.89396	1.83977	-0.00239	-3.67476	0.97013	6.00165	4.22084	4.21915
	0.15	1.85568	1.82534	5.85e-05	-3.65079	0.98368	6.00003	4.20491	4.20442
	0.05	1.66543	1.65798	2.04e-07	-3.31595	0.99553	6	4.34947	4.34944

Table 90: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.4$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.11175	0.11213	-0.00561	-0.21304	0.95321	6.02999	5.92870	5.92845
	0.10	0.17883	0.18148	-0.00235	-0.35825	1.00164	6.00574	5.82632	5.82632
	0.05	0.39901	0.40649	-0.00054	-0.81191	1.01741	6.00031	5.58741	5.58729
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.84874	0.82694	-0.04020	-1.57347	0.92694	6.09931	5.37459	5.37006
	0.35	1.31371	1.22689	-0.01938	-2.41502	0.91916	6.02401	4.87037	4.86549
	0.25	1.59409	1.53854	-0.00546	-3.06616	0.96173	6.00344	4.53137	4.52903
	0.15	1.89773	1.87282	-0.00047	-3.7447	0.98663	6.00017	4.15319	4.15285
	0.05	1.83543	1.84205	-3.90e-05	-3.68402	1.00358	6.00000	4.15142	4.15139
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.91154	1.81171	-0.00259	-3.61822	0.94642	6.00133	4.29465	4.28916
	0.25	1.97077	1.85729	-7.99e-05	-3.71442	0.94238	6.00001	4.25636	4.24982
	0.15	1.81133	1.80312	1.70e-05	-3.60627	0.99548	6.00000	4.20506	4.20502
	0.05	1.50891	1.51180	1.19e-06	-3.02361	1.00192	6	4.48530	4.48530

Table 91: $\mu_1(x) = x$, $\mu_2(x) = 2x + 1$, $\mu_3(x) = 5x$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01847	0.01980	-0.01073	-0.01813	0.49089	6.19804	6.19838	6.19359
	0.10	0.03003	0.02954	-0.00440	-0.05029	0.83723	6.03662	6.01637	6.01557
	0.05	0.06467	0.06391	-0.00093	-0.12595	0.97374	6.00191	5.94063	5.94059
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.06014	0.04173	0	-0.08346	0.69384	6	5.97669	5.97105
	0.35	0.08988	0.07524	0	-0.15048	0.83718	6	5.93939	5.93701
	0.25	0.21671	0.19817	0	-0.39635	0.91449	6	5.82036	5.81877
	0.15	0.93851	0.91024	0	-1.82048	0.96988	6	5.11803	5.11718
	0.05	1.81040	1.81515	0	-3.63029	1.00262	6	4.18010	4.18009
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.52980	2.21041	0	-4.42081	0.87375	6	4.10899	4.06866
	0.25	2.01624	1.88647	0	-3.77293	0.93564	6	4.24331	4.23495
	0.15	1.85355	1.85355	0	-3.70710	1	6	4.14645	4.14645
	0.05	1.67456	1.67826	0	-3.35652	1.00221	6	4.31804	4.31803

Table 92: $\mu_1(x) = x^2$, $\mu_2(x) = x^4$, $\mu_3(x) = \exp(x)$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15678	0.14147	-0.02290	-0.23713	0.75627	6.10333	6.02298	6.01367
	0.10	0.22115	0.21475	-0.00985	-0.40980	0.92654	6.02200	5.83334	5.83215
	0.05	0.41282	0.40566	-0.00186	-0.80760	0.97814	6.00132	5.60654	5.60634
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.40062	0.37833	-0.14910	-0.45845	0.57217	7.84738	7.78955	7.71622
	0.35	0.73565	0.71841	-0.10082	-1.23518	0.83951	6.51356	6.01403	5.99508
	0.25	1.22353	1.21172	-0.05102	-2.32141	0.94866	6.08447	4.98658	4.98336
	0.15	1.67288	1.60853	-0.00925	-3.19856	0.95606	6.00466	4.47898	4.47574
	0.05	1.80010	1.79530	-5.13e-05	-3.59049	0.99730	6.00001	4.20962	4.20961
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.78599	1.71566	-0.13502	-3.16127	0.88502	6.14485	4.76957	4.74596
	0.25	1.88636	1.79451	-0.00461	-3.57368	0.94724	6.00166	4.31434	4.30909
	0.15	1.92660	1.89673	-0.00017	-3.79310	0.98440	6.00003	4.13353	4.13306
	0.05	1.68734	1.66180	3.36e-06	-3.32362	0.98487	6	4.36373	4.36334

Table 93: $\mu_1(x) = \sin(x)$, $\mu_2(x) = \cos(x)$, $\mu_3(x) = \sin 5x$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.06416	0.05592	-0.02647	-0.05891	0.45903	6.02814	6.03340	6.01462
	0.10	0.09730	0.09044	-0.00985	-0.16118	0.82826	6.00321	5.93934	5.93647
	0.05	0.20303 0.19725	-0.00183	-0.39083	0.96250	6.00015	5.81235	5.81206	
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.16849	0.13180	-0.14553	0.02745	-0.08145	12.46979	12.66573	12.46867
	0.35	0.23322	0.20167	0.05919	-0.52172	1.11851	8.35948	8.07099	8.06771
	0.25	0.40224	0.39064	0.09503	-0.97135	1.20743	6.46865	5.89954	5.88224
	0.15	1.13812	1.14693	0.00284	-2.29955	1.01024	6.02598	4.86456	4.86444
	0.05	1.81261	1.80254	8.50e-05	-3.60526	0.99449	6.00003	4.20738	4.20732
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.52875	1.44014	-0.66691	-1.54646	0.50579	9.81032	9.79261	9.41922
	0.25	1.70241	1.52147	-0.05101	-2.94092	0.86375	6.01318	4.77467	4.74307
	0.15	1.79958	1.74605	-0.00119	-3.48970	0.96959	6.00004	4.30993	4.30826
	0.05	1.61288	1.61519	-1.30e-05	-3.23036	1.00143	6	4.38252	4.38252

Table 94: $\mu_1(x) = x$, $\mu_2(x) = x^2$, $\mu_3(x) = \exp(x)$ $(\rho = 0.8)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04204	0.04193	-0.00915	-0.06558	0.77989	6.04931	6.02577	6.02373
	0.10	0.06581	0.06375	-0.00389	-0.11972	0.90965	6.00964	5.95572	5.95518
	0.05	0.14383	0.14085	-0.00086	-0.27999	0.97334	6.00053	5.86437	5.86427
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.37452	0.36165	-0.11019	-0.50291	0.67140	6.22796	6.09957	6.05913
	0.35	0.66416	0.58929	-0.05722	-1.06414	0.80112	6.05147	5.65148	5.62521
	0.25	1.15446	1.11805	-0.01950	-2.19710	0.95157	6.00690	4.96426	4.96155
	0.15	1.65716	1.62149	-0.00255	-3.23788	0.97694	6.00031	4.41959	4.41871
	0.05	1.80467	1.79920	3.70e-05	-3.59848	0.99699	6.00000	4.20620	4.20618
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.82875	1.79243	-0.00326	-3.57833	0.97836	6.00108	4.25149	4.25064
	0.25	1.88448	1.86476	-9.76e-05	-3.72933	0.98948	6.00001	4.15517	4.15496
	0.15	1.86530	1.84717	-7.92e-06	-3.69432	0.99027	6.00000	4.17099	4.17081
	0.05	1.59560	1.59306	-9.08e-07	-3.18612	0.99841	6	4.40948	4.40948

Table 95: $\mu_1(x) = x$, $\mu_2(x) = 5x$, $\mu_3(x) = \exp(x)$ $(\rho = 0.8)$

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.01949	0.01958	-0.00995	-0.01925	0.49382	6.19250	6.19274	6.18774
	0.10	0.03035	0.03033	-0.00406	-0.05254	0.86539	6.03583	6.01365	6.01310
	0.05	0.06604	0.06297	-0.00089	-0.12415	0.93998	6.00188	5.94377	5.94353
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05577	0.03724	0.00169	-0.07786	0.69805	6.08346	6.06137	6.05629
	0.35	0.08196	0.06472	0.00062	-0.13067	0.79713	6.02008	5.97137	5.96800
	0.25	0.18360	0.17568	-7.48e-05	-0.35121	0.95647	6.00287	5.83526	5.83491
	0.15	0.83587	0.81286	-3.19e-05	-1.62565	0.97242	6.00014	5.21036	5.20973
	0.05	1.76065	1.75416	9.86e-05	-3.50852	0.99637	6.00000	4.25212	4.25210
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.32807	2.10083	-0.00410	-4.19346	0.90063	6.00108	4.13569	4.11270
	0.25	2.04508	1.90375	-9.23e-05	-3.80732	0.93085	6.00001	4.23778	4.22800
	0.15	1.87070	1.79869	-4.28e-05	-3.59729	0.96148	6.00000	4.27341	4.27063
	0.05	1.68897	1.68389	-1.56e-07	-3.36778	0.99699	6	4.32119	4.32118

Table 96: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12546	0.12373	-0.01034	-0.22678	0.90378	6.01324	5.91192	5.91076
	0.10	0.19062	0.18719	-0.00451	-0.36536	0.95834	6.00245	5.82771	5.82738
	0.05	0.38430	0.38447	-0.00084	-0.76727	0.99827	6.00013	5.61716	5.61716
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.60670	0.58222	-0.00541	-1.15363	0.95073	6.03581	5.48889	5.48742
	0.35	1.01283	0.99066	-0.00547	-1.97039	0.97272	6.00782	5.05026	5.04950
	0.25	1.46783	1.43356	-0.00327	-2.86058	0.97442	6.00101	4.60826	4.60730
	0.15	1.75172	1.72127	-0.00023	-3.44208	0.98249	6.00004	4.30968	4.30914
	0.05	1.85919	1.89007	-1.52e-05	-3.78001	1.01661	6.00000	4.07907	4.07856
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.89608	1.77648	-0.00048	-3.55200	0.93667	6.00046	4.34454	4.33694
	0.25	1.82468	1.79521	6.10e-05	-3.59054	0.98388	6.00000	4.23415	4.23368
	0.15	1.89298	1.86925	1.35e-06	-3.73850	0.98746	6	4.15448	4.15418
	0.05	1.63614	1.64058	-7.31e-07	-3.28115	1.00271	6	4.35498	4.35497

Table 97: $\mu_1(x) = \sin x$, $\mu_2(x) = x$, $\mu_3(x) = 5x$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.15058	0.12944	-0.01158	-0.23573	0.78273	6.17416	6.08901	6.08190
	0.10	0.20957	0.19080	-0.00480	-0.37201	0.88753	6.03217	5.86974	5.86708
	0.05	0.26074	0.23725	-0.00102	-0.47247	0.90601	6.00168	5.78995	5.78764
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.05921	0.03660	-0.01021	-0.05279	0.44580	6.01585	6.02226	6.00408
	0.35	0.08439	0.06518	-0.00537	-0.11962	0.70873	6.00393	5.96871	5.96155
	0.25	0.17633	0.16296	-0.00230	-0.32132	0.91112	6.00057	5.85558	5.85419
	0.15	0.81370	0.80878	-0.00056	-1.61645	0.99327	6.00003	5.19728	5.19724
	0.05	1.77336	1.81811	-5.22e-05	-3.63612	1.02520	6	4.13725	4.13612
$\widehat{\mu}_j^{(2)}(x)$	0.45	2.63311	2.32353	-0.00132	-4.64441	0.88193	6.00026	3.98895	3.95224
	0.25	1.89271	1.73896	4.80e-05	-3.47802	0.91879	6.00000	4.41469	4.40221
	0.15	1.99906	1.89474	-2.99e-05	-3.78941	0.94780	6	4.20965	4.20420
	0.05	1.64131	1.73125	1.04e-07	-3.46249	1.05480	6	4.17882	4.17389

Table 98: $\mu_1(x) = \sin x$, $\mu_2(x) = \cos x$, $\mu_3(x) = x^2$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.12285	0.12242	-0.01890	-0.20702	0.84255	6.02848	5.94431	5.94126
	0.10	0.19560	0.18183	-0.00857	-0.34652	0.88580	6.00555	5.85463	5.85208
	0.05	0.32694	0.31831	-0.00168	-0.63326	0.96847	6.00030	5.69399	5.69366
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.28389	0.28187	-0.13322	-0.29730	0.52362	6.18030	6.16689	6.10247
	0.35	0.58079	0.55859	-0.06828	-0.98061	0.84421	6.03921	5.63938	5.62528
	0.25	1.07823	1.03795	-0.02386	-2.20818	0.94051	6.00504	5.05509	5.05128
	0.15	1.69402	1.65589	-0.00294	-3.30589	0.97576	6.00022	4.38834	4.38734
	0.05	1.80821	1.80446	-5.49e-05	-3.60880	0.99789	6.00000	4.19941	4.19940
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.90718	1.81333	0.00208	-3.63082	0.95188	6.00046	4.27682	4.27240
	0.25	1.96188	1.92841	6.82e-05	-3.85695	0.98297	6.00000	4.10493	4.10437
	0.15	1.87716	1.86255	-1.72e-06	-3.72510	0.99222	6	4.15206	4.15195
	0.05	1.62884	1.64040	-5.3e-07	-3.28079	1.00709	6	4.34805	4.34797

Table 99: $\mu_1(x) = \sin x$, $\mu_2(x) = x^4$, $\mu_3(x) = x^2$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.64831	0.61398	-0.01697	-1.19401	0.92089	6.08499	5.53930	5.53524
	0.10	0.74849	0.71172	-0.00779	-1.40784	0.94046	6.01833	5.35898	5.35632
	0.05	1.03490	1.01052	-0.00162	-2.01780	0.97487	6.00111	5.01822	5.01756
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.58574	0.55336	-0.31227	-0.48218	0.41159	7.77976	7.88333	7.68053
	0.35	0.83693	0.84127	-0.	-1.33039	0.79480	6.49741	6.00395	5.96871
	0.25	1.29652	1.20418	-0.06756	-2.27324	0.87667	6.08217	5.10544	5.08572
	0.15	1.65040	1.59087	-0.00117	-3.15837	0.95685	6.00455	4.49658	4.49351
	0.05	1.78447	1.78458	-9.31e-05	-3.56896	1.00001	6.00001	4.21552	4.21552
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.70649	1.67563	-0.14549	-3.06029	0.89666	6.14403	4.79022	4.77200
	0.25	1.82939	1.72870	-0.00479	-3.44780	0.94234	6.00165	4.38324	4.37716
	0.15	1.86780	1.83254	-8.58e-05	-3.66492	0.98108	6.00003	4.20292	4.20225
	0.05	1.65831	1.64199	-1.07e-05	-3.28396	0.99015	6	4.37435	4.37419

Table 100: $\mu_1(x) = \sin x$, $\mu_2(x) = \exp(x)$, $\mu_3(x) = x$ ($\rho = 0.8$)

	α	A	B	D	F	$-F/2A$	MSE_{LR}	MSE_{naive}	MSE_{opt}
$\widehat{\mu}_j(x)$	0.15	0.04236	0.03890	-0.00466	-0.06848	0.80841	6.02999	6.00386	6.00231
	0.10	0.06607	0.06495	-0.00203	-0.12584	0.95225	6.00574	5.94598	5.94583
	0.05	0.14365	0.14065	-0.00044	-0.28041	0.97602	6.00031	5.86355	5.86347
$\widehat{\mu}_j^{(1)}(x)$	0.45	0.93839	0.85624	-0.08416	-1.54416	0.82277	6.09931	5.49354	5.46407
	0.35	1.32723	1.28382	-0.04357	-2.48049	0.93446	6.02401	4.87074	4.86504
	0.25	1.58090	1.51224	-0.01424	-2.99600	0.94756	6.00344	4.58833	4.58399
	0.15	1.83092	1.81539	-0.00135	-3.62808	0.99078	6.00000	4.20300	4.20285
	0.05	1.87607	1.84966	-6.03e-05	-3.69919	0.98589	6.00000	4.17688	4.17651
$\widehat{\mu}_j^{(2)}(x)$	0.45	1.90860	1.79446	-0.00555	-3.57781	0.93729	6.00133	4.33212	4.32462
	0.25	1.89901	1.86741	7.88e-05	-3.73497	0.98340	6.00001	4.16405	4.16353
	0.15	1.80365	1.79021	6.34e-06	-3.58043	0.99255	6.00000	4.22322	4.22322
	0.05	1.69901	1.68381	1.41e-06	-3.36762	0.99105	6	4.33139	4.33126

5 Application to nanoparticle data

In this section, we will revisit the nanoparticle example in Chapter 2 and apply James-Stein type estimation at fixed points discussed in Chapter 3. There are four response variables: M_{11}, M_{12}, M_{33} and M_{34}, which are normalized scattering elements S_{11}, S_{12}, S_{33} and S_{34}. Recall that scattering element is a function of configurations (for example, the diameters of nanoparticles (s), the proportion of agglomerated nanoparticles (p)) and the observation angle (θ). More details can be found in Charnigo et al (2011).

Consider data for nanoparticles with agglomeration level 0% and diameter 5nm at $\theta = 1, 2, \ldots, 179$. We can write scattering elements as a function of θ, i.e., $S_{11}(\theta)$, $S_{12}(\theta)$, $S_{33}(\theta)$ and $S_{34}(\theta)$. Define scattering profiles $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$ and $M_{34}(\theta)$ as following:

$$M_{11}(\theta) := \frac{S_{11}(\theta)}{\max_\theta S_{11}(\theta)},\ M_{12}(\theta) := \frac{S_{12}(\theta)}{S_{11}(\theta)},$$

$$M_{33}(\theta) := \frac{S_{33}(\theta)}{S_{11}(\theta)},\ M_{34}(\theta) := \frac{S_{34}(\theta)}{S_{11}(\theta)}.$$

Panels a through d of Figure 8 depict true values of scattering profiles as determined through a numerical model (Charnigo et al (2011)).

However, true values are usually unknown in practice. A practitioner is most likely to see noisy data as illustrated in Figure 9 (independent error terms from $N(0, 0.1^2)$ are added to $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$ and $M_{34}(\theta)$ for $\theta = 1, 2, \ldots, 179$). And our goal is to estimate $M_{11}(\theta)$, $M_{12}(\theta)$, $M_{33}(\theta)$, $M_{34}(\theta)$ and their derivatives at any fixed θ, in which case $p = 4$. To be consistent with notations in previous chapters, denote $x := \theta$ and for $k = 0, 1, , 2, \ldots, J$,

$$\mu_1^{(k)}(x) := \frac{d^k}{d\theta^k} M_{11}(\theta),\ \mu_2^{(k)}(x) := \frac{d^k}{d\theta^k} M_{12}(\theta), \mu_3^{(k)}(x) := \frac{d^k}{d\theta^k} M_{33}(\theta),\ \mu_4^{(k)}(x) := \frac{d^k}{d\theta^k} M_{34}(\theta).$$

James-Stein type estimation was employed to estimate $\mu_j^{(k)}(x)$ at $x \in \{75, 100, 125\}$, $j = 1, 2, 3, 4$, and $k = 0, 1, 2$. Let $\widehat{\mu}_j^{(k)}(x)$ be the local regression estimator of $\mu_j^{(k)}(x)$. Define James-Stein type estimator

$$\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2 \widehat{\sigma}_j^2}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2} \widehat{\mu}_j^{(k)}(x),$$

Figure 8: Scattering profiles: true data

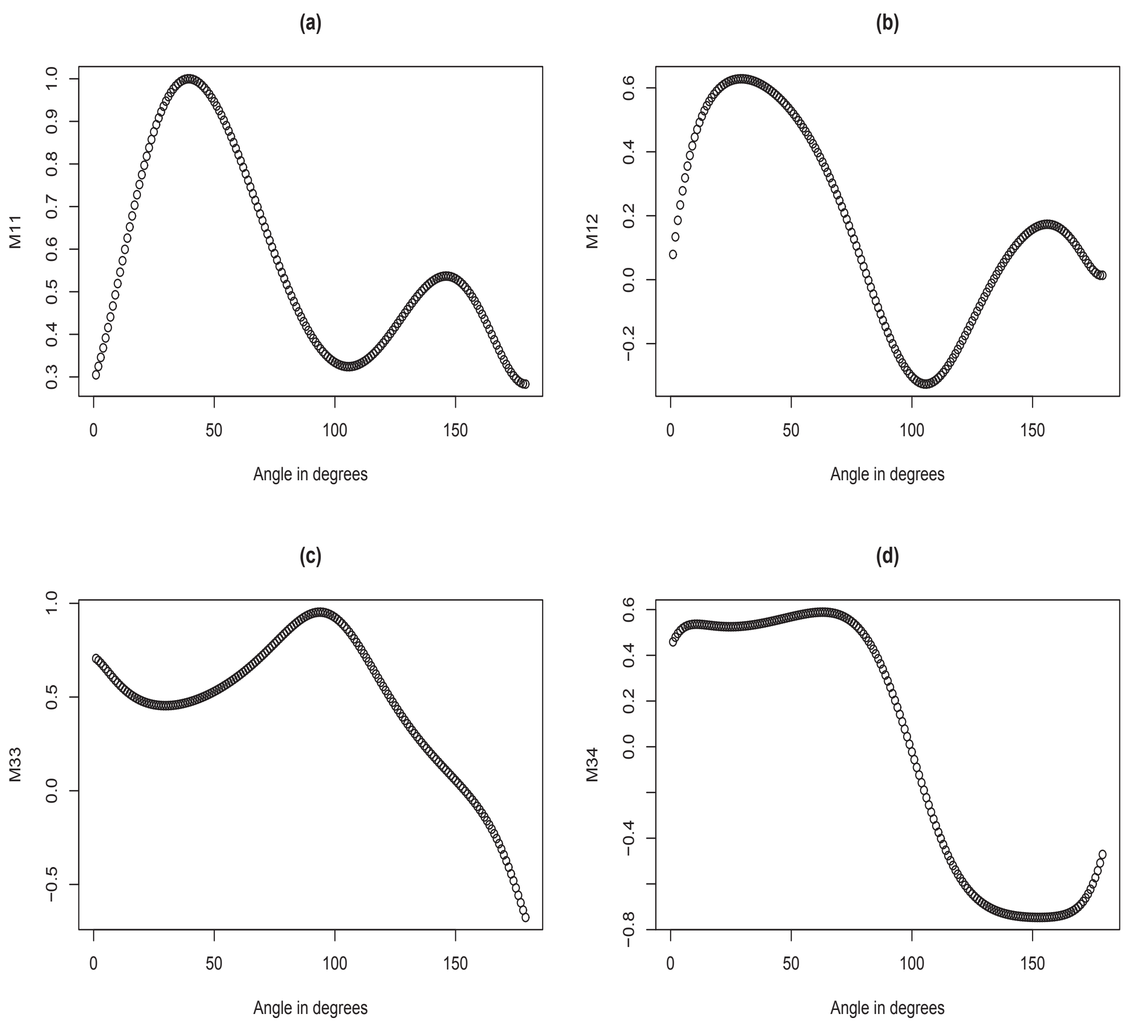
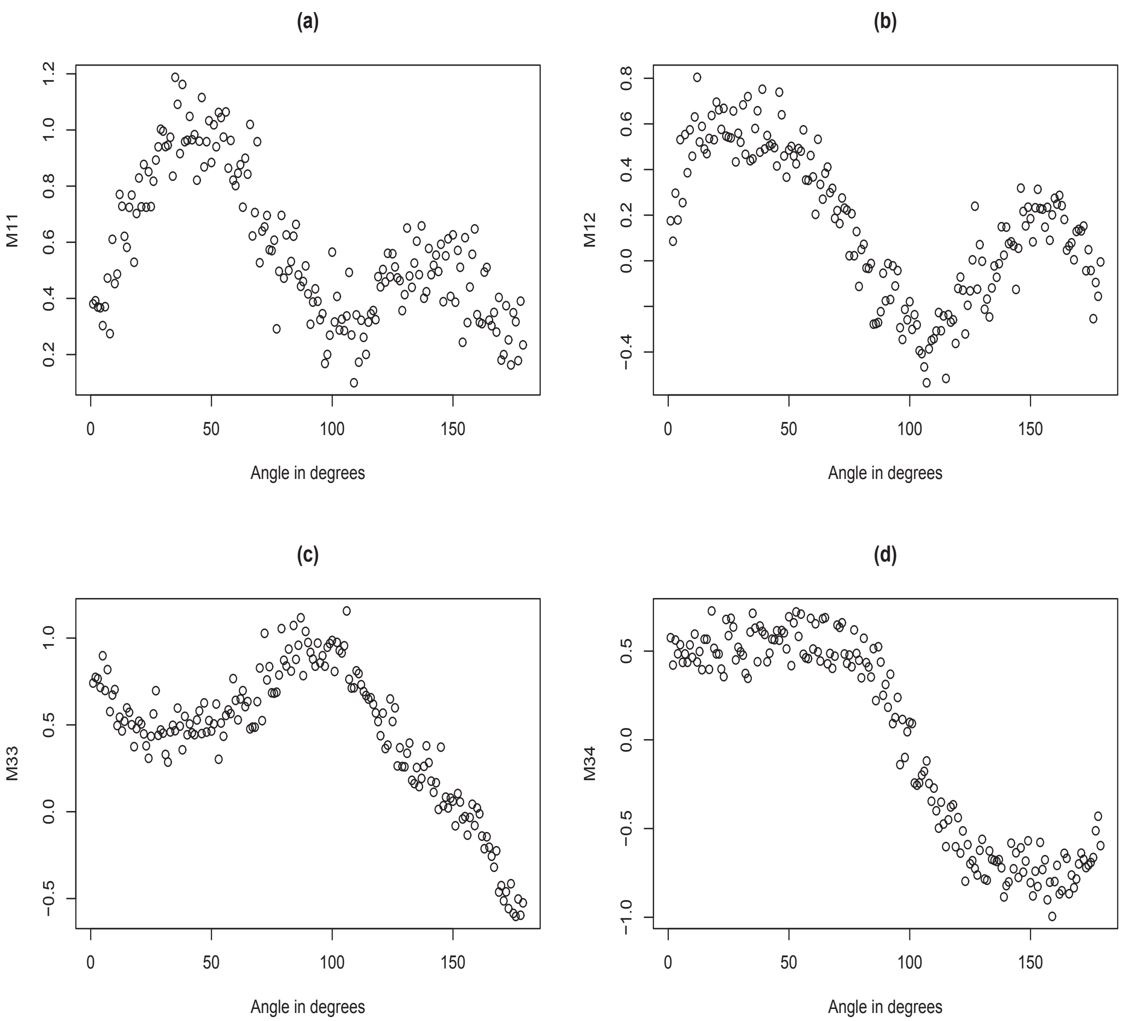


Figure 9: Scattering profiles: noisy data



where

$$\widehat{\sigma}_j^2 = \frac{1}{m} \sum_{i:|x_i - x| > 2h} [Y_{ji} - \widehat{\mu}_j(x_i)]^2$$

with

$$m = \sum_{i:|x_i - x| > 2h} [1 - 2l_i(x_i) + ||l(x_i)||^2]$$

is a consistent estimator of σ_j^2, and c is a positive constant. In local regression, tricube weight function is used and J is set to be 2, 3, 4 for estimating $\mu_j(x)$, $\mu_j'(x)$ and $\mu_j''(x)$, respectively. Consider the nearest neighbor parameter $\alpha \in \{0.05, 0.1, 0.15\}$.

To find the constant c in $\widetilde{\mu}_j^{(k)}(x)$, we use the results in Theorem 3.4. Since the same weight function is applied in estimating $\mu_j^{(k)}(x)$ for $j = 1, 2, 3, 4$, we have $L_0(x) = L_1(x) = ||l^{(k)}(x)||^2$. Values of F, G, H, m and n_x can be calculated in R. Set the upper and lower bounds of σ_j^2 to be $\sigma_{(p)}^2 = 2 * 0.1^2$ and $\sigma_{(1)}^2 = 0.1^2/2$. In this example, we make some modifications to Theorem 3.4. Instead of finding the global bounds for bias and mean response functions, i.e., find B, M_0 and M_1 such that $|b_j^{(k)}(x)| \leq B$ and $M_0 \leq |\mu_j^{(k)}(x)| \leq M_1$ for all $x = 1, , 2, \ldots, 179$, $k = 0, 1, 2$ and $j = 1, 2, 3, 4$, we look for local bounds such that $B_j(x) = 2 * \max_{|x_i - x| \leq h} b_j^{(k)}(x_i)$, $M_{1j}(x) = 2 * \max_{|x_i - x| \leq h} |\mu_j^{(k)}(x_i)|$ and $M_{0j}(x) = \min_{|x_i - x| \leq h} |\mu_j^{(k)}(x_i)|/2$, where h is the bandwidth. Local bounds are used so that B_j, M_{1j} are smaller and M_{0j} is bigger, which helps find c in James-Stein estimator. Modifying Theorem 3.4, we have

$$\begin{aligned}
& EL(\widetilde{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) - EL(\widehat{\boldsymbol{\mu}}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) \\
\leq \; & \frac{c^2}{\sigma_{(1)}^2} \left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2} \right) \mathrm{E} \frac{1}{p - 2 + 2K} - 2c \left[p - 2 \left(1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(p)}^2} \right) \right] \mathrm{E} \frac{1}{p - 2 + 2K} \\
& + 2c \left(1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(1)}^2} \right) \sum_{j=1}^{p} B_j \left(||l^{(k)}(x)|| \sigma_{(p)} \sqrt{\frac{2}{\pi}} + M_{1j} \right) \mathrm{E} \frac{1}{p - 3 + 2K_j},
\end{aligned} \tag{97}$$

where K and K_j follow Poisson distribution with parameters

$$\lambda = \sum_{s=1}^{p} \frac{\left[\mu_s^{(k)}(x) + b_s^{(k)}(x) \right]^2}{2\sigma_s^2 ||l^{(k)}(x)||^2}$$

and

$$\lambda_j = \sum_{s \neq j} \frac{\left[\mu_s^{(k)}(x) + b_s^{(k)}(x)\right]^2}{2\sigma_s^2 ||l^{(k)}(x)||^2},$$

respectively. Note that B in terms $1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(p)}^2}$ and $1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(1)}^2}$ is still the global upper bound for bias.

Upper and lower bounds of $\mathrm{E}\frac{1}{p-2+2K}$ and $\mathrm{E}\frac{1}{p-3+2K_j}$ can be found similar to Theorem 3.4. Therefore we have

$$\frac{1}{2} \cdot \frac{1}{1+\lambda_{max}} \leq \mathrm{E}\frac{1}{p-2+2K} \leq \frac{1}{2} \cdot \frac{1}{\lambda_{min}} \left(1 - \exp(-\lambda_{min})\right),$$

and

$$\mathrm{E}\frac{1}{p-3+2K_j} \leq \frac{1}{\lambda_{j(min)}} \left(1 - \exp(-\lambda_{j(min)})\right),$$

where $\lambda_{max} = \sum_{s=1}^{p} \frac{[M_{1s}+B]^2}{2\sigma_{(1)}^2 ||l^{(k)}(x)||^2}$, $\lambda_{min} = \sum_{s=1}^{p} \frac{[M_{0s}-2M_{1s}B]}{2\sigma_{(p)}^2 ||l^{(k)}(x)||^2}$, and $\lambda_{j(min)} = \sum_{s \neq j} \frac{[M_{0s}-2M_{1s}B]}{2\sigma_{(p)}^2 ||l^{(k)}(x)||^2}$. Then plug all the values in the inequality above, solve for c, which satisfies

$$\begin{aligned} & c\left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2}\right) \cdot \frac{1}{\sigma_{(1)}^2} \frac{1}{2} \cdot \frac{1}{\lambda_{min}} \left(1 - \exp(-\lambda_{min})\right) \\ \leq \; & 2c\left[p - 2\left(1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(p)}^2}\right)\right] \frac{1}{2} \cdot \frac{1}{1+\lambda_{max}} \\ & + 2c\left(1 + \frac{n_x}{m} \cdot \frac{B^2}{\sigma_{(1)}^2}\right) \sum_{j=1}^{p} B_j \left(||l^{(k)}(x)||\sigma_{(p)}\sqrt{\frac{2}{\pi}} + M_{1j}\right) \frac{1}{\lambda_{j(min)}} \left(1 - \exp(-\lambda_{j(min)})\right), \end{aligned}$$

and we obtain $\widetilde{\mu}_j^{(k)}(x)$. Since expected losses are compared, we repeat steps described above 500 times and take the average risks. Both James-Stein estimators where $c = 1$ and $c =$ optimum point which minimizes the right side of inequality (97) are considered. Results of estimating $\mu_j^{(k)}(x)$ are shown in Table (101). Columns MSE_{LR}, MSE_{JS}^{opt} and MSE_{JS}^{naive} refer to MSE of local regression estimator, James-Stein estimator where c is the optimum point, and James-Stein estimator where $c = 1$. Recall that the loss function is defined as

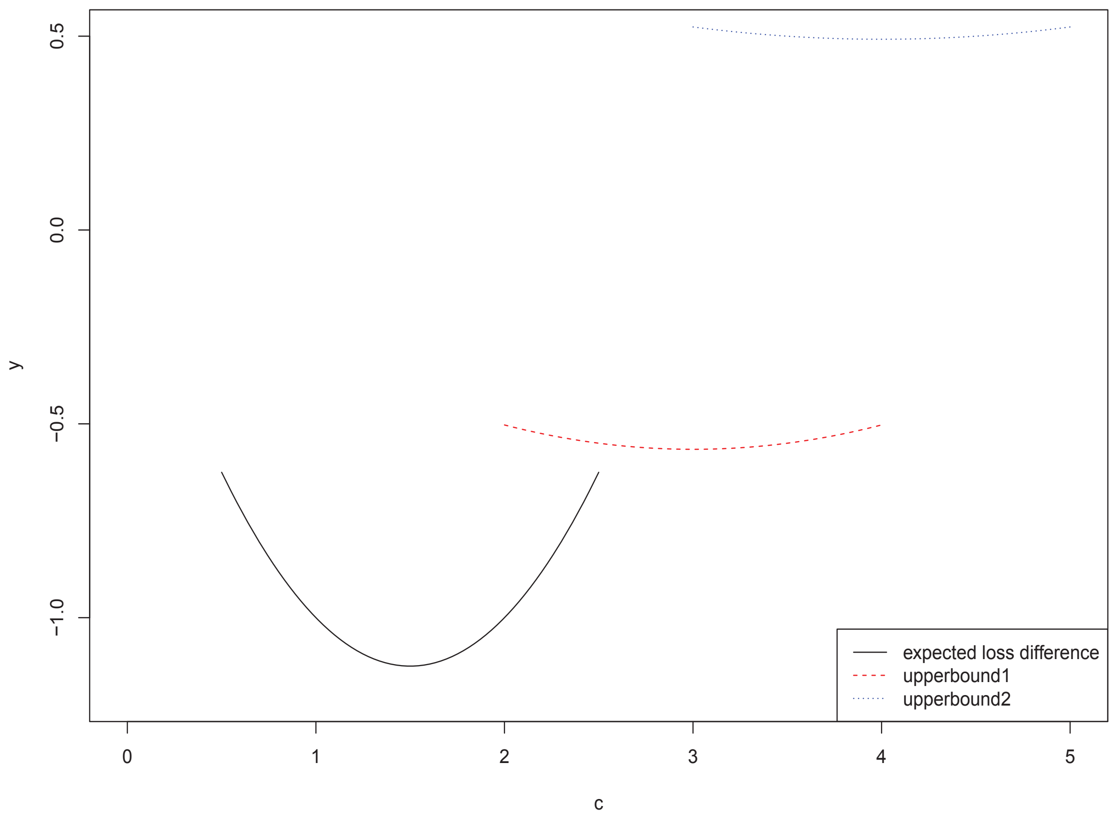
$$L(\mathbf{Z}^{(k)}(x), \boldsymbol{\mu}^{(k)}(x)) = (\mathbf{Z}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T ||l^{(k)}(x)||^{-2} \boldsymbol{\Sigma}^{-1} (\mathbf{Z}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x)).$$

Note that only the true values of mean responses are given and the calculations of MSE involve true values of $\mu'_j(x)$ and $\mu''_j(x)$, we applied local regression on the true mean responses and used the estimates as the true data for first and second order derivatives.

It is observed that when estimating the mean response functions, except for $\alpha = 0.05, \theta = 100$, James-Stein estimator where c is chosen as the optimum point has smaller MSE than local regression estimator. In most cases, naive JS estimator, i.e., JS estimator where $c = 1$, has the smallest MSE among all three estimators. However, Steinization with optimized c did not work well in estimating the first or the second order derivative. On the other hand, naive JS estimator does much better than local regression and reduces MSE up to 50%. Since the real difference between MSEs of $\widetilde{\boldsymbol{\mu}}^{(k)}(x)$ and $\widehat{\boldsymbol{\mu}}^{(k)}(x)$ is unknown, we look for an upper bound of the difference and find a positive c such that the upper bound is negative. It is possible that such c does not exist but JS estimators with some other c produce smaller MSE than local regression. This is because the c that optimizes the right hand side of (97) need not coincide with the c that optimizes the left hand side, which is illustrated in Figure 10. If upperbound1 (dashed) is found, then we can find c that makes both the expected loss difference (solid) and the upperbound negative. However, if upperbound2 (dotted) is established, there is no such c, which optimizes the right had side, that optimizes the left side.

Also, the bias of $\widehat{\mu}_j(x)$ is $b_j(x) = \sum_{s=1}^{n} l_s(x)\mu_j(x_s) - \mu_j(x)$ and it can be calculated since $\mu_j(x)$ is known for all $x = 1, 2, \ldots, 179$. In this example, $b_j(x)$ is of order 10^{-6}, which is very small. But bias of higher order derivatives can not be calculated directly since $b_j^{(k)}(x) = \sum_{s=1}^{n} l_s^{(k)}(x)\mu_j(x_s) - \mu_j^{(k)}(x)$ for $k = 1, 2, \ldots, J$, and $\mu_j^{(k)}(x)$ is unknown in the nanoparticle example. (We used $\sum_{s=1}^{n} l_s^{(k)}(x)\mu_j(x_s)$ as the true values for $\mu_j^{(k)}(x)$ in computing MSE, but we can not simply plug that in because obviously bias will be 0.) So we estimate the bias as $\widehat{b}_j^{(k)}(x) = \sum_{s=1}^{n} l_s^{(k)}(x)\widehat{\mu}_j(x_s) - \widehat{\mu}_j^{(k)}(x)$, which is of order 10^{-2}, larger than $b_j(x)$.

Figure 10: Illustration of upper bounds



To see the effect of bias in finding c, assume there is no bias, the right hand side of inequality (97) becomes

$$c^2\left(F\sigma_{(p)}^2 + G + \frac{H}{\sigma_{(1)}^2}\right)\cdot\frac{1}{\sigma_{(1)}^2}\mathrm{E}\frac{1}{p-2+2K} - 2c(p-2)\mathrm{E}\frac{1}{p-2+2K}.$$

And we can always find a c such that the quantity is negative, i.e., there exists a JS estimator with smaller MSE. However, while everything fixed, as the absolute value of bias increases, the third term on the right hand side of (97) becomes more positive and it will exceed the absolute value of second term at some point, which makes the upper bound positive and therefore we can not find a positive constant c such that the upper bound is negative. Hence, the key is to find a good estimate of bias, which is a main concern of nonparametric regression.

In Chapter 4, we conducted simulations to study JS estimators in reducing the integrated MSE over an interval. And it was observed that JS estimators with most values of $c \in [-2(p-2), 2(p-2)]$ reduced integrated MSE. Next we will apply Steinization on nanoparticle data and calculate the corresponding integrated MSEs. Define James-Stein estimator

$$\widetilde{\mu}_j^{(k)}(x) := \widehat{\mu}_j^{(k)}(x) - \frac{c||l^{(k)}(x)||^2\widehat{\sigma}_j^2}{\sum_{s=1}^p \widehat{\mu}_s^{(k)}(x)^2}\widehat{\mu}_j^{(k)}(x),$$

and the integrated MSE

$$\int_{\mathcal{X}} \mathrm{E}(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))^T(||l^{(k)}(x)||^2\boldsymbol{\Sigma})^{-1}(\widetilde{\boldsymbol{\mu}}^{(k)}(x) - \boldsymbol{\mu}^{(k)}(x))dx$$

where $\mathcal{X} = [1, 179]$. The integral can be approximated by the following steps:

1. At a fixed $c \in \{-4, -3.9, -3.8, \ldots, 3.8, 3.9, 4\}$, calculate $\widetilde{\mu}_j^{(k)}(x)$ where $k = 0, 1, 2$, $j = 1, 2, 3, 4$ and $x = 1, 2, \ldots, 179$.
2. Compute $\sum_{s=1}^{179}\sum_{j=1}^{p}\frac{\left[\widetilde{\mu}_j^{(k)}(x_s) - \mu_j^{(k)}(x_s)\right]^2}{||l^{(k)}(x_s)||^2\sigma_j^2}$ using results from step 1.
3. Repeat the steps above 500 times and take the average.

Note that when $c = 0$, no Steinization is applied and $\widetilde{\mu}_j^{(k)}(x)$ becomes the local regression estimator.

Table 101: MSE of local regression and James-Stein estimators: mean response functions and derivatives

		$\mu_j(x)$			$\mu'_j(x)$			$\mu''_j(x)$		
α	θ	MSE_{LR}	MSE_{JS}^{opt}	MSE_{JS}^{naive}	MSE_{LR}	MSE_{JS}^{opt}	MSE_{JS}^{naive}	MSE_{LR}	MSE_{JS}^{opt}	MSE_{JS}^{naive}
0.15	75	4.04771	4.04763	4.05264	3.78728	4.02700	2.07361	4.16989	4.67809	2.11943
0.15	100	4.02828	4.02779	4.02376	3.94387	4.19095	2.31801	3.95268	4.45594	2.47710
0.15	125	3.96406	3.96396	3.96742	4.10485	4.34675	2.28799	3.98844	4.49432	2.14095
0.10	75	4.06940	4.06918	4.06811	4.04519	4.30015	2.18736	3.97274	4.48387	2.09272
0.10	100	3.96521	3.96480	3.97802	3.73005	3.96094	2.12377	4.13767	4.62959	2.03202
0.10	125	3.98717	3.98694	3.98466	4.20927	4.45633	1.82384	3.83935	4.35079	1.86186
0.05	75	4.07604	4.07562	4.06765	4.09415	4.33782	2.59909	3.95166	4.45893	2.08988
0.05	100	3.99519	3.99525	3.99012	3.88239	4.12479	2.16759	4.00928	4.50816	1.95643
0.05	125	4.04821	4.04813	4.05020	3.71593	3.95638	2.15416	4.14963	4.66387	2.42300

Figures 11 - 13 show the relationship between integrated MSE and values of c, where the blue line indicates integrated MSE of local regression estimator. Notice that in Figure 11, when $\alpha = 0.15$, the curves are below the blue lines at negative values of c. And the optimum point which minimizes the integrated MSE of JS estimator moved towards the positive side as α decreases. We can also observe that in all three figures, as $\alpha \downarrow$ or thelocal polynomial degree $J \uparrow$, the optimum point shifts towards $p-2=2$. Recall that when there is no bias, James and Stein (1961) proposed that the optimum point is $p-2$. And as $\alpha \downarrow$ or $J \uparrow$, bias of local regression estimator goes to 0. And therefore, the optimum c moves towards $p-2$, which agrees with the simulation results in Chpater 4.

Remark. Note that in Figure 11, when $\alpha = 0.15$, James-Stein estimator with a negative c has smaller integrated MSE, which can be explained by Figure 14. Suppose the red dotted line represents true mean responses. If local regression with a big neighborhood parameter (for example, $\alpha = 0.15$) is applied to estimate the mean response function, we will get a curve like the blue solid line in the figure since the bias is high. Then a JS estimator with a negative c pulls local regression estimator away from zero, as shown in black dashed, which provides a better estimate.

In this nanoparticle example, true bias (for mean responses) and variance are used in finding the upper bound of expected loss difference. However, a prectitioner does not know the true values, and he/she needs to estimate the bias and variance, and set B, $\sigma^2_{(1)}$, $\sigma^2_{(p)}$ accordingly. A potential problem of doing so is that the practitioner may not be able to find a sharp upper bound. Recall that Figure 10 showed that a c optimizing the upper bound may not optimize the expected loss difference. As we have seen in estimating the frist and second order derivatives, where estimated bias was used to set B, since the estimation is not good enough, we were unable to find a c such that Steinization helps.

The end goal of the nanoparticle example is to improve ability of classifying nanoparticles based on the estimated scattering profiles and derivatives through Steinization. Figure 15 shows an example of characterization. Suppose the scattering profiles of nanoparticles with diameters 5nm (red dashed) and 10nm (blue dotted) are known. The estimated scattering profiles of nanoparticles with unknown diameter is depicted in black solid line. The

Figure 11: Integrated MSE: mean response

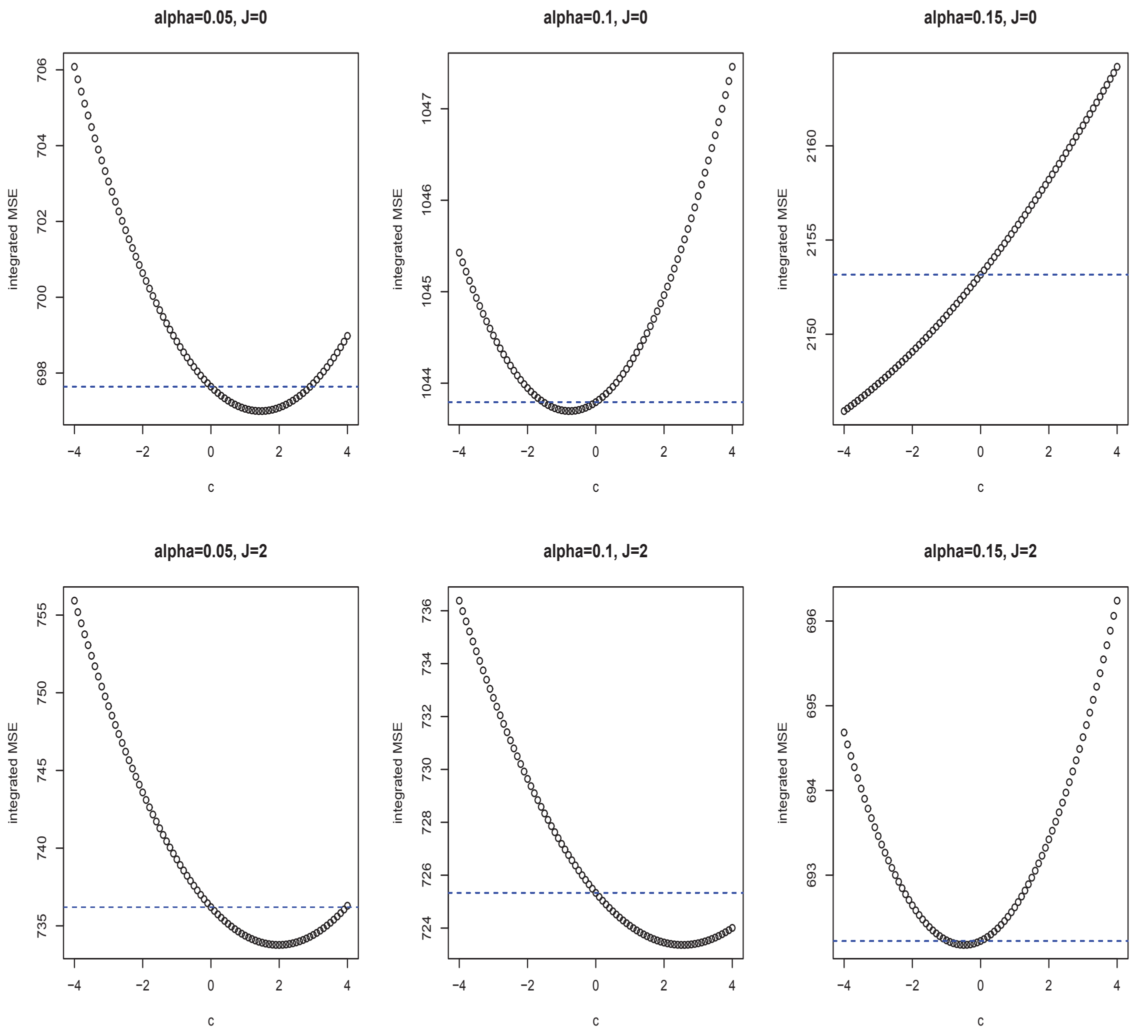


Figure 12: Integrated MSE: first order derivative

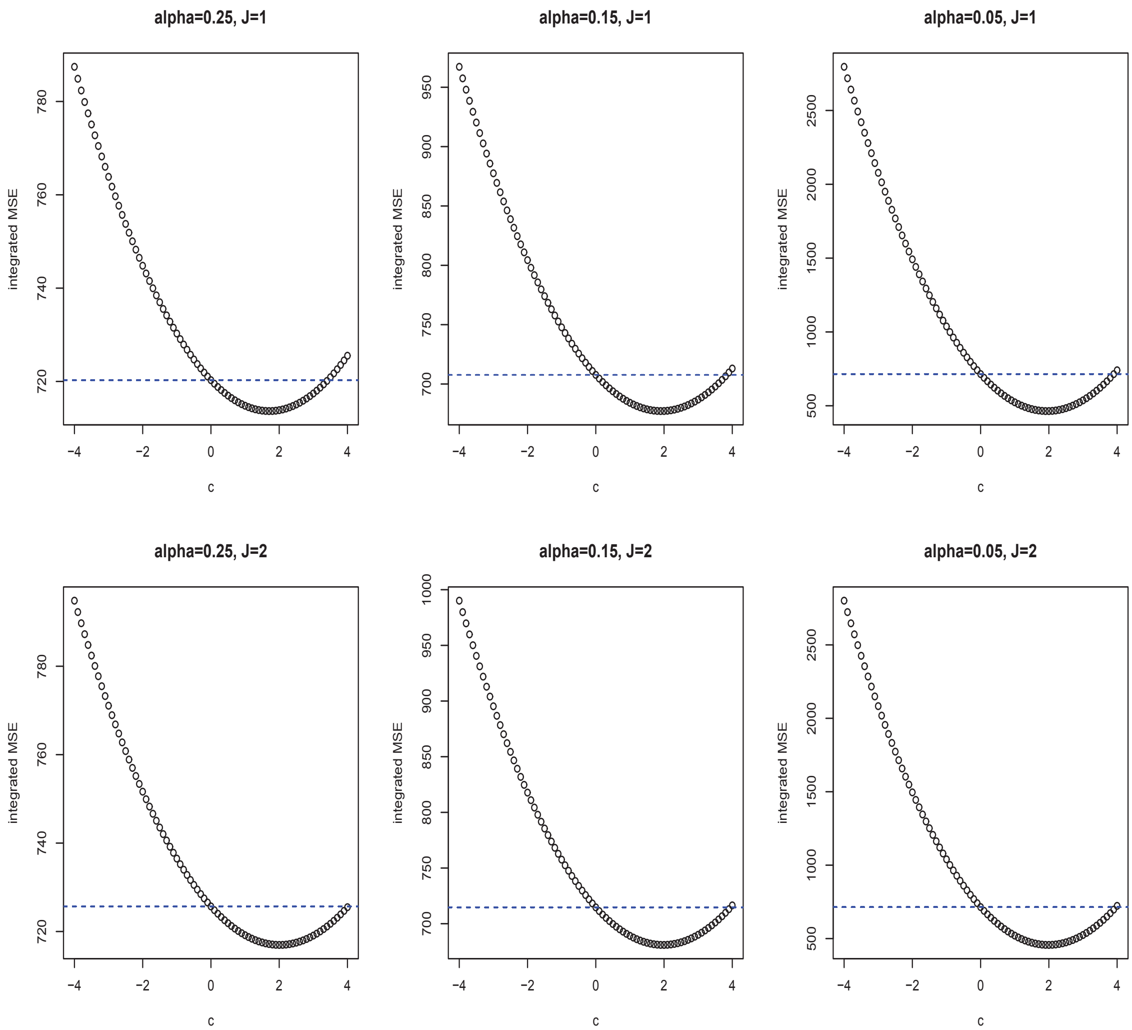


Figure 13: Integrated MSE: second order derivative

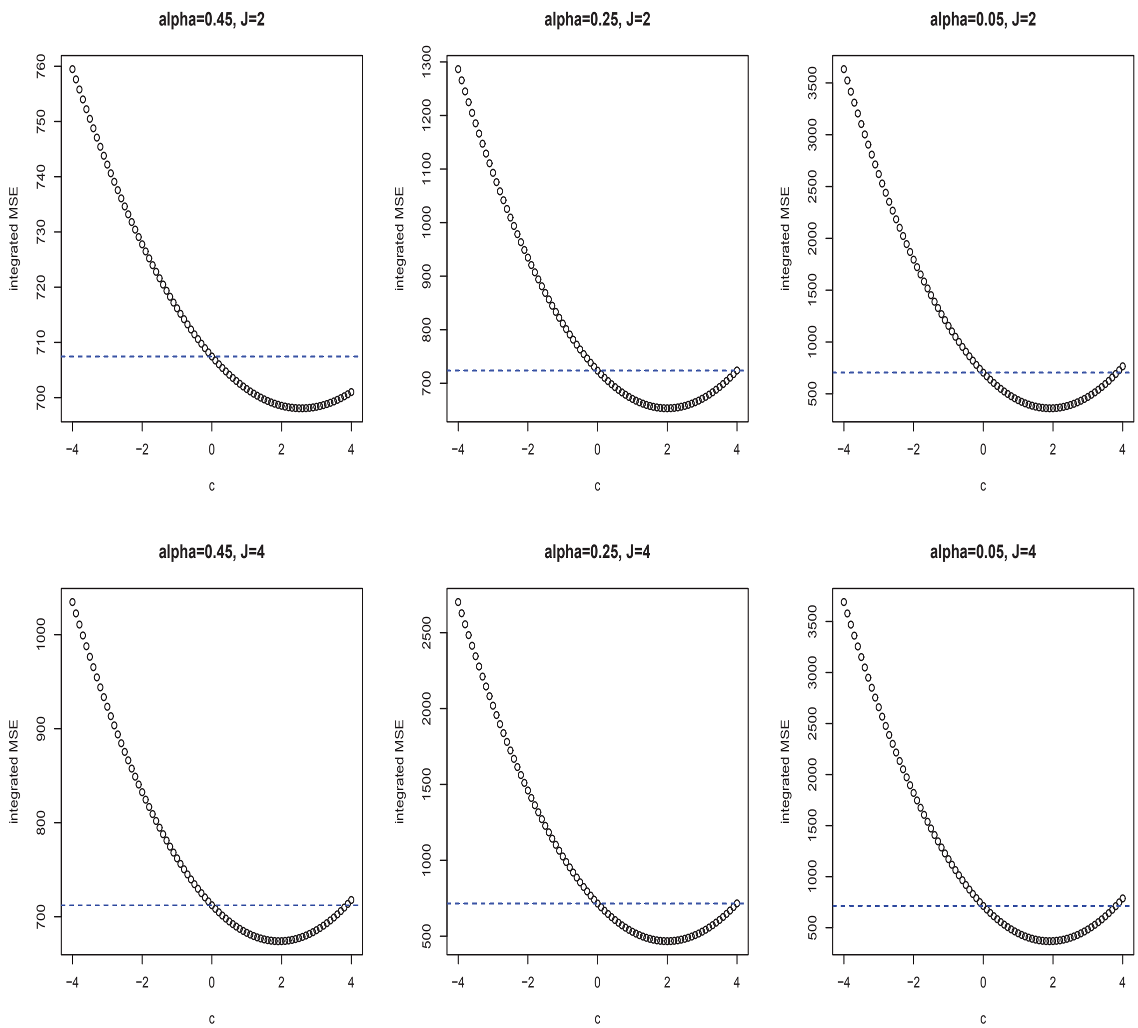
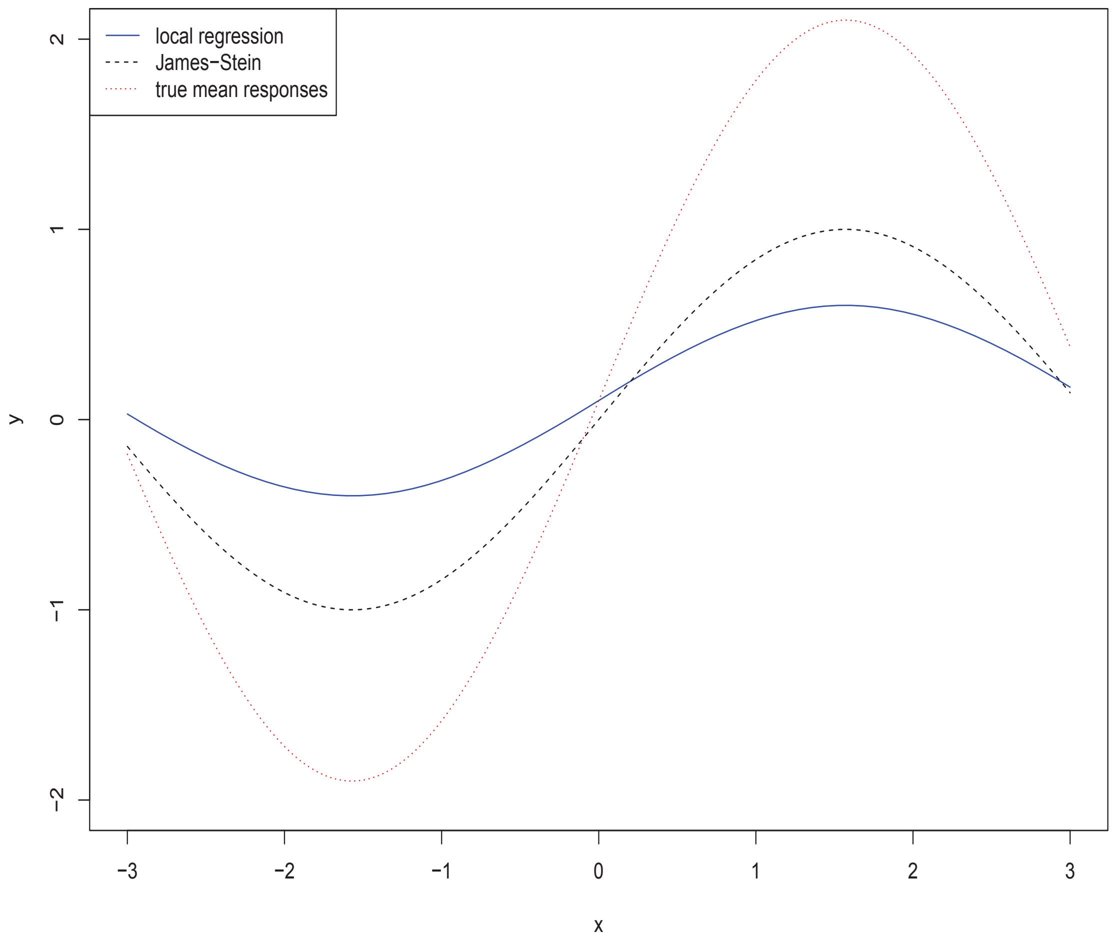
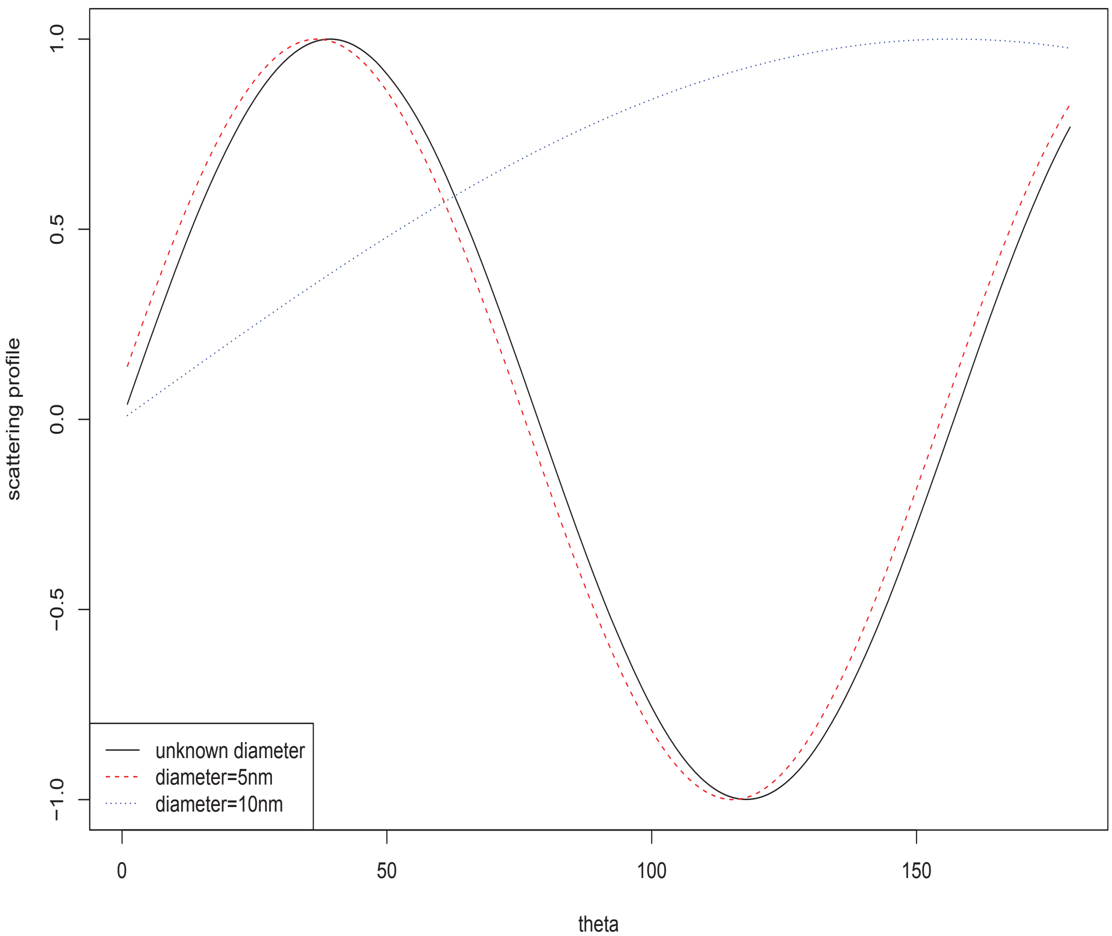


Figure 14: JS estimator with negative c



question is: is the unknown diameter close to 5nm or 10nm? Since the estimated curve is really close to the red dashed one, we can conclude that the unknown diameter is close to 5nm rather than 10nm. To address the question of whether Steinization improves characterization, a simulation study might be conducted. We generate noisy data at different noise levels, apply Steinization and other regression methods, for example, local regression, compund estimation, and estimate the configurations such as diameter. After repeating the steps many times, we can find out if Steinization actually improves ability to characterize nanoparticles.

Figure 15: Characterization



References

[1] Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, **41**, 642-645.

[2] Charnigo, R., Francoeur, M., Mengüç, M. P., Brock, A., Leichter, M., and Srinivasan, C. (2007). Derivatives of scattering profiles: tools for nanoparticle characterization. *Journal of the Optical Society of America A*, **24**: 2578–2589.

[3] Charnigo, R., Francoeur, M., Kenkel, P., Mengüç, M. P., Hall, B., Srinivasan, C. (2011). Estimating quantitative features of nanoparticles using multiple derivatives of scattering profiles. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **112**: 1369–1382.

[4] Charnigo, R., Hall, B. and Srinivasan, C. (2011). A generalized C_p criterion for derivative estimation. *Technometrics*, **53**, 238-253.

[5] Charnigo, R. and Srinivasan, C. (2011). Self-consistent estimation of mean response functions and their derivatives. *Canadian Journal of Statistics*, **39**, 280-299.

[6] Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.

[7] Cleveland, W. S. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83**, 596-610.

[8] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**, 377-403.

[9] Dykstra, R. L. (1970). Establishing the positive definiteness of the sample covariance matrix. *Annals of Mathematical Statistics*, **41**, 2153-2154.

[10] Efron, B. and Morric, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators – Part I: The Bayes case. *Journal of the American Statistical Association*, **66**, 807-815.

[11] Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators – Part II: The empirical Bayes case. *Journal of the American Statistical Association*, **67**, 130-139.

[12] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117-130.

[13] Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, **70**, 311-319.

[14] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics*, **4**, 11-21.

[15] Hastie, T. and Loader, C. (1993). Automatic kernel carpentry. *Statistical Science*, **8**, 120-129.

[16] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 361-379.

[17] Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters*, **9**, 129-132.

[18] Katkovnik, V. Y. (1979). Linear and nonlinear methods of nonparametric regression analysis. *Soviet Automatic Control*, **5**, 35-46.

[19] Lin, P. and Tsai, H. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Annals of Statistics*, **1**, 142-145.

[20] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., and Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, **56**: 1015–1022.

[21] Loader, C. (1999). Local regression and likelihood. *Springer, New York*.

[22] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**, 141-142.

[23] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.

[24] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference*, **35**, 65-75.

[25] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, **47**, 1-52.

[26] Stein, C. (1956). Inadmissibilty of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 197-206.

[27] Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, **5**, 595-620.

[28] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics*, **8**, 1348-1360.

[29] Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Transactions on Biomedical Engineering*, **57**: 884–893.

[30] Wahba, G. and Wold, S. (1975). A completely automatic French curve: Fiitting spline functions by cross-validation. *Communications in Statistics*, **4**, 1-17.

[31] Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A*, **26**, 359-372.

Vita

Author's Name - Limin Feng

Education

Master in Statistics
University of Kentucky
August 2008 - May 2010

Master in Statistics
University of Southern Maine
Septermber 2006 - July 2008

Bachelor in Information Security
University of Science and Technology of China
September 2002 - July 2006

Research Experience

University of Kentucky
Lexington, KY
August 2009 - present
Research Assistant

University of Kentucky
Lexington, KY
August 2008 - May 2009
Teaching Assistant

University of Southern Maine
Portland, ME
September 2006 - July 2008
Research Assistant