Weiping Ma, Yang Feng*, Kani Chen and Zhiliang Ying

Functional and Parametric Estimation in a Semi- and Nonparametric Model with Application to Mass-Spectrometry Data

Abstract: Motivated by modeling and analysis of mass-spectrometry data, a semi- and nonparametric model is proposed that consists of linear parametric components for individual location and scale and a nonparametric regression function for the common shape. A multi-step approach is developed that simultaneously estimates the parametric components and the nonparametric function. Under certain regularity conditions, it is shown that the resulting estimators is consistent and asymptotic normal for the parametric part and achieve the optimal rate of convergence for the nonparametric part when the bandwidth is suitably chosen. Simulation results are presented to demonstrate the effectiveness and finite-sample performance of the method. The method is also applied to a SELDI-TOF mass spectrometry data set from a study of liver cancer patients.

Keywords: local linear regression, bandwidth selection, nonparametric estimation

1 Introduction

We are concerned with the following semi- and nonparametric regression model

$$y_{it} = \alpha_i + \beta_i m(x_{it}) + \sigma_i(x_{it})\varepsilon_{it}, \tag{1}$$

where y_{it} is the observed response from i-th individual $(i = 1, \ldots, n)$ at time t for $(t = 1, \ldots, T)$, x_{it} is the corresponding explanatory variable, α_i and β_i are individual-specific location and scale parameters and $m(\cdot)$ is a baseline intensity function. Here, $\mathrm{E}(\varepsilon_{it}) = 0$, $\mathrm{Var}(\varepsilon_{it}) = 1$, and ε_{it} and x_{it} are independent. Of interest here is the simultaneous estimation of α_i, β_i and $m(\cdot)$. We shall assume throughout the paper that ε_{it} $(i = 1, \ldots, n; t = 1, \ldots, T)$ are independent and identically distributed (i.i.d.) with an unknown distribution function, though most results only require that the errors be independent with zero mean.

Model (1) is motivated by analyzing the data generated from mass spectrometer (MS), which is a powerful tool for the separation and large-scale detection of proteins present in a complex biological mixture. Figure 1 is an illustration of MS spectra, which can reveal proteomic patterns or features that might be related to specific characteristic of biological samples. They can also be used for prognosis and for monitoring disease progression, evaluating treatment or suggesting intervention. Two popular mass spectrometers are SELDI-TOF (surface enhanced laser desorption/ionization time-of-fight) and MALDI-TOF (matrix assisted laser desorption and ionization time-of-flight). The abundance of the protein fragments from a biological sample (such as serum) and their time of flight through a tunnel under certain electrical pressure can be measured by this procedure. The y-axis of a spectrum is the intensity (relative abundance) of protein/peptide, and the x-axis is the mass-to-charge ratio (m/z value) which can be calculated using time, length of flight, and the voltage applied. It is known that the SELDI intensity measures have errors up to 50% and that the m/z may shift its value by up to 0.1–0.2% [1]. Generally speaking, many pre-processing steps need to be done before the MS data can be analyzed. Some of the most important steps are noise filtering, baseline correction, alignment, normalization, etc. See, e.g., Guilhaus [2], Banks and Petricoin [3],

*Corresponding author: Yang Feng, Department of Statistics, Columbia University, New York, NY, USA,
E-mail: yang.feng@columbia.edu
Weiping Ma, Icahn School of Medicine at Mount Sinai, New York, NY, USA
Kani Chen, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong
Zhiliang Ying, Department of Statistics, Columbia University, New York, NY, USA
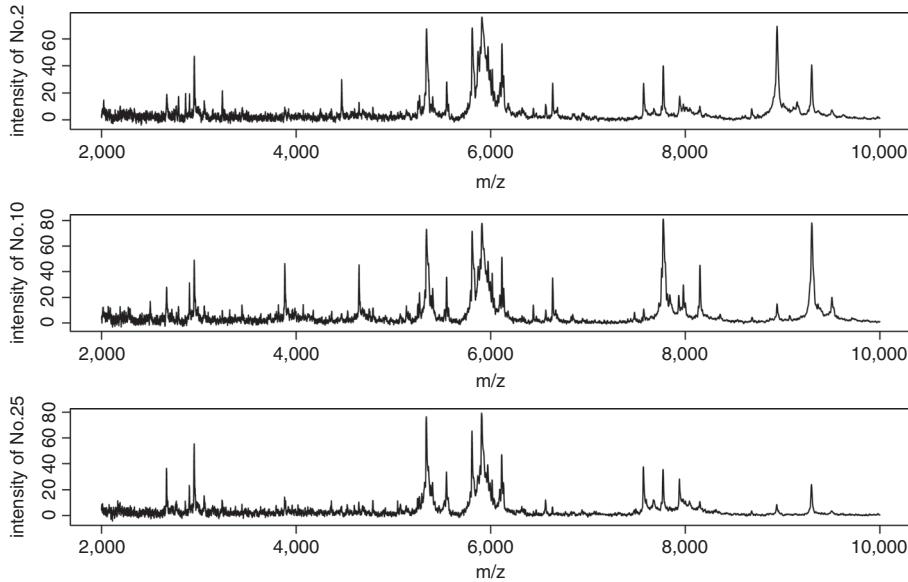
Figure 1: Illustration of MS spectra.

Baggerly et al. [4], Baggerly et al. [5], Diamandis [6], Feng et al. [7]. We refer readers to Roy et al. [8] for an extensive review about the recent advances in mass-spectrometry data analysis. Here, we assume all the pre-processing steps have already been taken.

In model (1), $m(\cdot)$ represents the common shape for all individuals while α_i and β_i represents the location and scale parameters for the i-th individual, respectively. Because $m(\cdot)$ is unspecified, model (1) may be viewed as a semiparametric model. However, it differs from the usual semi-parametric models in that for model (1), both the parametric and nonparametric components are of primary interest, while in a typical semiparametric setting, the nonparametric component is often viewed as a nuisance parameter. Model (1) contains many commonly encountered regression models as special cases. If all the parametric coefficients α_i and β_i are known, model (1) reduces to the classical nonparametric regression. On the other hand, if the function $m(\cdot)$ is known, then it reduces to the classical linear regression model with each subject having its own regression line. For the present case of α_i, β_i and function $m(\cdot)$ being unknown, the parameters are identifiable only up to a common location-scale change. Thus we assume, without loss of generality, that $\alpha_1 = 0$ and $\beta_1 = 1$. It is also clear that for α_i, β_i and $m(\cdot)$ to be consistently estimable, we need to require that both n and T go to ∞.

There is an extensive literature on semiparametric and nonparametric regression. For semiparametric regression, Begun et al. [9] derived semiparametric information bound while Robinson [10] developed a general approach to constructing $\sqrt{n}$-consistent estimation for the parametric component. We refer to Bickel et al. [11] and Ruppert et al. [12] for detailed discussions on the subject. For nonparametric regression, kernel and local polynomial smoothing methods are commonly used [13–16]. In particular, local polynomial smoothing has many attractive properties including the automatic boundary correction. We refer to Fan and Gijbels [17] and Hardle et al. [18] for comprehensive treatment of the subject.

The existing methods for dealing with nonparametric and semiparametric problems are not directly applicable to model (1). This is due to the mixing of the "finite dimensional" parameters and the nonparametric component. A natural way to handle such a situation is to de-link the two aspects of the estimation through a two-step approach. In this paper, we propose an efficient iterative procedure, alternating between estimation of the parametric component and the nonparametric component. We show that the proposed approach leads to consistent estimators for both the finite-dimensional parameter and the nonparametric function. We also establish the asymptotic normality for the parametric estimators, and the convergence rate for the nonparametric estimate that is then used to derive the optimal bandwidth selection.

2 Main results

In this section, we develop a multi-step approach to estimating both the finite-dimensional parameters α_i and β_i and the nonparametric baseline intensity $m(\cdot)$. Our approach is an iterative procedure which alternates between estimation of α_i and β_i and that of $m(\cdot)$. We show that under reasonable conditions, the estimation for the parametric component is consistent and asymptotically normal when the bandwidth selection are done appropriately. The estimation of the nonparametric component can also attain the optimal rate of convergence.

2.1 A multi-step estimation method

Recall that if α_i and β_i were known, the problem would reduce to the standard nonparametric regression setting; on the other hand, if $m(\cdot)$ were known, it would reduce to the simple linear regression for each i. For the nonparametric regression, we can apply the local linear regression with the weights $K_h(\cdot) = K(\cdot/h)/h$ for suitably chosen kernel function K and bandwidth h. For the simple linear regression, the least squares estimation may be applied.

To ensure identifiability, we shall set $\alpha_1 = 0$ and $\beta_1 = 1$. Thus, for $i = 1$, (1) becomes a standard nonparametric regression problem, from which an initial estimator of $m(\cdot)$ can be derived. Replacing $m(\cdot)$ in (1) by the initial estimator, we can apply the least squares method to get estimators of α_i, β_i for $i \geq 2$, which, together with $\alpha_1 = 0$ and $\beta_1 = 1$ and local polynomial smoothing, can then be used to get an updated estimator of $m(\cdot)$. This iterative estimation procedure is described as follows.

(a) Set $\alpha_1 = 0$ and $\beta_1 = 1$, so that $y_{1t} = m(x_{1t}) + \sigma_1(x_{1t})\varepsilon_{1t}$, $t = 1, \ldots, T$. Apply local linear regression to (x_{1t}, y_{1t}), $t = 1, \ldots, T$, to get initial estimator of $m(\cdot)$

$$\tilde{m}(x) = \frac{\sum_{t=1}^{T} \omega_{1t}(x) y_{1t}}{\sum_{t=1}^{T} \omega_{1t}(x)}, \tag{2}$$

where $\omega_{1t}(x) = K_h(x_{1t} - x)(S_{T,2} - (x_{1t} - x)S_{T,1})$ and

$$S_{T,k} = \sum_{t=1}^{T} K_h(x_{1t} - x)(x_{1t} - x)^k, \text{ for } k = 1, 2. \tag{3}$$

(b) With $m(\cdot)$ being replaced by $\tilde{m}(\cdot)$ as the true function, α_i, β_i, $i = 2, \ldots, n$ can be estimated by the least squares method, i.e.

$$\hat{\beta}_i = \frac{\sum_{t=1}^{T} [\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i\cdot})] y_{it}}{\sum_{t=1}^{T} [\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i\cdot})]^2}, \tag{4}$$

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \hat{\beta}_i \bar{\tilde{m}}(x_{i\cdot}) = \bar{y}_{i\cdot} - \frac{\sum_{t=1}^{T} (\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i\cdot})) y_{it}}{\sum_{t=1}^{T} (\tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i\cdot}))^2} \bar{\tilde{m}}(x_{i\cdot}), \tag{5}$$

where

$$\bar{y}_{i\cdot} = \frac{1}{T} \sum_{t=1}^{T} y_{it}, \text{ and } \bar{\tilde{m}}(x_{i\cdot}) = \frac{1}{T} \sum_{t=1}^{T} \tilde{m}(x_{it}).$$

(c) With the estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$, we can update the estimation of $m(\cdot)$ viewing $\hat{\alpha}_i$ and $\hat{\beta}_i$ as true values. Specifically, we apply the local linear regression with the same kernel function $K(\cdot)$ to get an updated estimator of $m(\cdot)$,

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^{\star}(x) y_{it}^{\star}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^{\star}(x)}, \tag{6}$$

where $y_{it}^{\star} = (y_{it} - \hat{\alpha}_i)/\hat{\beta}_i$,

$$\omega_{it}^{\star}(x) = \hat{\beta}_i^2 K_{h^{\star}}(x_{it} - x) \left[\sum_{i=1}^{n} \hat{\beta}_i^2 S_{T,2}^{\star(i)} - (x_{it} - x) \sum_{i=1}^{n} \hat{\beta}_i^2 S_{T,1}^{\star(i)} \right] \tag{7}$$

and

$$S_{T,k}^{\star(i)} = \sum_{t=1}^{T} K_{h^{\star}}(x_{it} - x)(x_{it} - x)^k, \; for \; k = 1, 2. \tag{8}$$

Note that the bandwidth for this step, $h^{\star}$, may be chosen differently from h in order to achieve better convergence rate. The optimal choices for h and $h^{\star}$ will become clear in the next subsection where large sample properties are studied.

(d) Given a pre-specified tolerance level $\eta > 0$, repeat steps (b) and (c) until the parametric and the nonparametric estimators for iterations l and $l+1$ meet the following convergence criteria.

$$\Sigma_{i=1}^{n} ||\hat{\alpha}_i^{(l+1)} - \hat{\alpha}_i^{(l)}||^2 + \Sigma_{i=1}^{n} ||\hat{\beta}_i^{(l+1)} - \hat{\beta}_i^{(l)}||^2 + \int (\hat{m}(x)^{(l+1)} - \hat{m}(x)^{(l)}) dx > \eta.$$

Our limited numerical experiences indicate that the final estimator is not sensitive to the initial estimate. However, as a safe guard, we may start the algorithm with different initial estimates by choosing different individuals as the baseline intensity. In step (c), the $\hat{\beta}_i$ is in the denominator, which, when close to 0, may cause instability. Thus, in practice, we can add a small constant to the denominator to make it stable, though we have not encountered this problem.

The iterative process often converges very quickly. In addition, our asymptotic analysis in the next subsection shows that no iteration is needed to reach the optimal convergence rate for the estimate of both parametric and nonparametric components when the bandwidths of each step are properly chosen. Therefore, we may stop after step (c) to save computation time for large problems.

2.2 Large sample properties

In this section, we study the large sample properties of the estimates for $m(\cdot)$, α_i and β_i. By large sample, we mean that both n and T are large. However, the size of n and that of T can be different. Indeed, for MS data, T is typically much larger than n. The optimal bandwidth selection in the nonparametric estimation will be determined by asymptotic expansions to achieve optimal rate of convergence. We will also investigate whether or not the accuracy of $\hat{\alpha}_i$ and $\hat{\beta}_i$ may affect the rate of convergence for the estimation of $m(\cdot)$.

The following conditions will be needed to establish the asymptotic theory.

C1. The baseline intensity $m(\cdot)$ is continuous and has a bounded second order derivative.

C2. There exist constants $\alpha > 0$ and $\delta > 0$, such that the marginal density $f(\cdot)$ of x_{it} satisfies $f(x) > \delta$, and $|f(x) - f(y)| \le c|x - y|^{\alpha}$ for any $\times$ and y in the support of $f(\cdot)$.

C3. The conditional variance $\sigma_i^2(x) = Var(y_{it}|x_{it} = x)$ is bounded and continuous in x, where $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

C4. The kernel $K(\cdot)$ is a symmetric probability density function with bounded support. Hence $K(\cdot)$ has the properties: $\int_{-\infty}^{\infty} K(u) du = 1$, $\int_{-\infty}^{\infty} u K(u) du = 0$, $\int_{-\infty}^{\infty} u^2 K(u) du \ne 0$ and bounded. Without loss of generality, we could further assume the support of $K(\cdot)$ lies in the interval $[-1, +1]$.

Condition C1 is a standard condition for nonparametric estimation. Condition C2 requires that the density of x_{it} is bounded away from 0, which may be a strong assumption in general but reasonable for mass

spectrometry data as x_{it} are approximately uniformly distributed on the support as shown in Figure 1. Condition C3 allows for heteroscedasticity while restricting the variances to be bounded. Condition C4 is a standard condition for kernel function used in the local linear regression.

The moments of K and K^2 are denoted respectively by $\mu_l = \int_{-\infty}^{\infty} u^l K(u) du$ and $v_l = \int_{-\infty}^{\infty} u^l K^2(u) du$ for $l \geq 0$.

Lemma 1 *Suppose that Conditions C1–C4 are satisfied. Then for $\tilde{m}(\cdot)$ defined by (2), we have, as $h \to 0$ and $Th \to \infty$,*

$$\tilde{m}(x) = m(x) + \frac{1}{2} m''(x) \mu_2 h^2 + o(h^2) + U_1(x), \tag{9}$$

where $U_1(x) = (\sum_{t=1}^{T} \omega_{1s}(x) \sigma_1(x_{1s}) \varepsilon_{1s}) / (\sum_{t=1}^{T} \omega_{1s}(x))$.

Lemma 1 allows us to derive the asymptotic bias, variance and mean squared error for the estimator $\tilde{m}(\cdot)$. This is summarized in the following corollary.

Corollary 1 *Let $\mathbf{X}$ denote all the observed covariates $\{x_{it}, i = 1, \ldots, n, t = 1, \ldots, T\}$. Under Conditions C1–C4, the bias, variance and mean squared error of $\tilde{m}(x)$ conditional on $\mathbf{X}$ have the following expressions.*

$$E(\tilde{m}(x) - m(x)|\mathbf{X}) = \frac{1}{2} m''(x) \mu_2 h^2 + o(h^2),$$

$$Var(\tilde{m}(x)|\mathbf{X}) = \frac{1}{Th} [f(x)]^{-1} \sigma_1^2(x) v_0 + o(\frac{1}{Th}),$$

$$E[\{\tilde{m}(x) - m(x)\}^2 |\mathbf{X}] = \frac{1}{4} (m''(x) \mu_2)^2 h^4 + \frac{1}{Th} [f(x)]^{-1} \sigma_1^2(x) v_0 + o(h^4 + \frac{1}{Th}).$$

It is clear from the above expansions that in order to minimize the mean squared error of $\tilde{m}(x)$, the bandwidth h should be chosen to be of order $T^{-1/5}$. However, we will show later that this is not necessarily optimal for our final estimator $\hat{m}(\cdot)$.

For estimation of scale parameters β_i, we can apply Lemma 1 together with the Taylor expansion to derive asymptotic bias and variance. In particular, we have the following theorem.

Theorem 1 *Suppose that Conditions C1–C4 are satisfied and that $h \to 0$ is chosen so that $Th \to \infty$. Then the following expansions hold for $i \geq 2$.*

$$E(\hat{\beta}_i - \beta_i|\mathbf{X}) = -\beta_i(h^2 P_i + \frac{1}{Th} Q_i) + o(h^2 + \frac{1}{Th}), \tag{10}$$

$$Var(\hat{\beta}_i|\mathbf{X}) = \frac{\sum_{t=1}^{T} W_{it}^2 \sigma_i^2(x_{it})}{(\sum_{t=1}^{T} W_{it}^2)^2} + \beta_i^2 \frac{\sum_{t=1}^{T} W_{1t}^2 \sigma_i^2(x_{1t})}{(\sum_{t=1}^{T} W_{it}^2)^2} + o(\frac{1}{T}), \tag{11}$$

where $W_{it} = m(x_{it}) - \bar{m}(x_{i \cdot})$, $\bar{m}(x_{i \cdot}) = T^{-1} \sum_{t=1}^{T} m(x_{it})$,

$$P_i = \frac{\mu_2}{2} \frac{\sum_{t=1}^{T} W_{it} m''(x_{it})}{\sum_{t=1}^{T} W_{it}^2}, \quad Q_i = \frac{v_0 \sum_{t=1}^{T} f^{-1}(x_{1t}) \sigma_i^2(x_{1t})}{\sum_{t=1}^{T} W_{it}^2}.$$

Remark 1 *The asymptotic bias and variance of parameter estimator $\hat{\alpha}_i$ can be similarly derived. In fact, they can be inferred from the bias and variance of $\hat{\beta}_i$ through its linear relationship with $\hat{\beta}_i$, thus having the same order as those of $\hat{\beta}_i$ in (10) and (11).*

Remark 2 *The bias of $\hat{\beta}_i$ is of the order $h^2 + (Th)^{-1}$ and the variance is of the order T^{-1}. To obtain the $\sqrt{T}$-consistency for $\hat{\beta}_i$, i.e. $\sqrt{T}(\hat{\beta}_i - \beta_i) = O_p(1)$, the order of bias should be $O(T^{-\frac{1}{2}})$. This is achieved by choosing the order of h between $O(T^{-\frac{1}{2}})$ and $O(T^{-\frac{1}{4}})$.*

From the asymptotic expansion for the mean and variance of the initial functional estimator $\tilde{m}(\cdot)$ and parameter estimator $\hat{\beta}_i$, we can obtain the asymptotic expansions for the bias and variance of the subsequent estimator of the baseline intensity, $\hat{m}(\cdot)$.

Theorem 2 *Suppose that Conditions C1–C4 are satisfied. Suppose also that h for $\tilde{m}(\cdot)$ and $h^{\star}$ for $\hat{m}(\cdot)$ are chosen so that $h \to 0$, $h^{\star} \to 0$, $Th \to \infty$, and $nTh^{\star} \to \infty$. Then the following expansions hold:*

$$\mathrm{E}(\hat{m}(x) - m(x)|\mathbf{X}) = \left(\sum_{i=2}^{n} \frac{\beta_i^2 (h^2 P_i + (Th)^{-1} Q_i)}{\sum_{i=1}^{n} \beta_i^2} \right) m(x) - \sum_{i=2}^{n} \frac{\beta_i^2 (h^2 R_i + (Th)^{-1} m(x_{i\cdot}) Q_i)}{\sum_{i=1}^{n} \beta_i^2}$$

$$+ \frac{m''(x)\mu_2}{2} h^{\star 2} + o\left(h^2 + \frac{1}{Th} + h^{\star 2} \right),$$

$$\mathrm{Var}(\hat{m}(x)|\mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n} \beta_i^2 \right)^2} \sum_{t=1}^{T} \left(\sum_{i=2}^{n} \beta_i^2 \left[\frac{1}{T} + Z_{it} \right] \right)^2 \sigma_1^2(x_{1t})$$

$$+ \frac{\nu_0 \sum_{i=2}^{n} \beta_i^2 f^{-1}(x) \sigma_i^2(x_{it})}{Th^{\star} \left(\sum_{i=1}^{n} \beta_i^2 \right)^2} + o\left(\frac{1}{T} + \frac{1}{nTh^{\star}} \right).$$

where P_i, Q_i, W_{it} are the same as those in Theorem 1, and $R_i = \bar{m}(x_{i\cdot}) P_i - 2^{-1} \mu_2 \bar{m}''(x_{i\cdot})$, $Z_{it} = (\sum_{s=1}^{T} W_{is}^2)^{-1} (m(x) - \bar{m}(x_{i\cdot})) W_{1t}$.

In the ideal case when the location-scale parameters are known, the bias and variance of the local linear estimator of baseline intensity $m(\cdot)$ should be of the order $O(h^{\star 2})$ and $O(\frac{1}{nTh^{\star}})$. And the optimal bandwidth in this ideal case should be of order $(nT)^{-\frac{1}{5}}$. Therefore the bias and variance of the nonparametric estimator are $O((nT)^{-\frac{2}{5}})$ and $O((nT)^{-\frac{4}{5}})$, respectively. In addition, the mean squared error is of order $O((nT)^{-\frac{4}{5}})$. Interestingly, by choosing the bandwidths h and $h^{\star}$ separately, we can achieve this optimal rate of convergence for the baseline intensity estimator $\hat{m}(\cdot)$ through the proposed multi-step estimation procedure when the orders of n and T satisfy certain requirement. Notice that the parametric components will have the optimal $\sqrt{T}$ convergence rate simultaneously. The conclusions are summarized in the following theorem.

Theorem 3 *Suppose that Conditions C1–C4 are satisfied. The optimal parametric convergence rate of location-scale estimators can be attained by setting h to be of order $T^{-\frac{1}{3}}$; the optimal nonparametric convergence rate of the baseline intensity estimator $\hat{m}(\cdot)$ can be attained by setting $h^{\star}$ to be of order $(nT)^{-\frac{1}{5}}$ and h of order $T^{-\frac{1}{3}}$, when $T \to \infty$, $n \to \infty$, and $n = O(T^{\frac{1}{4}})$.*

Remark 3 *In Theorem 3, if $T \to \infty$ while n is fixed, the optimal nonparametric convergence rate could still be reached by setting $h^{\star} = O(T^{-\frac{1}{5}})$ and h to between $O(T^{-\frac{1}{5}})$ and $O(T^{-\frac{3}{5}})$. By Remark 2, we still have the $\sqrt{T}$-consistency for parametric estimation for the fixed-n situation.*

Remark 4 *From Theorem 3, if the requirement $n = O(T^{\frac{1}{4}})$ is not satisfied, then the nonparametric estimator $\hat{m}(\cdot)$ will not achieve the optimal rate of convergence at any choice of the bandwidths. However, the choice of h and $h^{\star}$ is optimal even if $n = O(T^{\frac{1}{4}})$ does not hold.*

Theorem 4 *Suppose that Conditions C1–C4 are satisfied. In addition, assume $\mathrm{E}[m^2(x_{it})(\sigma_i^2(x_{it}) + 1)] < \infty$ and $\mathrm{E}[m^2(x_{it})] > 0$ for all $i = 1, ..., n$ and $t = 1, \ldots, T$. If we restrict the order of h to lie between $T^{-\frac{1}{2}}$ and $T^{-\frac{1}{4}}$, $\hat{\beta}_i$ is asymptotic normal:*

$$\sqrt{T}(\hat{\beta}_i - \beta) \to N(0, \sigma_i^{\star 2}), \tag{12}$$

where

$$\sigma_i^{\star 2} = \lim_{T \to \infty} \left(\frac{T^{-1} \sum_{t=1}^{T} W_{it}^2 \sigma_i^2(x_{it})}{\left(T^{-1} \sum_{t=1}^{T} W_{it}^2\right)^2} + \beta_i^2 \frac{T^{-1} \sum_{t=1}^{T} W_{1t}^2 \sigma_1^2(x_{1t})}{\left(T^{-1} \sum_{t=1}^{T} W_{it}^2\right)^2} \right).$$

Here, if we assume $\sigma_i^2(\cdot)$ to be a constant for each subject i, then its value can be consistently estimated by the plug-in estimator

$$\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} (y_{it} - \hat{\alpha}_i - \hat{\beta}_i \hat{m}(x_{it}))^2, \tag{13}$$

where $\hat{\alpha}_1 = 0$ and $\hat{\beta}_1 = 1$.

From (12), the asymptotic variance of $\hat{\beta}_i$ is of order $O(T^{-1})$, provided that the order of the bandwidth h is properly chosen. Since the asymptotic expansion for $\hat{\beta}_i$ does not involves the choice of $h^\star$, the specific choice of different $h^\star$ will not affect the order of the asymptotic variance of $\hat{\beta}_i$.

2.3 Bandwidth selection

In Section 2.2, we studied how the choice of bandwidths h and $h^\star$ may affect the asymptotic properties of the estimators. However, in practice, we need a data-driven approach to choosing the bandwidths. Our suggestion on this is to use a *K*-fold cross-validation bandwidth selection rule.

First, we divide the n individuals into K groups $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_K$ randomly. Here, $\mathbf{Z}_k$ is the k-th test set, and the k-th training set is $\mathbf{Z}_{-k} = \{\{1, \ldots, n\} \backslash \mathbf{Z}_k\}$. We estimate the baseline curve $m(\cdot)$ using the observations in the training set $\mathbf{Z}_{-k}$ and denote the estimator as $\hat{m}(\mathbf{Z}_{-k}, h, h^\star)$, where h and $h^\star$ are the bandwidths of the two nonparametric regression steps for $\tilde{m}(\cdot)$ and $\hat{m}(\cdot)$, respectively. Recall that at the beginning of the multi-step estimation procedure, we fix the first observation as the baseline to solve the identifiability issue. In the case of cross-validation, for each split, the baseline will corresponds to the first observation inside $\mathbf{Z}_{-k}$, which is different for different k. We circumvent the problem of comparing different baseline estimates by using them to predict the test data in $\mathbf{Z}_k$, i.e., after obtaining the estimator of baseline curve from $\mathbf{Z}_{-k}$. We then regress it on the data in Z_k, and compute the mean squared prediction error (MSPE).

$$MSPE(\mathbf{Z}_{-k}, h, h^\star) = \frac{1}{|\mathbf{Z}_k|} \sum_{i \in \mathbf{Z}_k} \sum_{t=1}^{T} [y_{it} - (\hat{\alpha}_{ki} + \hat{\beta}_{ki} \hat{m}_t(\mathbf{Z}_{-k}, h, h^\star))]^2, \tag{14}$$

where $\hat{\alpha}_{ki}$ and $\hat{\beta}_{ki}$ are the estimated regression coefficients. We repeat the calculation for $k = 1, \ldots, K$, and the optimal pair $(\hat{h}, \hat{h}^\star)$ is the one which minimizes the average MSPE, i.e.

$$(\hat{h}, \hat{h}^\star) = \arg \min_{(h, h^\star)} \frac{1}{K} \sum_{k=1}^{K} MSPE(\mathbf{Z}_{-k}, h, h^\star). \tag{15}$$

The effectiveness of the cross-validation will be evaluated in Sections 3 and 4.

3 Application to mass spectrometry data

We now apply the proposed multi-step method to a SELDI-TOF mass spectrometry data set from a study of 33 ($n = 33$) liver cancer patients conducted at Changzheng Hospital in Shanghai. For each patient, we extract the m/z values in the region 2,000–10,000 Da ($T = 21,000$), which is believed to contain all the useful information. Figure 2 contains the curves of 10 randomly picked patients.
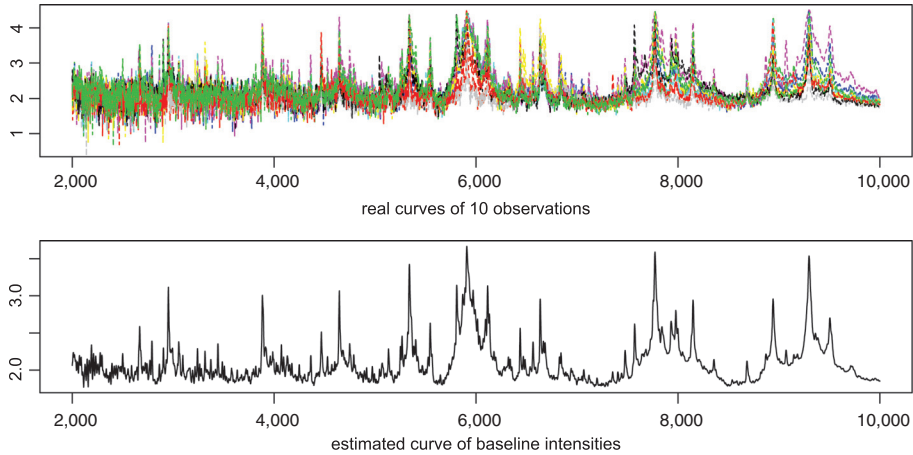
Figure 2: Curves of 10 observations and the baseline estimate.

There are some noticeable features in the data. All curves appear to be continuous. They peak simultaneously around certain locations; at each location, curves have the same shape but with different heights. All those features are captured well by our model. Since the observed values of m/z for each person may fluctuate, we need to perform registration to make the analysis easier. Here, we use the observations from the first individual and set his/her values as the reference. Then we use the linear interpolation method to compute the intensities of all the other individuals at the reference m/z locations. After that we get the preprocessed data which has the same m/z values for each observation.

We use the cross-validation method described in Section 2.3 to select the optimal bandwidths with $K = 33$, i.e., leave-one-out cross validation. We compute the MSPE at the grid of $h = 2, 4, 6, \ldots, 40$ and $h^* = 2, 4, 6, \ldots, 40$. Table 1 contains a portion of the result with $h = 30, 32, \ldots, 40$ and $h^* = 2, 4, 6, \ldots, 20$.

**Table 1:** MSPE of the leave-one-out prediction of real data. The minimum MSPE is shown in bold.

h h^*	30	32	34	36	38	40
2	1,104.946	1,104.941	1,104.936	1,104.934	1,104.931	1,104.930
4	1,104.483	1,104.482	**1,104.481**	1,104.483	1,104.484	1,,104.487
6	1,110.261	1,110.265	1,110.269	1,110.275	1,110.281	1,110.288
8	1,122.601	1,122.610	1,122.619	1,122.630	1,122.640	1,122.652
10	1,140.525	1,140.539	1,140.552	1,140.568	1,140.582	1,140.598
12	1,161.739	1,161.757	1,161.775	1,161.795	1,161.813	1,161.832
14	1,183.842	1,183.864	1,183.886	1,183.909	1,183.931	1,183.953
16	1,205.356	1,205.382	1,205.407	1,205.433	1,205.458	1,205.483
18	1,225.298	1,225.327	1,225.355	1,225.383	1,225.411	1,225.438
20	1,243.068	1,243.099	1,243.130	1,243.162	1,243.191	1,243.222

As we can see in Table 1, the minimum MSPE occurs at the location of $h = 34$ and $h^* = 4$, which agrees with our theory that h and h^* should not be chosen with the same rate for the purpose of estimating the nonparametric component.

Finally, we use the selected bandwidths to estimate the location and scale parameters as well as the nonparametric curve for the whole data set. The estimated parameters are reported in Table 2 and the baseline nonparametric curve estimation is shown in the lower part of Figure 2. From Table 2, we can see that each individual has very different regression coefficients, which was also verified by looking at Figure 2. In

Table 2: Regression parameters of real data.

ID	$\hat{\alpha}$	$\hat{\beta}$	ID	$\hat{\alpha}$	$\hat{\beta}$	ID	$\hat{\alpha}$	$\hat{\beta}$
1	0	1	12	−1.0302	1.5914	23	−0.7234	1.4448
2	−0.2086	1.1836	13	−0.1788	1.1366	24	0.2021	0.8915
3	−1.2208	1.6836	14	−0.3252	1.2586	25	0.5341	0.7957
4	−0.5630	1.3689	15	−0.6169	1.3599	26	0.5373	0.7203
5	−1.4761	1.8721	16	−0.3919	1.2418	27	−0.3748	1.2181
6	−1.2931	1.7142	17	0.0820	1.0178	28	0.0935	0.9642
7	0.7928	0.5925	18	0.7402	0.6569	29	0.7852	0.5971
8	0.0582	1.0387	19	−0.0609	1.0586	30	−0.0085	1.0503
9	−0.3338	1.1839	20	−0.9218	1.5053	31	−0.3827	1.2131
10	−0.9066	1.5397	21	−0.0580	1.0378	32	−0.7803	1.5011
11	−1.5054	1.8770	22	−0.3526	1.2149	33	−0.2115	1.1108
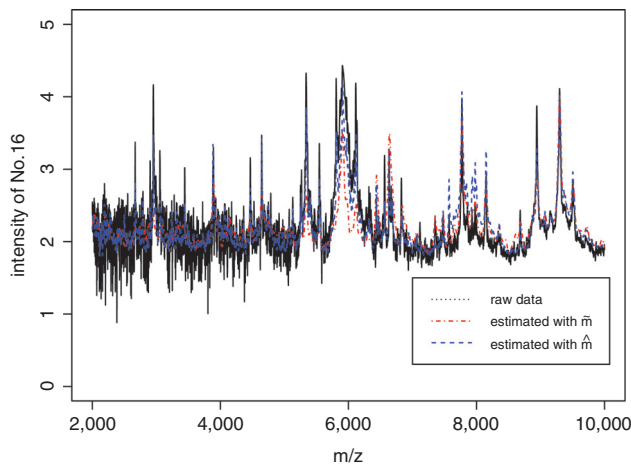
Figure 3: Nonparametric estimates of the curve from $\tilde{m}(\cdot)$ and $\hat{m}(\cdot)$ for the 16th object.

addition, comparing the estimated curve for the baseline intensities with the real curves of 10 observations, it is clear that the majority of the peaks and shapes are captured by the nonparametric estimate with appropriate degree of smoothing. A graphical representation of the raw curve of a single individual (16th subject) is also illustrated in Figure 3 along with estimates derived from $\tilde{m}(\cdot)$ and $\hat{m}(\cdot)$. We can see from the figure that the estimate from $\hat{m}(\cdot)$ is notably better than that from $\tilde{m}(\cdot)$, which shows that multi-step procedure is effective in improving the estimates for the baseline curve. We observed similar phenomenon for all the other subjects.

4 Simulation studies

We conduct simulations to assess the performance of the proposed method for parameter and curve estimation. The true curve $m(\cdot)$ is chosen from a moving average smoother of the cross-sectional mean of a fraction of real Mass Spectrometry data in Section 3 after log transformation. We set 10000 m/z values equally-spaced from 1 to 10,000 ($T = 10,000$) and the number of individuals $n = 30$. The true values of the parameters $\alpha_i, \beta_i, i = 1, 2, \ldots, n$ for each individual are shown in Table 3. And the error terms ε_{it} are sampled independently from $N(0, \sigma^2)$, where we set $\sigma = 0.25$.

We apply our multi-step procedure to the simulated data with different choices of the bandwidth. The number of runs is 100. The estimated parameters $\hat{\alpha}_i$ and $\hat{\beta}_i$ are shown in Table 3 along with the standard

Table 3: Regression parameter estimates ($h = 35$).

ID	α	β	$\hat{\alpha}$	$\hat{\beta}$	ID	α	β	$\hat{\alpha}$	$\hat{\beta}$
1	0	1	0.000(0.000)	1.000(0.000)	16	0.6	1	0.605(0.026)	0.997(0.019)
2	0.2	0.2	0.202(0.020)	0.198(0.013)	17	0.8	0.2	0.799(0.020)	0.201(0.014)
3	0.4	0.5	0.399(0.023)	0.501(0.016)	18	1	0.5	1.004(0.024)	0.497(0.016)
4	0.6	1.5	0.598(0.040)	1.502(0.027)	19	0	1.5	−0.001(0.044)	1.501(0.029)
5	0.8	2	0.801(0.047)	2.000(0.031)	20	0.2	2	0.206(0.043)	1.997(0.029)
6	1	1	0.999(0.029)	1.001(0.020)	21	0.4	1	0.403(0.030)	0.998(0.021)
7	0	0.2	−0.001(0.022)	0.201(0.015)	22	0.6	0.2	0.598(0.024)	0.201(0.016)
8	0.2	0.5	0.200(0.021)	0.500(0.014)	23	0.8	0.5	0.801(0.024)	0.500(0.016)
9	0.4	1.5	0.404(0.033)	1.498(0.023)	24	1	1.5	0.996(0.038)	1.503(0.025)
10	0.6	2	0.603(0.044)	1.999(0.030)	25	0	2	0.001(0.044)	2.000(0.029)
11	0.8	1	0.800(0.026)	1.001(0.018)	26	0.2	1	0.203(0.030)	0.998(0.020)
12	1	0.2	1.002(0.021)	0.198(0.014)	27	0.4	0.2	0.399(0.021)	0.201(0.015)
13	0	0.5	0.003(0.023)	0.499(0.016)	28	0.6	0.5	0.604(0.021)	0.497(0.014)
14	0.2	1.5	0.206(0.036)	1.497(0.024)	29	0.8	1.5	0.803(0.032)	1.499(0.023)
15	0.4	2	0.401(0.048)	2.001(0.033)	30	1	2	1.001(0.047)	2.000(0.031)

Note: Standard deviations are in parentheses.

Table 4: SSE of the initial and updated estimation of m. The minimum SSE is shown in bold.

SSE of $\tilde{m}$					
h	20	25	30	35	40
	9.1112	7.5509	6.7024	6.3418	6.3653

SSE of $\tilde{m}$					
h^* \\ h	20	25	30	35	40
20	0.6145	0.5936	0.5925	0.6078	0.6388
22	0.5762	0.5563	0.5561	0.5723	0.6042
24	0.5453	0.5265	0.5272	0.5443	0.5771
26	0.5204	0.5026	0.5043	0.5223	0.5561
28	0.5005	0.4838	0.4866	0.5056	0.5403
30	0.4850	0.4695	0.4733	0.4934	0.5291
32	0.4735	0.4592	0.4641	0.4852	0.5220
34	0.4657	0.4527	0.4587	0.4809	0.5188
36	0.4612	**0.4496**	0.4568	0.4802	0.5192
38	0.4601	0.4498	0.4583	0.4829	0.5231
40	0.4622	0.4533	0.4631	0.4889	0.5303

errors. We set $h = 35$, which leads to the smallest SSE of $\tilde{m}$ shown in Table 4. It is evident that the estimation is very accurate for all the location and scale parameters. In addition, for each h and h^*, the computation time is less than two minutes on a laptop.

From Table 4, we can see that the global optimal bandwidths are $h = 25, h^* = 36$. It is interesting to note that the optimal bandwidth for $\tilde{m}(\cdot)$ is $h = 35$, which is different from the optimal bandwidth for the final estimator.

To evaluate the quality of the our multi-step estimation method for the nonparametric baseline function, we consider a classical nonparametric estimation on another set of data where the same true function $m(\cdot)$ is used but $\alpha_i = 0, \beta_i = 1$ for all $i = 1, \ldots, n$. We applied the same local linear estimation with

Table 5: SSE of the estimation of m in the dataset with same parameters in each individual. The minimum SSE is shown in bold.

h	20	30	40	50	60
SSE	0.6881	0.4936	**0.4442**	0.4926	0.6337

different bandwidths. The mean SSE of the estimated $m(\cdot)$ from 100 repetitions for different hs are given in Table 5. When we applied the multi-step estimation procedure, the best mean SSE we achieved in Table 4 is very close to the minimal mean SSE 0.4442 for the oracle estimator. This comparison confirms that there is little loss of information in the proposed method when both parametric and nonparametric components are estimated simultaneously.

We use cross-validation to get a data-driven choice of the bandwidths. Here, we set $K = 5$ to get a mean MSPE of every different bandwidth choices of both steps over 100 runs, and the optimal bandwidths are those with the minimum mean MSPE. The mean MSPE values are shown in Table 6, from which we can see that the smallest value is located at $h = 25, h^* = 36$, which is quite close to the optimal bandwidths $h = 25$ and $h^* = 38$ in Table 4. Therefore, the cross-validation idea appears to work well in terms of selecting the best bandwidths.

Table 6: Mean MSPE of the 5-fold CV over 100 times. Here, we multiply T for all the MSPE values and subtract the minimum 625.61956.

h h*	20	25	30	35	40
20	0.293733	0.293728	0.293724	0.293722	0.293721
22	0.222948	0.222943	0.222940	0.222939	0.222939
24	0.165816	0.165813	0.165810	0.165809	0.165809
26	0.119253	0.119250	0.119248	0.119248	0.119248
28	0.081603	0.081600	0.081599	0.081598	0.081599
30	0.052297	0.052295	0.052294	0.052294	0.052295
32	0.030256	0.030255	0.030254	0.030255	0.030256
34	0.014621	0.014620	0.014620	0.014621	0.014622
36	0.004761	0.004760	0.004760	0.004761	0.004763
38	$3.4e-08$	0	$5.6e-07$	$2.0e-06$	$4.2e-06$
40	0.000590	0.000590	0.000592	0.000593	0.000596

5 Discussion

This paper proposes a semi- and nonparametric model suitable for analyzing the mass spectra data. The model is flexible and intuitive, capturing the main feature in the MS data. Both the parametric and nonparametric components have natural interpretation. A multi-step iterative algorithm is proposed for estimating both the individual location and scale regression coefficients and the nonparametric function. The algorithm combines local linear fitting and the least squares method, both of which are easy to implement and computationally efficient. Both simulation studies and real data analysis demonstrate that the proposed multi-step procedure works well.

The local linear fitting for the nonparametric function estimation maybe replaced with other nonparametric estimation techniques. Because the location and scale parameters are subject specific, the empirical

Bayes method [19] may be used. In addition, nonparametric Bayes may also be applicable with the nonparametric function being modeling as a realization of Gaussian process.

The proposed model and the associated iterative estimation method do not account for the random error in the measurement of X. It is desirable to incorporate the measurement error into the model [20].

Many studies involving MS data are aimed at classifying patients of different disease types. The information of peaks are usually applied as the basis of the classifier. The proposed method provides a natural way to identify the peaks for different group of patients by using the multi-step estimation procedure on each group and finding out the corresponding nonparametric baseline function. From the estimated baseline function, the information of peaks can be easily extracted, which can then be used for classification.

Appendix

The Appendix contains proofs of Theorems 1–4, where the proofs of Lemma 1 and Corollary 1 can be found in the supplementary materials. We begin with some notations, which will be used to streamline some of the proofs. Because all asymptotic expansions are derived with x_{it}'s being fixed, we will, for notational simplicity, use E to denote the conditional expectation and $\mathrm{V}ar$ to denote the conditional variance given x_{it}'s throughout the Appendix. For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, let

$$V_{it} = \frac{1}{T} - \frac{\sigma_i(X_{it}) W_{it} \bar{m}(x_{i\cdot})}{\sum_{s=1}^{T} W_{is}^2}.$$

A.1 Proof of Theorem 1

Proof. First of all, define $\tilde{W}_{it} = \tilde{m}(x_{it}) - \bar{\tilde{m}}(x_{i\cdot})$ to simplify the presentation. By definition, we have the following expansion for $\hat{\beta}_i$ when $i \geq 2$.

$$\hat{\beta}_i - \beta_i = \beta_i \frac{\sum_{t=1}^{T} \tilde{W}_{it}(m(x_{it}) - \tilde{m}(x_{it}))}{\sum_{t=1}^{T} \tilde{W}_{it}^2} + \frac{\sum_{t=1}^{T} \tilde{W}_{it} \sigma_i(x_{it}) \varepsilon_{it}}{\sum_{t=1}^{T} \tilde{W}_{it}^2} \tag{16}$$

$$\equiv \beta_i D_i + G_i.$$

From Lemma 1 and the proof of Corollary 1, we have

$$\tilde{W}_{it} - W_{it} = O_p(h^2). \tag{17}$$

Plugging (9) into D_i, we have

$$D_i = -\frac{\sum_{t=1}^{T}[(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))U_1(x_{it}) + \frac{1}{2}\mu_2 m''(x_{it})W_{it}h^2 + o(h^2)]}{\sum_{t=1}^{T}\tilde{W}_{it}^2}$$

$$-\frac{\sum_{t=1}^{T}\{[W_{it} + O(h^2)]U_1(x_{it}) + O(h^2)[U_1(x_{it}) - \bar{U}_1(x_{i\cdot})]\}}{\sum_{t=1}^{T}\tilde{W}_{it}^2} \tag{18}$$

$$= -\frac{\sum_{t=1}^{T}(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))U_1(x_{it})}{\sum_{t=1}^{T}W_{it}^2} - h^2 P_i - \frac{\sum_{t=1}^{T}W_{it}U_1(x_{it})}{\sum_{t=1}^{T}W_{it}^2}(1 + o_p(1)) + o(h^2 + \frac{1}{Th}),$$

where the last asymptotic expansion follows from (17). Similarly for G_i, we have

$$G_i = \frac{\sum_{t=1}^{T}W_{it}\sigma_i(x_{it})\varepsilon_{it}(1 + O(h^2))}{\sum_{t=1}^{T}\tilde{W}_{it}^2} + \frac{\sum_{t=1}^{T}(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T}\tilde{W}_{it}^2}$$

$$= \frac{\sum_{t=1}^{T}W_{it}\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T}W_{it}^2}(1 + o_p(1)). \tag{19}$$

We observe that for any $i \geq 2$, $U_1(x_{it})$ is a linear combination of $\{\varepsilon_{1t}, t = 1, \ldots, T\}$. Therefore, $U_1(x_{it})$ is independent of $\{\varepsilon_{it}, i = 2, \ldots, n, t = 1, \ldots, T\}$. By using the tower property, we have $EG_i = 0$. Therefore, $\beta_i D_i$ is the only part that contributes to the bias of $\hat{\beta}_i$. In view of these and Corollary 1, we have the following expansions for the bias and variance terms

$$E(\hat{\beta}_i - \beta_i) = -\beta_i h^2 P_i - \beta_i \frac{E\sum_{t=1}^{T}(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))^2}{\sum_{t=1}^{T}W_{it}^2} + o(h^2 + \frac{1}{Th})$$

$$= -\beta_i h^2 P_i - \beta_i \frac{\sum_{t=1}^{T}\text{Var}(U_1(x_{it}))(1 + o(1))}{\sum_{t=1}^{T}W_{it}^2} + o(h^2 + \frac{1}{Th})$$

$$= -\beta_i h^2 P_i - \beta_i \frac{1}{Th}Q_i + o(h^2 + \frac{1}{Th}),$$

and

$$\text{Var}(\hat{\beta}_i) = \text{Var}\left(-\beta_i \frac{\sum_{t=1}^{T}W_{it}U_1(x_{it})}{\sum_{t=1}^{T}W_{it}^2}(1 + o_p(1))\right)$$

$$+ \text{Var}\left(\frac{\sum_{t=1}^{T}W_{it}\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T}W_{it}^2}(1 + o_p(1))\right).$$

Straightforward variance calculation for an independent sum gives

$$\text{Var}\left(\sum_{t=1}^{T}W_{it}U_1(x_{it})\right) = \sum_{s=1}^{T}\left[\sum_{t=1}^{T}W_{it}\frac{\omega_{1s}(x_{it})}{\sum_{l=1}^{T}\omega_{1l}(x_{it})}\right]^2 \sigma_1^2(x_{1s}). \tag{20}$$

We have

$$\sum_{t=1}^{T}W_{it}\frac{\omega_{1s}(x_{it})}{\sum_{s=1}^{T}\omega_{1s}(x_{it})} = \sum_{t=1}^{T}(m(x_{it}) - \bar{m}(x_{i\cdot}))\frac{K_h(x_{it} - x_{1s})(S_{T,2} - (x_{it} - x_{1s})S_{T,1})}{S_{T,0}S_{T,2} - S_{T,1}^2}.$$

We expand $m(x)$ in the neighborhood of point x_{1s} using Taylor's expansion,

$$m(x_{it}) = m(x_{1s}) + (x_{it} - x_{1s})m'(x_{1s}) + \frac{1}{2}(x_{it} - x_{1s})^2 m''(x_{1s}) + o_p((x_{it} - x_{1s})^2).$$

Since the kernel function $K_h(x - x_{1s})$ vanishes out of the neighborhood of x_{1s} with diameter h, we can obtain the following

$$\sum_{t=1}^{T} m(x_{it}) \frac{K_h(x_{it} - x_{1s})(S_{T,2} - (x_{it} - x_{1s})S_{T,1})}{S_{T,0}S_{T,2} - S_{T,1}^2}$$

$$= m(x_{1s}) + \sum_{t=1}^{T} m'(x_{1s})(x_{it} - x_{1s})\frac{K_h(x_{it} - x_{1s})(S_{T,2} - (x_{it} - x_{1s})S_{T,1})}{S_{T,0}S_{T,2} - S_{T,1}^2}$$

$$+ \sum_{t=1}^{T}\left[\frac{1}{2}(x_{it} - x_{1s})^2 m''(x_{1s}) + o_p((x_{it} - x_{1s})^2)\right]\frac{K_h(x_{it} - x_{1s})(S_{T,2} - (x_{it} - x_{1s})S_{T,1})}{S_{T,0}S_{T,2} - S_{T,1}^2}$$

$$= m(x_{1s}) + O_p(h^2)\sum_{t=1}^{T}\frac{K_h(x_{it} - x_{1s})(S_{T,2} - (x_{it} - x_{1s})S_{T,1})}{S_{T,0}S_{T,2} - S_{T,1}^2}$$

$$= m(x_{1s}) + O_p(h^2),$$

where the functions $S_{T,k}, k = 0, 1, 2$ are evaluated at the point x_{it}. Combined with $\bar{m}(x_{i\cdot}) = \bar{m}(x_{1\cdot}) + O_p(T^{-1/2})$, we can have the expansion

$$\sum_{t=1}^{T} W_{it}\frac{\omega_{1s}(x_{it})}{\sum_{s=1}^{T}\omega_{1s}(x_{it})} = m(x_{1s}) + O_p(h^2) - \bar{m}(x_{1\cdot}) + O_p(T^{-1/2}) = W_{1s} + O_p(h^2 + T^{-1/2})$$

Then recall (20), we have $Var(\sum_{t=1}^{T} W_{it}U_1(x_{it})) = \sum_{t=1}^{T} W_{1t}^2\sigma_1^2(x_{1t}) + o_p(T)$, which leads to the variance expansion

$$Var(\hat{\beta}_i) = \beta_i^2\frac{\sum_{t=1}^{T} W_{1t}^2\sigma_1^2(x_{1t})}{\left(\sum_{t=1}^{T} W_{it}^2\right)^2} + \frac{\sum_{t=1}^{T} W_{it}^2\sigma_i^2(x_{it})}{\left(\sum_{t=1}^{T} W_{it}^2\right)^2} + o_p\left(\frac{1}{T}\right).$$

$\square$

A.2 Proof of Theorem 2

Proof. Recall (7) and (8), we have

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}(x)(x_{it} - x) = \sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)}\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)} - \sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)}\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)} = 0.$$

Then we have the asymptotic expansion of the updated estimator of baseline intensity $\hat{m}(\cdot)$ at time point $\times$ as follows.

By definition of $\hat{m}(\cdot)$ in (6), we can write

$$\hat{m}(x) - m(x) = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}(\alpha_i - \hat{\alpha}_i)/\hat{\beta}_i}{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}} + \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}\sigma(x_{it})\varepsilon_i/\hat{\beta}_i}{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}}$$

$$+ \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}m(x_{it})(\beta_i - \hat{\beta}_i)/\hat{\beta}_i}{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}} + \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}(\frac{1}{2}m''(x)(x_{it} - x)^2 + o((x_{it} - x)^2))}{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}}.$$

(21)

From the proof of Theorem 1, we have

$$
\hat{\beta}_i = \beta_i - \beta_i h^2 P_i - \beta_i \frac{\sum_{t=1}^{T}(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))U_1(x_{it})}{\sum_{t=1}^{T} W_{it}^2} - \beta_i \frac{\sum_{t=1}^{T} W_{it} U_1(x_{it})}{\sum_{t=1}^{T} W_{it}^2}(1 + o_p(1))
$$

$$
+ \frac{\sum_{t=1}^{T} W_{it}\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T} W_{it}^2}(1 + o_p(1)) + o\left(h^2 + \frac{1}{Th}\right). \tag{22}
$$

Then, from the least square expression, we have the asymptotic expansion for $\hat{\alpha}_i$ as follows.

$$
\hat{\alpha}_i = \bar{y}_{i\cdot} - \hat{\beta}_i \bar{\bar{m}}(x_{i\cdot})
$$

$$
= \alpha_i + \beta_i \bar{m}(x_{i\cdot}) + \bar{\varepsilon}_{i\cdot} - \hat{\beta}_i \left[\bar{m}(x_{i\cdot}) + \frac{\mu_2}{2}\bar{m}''(x_{i\cdot})h^2 + \bar{U}_1(x_{i\cdot}) + o(h^2)\right]
$$

$$
= \alpha_i + \beta_i h^2 R_i + \bar{m}(x_{i\cdot})\beta_i \frac{\sum_{t=1}^{T}(U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))U_1(x_{it})}{\sum_{t=1}^{T} W_{it}^2} + o\left(h^2 + \frac{1}{Th}\right) \tag{23}
$$

$$
+ \sum_{t=1}^{T} V_{it}\varepsilon_{it}(1 + o_p(1)) - \beta_i \sum_{t=1}^{T} V_{it} U_1(x_{it})(1 + o_p(1)).
$$

Now, we plug the above asymptotic expansions (22) and (23) into the right hand side of (21). The first part of (21) could be expanded as follows

$$
\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}(\alpha_i - \hat{\alpha}_i)/\hat{\beta}_i}{\sum_{i=1}^{n}\sum_{t=1}^{T}\omega_{it}^{\star}}
$$

$$
= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\alpha_i - \hat{\alpha}_i)/\hat{\beta}_i \star \hat{\beta}_i^2 K_{h^\star}(x_{it} - x)\left[\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)} - (x_{it} - x)\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)}\right]}{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\beta}_i^2 K_{h^\star}(x_{it} - x)\left[\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)} - (x_{it} - x)\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)}\right]} \tag{24}
$$

$$
= \frac{\sum_{i=1}^{n}(\alpha_i - \hat{\alpha}_i)\hat{\beta}_i \sum_{t=1}^{T} K_{h^\star}(x_{it} - x)\left[\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)} - (x_{it} - x)\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)}\right]}{\sum_{i=1}^{n}\hat{\beta}_i^2 \sum_{t=1}^{T} K_{h^\star}(x_{it} - x)\left[\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,2}^{\star(i)} - (x_{it} - x)\sum_{i=1}^{n}\hat{\beta}_i^2 S_{T,1}^{\star(i)}\right]}.
$$

The numerator of (24) has expansion

$$
\sum_{i=1}^{n}(\alpha_i - \hat{\alpha}_i)\hat{\beta}_i\left[Tf(x)\{1 + o_p(1)\}\sum_{i=1}^{n}\hat{\beta}_i^2 Th^{\star 2}f(x)\mu_2\{1 + o_p(1)\}\right.
$$

$$
\left. - Th^\star\left\{h^\star f'(x)\mu_2 + O_p\left(h^{\star 2} + \frac{1}{\sqrt{Th^\star}}\right)\right\}\sum_{i=1}^{n}\hat{\beta}_i^2 Th^\star\left\{h^\star f'(x)\mu_2 + O_p\left(h^{\star 2} + \frac{1}{\sqrt{Th^\star}}\right)\right\}\right]
$$

$$
= T^2 h^{\star 2}\sum_{i=1}^{n}(\alpha_i - \hat{\alpha}_i)\hat{\beta}_i\left[f(x)\{1 + o_p(1)\}\sum_{i=1}^{n}\hat{\beta}_i^2 f(x)\mu_2\{1 + o_p(1)\}\right. \tag{25}
$$

$$
\left. - \left\{h^\star f'(x)\mu_2 + O_p\left(h^{\star 2} + \frac{1}{\sqrt{Th^\star}}\right)\right\}\sum_{i=1}^{n}\hat{\beta}_i^2\left\{h^\star f'(x)\mu_2 + O_p\left(h^{\star 2} + \frac{1}{\sqrt{Th^\star}}\right)\right\}\right]
$$

$$
= T^2 h^{\star 2}\sum_{i=1}^{n}(\alpha_i - \hat{\alpha}_i)\hat{\beta}_i\left[f(x)\sum_{i=1}^{n}\beta_i^2 f(x)\mu_2\{1 + o_p(1)\}\right],
$$

where the last equation following from $\hat{\beta}_i = \beta_i + O(h^2) + O((Th)^{-1}) + O_p(T^{-1/2})$.

Similarly, the denominator of (24) has the following expansion

$$
\sum_{i=1}^{n} \hat{\beta}_i^2 \left[Tf(x)\{1+o_p(1)\} \sum_{i=1}^{n} \hat{\beta}_i^2 Th^{*2} f(x) \mu_2 \{1+o_p(1)\} \right.
$$

$$
\left. - Th^{*} \left\{ h^{*} f'(x) \mu_2 + O_p \left(h^{*2} + \frac{1}{\sqrt{Th^{*}}} \right) \right\} \sum_{i=1}^{n} \hat{\beta}_i^2 Th^{*} \left\{ h^{*} f'(x) \mu_2 + O_p \left(h^{*2} + \frac{1}{\sqrt{Th^{*}}} \right) \right\} \right]
$$

$$
= T^2 h^{*2} \sum_{i=1}^{n} \hat{\beta}_i^2 \left[f(x)\{1+o_p(1)\} \sum_{i=1}^{n} \hat{\beta}_i^2 f(x) \mu_2 \{1+o_p(1)\} \right. \tag{26}
$$

$$
\left. - \left\{ h^{*} f'(x) \mu_2 + O_p \left(h^{*2} + \frac{1}{\sqrt{Th^{*}}} \right) \right\} \sum_{i=1}^{n} \hat{\beta}_i^2 \left\{ h^{*} f'(x_i) \mu_2 + O_p \left(h^{*2} + \frac{1}{\sqrt{Th^{*}}} \right) \right\} \right]
$$

$$
= T^2 h^{*2} \sum_{i=1}^{n} \hat{\beta}_i^2 \left[f(x) \sum_{i=1}^{n} \hat{\beta}_i^2 f(x) \mu_2 \{1+o_p(1)\} \right].
$$

Then combining the expansions (25) and (26), we have the following expansion for the first part of (21).

$$
\frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^{*} (\alpha_i - \hat{\alpha}_i) / \hat{\beta}_i}{\sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it}^{*}} = \frac{T^2 h^{*2} \sum_{i=1}^{n} (\alpha_i - \hat{\alpha}_i) \hat{\beta}_i \left[f(x) \sum_{i=1}^{n} \hat{\beta}_i^2 f(x) \mu_2 \{1+o_p(1)\} \right]}{T^2 h^{*2} \sum_{i=1}^{n} \hat{\beta}_i^2 \left[f(x) \sum_{i=1}^{n} \hat{\beta}_i^2 f(x) \mu_2 \{1+o_p(1)\} \right]}
$$

$$
= \frac{\sum_{i=2}^{n} (\alpha_i - \hat{\alpha}_i) \beta_i}{\sum_{i=1}^{n} \beta_i^2} (1+o_p(1)).
$$

For other parts of (21), we can apply the same techniques for expansion. As a result, the following expansion of $\hat{m}$ holds.

$$
\hat{m}(x) - m(x) = \frac{\sum_{i=2}^{n} \beta_i (\alpha_i - \hat{\alpha}_i)}{\sum_{i=1}^{n} \beta_i^2} (1+o_p(1)) + \frac{\sum_{i=1}^{n} \beta_i \sum_{t=1}^{T} K_{h^{*}}(x_{it} - x) \varepsilon_{it} (1+o_p(1))}{\sum_{i=1}^{n} \beta_i^2 Tf(x)}
$$

$$
+ \frac{\sum_{i=1}^{n} \beta_i (\beta_i - \hat{\beta}_i) \sum_{t=1}^{T} K_{h^{*}}(x_{it} - x) \varepsilon_{it} (1+o_p(1))}{\sum_{i=1}^{n} \beta_i^2 Tf(x)}
$$

$$
= - \frac{\sum_{i=2}^{n} \beta_i^2 \left[h^2 R_i + \bar{m}(x_{i.}) \sum_{t=1}^{T} (U_1(x_{it}) - \bar{U}_1(x_{i.}))^2 / \sum_{t=1}^{T} W_{it}^2 - \sum_{t=1}^{T} V_{it} U_1(x_{it})(1+o_p(1)) \right]}{\sum_{i=1}^{n} \beta_i^2}
$$

$$
+ \frac{\sum_{i=1}^{n} \beta_i \sum_{t=1}^{T} K_{h^{*}}(x_{it} - x) \varepsilon_{it} (1+o_p(1))}{\sum_{i=1}^{n} \beta_i^2}
$$

$$
+ m(x) \left\{ \frac{\sum_{i=2}^{n} \beta_i^2 h^2 P_i}{\sum_{i=1}^{n} \beta_i^2} + \frac{\sum_{i=2}^{n} \beta_i^2 \sum_{t=1}^{T} (U_1(x_{it}) - \bar{U}_1(x_{i.}))^2 / \sum_{t=1}^{T} W_{it}^2}{\sum_{i=1}^{n} \beta_i^2 f(x)} \right\}
$$

$$
+ m(x) \left\{ \frac{\sum_{i=2}^{n} \beta_i^2 \sum_{t=1}^{T} W_{it} U_1(x_{it})(1+o_p(1)) / \sum_{t=1}^{T} W_{it}^2}{\sum_{i=1}^{n} \beta_i^2} \right\}
$$

$$
+ \frac{m''(x)}{2} \mu_2 h^{*2} + o(h^2 + \frac{1}{Th} + h^{*2}).
$$

Then it is straightforward to derive the bias of $\hat{m}(x)$ as follows

$$
E(\hat{m}(x) - m(x)) = - \frac{\sum_{i=2}^{n} \beta_i^2 (h^2 R_i + (Th)^{-1} \bar{m}(x_{i.}) Q_i)}{\sum_{i=1}^{n} \beta_i^2} + \left[\frac{\sum_{i=2}^{n} \beta_i^2 (h^2 P_i + (Th)^{-1} Q_i)}{\sum_{i=1}^{n} \beta_i^2} \right] m(x)
$$

$$
+ \frac{m''(x)}{2} \mu_2 h^{*2} + o \left(h^2 + \frac{1}{Th} + h^{*2} \right).
$$

For the variance of $\hat{m}(x)$, we notice that the error terms $\{\varepsilon_{it}, i = 1, \ldots, n, t = 1, \ldots, T\}$ are independent, which implies the independence of $\varepsilon_{it}, i = 2, \ldots, n$ and $U_1(x_{it})$. Therefore, we have the following asymptotic expansion for the variance.

$$\text{Var}(\hat{m}(x) - m(x)) = \text{Var}\left(\frac{\sum_{i=2}^n \beta_i^2 \sum_{t=1}^T V_{it} U_1(x_{it})}{\sum_{i=1}^n \beta_i^2} + \frac{m(x) \sum_{i=2}^n \beta_i^2 \sum_{t=1}^T W_{it} U_1(x_{it}) / \sum_{t=1}^T W_{it}^2}{\sum_{i=1}^n \beta_i^2}\right.$$

$$\left. + \frac{\sum_{i=1}^n \sum_{t=1}^T \beta_i K_{h^\star}(x_{it} - x)\varepsilon_{it}}{\sum_{i=1}^n \beta_i^2 \sum_{t=1}^T K_{h^\star}(x_{it} - x)}\right) + o\left(\frac{1}{T} + \frac{1}{nTh^\star}\right)$$

$$= \text{Var}\left(\frac{\sum_{i=2}^n \beta_i^2 \sum_{t=1}^T (V_{it} + m(x)W_{it} / \sum_{t=1}^T W_{it}^2) U_1(x_{it})}{\sum_{i=1}^n \beta_i^2}\right)$$

$$+ \frac{v_0 \sum_{i=2}^n \beta_i^2 f^{-1}(x)\sigma_i^2(x_{it})}{Th^\star(\sum_{i=1}^n \beta_i^2)^2} + o\left(\frac{1}{T} + \frac{1}{nTh^\star}\right),$$

where the expansions follow similar techniques as (25) and (26). Now, by the definition of U_1, we have

$$\text{Var}(\hat{m}(x) - m(x)) = \frac{1}{(\sum_{i=1}^n \beta_i^2)^2} \sum_{s=1}^T \left(\sum_{i=2}^n \beta_i^2 \left[\frac{1}{T} + Z_{is}\right]\right)^2 \sigma_1^2(x_{1s})$$

$$+ \frac{v_0 \sum_{i=2}^n \beta_i^2 f^{-1}(x)\sigma_i^2(x_{it})}{Th^\star(\sum_{i=1}^n \beta_i^2)^2} + o\left(\frac{1}{T} + \frac{1}{nTh^\star}\right).$$

$$\square$$

A.3 Proof of Theorem 3

Proof. From the results of Theorem 2, it is straightforward to show that the order of the mean squared error of $\hat{m}(x)$ is $h^4 + (T^2h^2)^{-1} + h^{\star 4} + T^{-1} + (nTh^\star)^{-1}$. To minimize the mean squared error, we can taken $h = O(T^{-\frac{1}{3}})$ and $h^\star = O((nT)^{-\frac{1}{5}})$. Under such choices of h and $h^\star$, the order of the mean squared error is $(nT)^{-\frac{4}{5}} + T^{-1}$.

Therefore, to match the optimal nonparametric convergence rate $(nT)^{-\frac{4}{5}}$ for mean squared error, the condition $n = O(T^{\frac{1}{4}})$ is required.

$$\square$$

A.4 Proof of Theorem 4

Proof. We start from the asymptotic expansion from (22) in the proof of Theorem 2. First, we investigate the asymptotic behavior of the third term on the right hand side of (22).

As a first step, we have

$$\text{Var}\left(\sum_{t=1}^T (U_1(x_{it}) - \bar{U}_1(x_{i\cdot}))^2\right) \leq 8\text{E}\left(\left[\sum_{t=1}^T U_1(x_{it})^2\right]^2\right). \tag{27}$$

Now, following the definition of $U_1(\cdot)$ and applying the same expansion of $\omega_{1s}(x_{it})$ as in the proof of Theorem 1,

$$E\left(\left[\sum_{t=1}^{T} U_1(x_{it})^2\right]^2\right) = E\left(\left[\sum_{t=1}^{T}\left[\frac{\sum_{s=1}^{T} K_h(x_{it}-x_{1s})\sigma_1(x_{1s})\varepsilon_{1s}}{Tf(x_{it})}\right]^2\right]^2\right)(1+o(1))$$

$$\leq \frac{1}{T^4}E\left(\sum_{s,u=1}^{T}\left(\sum_{t=1}^{T}\frac{K_h^2(x_{it}-x_{1s})}{f^2(x_{it})}I_{\{|x_{1s}-x_{1u}|<2h\}}\right)^2\sigma_1^2(x_{1s})\sigma_1^2(x_{1u})\right)(1+o(1)),$$

where the last inequality follows from exchanging the summation order and the property of the kernel function $K(\cdot)$. Observe that $f(\cdot)$ is bounded from below by Condition C2, the following inequality sequence is obtained.

$$E\left(\left[\sum_{t=1}^{T} U_1(x_{it})^2\right]^2\right) \leq \frac{1}{T^4\delta^4}E\left(\sum_{s,u=1}^{T}\left(\frac{Tv_0 f(x_{1s})}{h}\right)^2 I_{\{|x_{1s}-x_{1u}|<2h\}}\sigma_1^2(x_{1s})\sigma_1^2(x_{1u})\right)(1+o(1))$$

$$\leq \frac{O(1)}{T^2h^2}\sum_{s,u=1}^{T} I_{\{|x_{1s}-x_{1u}|<2h\}},$$

where the last term has the order of $O(h^{-1})$ by noticing

$$\sum_{s,u=1}^{T} I_{\{|x_{1s}-x_{1u}|<2h\}} = \sum_{s=1}^{T} 4Thf(x_{1s}).$$

We can also derive the order of the variance for the other two terms,

$$\text{Var}\left(-\beta_i\frac{\sum_{t=1}^{T} W_{it}U_1(x_{it})}{\sum_{t=1}^{T} W_{it}^2} + \frac{\sum_{t=1}^{T} W_{it}\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T} W_{it}^2}\right) = O(T^{-1}).$$

Due to the relationship of h and T, the third term is negligible when calculating the asymptotic variance. Then, the expansion for the bias of $\hat{\beta}_i$ can be rewritten as follows

$$\hat{\beta}_i - \beta_i + \beta_i(h^2 P_i + \frac{1}{Th}Q_i) = -\beta_i\sum_{s=1}^{T}\left(\frac{W_{1s}\sigma_1(x_{1s})\varepsilon_{1s}}{\sum_{t=1}^{T} W_{it}^2}\right) + \frac{\sum_{t=1}^{T} W_{it}\sigma_i(x_{it})\varepsilon_{it}}{\sum_{t=1}^{T} W_{it}^2}(1+o_p(1)) + o(h^2 + \frac{1}{Th}),$$

where the right hand side is an independent sum of random variables with their variances being of the same order, $O(T^{-1})$. As a result, the central limit theorem can be applied directly for $\hat{\beta}_i$.

$$\sqrt{T}[\hat{\beta}_i - \beta_i - \beta_i(h^2 P_i + \frac{1}{Th}Q_i)] \to^d N(0, \sigma_i^{\star 2}),$$

where the asymptotic variance $\sigma_i^{\star 2}$ is finite with the following expression.

$$\sigma^{\star 2} = \lim_{T\to\infty}\left[\frac{T^{-1}\sum_{t=1}^{T} W_{it}^2\sigma_i^2(x_{it})}{(T^{-1}\sum_{t=1}^{T} W_{it}^2)^2} + \beta_i^2\frac{T^{-1}\sum_{t=1}^{T} W_{1t}^2\sigma_1^2(x_{1t})}{(T^{-1}\sum_{t=1}^{T} W_{it}^2)^2}\right].$$

Notice that if the order of h is between $T^{-\frac{1}{2}}$ and $T^{-\frac{1}{4}}$, then $\hat{\beta}_i$ is asymptotic unbiased since $\sqrt{T}\beta_i(h^2 P_i + \frac{1}{Th}Q_i) \to^d 0$.

From Theorems 1 and 2 we have $\hat{\beta}_i, \hat{\alpha}_i, \hat{m}(\cdot)$ are consistent estimators of $\beta_i, \alpha_i, m(\cdot)$, respectively. Thus, $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\alpha}_i - \hat{\beta}_i\hat{m}(x_{it}))^2$ is also consistent for the variance under the assumption that $\sigma_i(\cdot)$ is a constant function for each subject i. $\square$

References

1. Yasui Y, Pepe M, Thompson M, Adam B-L, Wright G, Qu Y, et al. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. Biostatistics 2003;4(3):449–63.
2. Guilhaus M. Principles and instrumentation in time-of-flight mass spectrometry. J Mass Spectrom 1995;30:1519–32.
3. Banks D, Petricoin E. Finding cancer signals in mass spectrometry data. Chance 2003;16:8–57.
4. Baggerly KA, Morris JS, Wang J, Gold D, Xiao L-C, Coombes KR. A comprehensive approach to the analysis of MALDI-TOF proteomics spectra from serum samples. Proteomics 2003;3:1667–72.
5. Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics 2004;20:777–85.
6. Diamandis EP. Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol Cell Proteomics 2004;3:367–78.
7. Feng Y, Ma W, Wang Z, Ying Z, Yang Y. Alignment of protein mass spectrometry data by integrated Markov chain shifting method. Stat Interface 2009;2:329–40.
8. Roy P, Truntzer C, Maucort-Boulch D, Jouve T, Molinari N. Protein mass spectra data analysis for clinical biomarker discovery: a global review. Briefings Bioinf 2011;12:176–86.
9. Begun J, Hall WJ, Huang WM, Wellner JA. Information and asymptotic efficiency in parametric-nonparametric models. Ann Statist 1983;11:432–52.
10. Robinson PM. Root-N consistent semiparametric regression. Econometrica 1988;55:931–54.
11. Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA. Efficient and adaptive estimation for semiparametric models. Berlin: Springer, 1998.
12. Ruppert D, Wand M, Carroll R. Semiparametric regression. Cambridge: Cambridge University Press, 2003.
13. Fan J. Local linear regression smoothers and their minimax efficiency. Ann Statist 1993;21:196–216.
14. Rosenblatt M. Remarks on Some nonparametric estimates of a density function. Ann Math Statist 1956;27:832–7.
15. Stone CJ. Consistent nonparametric regression, with discussion. Ann Statist 1977;5:549–645.
16. Stone CJ. Optimal global rates of convergence for nonparametric regression. Ann Stat 1982;10:1040––1053.
17. Fan J, Gijbels I. Local polynomial modelling and its applications. London: Chapman and Hall, 1996.
18. Hardle W, Muller M, Sperlich S, Werwatz A. Nonparametric and semiparametric models. Springer Series in Statistics. Berlin: Springer, 2004.
19. Carlin B, Louis T. Bayesian methods for data analysis, 3rd ed. London: Chapman & Hall, 2008.
20. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective, 2nd ed. London: Chapman and Hall/CRC, June 2006.