

A semi-local likelihood regression estimator of the proportion based on group testing data

Dewei Wang^a, Haiming Zhou^b and K.B. Kulasekera^{c*}

^aDepartment of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA; ^bDepartment of Statistics, University of South Carolina, Columbia, SC 29208, USA; ^cDepartment of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40202, USA

(Received 26 March 2012; final version received 10 November 2012)

In this paper, we are concerned with the estimation of a proportion based on group testing data where the prevalence of a disease is observed for groups of individuals. Based on the covariates measured on all individuals in every group, we propose a local likelihood estimator of the prevalence probability as a function of the covariate. We show that the proposed estimator has an asymptotic normal distribution. Finite sample performance of the method is exhibited via some simulated examples and a real data analysis.

Keywords: nonparametric binary regression; local smoothing; local likelihood

AMS Subject Classifications: 62G05; 62G20

1. Introduction

Group (pooled) testing arises frequently in scientific studies. Pooling specimens for the purpose of estimating the prevalence of disease has proven to be an efficient method of reducing time and cost associated with sampling. For example, rather than testing blood specimens collected from individuals separately, group testing specifies that the specimens are first pooled and the resulting pooled specimen is then tested for the existence of the characteristic. This type of testing has also been used in pollution detection (Nagi and Raggi 1972; Wahed et al. 2006) and contamination and toxicity studies (Lennon 2007).

In group testing studies, experimenters often collect data on auxiliary variables that are easy and cost effective to measure. In most of these studies, the probability curve $p(x) = P[Y = 1|X = x]$ is of interest, where Y is the binary response and X is a covariate. Delaigle and Hall (2012) proposed a nonparametric estimator of $p(x)$ when the grouping mechanism is homogeneous, i.e. groups are constructed using similar values of the covariate. In practice, constructing pools in this manner may not be feasible. In this paper, we consider the case where individuals are grouped randomly with observed binary responses of the form Y_j^* , $j = 1, \dots, J$, where $Y_j^* = \max_{1 \leq i \leq n_j} Y_{ij}$, where Y_{ij} is the status of the i th individual in the j th pool. Y_{ij} 's are not observed although all the accompanying covariates X_{ij} , $i = 1, \dots, n_j$, $j = 1, \dots, J$, are measured. Parametric analysis of binary data of this type has been addressed by Bilder and Tebbs (2009) and Chen, Tebbs, and

*Corresponding author. Emails: kb.kulasekera@louisville.edu, kk@clemson.edu

Bilder (2009) among others. A thorough literature review followed by a nonparametric estimation method for $p(x)$ based on a method of moment argument is presented by Delaigle and Meister (2011). They obtained an expression for the pointwise asymptotic mean square error of their estimator and provided a detailed bandwidth selection method. However, these authors did not provide the asymptotic distribution of their estimator of $p(x)$.

In this paper, we address the estimation of $p(x)$ using a likelihood approach. We examine a localisation procedure that provides an asymptotically normal estimator of $p(x)$ which maintains high finite sample accuracy. Our numerical results show that, for the examined examples, the proposed procedure has the same type of finite sample properties compared with Delaigle and Meister (2011) when group sizes are equal and the regression function p is smooth. However, when the group sizes are not equal, our method appears to have better finite sample properties compared with theirs. The same trend seems to be true when $p(x)$ is more fluctuant. It is noteworthy that the function $p(x)$ may not be smooth in all situations. For example, when examining the probability of an adverse reaction based on a drug dosage, the reaction probability can sharply increase or even jump at certain dosage thresholds. In addition, unequal grouping is not uncommon when one uses individuals in clusters of units in a system. For example, one may consider each class as a group when there are multiple schools of different sizes and levels in a school system from which the data are collected.

The remainder of this paper is organised as follows. In Section 2, we describe our procedure and state the main asymptotic results. Section 3 is devoted to a simulation study and a real data analysis followed by a short discussion. All the assumptions are listed in the appendix along with the sketches of the proofs.

2. A semi-local likelihood method

We describe the proposed estimator followed by its properties in this section. We assume (Y_{ij}, X_{ij}) , $i = 1, \dots, n_j$; $j = 1, \dots, J$, are identical and independently distributed (i.i.d.) random vectors. In what follows, we assume that all of the covariates, X_{ij} 's, and the pool testing responses Y_j^* 's, as defined in the previous section, are available. For any fixed x , and a user-defined finite bandwidth h , we define $I_x = [x - h, x + h]$ and $Z_{ij} = X_{ij}I_x(X_{ij})$, where $I_x(X_{ij}) = 1$ if $X_{ij} \in I_x$, and $I_x(X_{ij}) = 0$ otherwise. Then the mixed probability density function of the Z s is given by

$$f_Z(z) = \begin{cases} \int_{I_x^c} f(u) du & \text{if } z = 0, \\ f(z) & \text{if } z \in I_x \setminus 0, \end{cases}$$

where I_x^c is the complement of the set I_x , $I_x \setminus 0$ is the set I_x excluding 0 and $f(\cdot)$ is the density function of an X . Then $Y_{ij}|Z_{ij} = z$ is a Bernoulli random variable with $P[Y_{ij} = 0|Z_{ij} = z] = r(z)$ where

$$r(z) = \begin{cases} r_1 & \text{if } z = 0, \\ q(z) & \text{if } z \in I_x \setminus 0, \end{cases}$$

with $r_1 = \int_{I_x^c} q(u)f(u) du / \int_{I_x^c} f(u) du$ and $q(z) = 1 - p(z)$. It is easy to see that $0 < r_1 \leq \sup_x q(x)$, and $r_1 \rightarrow q_*$ where $q_* = E[q(X)]$, as $h \rightarrow 0$. Note that $r(z)$ can also be written as

$$r(z) = r_1^{1-I_x(z)} \times q(z)^{I_x(z)}.$$

Now, we can write the log-likelihood of $Y_j^*, j = 1, \dots, J$, conditional on Z_{ij} 's as

$$\frac{1}{N} \sum_{j=1}^J \left\{ (1 - Y_j^*) \sum_{i=1}^{n_j} \log r(Z_{ij}) + Y_j^* \log \left[1 - \exp \left(\sum_{i=1}^{n_j} \log r(Z_{ij}) \right) \right] \right\}.$$

For small h and a fixed x , a Taylor expansion gives the following approximation:

$$\begin{aligned} \log r(Z_{ij}) &\approx I_x(X_{ij})g(x) + I_x(X_{ij})g'(x)(X_{ij} - x) + (1 - I_x(X_{ij})) \log r_1 \\ &= I_x(X_{ij})\theta_1 + I_x(X_{ij})(X_{ij} - x)\theta_2 + (1 - I_x(X_{ij}))\theta_3, \end{aligned} \tag{1}$$

where $g(\cdot) = \log q(\cdot)$, $\theta_1 = g(x)$, $\theta_2 = g'(x)$, $\theta_3 = \log r_1$. Define $\theta = (\theta_1, \theta_2, \theta_3)^\top$, $\mathbf{X}_j = (X_{1j}, \dots, X_{n_jj})^\top$ and $\tilde{\mathbf{X}}_j = \sum_{i=1}^{n_j} (I_x(X_{ij}), I_x(X_{ij})(X_{ij} - x), 1 - I_x(X_{ij}))^\top$. Equation (1) provides a local linear approximation of $\log r(\cdot)$ using the X_{ij} in I_x , in the event that no X_{ij} in I_x , then no local linear approximation would be performed. Then, we can write the local log-likelihood as

$$l(\theta) = \frac{1}{N} \sum_{j=1}^J \{ (1 - Y_j^*) \theta^\top \tilde{\mathbf{X}}_j + Y_j^* \log [1 - \exp(\theta^\top \tilde{\mathbf{X}}_j)] \} \omega_h(\mathbf{X}_j, x), \tag{2}$$

where $\omega_h(\mathbf{X}_j, x) = \prod_{i=1}^{n_j} K_h(X_{ij} - x)^{\delta_x(X_{ij})}$, $\delta_x(X_{ij}) = I_x(X_{ij}) / \sum_{i=1}^{n_j} I_x(X_{ij})$, which is defined to be 0 if the denominator is 0, and $K_h(\cdot) = h^{-1}K(\cdot/h)$ for a symmetric and continuous density function $K(\cdot)$.

Note that if $p(\cdot)$ has sufficient smoothness, we can use a local polynomial approximation for $g(x)$ in Equation (1) and estimate the derivatives of g up to a desired order. However, since in practice the order of the smoothness of $p(\cdot)$ is usually unknown and the local linear estimator behaves better than the local constant estimator (Fan and Gijbels 1996), we present the local linear approximation case here.

The Hessian matrix of $l(\theta)$ is given by

$$l''(\theta) = -\frac{1}{N} \sum_{j=1}^J \frac{Y_j^* \exp(\tilde{\mathbf{X}}_j^\top \theta)}{(1 - \exp(\tilde{\mathbf{X}}_j^\top \theta))^2} \tilde{\mathbf{X}}_j \tilde{\mathbf{X}}_j^\top \omega_h(\mathbf{X}_j, x).$$

Since $l''(\theta)$ is negative definite with probability 1 when $N \rightarrow \infty$, the local log-likelihood Equation (2) has a unique maximiser with respect to θ with probability 1. Let $\hat{\theta}$ be the maximiser of l . Then the first component of $\hat{\theta}$, $\hat{g}(x)$, is our proposed estimator of $g(x)$. Subsequently, our estimator of $p(x)$ is given by $1 - \exp(\hat{g}(x))$.

Remark 1 The log-likelihood of Y_j^* 's conditional on X_{ij} 's instead of Z_{ij} 's is

$$\frac{1}{N} \sum_{j=1}^J \left\{ (1 - Y_j^*) \sum_{i=1}^{n_j} \log q(X_{ij}) + Y_j^* \log \left[1 - \exp \left(\sum_{i=1}^{n_j} \log q(X_{ij}) \right) \right] \right\}.$$

One could suggest to estimate $\log q(x)$ by applying a local Poisson function, i.e. $q(X_{ij}) \approx \exp(\theta_1 + \theta_2(X_{ij} - x))$ and then maximising the following local log-likelihood with respect to (θ_1, θ_2) :

$$\begin{aligned} \tilde{l}(\theta_1, \theta_2) &= \frac{1}{N} \sum_{j=1}^J \left\{ (1 - Y_j^*) \sum_{i=1}^{n_j} (\theta_1 + \theta_2(X_{ij} - x)) \right. \\ &\quad \left. + Y_j^* \log \left[1 - \exp \left(\sum_{i=1}^{n_j} (\theta_1 + \theta_2(X_{ij} - x)) \right) \right] \right\} \prod_{i=1}^{n_j} K_h(X_{ij} - x). \end{aligned}$$

When group sizes are larger than one, the product of kernel functions acts like a multivariate kernel which results in a degraded estimation rate (Fan and Gijbels 1996). Moreover, if we take $K(\cdot)$ to be a kernel function of compact support, such as the Epanechnikov kernel, once one $K_h(X_{ij} - x)$ is zero, the whole product part is zero which impacts the contribution of other X_{ij} 's with nonzero values of $K_h(X_{ij} - x)$. Our truncated version rectifies this problem by counting every X_{ij} in the neighbourhood I_x . Thus, the use of Z_{ij} 's is more informative. One might argue to use $\omega_h(\mathbf{X}_j, x) = K(\|\mathbf{X}_j - x\|/h)$ in place of $\prod_{i=1}^{n_j} K_h(X_{ij} - x)$ above. However, this still acts like a multivariate kernel limiting its use.

The estimator $\hat{\theta}$ has very desirable (and expected) large sample properties under a set of mild regularity conditions. We state these conditions in the appendix. In the following, we denote by $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)^\top$ the value of θ calculated by the true $p(x)$. The first theorem provides the consistency of the estimator and the second theorem provides the large sample distribution of $\hat{\theta}$.

THEOREM 2.1 *Under conditions 1–3 in the appendix we have*

$$H(\hat{\theta} - \theta^*) \rightarrow_p 0,$$

where $H = \text{diag}\{1, h, 1\}$ and \rightarrow_p means converges in probability.

THEOREM 2.2 *Under the same conditions of the theorem above,*

$$\sqrt{Nh}(H(\hat{\theta} - \theta^*) - \text{Bias}_\theta) \rightarrow_d N(0, V_0^{-1}V_1V_0^{-1}),$$

where \rightarrow_d means converges in distribution and Bias_θ , V_0 and V_1 are defined in the appendix.

For any vector γ , let $[\gamma]_1$ be its first element, and for any matrix Γ , let $[\Gamma]_{11}$ be its $(1, 1)$ th element. Then, we have

$$\sqrt{Nh}(\hat{\theta}_1 - \theta_1^* - [\text{Bias}_\theta]_1) \rightarrow_d N(0, [V_0^{-1}V_1V_0^{-1}]_{11}).$$

Our estimate of $p(x)$ is $\hat{p}(x) = 1 - \exp(\hat{\theta}_1)$. Then $\hat{p}(x) - p(x) = -[\exp(\hat{\theta}_1) - \exp(\theta_1^*)]$. By the delta method, we have

COROLLARY 2.3 *Under conditions of Theorem 2, we have*

$$\sqrt{Nh}(\hat{p}(x) - p(x) - B(x)) \rightarrow_d N(0, V(x)),$$

where $B(x) = -(1 - p(x))[\text{Bias}_\theta]_1$ and $V(x) = [1 - p(x)]^2[V_0^{-1}V_1V_0^{-1}]_{11}$ with

$$[\text{Bias}_\theta]_1 = \frac{g^{(2)}(x)\mu_2}{2\mu_0}h^2,$$

$$[V_0^{-1}V_1V_0^{-1}]_{11} = \frac{v_0}{\mu_0} \left[\frac{c\mu_0 + d}{ac\mu_0^2 + ad\mu_0 - b^2\mu_0^2} - \frac{b^2d\mu_0^2}{(ac\mu_0^2 + ad\mu_0 - b^2\mu_0^2)^2} \right],$$

where $\mu_0, \mu_2, v_0, a, b, c$ and d are defined in the appendix.

3. Empirical studies

In this section, we provide a simulation study followed by the analysis of a real data set to illustrate our proposed method.

3.1. Bandwidth selection

It is well known that bandwidth selection is crucial in nonparametric estimation. To save computational cost, we follow Delaigle and Meister (2011) to investigate a plug-in method. Based on Theorem 2, we can write $B_\theta(x) = [\text{Bias}_\theta]_1$ and $V_\theta(x) = [V_0^{-1}V_1V_0^{-1}]_{11}$ to emphasise the dependence of these quantities on x , which are the bias and the asymptotic variance of estimating θ_1 by $\hat{\theta}_1$. A reasonable way to pick the bandwidth h is by minimising a weighted ‘asymptotic mean-integrated-squared error’ given by

$$\text{AMISE}(h) = \int \left[B_\theta^2(u) + \frac{V_\theta(u)}{Nh} \right] \omega(u) \, du,$$

with respect to h for a suitable weight function w . Here, we take $w(u) = f(u)$. This gives

$$\text{AMISE} = \frac{\mu_2^2}{4\mu_0^2} B_\theta h^4 + \frac{V_\theta^*}{Nh},$$

where $B_\theta = \int g^{(2)}(x)f(x) \, dx$, $V_\theta^* = \int V_\theta(x)f(x) \, dx$. Then, the optimal bandwidth is given by

$$h^* = \left(\frac{V_\theta^* \mu_0^2}{B_\theta \mu_2^2} \right)^{-1/5} N^{-1/5}.$$

However, h^* cannot be directly calculated since B_θ and V_θ^* are unknown. We can use

$$\hat{h} = \left(\frac{\hat{V}_\theta \mu_0^2}{\hat{B}_\theta \mu_2^2} \right)^{-1/5} N^{-1/5},$$

by replacing V_θ^* and B_θ with the estimates \hat{V}_θ^* and \hat{B}_θ given below.

We denote G_i as the number of groups of size $\geq i$, where $i = 1, \dots, \max_j n_j$. For each fixed i , we pick $X_{i,j}$, $j = 1, \dots, G_i$ from each group and denote the order statistics by $X_{i,(1)} < X_{i,(2)} < \dots < X_{i,(G_i)}$. For a given estimator $\hat{V}_\theta(X_{i,(j)})$ of $V_\theta(X_{i,(j)})$, let $\hat{V}_i = \sum_{j=1}^{G_i-1} \hat{V}_\theta(X_{i,(j)}) \hat{f}(X_{i,(j)})(X_{i,(j+1)} - X_{i,(j)})$, where $\hat{f}(x)$ is a kernel density estimate of $f(x)$. Then we can estimate V_θ^* by

$$\hat{V}_\theta^* = \sum_{i=1}^{\max_j n_j} w_i \hat{V}_i,$$

where $w_i = \sqrt{G_i} / \sum_{l=1}^{\max_j n_j} \sqrt{G_l}$. Now, it suffices to find a $\hat{V}_\theta(X_{i,(j)})$. We start by deriving a consistent estimate \hat{q}_* of q_* by maximising the full likelihood of Y_j^* 's given by

$$\prod_{j=1}^J \{Y_j^*(1 - q_*^{n_j}) + (1 - Y_j^*)q_*^{n_j}\}.$$

Note that γ_k (defined in the appendix) can be estimated by the proportion of the groups of size n_k among all the groups. In estimating the ratio $q(X_{i,(j)}) / (1 - q_*^{(k)-1} q(X_{i,(j)}))$ which appears in V_{k1} (defined in the appendix), we use $J_k^{-1} \sum_{j:n_j=n^{(k)}} (1 - Y_j^*)$ to estimate the denominator since it is required to be less than 1. Using the arguments of Delaigle and Meister (2011), the numerator can be estimated by $N \hat{\mu}^* \hat{q}_*^{-n_j} (1 - Y_j^*) / \sum_{j=1}^J n_j \hat{q}_*^{n_j-1}$ where $\hat{\mu}^* = N^{-1} \sum_{j=1}^J n_j (1 - Y_j^*)$.

Furthermore, B_θ can be estimated nonparametrically by

$$\hat{B}_\theta = \sum_{i=1}^{\max_j n_j} G_i^{-1} w_i \sum_{j=1}^{G_i} \{\hat{g}_i^{(2)}(X_{ij})\}^2,$$

where the construction of $\hat{g}_i^{(2)}(x)$ is similar to Delaigle and Meister (2011) which is omitted here.

It is well known that nonparametric estimators are in general not stable near boundaries. We replace \hat{B}_θ and \hat{V}_θ by weighted versions (Gasser, Kneip, and Kohler 1991) as $\hat{B}_\theta = \sum_{i=1}^{\max_j n_j} G_i^{-1} w_i \sum_{j=1}^{G_i} \{\hat{g}_i^{(2)}(X_{ij})\}^2 \omega_B(X_{ij})$ and $\hat{V}_\theta = \sum_{j=1}^{G_i-1} \hat{V}_\theta(X_{i,(j)}) \hat{f}(X_{i,(j)})(X_{i,(j+1)} - X_{i,(j)}) \omega_V(X_{i,(j)})$, where $\omega_B(x)$ and $\omega_V(x)$ are two weight functions. Our suggestion is to take $\omega_B(x) = 1_{(q_{0.1}, q_{0.9})}(x)$ and $\omega_V(x) = 1_{(q_{0.3}, q_{0.7})}(x)$, where $1_{(a,b)}(x)$ is the indicator function (it equals to 1 if $a \leq x \leq b$; otherwise 0), and q_α is the α quantile of all the X_{ij} s.

3.2. Numerical simulation

Our numerical studies were conducted to check the finite sample performance of the proposed semi-local likelihood estimator of $p(x)$. We considered the following models each with $X \sim U[-1, 1]$ and $X \sim N(0, 0.5^2)$:

- (1) $p(x) = \{\sin(3\pi x/2) + 1.2\} / [20 + 360x^2 \{\text{sign}(x) + 1\}]$,
- (2) $p(x) = \sin^2(\pi(x-1)/2) \cos^2(1.5\pi(x-1))/6$,
- (3) $p(x) = \cos^2(\pi x)/8$,
- (4) $p(x) = \cos^2(\pi x)/16 + x^2/20$.

The first model is similar to the model used in Delaigle and Meister (2011). The others are designed to have relatively high fluctuant structure. For each model above, we considered both $N = 5000$ and 10,000. The group sizes for equal group size case were $n_j = 5$ or 10. For the unequal group sizes case, n_j 's were randomly and uniformly chosen from $\{1, \dots, 5\}$ or $\{1, \dots, 10\}$. We simulated 200 random samples of $\{(X_{ij}, Y_j^*), i = 1, \dots, n_j, j = 1, \dots, J\}$ for each setting of N, n_j, p and the distribution of X , where $N = \sum_{j=1}^J n_j$, $Y_j^* = \max_{1 \leq i \leq n_j} Y_{ij}$ and Y_{ij} 's are generated according to a Bernoulli distribution with success probability $p(X_{ij})$. The bandwidth, h , was selected using the procedure outlined in Section 3.1. Based on this h , the estimator $\hat{p}(x)$, written LL (local likelihood estimator), of $p(x)$ was calculated. We also applied the method from Delaigle and Meister (2011). These authors provided four ways for selecting the bandwidth, ROT, ROT $_{\omega_0}$, PI $_{\omega_1}$ and PI $_{\omega_0}$. The kernel $K(\cdot)$ was taken to be the standard normal density in all cases and resulting estimates were truncated to be in $[0, 1]$ since $p(x)$ is a probability curve. We compared our estimate with each of their four estimates based on the integrated-squared error (ISE) = $\int_a^b \{\hat{p}(x) - p(x)\}^2 dx \approx M^{-1}(b-a) \sum_{i=1}^M \{\hat{p}(t_i) - p(t_i)\}^2$ for 200 replications, where $[a, b]$ is the interval of interest, and $\{t_i, i = 1, \dots, M\}$ is an even partition of $[a, b]$. Furthermore, to get a feel for the pointwise behaviour of each estimator, we calculated the following pointwise mean square error ratio (PMSER),

$$\text{PMSER}(t_i) = \frac{\sum_{k=1}^{200} \{\hat{p}_k(t_i) - p(t_i)\}^2}{\sum_{k=1}^{200} \{\tilde{p}_k(t_i) - p(t_i)\}^2}, \quad i = 1, \dots, M,$$

where \hat{p}_k is our estimator of p for the k th sample and \tilde{p}_k denotes the estimators proposed by Delaigle and Meister (2011).

In Tables 1–3, we provide a subset of our findings. The average and the standard deviation of the 200 ISEs corresponding to each estimator and the proportion of PMSER(t_i) values < 1 among all the t_i 's for $M = 300$ for $N = 10,000$ are also given. The results for $N = 5000$ followed an almost

Table 1. Simulation results for models 1–4 when group sizes are unequal and X follows uniform, $N = 10^4$.

n_j	Model	LL	ROT	ROT $_{\omega_0}$	PI $_{\omega_1}$	PI $_{\omega_0}$
1–5	1	1.56 (0.76)	3.63 (1.22, 0.76)	4.21 (1.20, 0.77)	2.35 (0.94, 0.66)	1.95 (0.89, 0.67)
	2	7.92 (2.21)	23.7 (1.43, 0.80)	24.4 (1.34, 0.80)	14.9 (2.15, 0.90)	11.8 (2.32, 0.92)
	3	3.61 (1.82)	28.6 (4.79, 0.90)	30.7 (4.25, 0.91)	6.72 (2.60, 0.71)	4.56 (2.15, 0.72)
	4	2.53 (1.16)	7.13 (1.29, 0.80)	7.65 (1.34, 0.81)	3.04 (1.23, 0.63)	2.62 (1.21, 0.63)
1–10	1	2.60 (1.40)	4.61 (1.66, 0.62)	5.12 (1.60, 0.65)	2.97 (1.38, 0.58)	2.80 (1.40, 0.65)
	2	8.93 (3.01)	26.0 (2.30, 0.76)	26.6 (2.26, 0.76)	15.8 (3.28, 0.87)	12.3 (3.32, 0.88)
	3	6.13 (2.86)	30.5 (5.52, 0.87)	32.8 (4.61, 0.87)	9.65 (3.59, 0.71)	7.64 (3.37, 0.71)
	4	4.37 (2.19)	8.89 (2.24, 0.74)	9.47 (2.26, 0.76)	4.70 (2.26, 0.56)	4.52 (2.34, 0.61)

Note: The presented results are: $10^4 \times \text{MISE}$ ($10^4 \times \text{stdev}$, proportion of PMSER < 1).

Table 2. Simulation results for models 1–4 when group sizes are unequal and X follows normal, $N = 10^4$.

n_j	Model	LL	ROT	ROT $_{\omega_0}$	PI $_{\omega_1}$	PI $_{\omega_0}$
1–5	1	3.57 (1.37)	8.80 (1.87, 0.87)	9.40 (1.68, 0.87)	4.58 (1.53, 0.85)	3.88 (1.49, 0.66)
	2	19.7 (5.41)	33.1 (5.29, 0.82)	33.9 (5.40, 0.82)	23.0 (5.08, 0.77)	20.9 (5.40, 0.70)
	3	15.1 (4.64)	40.7 (2.44, 0.95)	41.3 (1.90, 0.94)	22.0 (4.52, 0.88)	17.7 (4.64, 0.62)
	4	8.26 (4.52)	14.6 (3.90, 0.74)	15.0 (3.75, 0.74)	8.79 (4.33, 0.56)	8.44 (4.52, 0.55)
1–10	1	3.47 (1.84)	9.17 (2.68, 0.82)	9.75 (2.56, 0.82)	4.26 (2.18, 0.81)	3.86 (2.19, 0.80)
	2	19.0 (4.19)	30.9 (3.65, 0.84)	31.6 (3.79, 0.82)	21.3 (3.27, 0.74)	19.3 (3.35, 0.63)
	3	11.2 (3.90)	41.5 (3.06, 0.90)	41.9 (2.77, 0.90)	17.7 (4.30, 0.87)	14.6 (4.18, 0.87)
	4	4.99 (2.54)	12.53 (3.01, 0.83)	13.0 (2.9, 0.83)	6.28 (2.76, 0.79)	5.69 (2.55, 0.76)

Note: The presented results are: $10^4 \times \text{MISE}$ ($10^4 \times \text{stdev}$, proportion of PMSER < 1).

Table 3. Simulation results for models 1 and 3 when group sizes are equal, $N = 10^4$.

n_j	Model	$f(x)$	LL	ROT	ROT $_{\omega_0}$	PI $_{\omega_1}$	PI $_{\omega_0}$
5	1	U	2.22 (1.08)	3.87 (1.23, 0.50)	4.43 (1.22, 0.53)	2.67 (1.0, 0.43)	2.28 (1.0, 0.44)
		N	4.59 (1.51)	8.92 (2.08, 0.78)	9.54 (1.94, 0.79)	5.22 (1.55, 0.58)	4.46 (1.47, 0.23)
	3	U	5.56 (2.57)	29.8 (5.06, 0.87)	31.9 (4.38, 0.88)	8.76 (3.36, 0.59)	5.91 (2.79, 0.50)
		N	19.8 (5.27)	41.2 (2.74, 0.91)	41.7 (2.47, 0.91)	26.3 (4.42, 0.81)	21.9 (4.58, 0.49)
10	1	U	3.89 (2.04)	5.28 (1.90, 0.40)	5.72 (1.80, 0.42)	3.71 (1.81, 0.34)	3.48 (1.91, 0.34)
		N	5.94 (3.13)	10.2 (3.33, 0.72)	10.7 (3.27, 0.74)	6.01 (3.08, 0.47)	5.56 (3.13, 0.31)
	3	U	11.7 (5.73)	32.8 (5.6, 0.77)	34.8 (4.99, 0.79)	13.9 (5.42, 0.50)	11.3 (5.38, 0.43)
		N	16.9 (6.84)	42.8 (5.0, 0.85)	43.3 (4.74, 0.86)	24.5 (5.94, 0.80)	20.9 (6.21, 0.76)

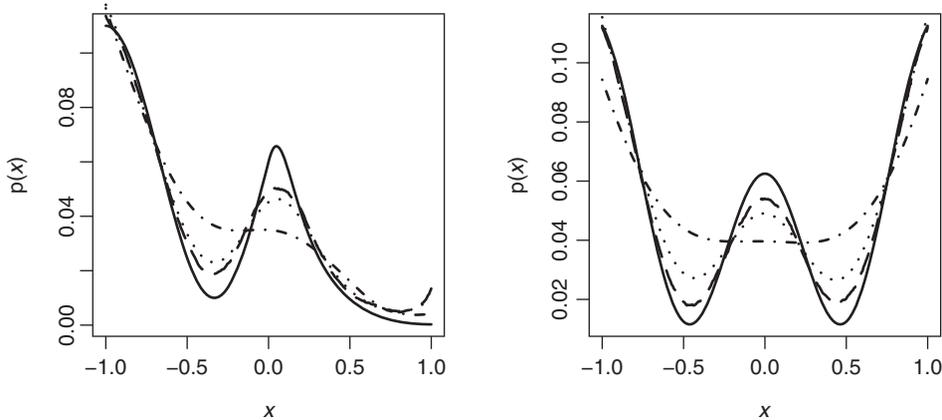
Notes: The presented results are: $10^4 \times \text{MISE}$ ($10^4 \times \text{stdev}$, proportion of PMSER < 1). U and N denote uniform and normal, respectively.

identical pattern and are therefore not presented here. Additionally, global-integrated-squared errors (GISE) = $\int_a^b \{\bar{p}(x) - p(x)\}^2 dx$ were compared, where $\bar{p}(x) = \sum_{k=1}^{200} \hat{p}_k(x)/200$ which is referred to as the average curve for each method.

From Tables 1 and 2, we can see that all means and standard deviations of 200 replications using our method are smaller than the corresponding values for the methods in Delaigle and Meister (2011) for the case of unequal groups. Moreover, the proportion of PMSER value below 1 is greater than 50% in all such cases. For the case of equal groups, a summary is presented in Table 3. The average ISE values and the pointwise mean square error values indicate that the two methods are very similar in the case of equal group sizes. By comparing the GISEs, our method seems to outperform the moment type estimator in all examined cases, a few results are listed in Table 4. Plots of the averaged estimates of $p(x)$, $\bar{p}(x)$, for

Table 4. $10^4 \times$ GISE for models 1–4 when all $n_j = 10$, X is normal and $N = 10^4$.

Model	LL	ROT	ROT $_{\omega_0}$	PI $_{\omega_1}$	PI $_{\omega_0}$
1	2.58	8.56	9.15	2.87	3.55
2	19.9	30.0	30.8	20.0	21.9
3	9.03	39.4	40.1	15.2	19.4
4	2.21	11.0	11.6	3.31	4.31

Figure 1. Average curves: LL (— —), best between ROT and ROT $_{\omega_0}$ (- · - · -), best between PI $_{\omega_1}$ and PI $_{\omega_0}$ (· · ·). Left to right: model 1, $X \sim U[-1, 1]$, $n_j = 10$, $N = 10,000$; model 4, $X \sim U[-1, 1]$, $n_j \sim U\{1, \dots, 10\}$, $N = 5000$.

all models reveal (see Figure 1 for models 1 and 4) that our estimator appears to be significantly less biased over almost the entire support of X . When compared to the estimator proposed in Delaigle and Meister (2011), it is worthwhile to point out that the bias in their estimators becomes more prominent when $p(x)$ is less smooth or more fluctuant. This suggests that our method generally outperforms those proposed in Delaigle and Meister (2011) both globally and locally.

3.3. Real data analysis

We also applied our method to two real data sets from 1999 to 2000 in the NHANES study which were previously analysed by Delaigle and Meister (2011) and are available at www.cdc.gov/nchs/nhanes/nhanes1999-2000/nhanes99_00.htm. The first data set contained two variables: the age variable X and the test result Y_{HBC} which is a binary response taking values 0 and 1, indicating that the antibody to hepatitis B virus core antigen is absent or present in the patient's serum or plasma, respectively. The sample size was 7121, and X ranged from 6 to 85 years after removing the individuals with missing X or Y_{HBC} . The second data set contained the age variable X , and a response variable $Y_{\text{CL}} = 0$ or 1, indicating the absence or presence of genital chlamydia trachomatis infection in the urine of the patient, respectively. After removing the missing values, X ranged from 12 to 40 years, and the sample size was 2042. Our goal is to estimate the following two conditional probability curves: $p_{\text{HBC}}(x) = P(Y_{\text{HBC}}=1|X=x)$ and $p_{\text{CL}}(x) = P(Y_{\text{CL}}=1|X=x)$.

To evaluate the performance of our method, in each case, we first applied the local linear estimation based on all the (X, Y) . The resulting estimator is denoted by \hat{p} and is treated as our

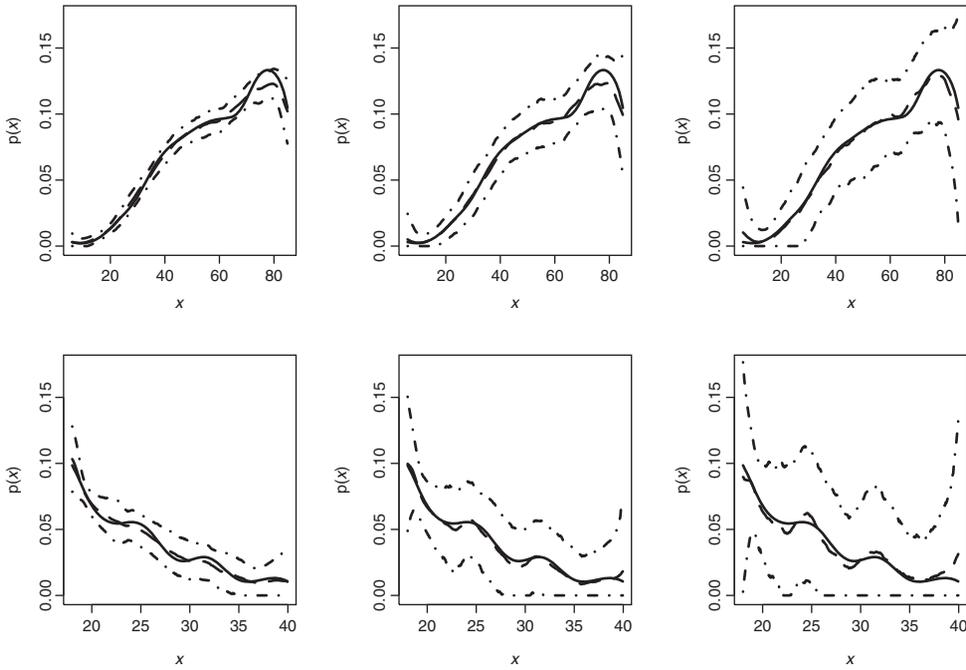


Figure 2. NHANES study: average curves (—) and confidence bands (---) for Y_{HBC} (Top) and Y_{CL} (Bottom). Left to right: $n_j \sim U\{1, 2\}$, $n_j \sim U\{1, \dots, 5\}$, $n_j \sim U\{1, \dots, 10\}$.

reference curve. Then we artificially pooled the data randomly assigning individuals to groups of size $n_j \sim U\{1, 2\}$, $n_j \sim U\{1, \dots, 5\}$, or $n_j \sim U\{1, \dots, 10\}$. In each of these aforementioned cases, we calculated our estimator \hat{p} using the individual-level covariates and the simulated pool responses. This process was then repeated 200 times for both infections on pooling strategy. The average curve along with a two standard deviation pointwise confidence bands based on the 200 replications are presented in Figure 2. Here, the lower band was truncated at 0. From these graphs, it appears that there is a large degree of agreement between our estimator and the reference estimator.

Delaigle and Meister (2011) evaluated their estimator using the estimates corresponding to quantiles of the integrated squared difference (ISD) values, and the estimate corresponding to the median ISD value showed boundary bias. Since we have established the asymptotic normality of our estimator, we prefer to use the average of the estimates with pointwise confidence bands in assessing the estimation accuracy. The average of the 200 estimates shows minimal boundary bias and the ideal curve is well within the pointwise confidence bounds.

4. Discussion

We have provided an effective way of estimating the regression function $P[Y = 1|X]$ based on group data. Our estimator seems to perform well in all possible sampling situations for a variety of model functions. The proposed bandwidth selection procedure seems to provide very satisfactory estimation results. An interesting extension of these ideas would be to test the equality of the regression curves for different populations.

References

- Bilder, C.R., and Tebbs, J.M. (2009), 'Bias, Efficiency, and Agreement for Group-Testing Regression Models', *Journal of Statistical Computation and Simulation*, 79, 67–80.
- Chen, P., Tebbs, J.M., and Bilder, C.R. (2009), 'Group Testing Regression Models with Fixed and Random Effects', *Biometrics*, 65, 1270–1278.
- Delaigle, A., and Hall, P. (2012), 'Nonparametric Regression with Homogeneous Group Testing Data', *Annals of Statistics*, 40, 131–158.
- Delaigle, A., and Meister, A. (2011), 'Nonparametric regression analysis for group testing data', *Journal of the American Statistical Association*, 106, 640–650.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman & Hall, pp. 641, 643, 650.
- Gasser, T., Kneip, A., and Kohler, W. (1991), 'A Flexible and Fast Method for Automatic Smoothing', *Journal of the American Statistical Association*, 86, 643–652.
- Lennon, J.T. (2007), 'Diversity and Metabolism of Marine Bacteria Cultivated on Dissolved DNA', *Applied and Environmental Microbiology*, 73, 2799–2805.
- Nagi, M.S., and Raggi, L.G. (1972), 'Importance to "Airsac" Disease of Water Supplies Contaminated with Pathogenic Escherichia Coli', *Avian Diseases*, 16, 718–723.
- Wahed, M.A., Chowdhury, D., Nermell, B., Khan, S.I., Ilias, M., Rahman, M., Persson, L.A., and Vahter, M. (2006), 'A Modified Routine Analysis of Arsenic Content in Drinking-Water in Bangladesh by Hydride Generation-Atomic Absorption Spectrophotometry', *Journal of Health, Population and Nutrition*, 24, 36–41.

Appendix 1. Regularity conditions

We first state some regularity conditions:

- (1) $\sup_j n_j < \infty$.
- (2) $\log q(x)$ has bounded second-order derivative in a neighbourhood of x , and f is positive and continuous in that neighbourhood.
- (3) $Nh \rightarrow \infty$ and Nh^5 is bounded.

This first condition is also used in Delaigle and Meister (2011). The next two are commonly used conditions on smoothness.

Appendix 2. Notation

We now introduce some notations. Under condition 1, suppose there are only K different group sizes, denoted by $n^{(1)}, \dots, n^{(K)}$. Let J_k be the number of groups of size $n^{(k)}$ and $\lim_{N \rightarrow \infty} n^{(k)} J_k / N = \gamma_k$. Then $\sum_{k=1}^K \gamma_k = 1$. For easy notation, we suppose the data are ordered as follows: the first J_1 groups are of size n_1 , the next J_2 groups are of size n_2 , and so on until the last J_K groups are of size n_K . Now, let $a = \sum_{k=1}^K \gamma_k V_{k0}$, $b = \sum_{k=1}^K \gamma_k (n^{(k)} - 1) V_{k1}$, $c = \sum_{k=1}^K \gamma_k (n^{(k)} - 1)^2 V_{k1}$, $d = \sum_{k=1}^K \gamma_k V_{k0}$, and $e = \sum_{k=1}^K \gamma_k V_{k2}$, where

$$V_{k0} = \frac{n^{(k)} \exp(E_{k0})}{1 - \exp(E_{k0})},$$

$$V_{k1} = \frac{f(x) \exp(E_{k1})}{1 - \exp(E_{k1})},$$

$$V_{k2} = \left[\frac{\exp(E_{k1}) \theta_2^* f(x)}{(1 - \exp(E_{k1}))^2} + \frac{\exp(E_{k1}) f'(x)}{1 - \exp(E_{k1})} \right],$$

with $E_{km} = m\theta_1^* + (n^{(k)} - m) \log q_*$. Further denote

$$V_0 = \begin{pmatrix} a\mu_0 & 0 & b\mu_0 \\ 0 & a\mu_2 & 0 \\ b\mu_0 & 0 & c\mu_0 + d \end{pmatrix}, \quad V_1 = \begin{pmatrix} av_0 & 0 & bv_0 \\ 0 & av_2 & 0 \\ bv_0 & 0 & cv_0 \end{pmatrix} \quad \text{and} \quad \text{Bias}_\theta = V_0^{-1} b_\theta,$$

where $b_\theta = 2^{-1} g^{(2)}(x) \cdot (a\mu_2 h^2, e\mu_4 h^3, b\mu_2 h^2)^\top$, $\mu_i = \int_{-1}^1 u^i K(u) du$ and $v_i = \int_{-1}^1 u^i K^2(u) du$.

Appendix 3. Proof of Theorem 1

Let $\alpha = H(\theta - \theta^*)$, $\hat{\alpha} = H(\hat{\theta} - \theta^*)$ and $\tilde{\mathbf{U}}_j = H^{-1}\tilde{\mathbf{X}}_j$. Put

$$\begin{aligned} l(\alpha) &= \frac{1}{N} \sum_{j=1}^J \{ (1 - Y_j^*) (\tilde{\mathbf{X}}_j^\top \theta^* + \tilde{\mathbf{U}}_j^\top \alpha) + Y_j^* \log [1 - \exp(\tilde{\mathbf{X}}_j^\top \theta^* + \tilde{\mathbf{U}}_j^\top \alpha)] \} \omega_h(\mathbf{X}_j, x) \\ &= \sum_{k=1}^K \frac{J_k}{N} \cdot \frac{1}{J_k} \sum_{j=J_{k-1}+1}^{J_{k-1}+J_k} l_j(\alpha; k), \end{aligned}$$

where $J_0 = 0$ and $l_j(\alpha; k)$ is the kernel-weighted likelihood corresponding to a pooled data of size $n^{(k)}$. Since $l(\alpha)$ is strictly concave, it is sufficient to show that, for any given $\eta > 0$, there exists a small constant ε , such that

$$\liminf_N P \left\{ \sup_{\|\alpha\|=\varepsilon} l(\alpha) < l(0) \right\} = 1 - \eta.$$

By Taylor's expansion around the origin, for any α with $\|\alpha\| = \varepsilon$,

$$l(\alpha) - l(0) = l'(0)^\top \alpha + \frac{1}{2} \alpha^\top l''(0) \alpha + R(\alpha'), \tag{A1}$$

with α' lying between α and 0, and where

$$R(\alpha') = \frac{1}{6} \sum_{j,k,l} \alpha'_j \alpha'_k \alpha'_l \frac{\partial^3 l(\alpha')}{\partial \alpha_j \partial \alpha_k \partial \alpha_l}.$$

First, since for fixed k , $l_j(\alpha; k)$, $j = J_{k-1} + 1, \dots, J_{k-1} + J_k$ are i.i.d., we have

$$l'(0) = \sum_{k=1}^K \frac{J_k}{N} \cdot \frac{1}{J_k} \sum_{j=J_{k-1}+1}^{J_{k-1}+J_k} l'_j(0; k) \rightarrow_p \sum_{k=1}^K \frac{\gamma_k}{n^{(k)}} E[l'_j(0; k)], \tag{A2}$$

where

$$l'_j(0; k) = \left(1 - \frac{Y_j^*}{1 - \exp(\tilde{\mathbf{X}}_j^\top \theta^*)} \right) \tilde{\mathbf{U}}_j \omega_h(\mathbf{X}_j, x).$$

We know $E[l'_j(0; k)] = E_{\mathbf{Z}_j} \{ E[l'_j(0; k) | \mathbf{Z}_j] \}$ with $\mathbf{Z}_j = (Z_{1j}, \dots, Z_{n^{(k)}j})^\top$. It is easy to see that $E[l'_j(0; k) | \mathbf{Z}_j = 0] = 0$. When some of the Z_{ij} 's are not zero, local linear approximation provides that

$$E[l'_j(0; k) | \mathbf{Z}_j] = 1 - \exp(\tilde{\mathbf{X}}_j^\top \theta^*) - \exp(\tilde{\mathbf{X}}_j^\top \theta^*) \frac{g^{(2)}(x)}{2} \sum_{i=1}^{n^{(k)}} I_x(X_{ij})(X_{ij} - x)^2 \{1 + o(1)\}.$$

Applying Taylor's expansion, $E[l'_j(0; k)]$ can be written as

$$E_{\mathbf{Z}_j} \left[\frac{g^{(2)}(x)}{2} \cdot \left\{ \frac{\exp(A)}{1 - \exp(A)} + \frac{\exp(A)\theta_2^*}{(1 - \exp(A))^2} \cdot B \{1 + o(1)\} \right\} \tilde{\mathbf{U}}_j h^{-\sum_{i=1}^{n^{(k)}} I_x(X_{ij})} C \right],$$

where $A = \sum_{i=1}^{n^{(k)}} \{ I_x(X_{ij})\theta_1^* + (1 - I_x(X_{ij}))\theta_3^* \}$, $B = \sum_{i=1}^{n^{(k)}} I_x(X_{ij})(X_{ij} - x)$ and $C = \sum_{i=1}^{n^{(k)}} (X_{ij} - x)^2 \{1 + o(1)\}$. Let M_m be the event that only m of the $I_x(X_{ij})$ s are zero (since X s are i.i.d., without loss of generality, we assume $I_x(X_{1j}) = \dots = I_x(X_{mj}) = 0$). Then conditioning on M_m 's,

$$E[l'_j(0; k)] = \frac{g^{(2)}(x)}{2} \sum_{m=1}^{n^{(k)}} \binom{n^{(k)}}{m} P_x^{(n^{(k)}-m)} P_k^{(m)},$$

where P_x is the probability of an X falling out of I_x , i.e. $P_x = \int_{I_c} f(u) du$, and

$$\begin{aligned} I_k^{(m)} &= \int_{x-h}^{x+h} \dots \int_{x-h}^{x+h} \left\{ \frac{\exp(A_m)}{1 - \exp(A_m)} + \frac{\exp(A_m)\theta_2^*}{(1 - \exp(A_m))^2} \cdot B_m \{1 + o(1)\} \right\} \\ &\quad \times C_m \binom{m}{\binom{m}{n^{(k)}-m}} \frac{1}{h} \prod_{i=1}^m f(X_{ij}) K \left(\frac{X_{ij} - x}{h} \right)^{1/m} dX_{1j} \dots dX_{mj}, \end{aligned}$$

with $A_m = m\theta_1^* + (n^{(k)} - m)\theta_3^*$, $B_m = \sum_{i=1}^m (X_{ij} - x)$ and $C_m = \sum_{i=1}^m (X_{ij} - x)^2 \{1 + o(1)\}$. By conditions 2 and 3, $h \rightarrow 0$. Then $\theta_3^* \rightarrow \log q_*$ and $P_x \rightarrow 1$. We can write $I_k^{(1)} = (\mu_2 V_{k1} h^2, \mu_4 V_{k2} h^3, (n^{(k)} - 1)\mu_2 V_{k1} h^2)^\top$. Simple integration

provides that for $m > 1$, $I_k^{(m)} = o(I_k^{(1)})$. Hence

$$E[l_j'(0; k)] = \frac{n^{(k)} g^{(2)}(x) I_k^{(1)}}{2} \{1 + o(1)\}. \quad (\text{A3})$$

By the assumption $n^{(k)} J_k / N \rightarrow \gamma_k$ and Equation (A2), we can conclude that

$$l'(0) = b_\theta + o_p(1) = o_p(1). \quad (\text{A4})$$

Thus, with probability tending to 1,

$$|l'(0)^\top \alpha| \leq \varepsilon^3. \quad (\text{A5})$$

For $l''(0)$, similarly

$$l''(0) = \sum_{k=1}^K \frac{J_k}{N} \cdot \frac{1}{J_k} \sum_{j=J_{k-1}+1}^{J_{k-1}+J_k} l_j''(0; k) \rightarrow_p \sum_{k=1}^K \frac{\gamma_k}{n^{(k)}} E[l_j''(0; k)], \quad (\text{A6})$$

where

$$l_j''(0; k) = \frac{Y_j^* \exp(\tilde{\mathbf{X}}_j^\top \theta^*)}{(1 - \exp(\tilde{\mathbf{X}}_j^\top \theta^*))^2} \tilde{\mathbf{U}}_j \tilde{\mathbf{U}}_j^\top \omega_h(\mathbf{X}_j, x).$$

When $\mathbf{Z}_j = 0$, $\tilde{\mathbf{U}}_j \tilde{\mathbf{U}}_j^\top = \text{diag}\{0, 0, n^{(k)2}\}$ and $\omega_h(\mathbf{X}_j, x) = 1$. Thus, as $h \rightarrow 0$,

$$E[l_j''(0; k) | \mathbf{Z}_j = 0] \rightarrow \text{diag} \left\{ 0, 0, \frac{n^{(k)2} \exp(E_{k0})}{1 - \exp(E_{k0})} \right\}.$$

When some of Z_{ij} 's are not zero, using the same argument as above, we have

$$E[l_j''(0; k)] = n^{(k)} \begin{pmatrix} V_{k1} \mu_0 & 0 & (n^{(k)} - 1) V_{k1} \mu_0 \\ 0 & V_{k1} \mu_2 & 0 \\ (n^{(k)} - 1) V_{k1} \mu_0 & 0 & (n^{(k)} - 1)^2 V_{k1} \mu_0 + V_{k0} \end{pmatrix} \{1 + o(1)\}.$$

By Equation (A2),

$$l''(0) = -V_0 + o_p(1). \quad (\text{A7})$$

Let $\lambda_{\min}(V_0)$ be the smallest eigenvalue of V_0 . Since V_0 is positive definite, $\lambda_{\min}(V_0)$ is a positive number. Thus, with probability tending to 1,

$$\alpha^\top l''(0) \alpha \leq -\lambda_{\min}(V_0) \varepsilon^2. \quad (\text{A8})$$

Similarly, we can find that

$$|R(\alpha)| = \varepsilon^3 \mathcal{O}_p(1). \quad (\text{A9})$$

Substituting Equations (A5), (A8) and (A9) into (A1), its sign is completely decided by the term of ε^2 when ε is small enough. This completes the proof of Theorem 1.

Appendix 4. Proof of Theorem 2

Continuing to use the notation introduced in the proof of Theorem 1, by Taylor's expansion, we have $0 = l'(\hat{\alpha}) = l'(0) + l''(0) \hat{\alpha} + \mathcal{O}_p(\|\hat{\alpha}\|^2)$. Hence, by Equation (A7),

$$\hat{\alpha} = -\{-V_0 + o_p(1)\}^{-1} l'(0). \quad (\text{A10})$$

It suffices to establish the asymptotic normality of $l'(0)$. By Equation (A4), $E[l'(0)] = b_\theta + o(1)$. For $\text{Var}[l'(0)]$, we have $\text{Var}[l'(0)] = N^{-1} \sum_{k=1}^K (J_k / N) \text{Var}[l_j'(0; k)]$. Since $\text{Var}[l_j'(0; k)] = E[l_j'(0; k) l_j'(0; k)^\top] - E[l_j'(0; k)] E[l_j'(0; k)]^\top$, and

Equation (A3) shows the rate of $E[l'_j(0; k)]$, we only need to find $E[l'_j(0; k)l'_j(0; k)^\top]$ where

$$l'_j(0; k)l'_j(0; k)^\top = \left(1 - \frac{Y_j^*}{1 - \exp(\tilde{\mathbf{X}}_j^\top \theta^*)}\right)^2 \tilde{\mathbf{U}}_j \tilde{\mathbf{U}}_j^\top \omega_h^2(\mathbf{X}_j, x).$$

Using similar argument as in the proof of Theorem 1, we have

$$\text{Var}[l'_j(0; k)] = n^{(k)} \frac{V_{k1}}{h} \begin{pmatrix} v_0 & 0 & (n^{(k)} - 1)v_0 \\ 0 & v_2 & 0 \\ (n^{(k)} - 1)v_0 & 0 & (n^{(k)} - 1)^2 v_0 \end{pmatrix} \{1 + o(1)\}.$$

Combining with the assumption that $J_k/N \rightarrow \gamma_k/n^{(k)}$, we have

$$\text{Var}[l'(0)] = N^{-1}h^{-1}V_1 + o(N^{-1}h^{-1}). \tag{A11}$$

By the Cauchy–Schwarz inequality, V_1 is a singular matrix only when $K = 1$. To make the notation consistent, we treat a constant as a degraded normal random variable with mean being itself and variance being 0. Applying the Cramér–Wold device, we need to show that for any constant vector $b \neq 0$,

$$\sqrt{Nh}\{b^\top l'(0) - b^\top E l'(0)\} \rightarrow_D N\{0, b^\top V_1 b\}. \tag{A12}$$

When $K = 1$ and b is linear to $(-n_1 - 1, 0, 1)^\top$, $\text{Var}[\sqrt{Nh}b^\top l'(0)] \rightarrow b^\top V_1 b = 0$. Otherwise $b^\top V_1 b$ is a positive number. From Equation (A2) and for any fixed k , $b^\top l'_j(0; k)$'s are i.i.d., the normality of Equation (A12) of $b^\top l'(0)$ follows from the central limit theory combining with Equations (A3) and (A11). Consequently, by Equation (A10), it completes the proof.