Statistica Sinica **15**(2005), 597-618

ASYMPTOTIC ANALYSIS OF A TWO-WAY SEMILINEAR MODEL FOR MICROARRAY DATA

Jian Huang and Cun-Hui Zhang

University of Iowa and Rutgers University

Abstract: The cDNA microarray technology is a tool for monitoring gene expression levels on a large scale and has benn widely used in functional genomics. A basic question in analyzing microarray data is proper normalization to ensure meaningful down-stream analyses. We propose a two-way semilinear model for microarray data with two important features. First, it does not require pre-selection of constantly expressed genes or the assumptions that either the percentage of differentially expressed genes is small or there is symmetry in the expression levels of up- and down-regulated genes. Second, when used for deciton of differentially expressed genes, it incorporates variations due to normalization in the assessment of uncertainty in the estimated differences in gene expressions. The proposed model presents novel and challenging theoretical questions in the area of semiparametric statistics due to the presence of infinitely many nonparametric components. We provide theoretical justification that unbiased statistical inference is possible in the two-way semilinear model when self calibration is needed with a large number of parameters. We also prove that the nonparametric optimal rate of convergence can be achieved in estimating the normalization curves under appropriate conditions.

Key words and phrases: Asymptotic theory, gene expression, high-dimensional model, microarray, normalization, semiparametric statistics.

1. Introduction

The cDNA microarray technology is a tool for monitoring gene expression levels on a large scale and has been widely used in functional genomics (Brown and Botstein (1999)). A basic question in analyzing microarray data is normalization for the purpose of removing bias in the observed gene expression levels. Many experimental factors can cause bias in the observed intensity levels, such as differential efficiency of dye incorporation, differences in concentration of DNA on arrays, differences in the amount of RNA labeled between the two channels, uneven hybridizations, differences in the printing pin heads, and so on. Proper normalization is crucially important in ensuring meaningful down-stream analyses, such as detecting differentially expressed genes, clustering co-regulated genes, and classifying biosamples using gene expression profiles.

Normalization is accomplished by establishing a baseline intensity ratio curve from florescent dyes Cy3 and Cy5 across the whole dynamic range for each array. Researchers have considered various normalization methods. For example, the analysis of variance (ANOVA) method has been used for joint normalization and detection of differentially expressed genes (Kerr and Churchill (2000)). This method takes into account the variations due to normalization, but it assumes that normalization is a linear factor in the overall ANOVA model. Another method employs local regression (*loess*, Cleveland (1979, 1986) and Fan and Gijbels (1996)) to first regress the log-intensity ratio on the log-intensity product using all the genes printed on a slide, and then uses the residuals of the regression as the normalized data in the subsequent analysis (Yang, Dudoit, Luu and Speed (2001)). Thus this method takes into account nonlinear normalization effects. However, because one uses all the genes, including those with differential expressions, the resulting normalization curves can be biased, and variations due to normalization are not considered in the subsequent analysis. To alleviate this problem, the *loess* normalization method requires the assumption that either the number of differentially expressed genes is relatively small, or there is symmetry in the expression values of up- and down-regulated genes. If it is expected that many genes have differential expressions, Yang et al. (2000) suggest using dye-swap for normalization. This approach makes the assumption that the normalization curves in the two dye-swapped slides are symmetric. Because of slide-to-slide variation, this assumption may not be satisfied. Fan, Tam, Woude and Ren (2004) introduce a Semi-Linear In-slide Model (SLIM) method for normalization, which is expected to work well with balanced replications of genes within individual arrays (Fan, Peng and Huang (2004)).

By definition, an unbiased normalization curve should be estimated using genes whose expression levels remain constant and cover the whole range of the intensity. Thus Tseng, Oh, Rohlin, Liao and Wong (2001) first used a rank-based procedure to select a set of 'invariant genes' that are likely to be non-differentially expressed, then used these genes in *loess* normalization. However, the set of selected non-differentially expressed genes may not cover the whole dynamic range of the intensity levels. In addition, a threshold value is required in this rank-based selection procedure. How sensitive the final results are to the threshold value may need to be evaluated on a case by case basis.

We propose a two-way semilinear model (TW-SLM) for microarray data. We have introduced the TW-SLM in Huang, Kuo, Koroleva, Zhang, and Soares (2003) and applied our method to a number of microarray datasets. In this paper, we further develop TW-SLM and provide theoretical justification of our method. More experiments with real and simulated data and an alternative theory is provided in Huang, Wang and Zhang (2004), an updated version of Huang et al. (2003).

The TW-SLM grows out of the idea of the *loess* and ANOVA normalization methods. In essence, the TW-SLM is a semiparametric analysis of covariance model that includes nonlinear normalization factors. In addition, as can be seen below, the *loess* method can be considered as the first step in an iterative fitting algorithm in solving the proposed TW-SLM. There are two important features of the TW-SLM. First, when TW-SLM is used for normalization, it does not require the assumption that the percentage of differentially expressed genes is small, nor does it require pre-selection of constantly expressed genes. Second, when TW-SLM is used for detection of differentially expressed genes, it incorporates variations due to normalization in the assessment of uncertainty in the estimated differences in gene expressions.

The TW-SLM presents novel and challenging theoretical questions in the area of semiparametric statistics. In the TW-SLM, the number of genes J is always much greater than the number of arrays n. This fits the description of the well-known "small n, large p" problem (we use p instead of J to be consistent with the phrase used in the literature). In addition, both n and J play the dual role of sample size and number of parameters. For estimating gene effects J is the number of parameters and n is the sample size, but for estimating the normalization curves n is the number of infinite-dimensional parameters and J is the sample size. On one hand, sufficiently large n is needed for the inference of gene effects, but a large n makes normalization more difficult since more nonparametric curves need to be estimated. On the other hand, sufficiently large J is needed for accurate normalization, but then estimation of $\boldsymbol{\beta}$ becomes more difficult. Although there has been intensive research in semiparametric statistics (Bickel, Klaassen, Ritov and Wellner (1993)), we are not aware of any other semiparametric models in which n and J play such dual roles of sample size and number of parameters. Indeed, here the difference between the sample size and the number of parameters is no longer as clear as that in a conventional statistical model. This reflects a basic feature of the microarray data in which self calibration in the data is required when making statistical inference. Our results in this paper provide theoretical justification that, for large J and n, unbiased statistical inference is possible when self calibration is needed. An alternative theory for large J (with large or fixed $n \geq 2$) is provided in Huang, Wang and Zhang (2004) under a different set of regularity conditions.

2. The Two-Way Semiparametric Regression Model

To motivate the model, we first consider the important special case of direct comparison of two cell populations, in which two cDNA samples from the respective cell populations are competitively hybridized on the same slide. Suppose there are J genes and n slides in the study. Let u_{ij} and v_{ij} be the intensity levels of gene j in slide i from the type 1 and the type 2 samples, respectively. Let y_{ij} be the log-intensity ratio of the jth gene in the ith slide, and let x_{ij} be the corresponding average of the log-intensities. That is,

$$y_{ij} = \log_2 \frac{u_{ij}}{v_{ij}}, \quad x_{ij} = \frac{1}{2} \log_2(u_{ij}v_{ij}), \ i = 1, \ldots, n, \ j = 1, \ldots, J. \tag{2.1}$$

The proposed TW-SLM is

$$y_{ij} = f_i(x_{ij}) + \beta_j + \varepsilon_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, J, \tag{2.2}$$

where f_i is the normalization curve for the ith array, $\beta_j \in \mathbb{R}$ is the difference in the expression levels of gene j after normalization, and ε_{ij} is the residual error term. The function f_i is the normalization curve for the ith slide, because it is the difference in the log intensities of red and green channels in the absence of gene effects. Therefore, f_i's represent the experimental effects, and should be removed from the log-intensity ratios. The β_j's are the biologically meaningful effects. We note that in (2.2), it is only made explicit that the normalization curve f_i is slide-dependent. It can also be made dependent upon regions of a slide to account for spatial effects. For example, it is straightforward to extend the model with an additional subscript in (y_{ij}, x_{ij}) and f_i and make f_i also depend on the printing-pin blocks within a slide.

In general, let $z_i \in R^d$ be a covariate vector associated with the ith slide. The general form of the TW-SLM is:

$$y_{ij} = f_i(x_{ij}) + z_i'\beta_j + \varepsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, J, \tag{2.3}$$

where $\beta_j \in R^d$ is the effect associated with the jth gene, z_i' is the transpose of z_i, and f_i and ε_{ij} are as in (2.2).

The covariate vectors z_i can be used to code various types of designs and can include other types of covariates. For example, for the two sample direct comparison design, $z_i = 1, i = 1, \ldots, n$, which is (2.2). For an indirect comparison design using a common reference, we can introduce a two-dimensional covariate vector $z_i = (z_{i1}, z_{i2})'$. Let $z_i = (1, 0)'$ if the ith array is for the type 1 sample versus the reference, and $z_i = (0, 1)'$ if the ith array is for the type 2 sample versus the reference. Now $\beta_j = (\beta_{j1}, \beta_{j2})'$ is a two-dimensional vector and $\beta_{j1} - \beta_{j2}$ represents the difference in the expression levels of gene j after normalization.

We denote the collection of the normalization curves by $\boldsymbol{f} = \{f_1, \ldots, f_n\}$ and the matrix of the gene effects by $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)'$. The TW-SLM is an extension of the semiparametric regression model (SRM) proposed by Engle, Granger, Rice and Weiss (1986) in a study of relationship between weather and electricity sales, while adjusting for other factors. Specifically, if $f_1 = \cdots = f_n \equiv f$ and $J = 1$, then the TW-SLM simplifies to the model of Engle et al. (1986), which has one infinite-dimensional component and one finite-dimensional regression parameter. Much work has been done concerning the properties of the semiparametric least squares estimator (SLSE) in the SRM, see e.g., Heckman (1986), Rice (1986), Chen (1988) and Härdle, Liang and Gao (2000). For example, it has been shown that, under appropriate regularity conditions, the SLSE of the finite-dimensional parameter in the SRM is asymptotically normal, although the rate of convergence of the estimator of the nonparametric component is slower than $n^{1/2}$.

We study the computation, the asymptotic properties of the SLSE of $\boldsymbol{\beta}$ as $(n, J) \to (\infty, \infty)$, and error bounds for normalization. Our results cover the important case of $n/J \to 0$ for the analysis of microarray data in which the number J of genes is always several magnitude greater than the number n of arrays. Because the cost of making cDNA arrays is getting less and less expensive, and because many investigators now use adequate replication to ensure that the analysis results are biologically meaningful, many microarray data sets now have a respectable number of replicated arrays. Therefore, we consider the case $n/J \to 0$ as $(n, J) \to (\infty, \infty)$ as an approximation to the finite sample situation. Our results provide theoretical justifications for the normalization and detection of differentially expressed genes using microarray data in the framework of the proposed TW-SLM.

3. Semiparametric Least Squares Estimation and Computation in the TW-SLM

We assume that the normalization curves can be adequately approximated by linear combinations of certain basis functions. Specifically. let

$$S_i \equiv \overline{\{\psi_{i1}(x) = 1, \psi_{ik}(x), k = 2, \ldots, K_i\}} \tag{3.1}$$

be the spaces of all linear combinations of the basis functions $\psi_{ik}, k \le K_i$. For example, these basis functions can be splines, wavelets, trigonometric functions, or polynomials. We use members of S_i to approximate the normalization curves f_i. Let $\Omega_0^{J\times d}$ be the space of all $J \times d$ matrices $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_J)'$ satisfying $\sum_{j=1}^J \beta_j = 0$. It is clear from the definition of the TW-SLM model (2.2) that $\boldsymbol{\beta}$ is identifiable only up to a member in $\Omega_0^{J\times d}$, since we may simply replace β_j by $\beta_j - \sum_{k=1}^J \beta_k/J$ and $f_i(x)$ by $f_i(x) + \sum_{k=1}^J z_i'\beta_k/J$ in (2.3). In what follows, we assume

$$\boldsymbol{\beta} \in \Omega_0^{J\times d} \equiv \Big\{\boldsymbol{\beta} : \sum_{j=1}^J \beta_j = 0\Big\}. \tag{3.2}$$

Let

$$D^2(\boldsymbol{\beta}, \boldsymbol{f}) = \sum_{i=1}^n \sum_{j=1}^J \Big(y_{ij} - \beta_j' z_i - f_i(x_{ij})\Big)^2. \tag{3.3}$$

We define the SLSE of $\{\boldsymbol{\beta}, \boldsymbol{f}\}$ to be the $\{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}\} \in \Omega_0^{J\times d} \times \prod_{i=1}^n S_i$ that minimizes $D^2(\boldsymbol{\beta}, \boldsymbol{f})$:

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{f}}) = \underset{(\boldsymbol{\beta},\boldsymbol{f})\in\Omega_0^{J\times d}\times\prod_{i=1} S_i}{\arg\min} D^2(\boldsymbol{\beta}, \boldsymbol{f}). \tag{3.4}$$

For computational reasons, it is helpful to write out the definitions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{f}}$ in terms of each other. When $\widehat{\boldsymbol{\beta}}$ is given, the SLSE of f_i is

$$\widehat{f}_i \equiv \underset{f\in S_i}{\arg\min} \sum_{j=1}^J \Big(y_{ij} - \widehat{\beta}_j' z_i - f(x_{ij})\Big)^2,\ i = 1, \ldots n. \tag{3.5}$$

When the normalization curves $\widehat{\boldsymbol{f}}$ are given, the explicit form of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\beta}_j = \Big(\sum_{i=1}^n z_i z_i'\Big)^{-1} \Big(\sum_{i=1}^n z_i\Big(y_{ij} - \widehat{f}_i(x_{ij})\Big) - \frac{1}{J}\sum_{k=1}^J\sum_{i=1}^n z_i\Big(y_{ik} - \widehat{f}_i(x_{ik})\Big)\Big) \tag{3.6}$$

as in standard linear models, provided that $\sum_{i=1}^n z_i z_i'$ is positive definite. Therefore the joint SLSE of $\{\boldsymbol{\beta}, \boldsymbol{f}\}$ can be computed by iterating (3.6) and (3.5) until convergence, with a simple initialization such as $\widehat{f}_i = 0$. Since the square function is strictly convex, the iteration between (3.6) and (3.5) converges monotonically to the sum of residual squares.

We now consider orthogonalization of the design vectors in the TW-SLM. The purpose is to define the observed information matrix for $\boldsymbol{\beta}$ in the presence of the normalization curves $\boldsymbol{f}$. In the cases of smaller values of J, if we use local basis functions for approximating $\boldsymbol{f}$, the orthogonalization can lead to direct computation of the SLSE of $\boldsymbol{\beta}$ without resorting to the iterative procedure described above.

Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iJ})'$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iJ})'$ and $f(\boldsymbol{x}_i) \equiv (f(x_{i1}), \ldots, f(x_{iJ}))'$ for a univariate function f. We write (2.3) in vector notation as

$$\boldsymbol{y}_i = \boldsymbol{\beta} z_i + f_i(\boldsymbol{x}_i) + \epsilon_i, \quad i = 1, \ldots, n. \tag{3.7}$$

We orthogonalize the design vectors for the SLSE as follows. Let

$$V_i \equiv \{f(\boldsymbol{x}_i) : f \in S_i\} = \overline{\{\psi_{ik}(\boldsymbol{x}_i) : k \le K_i\}} \tag{3.8}$$

be the linear spans of the bases in $\mathbb{R}^J$ for approximating vectors $f_i(\boldsymbol{x}_i)$, where S_i are as in (3.1). Let Q_i be the projection matrices from $\mathbb{R}^J$ to V_i with

$$(I_J - Q_i) f(\boldsymbol{x}_i) = 0,\ \forall f \in S_i, \qquad \widehat{K_i} \equiv \mathrm{rank}\Big(Q_i\Big) = \dim\Big(V_i\Big), \tag{3.9}$$

where I_J is the $J \times J$ identity matrix. We show in the Appendix that in (3.4),

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^n \Big\|\boldsymbol{y}_i - (I_J - Q_i)\boldsymbol{\beta} z_i\Big\|^2. \tag{3.10}$$

For $d = 1$ (scalar β_j), $\boldsymbol{\beta}$ is a vector in $\mathbb{R}^J$ and (3.10) is explicitly

$$\widehat{\boldsymbol{\beta}} = \widehat{\Lambda}_{J,n}^{-1} \sum_{i=1}^{n} (I_J - Q_i) \boldsymbol{y}_i z_i', \tag{3.11}$$

since $I_J - Q_i$ are projections in $\mathbb{R}^J$, where $\widehat{\Lambda}_{J,n}$ is the 'observed information matrix'

$$\widehat{\Lambda}_{J,n} \equiv \sum_{i=1}^{n} (I_J - Q_i) z_i^2. \tag{3.12}$$

For $d > 1$, (3.10) is still given by (3.11) with

$$\widehat{\Lambda}_{J,n} \equiv \sum_{i=1}^{n} (I_J - Q_i) \otimes z_i z_i'. \tag{3.13}$$

The information operator (3.13) is a sum of tensor products, i.e., a linear mapping from $\Omega_0^{J\times d}$ to $\Omega_0^{J\times d}$ defined by $\widehat{\Lambda}_{J,n}\boldsymbol{\beta} \equiv \sum_{i=1}^{n}(I_J - Q_i)\boldsymbol{\beta} z_i z_i'$.

Here and in the sequel, $\|\cdot\|$ denotes the Euclidean norm, and A^{-1} denotes the generalized inverse of matrix A, defined by $A^{-1}\boldsymbol{x} \equiv \arg\min\{\|\boldsymbol{b}\| : A\boldsymbol{b} = \boldsymbol{x}\}$. If A is a symmetric matrix with eigenvalues λ_j and eigenvectors $\mathbf{v}_j$, then $A = \sum_j \lambda_j \mathbf{v}_j \mathbf{v}_j'$ and $A^{-1} = \sum_{\lambda_j \neq 0} \lambda_j^{-1} \mathbf{v}_j \mathbf{v}_j'$. The generalized inverse of $\widehat{\Lambda}_{J,n}$ is defined by treating $\Omega_0^{J\times d}$ as a subspace of the Euclidean vector space $\mathbb{R}^{dJ}$.

There is a clear interpretation for the above expressions from the semiparametric information calculation point of view. The projection Q_i is from the sample space of $\boldsymbol{y}_i$ to the approximation space S_i for f_i. It 'spends' part of the information in the data for estimating the unknown normalization curve f_i. Thus the remaining information for estimating $\boldsymbol{\beta}$ is the total information minus the information spent on $\boldsymbol{f}$, which is reflected in $I_J - Q_i$ (3.13).

Example 1. Polynomial spline SLSE: Let $b_1, \ldots, b_{K_i}$ be K_i B-spline base functions (Schumaker (1981)). We approximate f_i by $s_i(x) = \alpha_{i0} + \sum_{k=1}^{K_i} b_k(x)\alpha_{ik} \equiv \boldsymbol{b}_i(x)'\alpha_i \in S_i$, where $\boldsymbol{b}_i(x) = (1, b_1(x), \ldots, b_{K_i}(x))'$, and $\alpha_i = (\alpha_{i0}, \alpha_{i1}, \ldots, \alpha_{iK_i})'$ are coefficients to be estimated from the data. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. The LS objective function is

$$D(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{J} [y_{ij} - \boldsymbol{b}_i(x_{ij})'\alpha_i - z_i'\beta_j]^2. \tag{3.14}$$

Let $B_{ij} = (1, b_1(x_{ij}), \ldots, b_{K_i}(x_{ij}))'$ be the spline basis functions evaluated at $x_{ij}, 1 \le i \le n, 1 \le j \le J$. The spline basis matrix for the ith array is

$$B_i = \begin{pmatrix} B_{i1}' \\ \vdots \\ B_{iJ}' \end{pmatrix} = \begin{pmatrix} 1 & b_1(x_{i1}) & \ldots & b_{K_i}(x_{i1}) \\ \vdots & & \vdots & \\ 1 & b_1(x_{iJ}) & \ldots & b_{K_i}(x_{iJ}) \end{pmatrix}.$$

The projection matrix Q_i defined in (3.9) is $Q_i = B_i(B_i'B_i)^{-1}B_i'$, $i = 1, \ldots, n$. The iterative algorithm described earlier becomes the following. Set $\boldsymbol{\beta}^{(0)} = 0$.

Step 1: Compute $\boldsymbol{\alpha}^{(k)}$ by minimizing $D_w(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(k)})$ with respect to $\boldsymbol{\alpha}$ (the explicit solution is $\alpha_i^{(k)} = (B_i'B_i)^{-1}B_i'(\boldsymbol{y}_i - \boldsymbol{\beta}^{(k)} z_i)$, $i = 1, \ldots, n$).

Step 2: For the $\boldsymbol{\alpha}^{(k)}$ computed above, obtain $\boldsymbol{\beta}^{(k+1)}$ by minimizing $D_w(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\alpha}$ (the explicit solution is given by (3.6)).

Iterate between Steps 1 and 2 until the desired convergence criterion is satisfied. Because the objective function is strictly convex, the algorithm converges to the unique global optimal point. Suppose that the algorithm meets the convergence criterion at step K. Then the estimated values of β_j are $\widehat{\beta}_j = \beta_j^{(K)}$, $j = 1, \ldots, J$, and the estimated normalization curves are $\widehat{f}_i(x) = \boldsymbol{b}_i(x)'\alpha_i^{(K)}$, $i = 1, \ldots, n$.

Local regression (loess) method: The loess method can also be used to estimate the TW-SLM. Let W_λ be a kernel function with window width λ. Let $s_p(t; \boldsymbol{\alpha}, x) = \alpha_0(x) + \alpha_1(x)t + \cdots + \alpha_p(x)t^p$ be a polynomial in t with order p, where $p = 1$ or 2 are common choices. The objective function of the *loess* method for the TW-SLM is

$$M_L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{J} W_\lambda(x_{ij}, x_{ik}) \left(y_{ij} - s_p(x_{ik}, \boldsymbol{\alpha}, x_{ij}) - z_i'\beta_j \right)^2 . \tag{3.15}$$

Let $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ be the value that minimizes M_L. The *loess* estimator of f_i at x_{ij} is $\widehat{f}_i(x_{ij}) = s_p(x_{ij}, \widehat{\boldsymbol{\alpha}}, x_{ij})$.

Again, the back-fitting algorithm can be used to compute the *loess* estimators. We expect that the performance and asymptotic properties of the loess estimator and the spline estimator in the TW-SLM are similar. However, it appears to be more difficult to work out the technical details for the loess estimators. It is clear that if we set the initial value $\boldsymbol{\beta} = 0$, then the values of $\boldsymbol{\alpha}$ resulting from the *first iteration* in the back-fitting gives the *loess* normalization curves (Yang et al. (2001); Tseng et al. (2001)).

4. Distributional Theory and Normalization Error Bounds

In this section we develop methodologies for statistical inference about $\boldsymbol{\beta}$ based on (3.4) and (3.13), and provide error bounds for normalization. We provide limiting distributions for certain pivotal quantities involving $\boldsymbol{\beta}$ and individual β_j and the resulting approximate confidence regions and intervals for large n and J. Our results allow the nonconventional situation of $n/J \to 0$, which is especially appropriate for microarray data. We assume throughout the sequel that z_i are deterministic covariates. The proofs of our results are given in the Appendix.

4.1. Distributions of pivotal quantities and approximate confidence intervals

Unless otherwise stated, we assume in this section that ϵ_{ij} are i.i.d. $N(0,\sigma^2)$ variables given all the covariate variables. The normality condition can be weakened, but it is not a main concern in this paper. The unknown error variance σ^2 can be estimated by the residual mean squares in (3.4):

$$\widehat{\sigma}^2 = \Big(Jn - \widehat{\nu} - \sum_{i=1}^{n} \widehat{K_i}\Big)^{-1} \min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \min_{f\in S_i} \Big\|\boldsymbol{y}_i - \boldsymbol{\beta} z_i - f(\boldsymbol{x}_i)\Big\|^2, \tag{4.1}$$

where $\widehat{K_i}$ are the dimensions of V_i in (3.9) and $\widehat{\nu}$ is the rank of the observed information operator $\widehat{\Lambda}_{J,n}$ as a linear mapping in $\Omega_0^{J\times d}$. Conditionally on the covariates, $\widehat{\sigma}^2/\sigma^2$ is the ratio of a non-central chi-square variable and its degrees of freedom.

Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_{J,n}$ be as in (3.4) and (3.13), or equivalently (3.10) and (3.13). Define

$$\Sigma_n \equiv \sum_{i=1}^{n} z_i z_i', \quad \sigma_n \equiv \sum_{i=1}^{n} \|z_i\|^2, \quad \widehat{\Sigma}_{J,n} \equiv \sum_{i=1}^{n} \frac{(J-\widehat{K_i}) z_i z_i'}{J-1}, \tag{4.2}$$

with the $\widehat{K_i}$ in (3.9). Let $\chi^2_{1-\alpha,\nu}$ be the $(1-\alpha)$-quantile of the χ^2-distribution with ν degrees of freedom. Our confidence regions for $\boldsymbol{\beta}$ are based on the distributional approximations

$$P\Big\{\sum_{i=1}^{n} \|(I_J - Q_i)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) z_i\|^2/\widehat{\sigma}^2 \le \chi^2_{1-\alpha,\widehat{\nu}}\Big\} \approx 1-\alpha, \tag{4.3}$$

with $\widehat{\nu}$ being the rank of $\widehat{\Lambda}_{J,n}$, together with

$$P\Big\{\sum_{j=1}^{J} (\widehat{\beta}_j - \beta_j)' \widehat{\Sigma}_{J,n} (\widehat{\beta}_j - \beta_j)/\widehat{\sigma}^2 \le \chi^2_{1-\alpha,d(J-1)}\Big\} \approx 1-\alpha, \tag{4.4}$$

$$\frac{\big(\sum_{j=1}^{J} \|\widehat{\beta}_j - \beta_j\|^2\big)/\widehat{\sigma}^2 - (J-1)\mathrm{trace}(\widehat{\Sigma}^{-1}_{J,n})}{\{2J\,\mathrm{trace}(\widehat{\Sigma}^{-2}_{J,n})\}^{1/2}} \xrightarrow{\mathrm{D}} N(0,1). \tag{4.5}$$

Our inference procedures about individual β_j are based on

$$\Big(\widehat{\Sigma}^{1/2}_{J,n}/\widehat{\sigma}\Big)\Big(\widehat{\beta}_j - \beta_j\Big) \xrightarrow{\mathrm{D}} N(0, I_d). \tag{4.6}$$

This should be compared with the case of known f_i in which the LSE's of β_j are multivariate normal vectors with means β_j and covariance $\sigma^2\Sigma_n^{-1}$. In practice, $\max_{i\le n}\widehat{K}_i \le \max_{i\le n} K_i = o(J)$ so

$$(1+o(1))\Sigma_n \le \Sigma_n \min_{i\le n}(J-\widehat{K}_i)/(J-1) \le \widehat{\Sigma}_{J,n} \le \Sigma_n.$$

The above approximations are obtained under "smoothness" conditions on f_i of the form

$$J^{q/2}\sum_{i=1}^n \frac{\rho_{J,i}^2}{n} \to 0, \quad J^{q/2}\frac{\sum_{i=1}^n \|z_i\|^{2p}\rho_{J,i}^2}{\left(\sum_{i=1}^n \|z_i\|^2\right)^p} \to 0, \qquad \text{as } (J,n)\to(\infty,\infty), \tag{4.7}$$

where $\rho_{J,i}^2$ are the distances between the vectors $f_i(\boldsymbol{x}_i)$ and the approximating spaces (3.8),

$$\rho_{J,i} \equiv \Big[E\min_{f\in S_i}\Big\{\frac{1}{J}\sum_{j=1}^J \Big|f_i(x_{ij})-f(x_{ij})\Big|^2\Big\}\Big]^{1/2}, \tag{4.8}$$

with the S_i in (3.1). For (4.4), (4.5) and (4.6), we further assume the following.

Condition A. *The random vectors $\boldsymbol{x}_i, i\le n$, are pairwise independent and, for each i, the $\{x_{ij}, j\le J\}$ are exchangeable. Moreover, for each i the space S_i in* (3.1) *depends on the data only through the covariates $\{z_i, i\le n\}$ and the set $\{x_{ij}, j\le J\}$.*

Theorem 1. *Suppose $K_i \le \kappa^* J$ in* (3.1) *for certain fixed $0<\kappa^*<1$.*
(i) *If* (4.7) *holds for $(p,q)=(0,1)$, then* (4.3) *holds as $(J,n)\to(\infty,\infty)$.*
(ii) *Let Σ_n and σ_n be as in* (4.2). *Suppose Condition A holds and*

$$\frac{\lambda_{\max}(\Sigma_n)}{\lambda_{\min}(\Sigma_n)} = O(1), \quad \sigma_n^{-2}\sum_{i=1}^n \|z_i\|^4 E\widehat{K}_i = o(1), \tag{4.9}$$

where $\widehat{K}_i$ is as in (3.9). *If* (4.7) *holds with $(p,q)=(1,1)$, then* (4.3), (4.4) *and* (4.5) *hold as $(J,n)\to(\infty,\infty)$. If* (4.7) *holds with $(p,q)=(1,0)$, then* (4.6) *holds uniformly in j:*

$$\sup_{j\le J}\sup_{\|b\|=1}\sup_t \Big|P\Big\{b'(\widehat{\Sigma}_{J,n}^{1/2}/\widehat{\sigma})\Big(\widehat{\beta}_j-\beta_j\Big)\le t\Big\} - P\Big\{N(0,1)\le t\Big\}\Big| = o(1), \tag{4.10}$$

where $b\in\mathbb{R}^d$. Moreover, $P\{\widehat{\nu}=(J-1)d\}\to 1$, where $\widehat{\nu}$ is the rank of $\widehat{\Lambda}_{J,n}$.

Corollary 1. *Theorem* 1 *holds with* (4.7) *replaced by $J^{q/2}\sum_{i=1}^n \rho_{J,i}^2/n^{p\wedge 1} = o(1)$ and* (4.9) *replaced by $\sum_{i=1}^n K_i/n^2 \to 0$, where K_i are as in* (3.1), *provided that*

$$\limsup_{n\to\infty}\Big(\max_{i\le n}\|z_i\|\Big) < \infty, \quad \liminf_{n\to\infty}\Big(\frac{\lambda_{\min}(\Sigma_n)}{n}\Big) > 0. \tag{4.11}$$

Suppose spline, wavelet or certain other bases are used in (3.1) with a cut-off K_i for functions with smoothness index $\alpha > 0$, so that with $\|Q_i\|_2^2 \equiv \mathrm{trace}(Q_i'Q_i)$ and M^* depending on (α, γ, M),

$$\max_{i\le n} \|Q_i\|_2^2 \sim J^{1/(2\alpha+1)}, \quad \sup_{f\in\mathcal{F}_{\gamma,M}} \max_{i\le n} E\|(I_J - Q_i)f(\boldsymbol{x}_i)\|^2 \le \frac{M^*J}{J^{2\gamma/(2\alpha+1)}}, \tag{4.12}$$

for certain (e.g., Lipchitz or Sobolev) classes $\mathcal{F}_{\gamma,M}$ with smoothness indices $\gamma \le \alpha$.

Corollary 2. *Suppose Condition A, the first part of* (4.9), (4.11) *and* (4.12) *hold for certain* $\alpha > 0$ *with* $J/n^{2\alpha+1} \to 0$*. If* $\{f_i\} \subseteq \mathcal{F}_{\gamma,M}$ *for certain* γ *satisfying* $\alpha/2 + 1/4 < \gamma \le \alpha$*, then* (4.3), (4.4) *and* (4.5) *hold as* $(J, n) \to (\infty, \infty)$*. If* f_i *are uniformly continuous with a common compact support, then* (4.10) *holds.*

Approximate confidence regions and intervals can be easily constructed based on Theorem 1 and Corollaries 1 and 2. By the Central Limit Theorem applied to i.i.d. χ_1^2 variables, $\chi^2_{1-\alpha,\nu}$ can be replaced by $\nu + \chi_{1-2\alpha,1}\sqrt{2\nu}$ for $\alpha \le 1/2$. For $d = 1$, (4.4) or (4.5) give 95% confidence regions $\widehat{\Sigma}_{J,n}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2/\widehat{\sigma}^2 \le J + 1.645\sqrt{2J}$, and (4.10) gives 95% confidence intervals $\sqrt{\widehat{\Sigma}_{J,n}}|\widehat{\beta}_j - \beta_j|/\widehat{\sigma} \le 1.96$. If $K_i \le \kappa^* J$ and (4.9) holds, then all the eigenvalues of Σ_n and $\widehat{\Sigma}_{J,n}$ are of the order σ_n. If (4.11) holds as in Corollaries 1 and 2, then σ_n is of the order n.

4.2. Bounds on the normalization error

It is of interest to assess the quality of normalization provided by the TW-SLM, we turn to that here.

If $\boldsymbol{\beta}$ were known, we could have used many suitable smoothing method to estimate f_i, cf. Fan and Gijbels (1996), Efromovich (1999) and Hastie, Tibshirani and Friedman (2001), to generate ideal normalizing curves. Consider linear ideal normalizing curves of the form

$$\widetilde{f}_i(\boldsymbol{x}_i) \equiv Q_i^*(\boldsymbol{y}_i - \boldsymbol{\beta} z_i), \tag{4.13}$$

where Q_i^* are linear mappings depending on covariates. Since $\boldsymbol{\beta}$ is not available to us, we could use

$$\widehat{f}_i^*(\boldsymbol{x}_i) \equiv Q_i^*(\boldsymbol{y}_i - \widehat{\boldsymbol{\beta}} z_i) \tag{4.14}$$

instead of (4.13). In this case the normalized data are $\widehat{y}^*_{ij} \equiv y_{ij} - \widehat{f}_i^*(x_{ij})$, $1 \le i \le n, 1 \le j \le J$, while the unobservable ideally normalized data are $\widetilde{y}_{ij} \equiv y_{ij} - \widetilde{f}_i(x_{ij})$, $1 \le i \le n, 1 \le j \le J$. The problem of comparing the normalized data $\widehat{y}^*_{ij}$ and the unobservable ideally normalized data $\widetilde{y}_{ij}$ becomes that of comparing $\widehat{f}_i^*$ and $\widetilde{f}_i$. Theorem 4 below provides upper bounds for the differences $\widehat{y}^*_{ij} - \widetilde{y}_{ij} = \widetilde{f}_i(x_{ij}) - \widehat{f}_i^*(x_{ij})$ between the actual and ideal normalized data. We use two norms for linear mappings A (including matrices and tensor products): the operator norm $\|A\| \equiv \{\|A\mathbf{v}\| : \|\mathbf{v}\| = 1\} = \lambda_{\max}(A'A)$, and the Hilbert-Schmidt norm $\|A\|_2 \equiv \{\text{trace}(A'A)\}^{1/2}$.

Theorem 2. *Suppose $\{\varepsilon_{ij}\}$ are uncorrelated with $E\varepsilon_{ij}^2 \le \sigma^2$.*

(i) *Suppose $K_i \le \kappa^* J$ for certain $\kappa^* < 1$, and that* (4.9) *and Condition A hold. Then,*

$$\frac{1}{J}\Big\|\widehat{f}_i^*(\boldsymbol{x}_i) - \widetilde{f}_i(\boldsymbol{x}_i)\Big\|^2 \le O_P(\|z_i\|^2)\Big(\frac{\|Q_i^*\|^2}{\sigma_n^2}\sum_{k=1}^n \|z_k\|^2\rho_{J,k}^2 + \frac{\|Q_i^*\|_2^2}{J\sigma_n}\Big). \quad (4.15)$$

(ii) *Suppose Condition A, the first part of* (4.9), (4.11) and (4.12) *hold for certain $\alpha > 0$ with $J/n^{2\alpha+1} \to 0$. Suppose $\|z_i\|\|Q_i^*\| = O_P(1)$ and $\|z_i\|^2\|Q_i^*\|_2^2 = O_P(J^{1-2\gamma/(2\alpha+1)})$ uniformly in $i \le n$. If $\{f_i\} \subset \mathcal{F}_{\gamma,M}$ then, uniformly in $i \le n$,*

$$\frac{1}{J}\Big\|\widehat{f}_i^*(\boldsymbol{x}_i) - \widetilde{f}_i(\boldsymbol{x}_i)\Big\|^2 = o_P(J^{-(2\gamma+1)/(2\alpha+1)}). \quad (4.16)$$

Consequently, uniformly in i, $\widehat{f}_i^*$ converges at the optimal rate

$$\frac{1}{J}\Big\|\widehat{f}_i^*(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_i)\Big\|^2 = O_P(J^{-2\gamma/(2\gamma+1)}), \quad (4.17)$$

provided that (4.12) *holds with (α, Q_i) replaced by (γ, Q_i^*) and $(2\gamma+1)^2 \ge 2\gamma(2\alpha+1)$.*

Remark. Suppose conditions of Theorem 2 (ii) hold. If (4.12) holds with (α, Q_i) replaced by (γ, Q_i^*) and $(2\gamma+1)^2 \ge 2\gamma(2\alpha+1)$, then in the case of $\boldsymbol{\beta} = 0$, Q_i^* provides the optimal rate of convergence in mean for $f_i \in \mathcal{F}_{\gamma,M}$:

$$\begin{aligned}\frac{1}{J}E\Big\|\widetilde{f}_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_i)\Big\|^2 &= \frac{1}{J}E\Big\|(I_J - Q_i^*)f_i(\boldsymbol{x}_i)\Big\|^2 + \frac{\sigma^2}{J}E\|Q_i^*\|_2^2 \\ &= O(J^{-2\gamma/(2\gamma+1)}). \quad (4.18)\end{aligned}$$

Moreover, if $J/n^{(2\gamma+1)^2/(2\gamma)} \to 0$, then the estimators $\widehat{f}_i^*$ achieve the optimal convergence rate in probability for all $\boldsymbol{\beta}$ with $K_i \sim J^{1/(2\alpha+1)}$ and $\|Q_i^*\|_2^2 \sim J^{1/(2\gamma+1)}$, where $\alpha = (2\gamma+1)^2/(4\gamma) - 1$. If (4.12) holds with $J/n^{2\alpha+1} \to 0$ and $\{f_i\} \subset \mathcal{F}_{\alpha,M}$, then $|\widehat{f}_i^*(x_{ij}) - \widetilde{f}_i(x_{ij})|^2$ is small, $o(1/J)$ in average, with large probability, so that $Q_i^* = Q_i$ can be directly used to achieve the optimal convergence rate

$$\frac{1}{J}\Big\|\widehat{f}_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_i)\Big\|^2 = O_P(J^{-2\alpha/(2\alpha+1)}). \quad (4.19)$$

4.3. Inequalities about the SLSE and the observed information operator

A crucial step in proving Theorems 1 and 2 is to understand the effects of approximation errors for the unknown functions f_i on the SLSE of $\boldsymbol{\beta}$ in (3.10) and the distribution of eigenvalues of the observed information operator (3.13). Here we provide upper bounds for the effects of the approximation errors and the variance of the observed information operator.

We measure the variance of $\widehat{\Lambda}_{J,n}$ by $E\|\widehat{\Lambda}_{J,n} - E\widehat{\Lambda}_{J,n}\|_2^2$, where $\|\cdot\|_2$ is the Hilbert-Schmidt norm. Let $\langle A, B\rangle_2 \equiv \text{trace}(B'A)$ be the inner product of matrices A and B of common finite-dimensions, so that the Hilbert-Schmidt norm of A is $\|A\|_2 \equiv \langle A, B\rangle_2^{1/2}$. The inner-product $\langle\cdot,\cdot\rangle_2$ and the Hilbert-Schmidt norm $\|\cdot\|_2$ for tensor products are defined by treating them as linear mappings. For example, for tensor products $A_j \otimes B_j$ of common dimensions, $\langle A_1 \otimes B_1, A_2 \otimes B_2\rangle_2 = \langle A_1, A_2\rangle_2 \langle B_1, B_2\rangle_2$. Let

$$\widehat{\Omega}_{J,n} \equiv \left\{\widehat{\Lambda}_{J,n}\boldsymbol{\beta} : \boldsymbol{\beta} \in \Omega_0^{J\times d}\right\} \subseteq \Omega_0^{J\times d} \tag{4.20}$$

be the range of (3.13). Given $X \equiv (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\widehat{\Omega}_{J,n}$ is a linear space of finite dimension, thus we can define a standard normal random matrix $Z \in \widehat{\Omega}_{J,n}$ such that, conditionally on X,

$$\langle A, Z\rangle_2 \sim N\Big(0, \|A\|_2^2\Big), \quad \forall A \in \widehat{\Omega}_{J,n}. \tag{4.21}$$

Theorem 3. *Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_{J,n}$ be as in* (3.10) *and* (3.13). *Set $Z_n \equiv \widehat{\Lambda}_{J,n}^{-1/2} \sum_{i=1}^n (I_J - Q_i)\varepsilon_i z_i'$. Then, $\widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - Z_n$ is an unobservable matrix-valued function of covariates X and $\{z_i, i \le n\}$ such that*

$$E\Big\|\widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - Z_n\Big\|_2^2 \le J \sum_{i=1}^n \rho_{J,i}^2 \,. \tag{4.22}$$

Moreover, if the errors $\{\varepsilon_{ij}, i \le n, j \le J\}$ in (2.2) *are i.i.d. $N(0, \sigma^2)$ given X, then Z_n/σ given X is a standard normal matrix in $\widehat{\Omega}_{J,n}$ as in* (4.21).

Theorem 3 is derived from the standard theory of linear models. Theorem 4 below provides much stronger results under Condition A on the distribution of covariate vectors $\boldsymbol{x}_i$. This theorem is a key step in establishing Theorem 1 (ii) and Theorem 2.

Theorem 4. *Suppose $K_i \le \kappa^* J$ for certain $\kappa^* < 1$ and* (4.9) *and Condition A hold. Then*

$$\Big\|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \widehat{\Lambda}_{J,n}^{-1/2} Z_n\Big\|_2^2 \le O_P(J/\sigma_n^2) \sum_{i=1}^n \|z_i\|^2 \rho_{J,i}^2, \tag{4.23}$$

with $\sigma_n \equiv \sum_{i=1}^n \|z_i\|^2$ *as in* (4.2). *With* $\mathbf{e}_j = (0,\ldots,0,1,0\ldots,0)'$,

$$\max_{1\le j\le J} E\min\Big\{1,\sigma_n\Big\|(\widehat\beta_j-\beta_j)-\Big(\widehat\Lambda_{J,n}^{-1/2}Z_n\Big)'\mathbf{e}_j\Big\|^2\Big\}\le o(1)+O(\sigma_n^{-1})\sum_{i=1}^n\|z_i\|^2\rho_{J,i}^2, \tag{4.24}$$

as $(J,n)\to(\infty,\infty)$. *Moreover,* $P\{\widehat\nu = d(J-1)\}\to 1$ *with* $\widehat\nu$ *being the rank of* $\widehat\Lambda_{J,n}$ *and*

$$E\Big\|\widehat\Lambda_{J,n}-I_{J,0}\otimes\widehat\Sigma_{J,n}\Big\|_2^2\le\sum_{i=1}^n\|z_i\|^4E\widehat{K}_i=o(\sigma_n^2), \tag{4.25}$$

where $I_{J,0}\equiv I_J - J^{-1}\mathbf{e}\mathbf{e}'$ *with* $\mathbf{e}\equiv(1,\ldots,1)'$.

5. Concluding Remarks

We have shown that, under appropriate conditions, statistical inference about $\boldsymbol\beta$ in the TW-SLM can be carried out in the same order of precision as in a regular semiparametric model. This suggests that some important inference tools, such as the bootstrap, can be consistently applied to the TW-SLM.

It was not intuitively clear to us at the outset whether many β_j might be inestimable due to the singularities of (3.12) or (3.13) if reasonably rich approximation spaces S_i are used to estimate the f_i in (3.4). We were particularly intrigued by the presence of the large number of nonparametric components $f_i, i=1,\ldots,n$, where n is the sample size for estimating $\boldsymbol\beta$. Thus, from the asymptotic point of view, the TW-SLM is an *infinite*-dimensional semiparametric model. In contrast, in the standard semiparametric models, such as the semiparametric regression model and the proportional hazards model, there is only one or a fixed number of nonparametric components. This appears to be a key distinction between the TW-SLM and the standard semiparametric models and renders that the existing theory for semiparametric models (Bickel et al. (1993)) inapplicable here.

There are several interesting and challenging questions that have not been addressed in this paper. For example, it is of interest to extend our results to robust estimators (Huber (1981)) of the TW-SLM. Our analysis makes essential use of the fact that the least squares estimators can be considered as orthogonal projections. The second extension is to allow heteroscedasticity in the TW-SLM. For microarray data, this is desirable since the variability of the intensity ratios usually tend to be higher in the low intensity range than in the high intensity range. However, both the computation and the theoretical analysis of a heteroscedasticity TW-SLM will be more complicated. The third question is how to incorporate correlation into the TW-SLM. This may provide a way of identifying groups of genes that have differential expressions, instead of a single gene at a time. This is useful, because genes tend to express in a coordinated fashion corresponding to different functional groups. This appears to be a difficult modeling problem because of the high dimension of any typical gene expression data set. Integration of known biological functions of the genes under study will be essential to make such modeling exercises successful.

Acknowledgements

The research of Jian Huang is supported in part by the NIH grants MH001541 and HL72288-01 and an Iowa Informatics Initiative grant. The research of Cun-Hui Zhang is partially supported by the NSF grants DMS-0102529 and DMS-0203086.

Appendix. Proofs

Proof of (3.10). Since Q_i in (3.9) is a projection to V_i, $Q_i\boldsymbol{\beta} z_i + \mathbf{v} \in V_i$ for $\mathbf{v} \in V_i$ and

$$\begin{aligned}\min_{f\in S_i}\Big\|\boldsymbol{y}_i - \boldsymbol{\beta} z_i - f(\boldsymbol{x}_i)\Big\|^2 &= \min_{\mathbf{v}\in V_i}\Big\|\boldsymbol{y}_i - \boldsymbol{\beta} z_i - \mathbf{v}\Big\|^2 \\ &= \min_{\mathbf{v}\in V_i}\Big\|\boldsymbol{y}_i - (I_J - Q_i)\boldsymbol{\beta} z_i - \mathbf{v}\Big\|^2 \\ &= \Big\|\boldsymbol{y}_i - (I_J - Q_i)\boldsymbol{\beta} z_i\Big\|^2 - \|\boldsymbol{y}_i\|^2 + \min_{\mathbf{v}\in V_i}\Big\|\boldsymbol{y}_i - \mathbf{v}\Big\|^2,\end{aligned}$$

due to $(I - Q_i)\mathbf{v} = 0,\ \forall \mathbf{v} \in V_i$. Thus, (3.4) and (3.10) are equivalent.

Proof of Theorem 3. Let $\mathbf{r}_i \equiv (I_J - Q_i)f_i(\boldsymbol{x}_i)$. As in (3.11),

$$B_n \equiv \operatorname*{arg\,min}_{\boldsymbol{\beta}\in\Omega_0^{J\times d}} \sum_{i=1}^n \Big\|\mathbf{r}_i - (I_J - Q_i)\boldsymbol{\beta} z_i\Big\|^2 = \widehat{\Lambda}_{J,n}^{-1}\sum_{i=1}^n (I_J - Q_i)\mathbf{r}_i z_i'. \tag{A.1}$$

Since $(I_J - Q_i)f_i(\boldsymbol{x}_i) = (I_J - Q_i)\mathbf{r}_i$ and $\boldsymbol{\beta}$ is a LSE of $\boldsymbol{\beta}$ for $\varepsilon_{ij} = f_i(x_{ij}) = 0$, by (3.10) and (3.11)

$$\widehat{\Lambda}_{J,n}^{1/2} B_n = \widehat{\Lambda}_{J,n}^{-1/2}\sum_{i=1}^n (I_J - Q_i) f_i(\boldsymbol{x}_i) z_i' = \widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - Z_n \tag{A.2}$$

are functions of covariates. Since $\|\widehat{\Lambda}_{J,n}^{1/2} B_n\|_2^2 = \langle B_n, \widehat{\Lambda}_{J,n} B_n\rangle_2$, it follows from (3.13) that

$$\begin{aligned}\Big\|\widehat{\Lambda}_{J,n}^{1/2} B_n\Big\|_2^2 &= \sum_{i=1}^n \langle B_n, (I_J - Q_i) B_n z_i z_i'\rangle_2 \\ &= \sum_{i=1}^n z_i' B_n' (I_J - Q_i) B_n z_i = \sum_{i=1}^n \|(I_J - Q_i) B_n z_i\|^2.\end{aligned}$$

This and the definition of B_n in (A.1) imply that $E\|\widehat{\Lambda}_{J,n}^{1/2}B_n\|_2^2 \le E\sum_{i=1}^n \|\mathbf{r}_i\|^2$, so that by (4.8) $E\|\widehat{\Lambda}_{J,n}^{1/2}B_n\|_2^2 \le J\sum_{i=1}^n \rho_{J,i}^2$. Thus, (4.22) holds in view of (A.2). Under the i.i.d. normality assumption on ε_{ij}, Z_n is a linear combination of ε_{ij} standardized by the root of its covariance operator, so that Z_n is a standard normal matrix. The proof of Theorem 3 is complete.

We need the following proposition for the proof of Theorem 4.

Proposition 1. (i) *Let* $\widehat{\Sigma}_{J,n}$ *be as in* (4.2). *Then,*

$$\widehat{\tau}_{J,n} \equiv \left\|\widehat{\Lambda}_{J,n} - I_{J,0}\otimes\widehat{\Sigma}_{J,n}\right\|_2 = \min_C \left\|\widehat{\Lambda}_{J,n} - I_{J,0}\otimes C\right\|_2. \tag{A.3}$$

(ii) *If* $\{x_{ij}, j\le J\}$ *is exchangeable for each* i *and* S_i *depends on data only through* $\{z_i, i\le n\}$ *and the set* $\{x_{ij}, j\le J\}$ *for each* i, *then*

$$E\widehat{\Lambda}_{J,n} = I_{J,0}\otimes E\widehat{\Sigma}_{J,n} = I_{J,0}\otimes\sum_{i=1}^n \frac{J - E\widehat{K}_i}{J-1} z_i z_i'. \tag{A.4}$$

(iii) *If* $\boldsymbol{x}_i$ *are pairwise independent random vectors, then*

$$E\left\|\widehat{\Lambda}_{J,n} - E\widehat{\Lambda}_{J,n}\right\|_2^2 \le \sum_{i=1}^n \|z_i\|^4 E\widehat{K}_i. \tag{A.5}$$

Proof. (i) Setting $(\partial/\partial t)\|\widehat{\Lambda}_{J,n} - I_{J,0}\otimes(C+tA)\|_2^2 = 0$ at $t=0$, we find

$$\langle I_{J,0}, I_{J,0}\rangle_2\langle C, A\rangle_2 = \langle\widehat{\Lambda}_{J,n}, I_{J,0}\otimes A\rangle_2 = \sum_{i=1}^n \langle I_J - Q_i, I_{J,0}\rangle_2\langle z_i z_i', A\rangle_2$$

by (3.13). Since $\psi_{i1}(x) = 1$, $\mathbf{e}\equiv(1,\ldots,1)'$ is an element of V_i in (3.8), so that $(I_J - Q_i)\mathbf{e} = 0$. Thus, $\langle I_J - Q_i, I_{J,0}\rangle_2 = \mathrm{trace}(I_J - Q_i) = J - \widehat{K}_i$, which implies $(J-1)\langle C, A\rangle_2 = \langle I_{J,0}, I_{J,0}\rangle_2\langle C, A\rangle_2 = (J-1)\langle\widehat{\Sigma}_{J,n}, A\rangle_2$ by (4.2). This proves (A.3) since A is an arbitrary $d\times d$ matrix.

(ii) By the exchangeability, the diagonal elements of $E(I_J - Q_i)$ must all equal $c_1 \equiv E\,\mathrm{trace}(I_J - Q_i)/J = (J - E\widehat{K}_i)/J$. Similarly, the off-diagonal elements of $E(I - Q_i)$ must share a common value c_2. Since $(I_J - Q_i)\mathbf{e} = 0$, the constant c_2 satisfies $J(J-1)c_2 + Jc_1 = E\mathbf{e}'(I_J - Q_i)\mathbf{e} = 0$, which implies $c_2 = -c_1/(J-1)$. Thus,

$$E(I_J - Q_i) = (c_1 - c_2)I_J + c_2\mathbf{e}\mathbf{e}' = \frac{Jc_1}{J-1}\Big(I_J - J^{-1}\mathbf{e}\mathbf{e}'\Big) = \frac{J - E\widehat{K}_i}{J-1} I_{J,0}.$$

This proves (A.4) in view of (3.13) and (4.2).

(iii) The pairwise independence of $\boldsymbol{x}_i$ implies that of Q_i. Since the square of the Hilbert-Schmidt norm of a matrix is just the sum of squares of all its elements and tensor products are linear mappings, by (3.13)

$$E\left\|\widehat{\Lambda}_{J,n}-E\widehat{\Lambda}_{J,n}\right\|_2^2=\sum_{i=1}^n E\left\|(Q_i-EQ_i)\otimes z_iz_i'\right\|_2^2=\sum_{i=1}^n E\left\|(Q_i-EQ_i)\right\|_2^2\|z_iz_i'\|_2^2.$$

This implies (A.5) since $\|z_iz_i'\|_2^2=\|z_i\|^4$ and $E\|(Q_i-EQ_i)\|_2^2\le E\|Q_i\|_2^2=E\widehat{K}_i$. The proof of Proposition 1 is complete.

Proof of Theorem 4. For linear mappings A and B, e.g., $A=\widehat{\Lambda}_{J,n}$, the following relationships hold with their operator norm $\|A\|=\max\{\|A\boldsymbol{\beta}\|:\|\boldsymbol{\beta}\|=1\}$ and the Hilbert-Schmidt norm $\|A\|_2$:

$$\|A\|\le\|A\|_2,\quad \|AB\|_2\le\|A\|\|B\|_2,\quad \|(A+B)^{-1}\|\le\frac{\|A^{-1}\|}{1-\|A^{-1}\|\|B\|},\tag{A.6}$$

since A and B are matrices operating in linear spaces (e.g., $\Omega_0^{J\times d}$). The first two inequalities follow easily from the fact that $\|A\|_2^2=\sum_k\|A\mathbf{v}_k\|^2$ for any orthonormal basis $\{\mathbf{v}_k\}$, while the third follows from $\|A^{-1}\|^{-1}=\min\{\|A\boldsymbol{\beta}\|:\|\boldsymbol{\beta}\|=1,\boldsymbol{\beta}\in\Omega_0^{J\times d}\}$. If A is positive-definite, $\|A^{-1}\|$ is simply the reciprocal of the smallest eigenvalue of A.

Now, since $\lambda_{\max}(\Sigma_n)/\lambda_{\min}(\Sigma_n)=O(1)$ and d is fixed, all the eigen-values of Σ_n are of the order $\text{trace}(\Sigma_n)=\sigma_n\equiv\sum_{i=1}^n\|z_i\|^2$. The same is true for $\widehat{\Sigma}_{J,n}$ as matrices and $I_{J,0}\otimes\widehat{\Sigma}_{J,n}$ as operators in $\Omega_o^{J\times d}$, since $(1-\kappa^*)\Sigma_n\le\widehat{\Sigma}_{J,n}\le\Sigma_n$ due to $\widehat{K}_i\le K_i\le\kappa^*J$. In particular $\|I_{J,0}\otimes\widehat{\Sigma}_{J,n}\|=O(\sigma_n)$ and $\|(I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1}\|=O(\sigma_n^{-1})$. It follows from Proposition 1 and (4.9) that

$$\left\|\widehat{\Lambda}_{J,n}-I_{J,0}\otimes\widehat{\Sigma}_{J,n}\right\|_2^2\le\left\|\widehat{\Lambda}_{J,n}-E\widehat{\Lambda}_{J,n}\right\|_2^2=O_P(1)\sum_{i=1}^n\|z_i\|^4E\widehat{K}_i=o_P(\sigma_n^2),$$

so that by (A.6) and algebra, e.g., $A^{-1}-B^{-1}=A^{-1}(B-A)B^{-1}$, for all integers k,

$$\|\widehat{\Lambda}_{J,n}^k\|=O_P(\sigma_n^k),\quad \|\widehat{\Lambda}_{J,n}^k-(I_{J,0}\otimes\widehat{\Sigma}_{J,n})^k\|_2^2=o_P(\sigma_n^{2k}).\tag{A.7}$$

It follows from (A.1), (A.2), (A.6) and (A.7) that $(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\widehat{\Lambda}_{J,n}^{-1/2}Z_n=B_n$ and

$$\left\|B_n\right\|_2^2=\left\|\widehat{\Lambda}_{J,n}^{-1}\sum_{i=1}^n\mathbf{r}_iz_i'\right\|_2^2\le\left\|\widehat{\Lambda}_{J,n}^{-1}\right\|^2\left\|\sum_{i=1}^n\mathbf{r}_iz_i'\right\|_2^2\le O_P(\sigma_n^{-2})\left\|\sum_{i=1}^n\mathbf{r}_iz_i'\right\|_2^2.$$

Since $\mathbf{r}_i$ is a permutation symmetric $\mathbb{R}^J$-valued function of exchange variables $x_{ij}, j=1,\dots,J$, its components are also exchangeable. Thus, $JE\mathbf{r}_i=\mathbf{e}\mathbf{e}'E\mathbf{r}_i=$

$\mathbf{e}E\mathbf{e}'(I_J - Q_i)f_i(\boldsymbol{x}_i) = 0$. Since the $\mathbf{r}_i$ are pairwise independent, $E\|\sum_{i=1}^n \mathbf{r}_i z_i'\|_2^2 = \sum_{i=1}^n E\|\mathbf{r}_i z_i'\|_2^2 = \sum_{i=1}^n \|z_i\|^2 E\|\mathbf{r}_i\|^2 = J\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2$. This and the bound for $\|B_n\|_2^2$ above imply (4.23).

Finally, let us prove (4.24). As in the proof of (4.23) above, we find by (A.2) that $(\widehat{\beta}_j - \beta_j) - (\widehat{\Lambda}_{J,n}^{-1/2} Z_n)'\mathbf{e}_j = B_n'\mathbf{e}_j = \Big(\widehat{\Lambda}_{J,n}^{-1}\sum_{i=1}^n \mathbf{r}_i z_i\Big)'\mathbf{e}_j$ are exchangeable. Thus, by (4.23), $\max_j E\min\Big(1, \sigma_n\|B_n'\mathbf{e}_j\|^2\Big) = E\sum_j \min(1, \sigma_n\|B_n'\mathbf{e}_j\|^2)/J$ is bounded by

$$
\begin{aligned}
E\min\Big(1, \sigma_n\sum_{j=1}^J \frac{\|B_n'\mathbf{e}_j\|^2}{J}\Big) &= E\min\Big(1, \sigma_n\frac{\|B_n\|_2^2}{J}\Big) \\
&= E\min\Big(1, O_P(\sigma_n^{-1})\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2\Big)
\end{aligned}
$$

which, in turn, is bounded by $o(1) + O(\sigma_n^{-1})\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2$. Therefore, (4.24) holds. Since $\|\widehat{\Lambda}_{J,n}^{-1}\| < \infty$ implies that it is of full rank as an operator in $\Omega_0^{J\times d}$ with rank $d(J-1)$, the first part of (A.7) with $k=-1$ implies $P\{\widehat{\nu} = d(J-1)\} \to 1$. Moreover, Proposition 1 implies (4.25) directly. Hence, the proof of Theorem 4 is complete.

Proof of Theorem 1. By (3.13), $\widehat{\Lambda}_{J,n}$ is a sum of nonnegative definite tensor products, so that $\widehat{\Lambda}_{J,n}\boldsymbol{\beta} = 0$ iff $(I - Q_i)\boldsymbol{\beta} z_i = 0$ for all i. Thus, $m \equiv nJ - \widehat{\nu} - \sum_{i=1}^n \widehat{K}_i$ in (4.1) is indeed the residual degrees of freedom, and $\widehat{\sigma}^2$ the mean residual sum of squares. Furthermore $m \ge J\{n(1-\kappa^*) - d\}$ due to $\widehat{\nu} \le (J-1)d$ and, conditionally on X, the noncentrality parameter, say θ_0, of the residual sum of squares is bounded by $\sum_{i=1}^n \|\mathbf{r}_i\|^2/\sigma^2$ with the $\mathbf{r}_i$ in (A.1). Thus, since $E|N(\theta_0,\theta_0)| \le \sqrt{\theta_0^2 + \theta_0} \le \theta_0 + 1/2$ and $\mathrm{Var}(\chi_k^2) = 2k$, the first part of (4.7) with $q = 1$ implies

$$
\begin{aligned}
E\Big|\frac{\widehat{\sigma}^2}{\sigma^2} - 1\Big| &\le E\Big|\frac{N(\theta_0,\theta_0)}{m}\Big| + E\Big|\frac{\chi_m^2}{m} - 1\Big| \\
&\le E\Big\{\frac{1}{2m} + \sum_{i=1}^n \frac{\|\mathbf{r}_i\|^2}{\sigma^2 m} + \Big(\frac{2}{m}\Big)^{1/2}\Big\} \\
&\le O(1)\Big\{\sum_{i=1}^n \frac{\rho_{J,i}^2}{\sigma^2 n} + \frac{1}{\sqrt{nJ}}\Big\} = \frac{o(1)}{\sqrt{J}}. \qquad \text{(A.8)}
\end{aligned}
$$

(i) Let $B_n \equiv \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} - \widehat{\Lambda}_{J,n}^{-1/2} Z_n$ as in (A.1) and (A.2). By Theorem 3, (A.6), (A.7) and (4.7) with $(p,q) = (0,1)$, $E[|\langle\widehat{\Lambda}_{J,n}^{1/2} B_n, Z_n\rangle_2|^2|X] = \|\widehat{\Lambda}_{J,n}^{1/2} B_n\|_2^2 = o_P(\sqrt{J})$, so that

$$
\sum_{i=1}^n \|(I_J - Q_i)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})z_i\|^2 = \|\widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2 = \|Z_n\|_2^2 + o_P(\sqrt{J}). \qquad \text{(A.9)}
$$

Now, given X, $\|Z_n\|_2^2/\sigma^2$ has the chi-square distribution with $\widehat{\nu} \le d(J-1)$ degrees of freedom, so that $\|Z_n\|_2^2 = O_P(J)$. Moreover, by (3.13) the eigenvalues of $\widehat{\Lambda}_{J,n}$ is no greater than $\sigma_n \equiv \sum_{i=1}^n \|z_i\|^2$, so that $\sigma_n \widehat{\nu} \ge \mathrm{trace}(\widehat{\Lambda}_{J,n}) = \sum_{i=1}^n \|z_i\|^2(J-\widehat{K}_i)$, which implies $\widehat{\nu} \ge J(1-\kappa^*)$. Thus, by (A.8), (A.9) and the Central Limit Theorem applied to i.i.d. χ_1^2-variables,

$$\frac{\sum_{i=1}^n \|(I_J-Q_i)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})z_i\|^2-\widehat{\sigma}^2\widehat{\nu}}{\widehat{\sigma}^2\sqrt{2\widehat{\nu}}} = \frac{\|Z_n\|_2^2/\sigma^2-\widehat{\nu}}{\sqrt{2\widehat{\nu}}} + o_P((J/\widehat{\nu})^{1/2}) \xrightarrow{\mathrm{D}} N(0,1).$$

Since $(\chi^2_{1-\alpha,m}-m)/\sqrt{2m} \to z_{*,\alpha}$ as $m\to\infty$ with $P(N(0,1)\le z_{*,\alpha}) = 1-\alpha$, the left-hand side of (4.3) is approximated by

$$P\Big\{\frac{\|Z_n\|_2^2/\sigma^2-\widehat{\nu}}{\sqrt{2\widehat{\nu}}} + o_P((J/\widehat{\nu})^{1/2}) \le \frac{\chi^2_{1-\alpha,\widehat{\nu}}-\widehat{\nu}}{\sqrt{2\widehat{\nu}}}\Big\} \to P\Big\{N(0,1)\le z_{*,\alpha}\Big\} = 1-\alpha. \quad \text{(A.10)}$$

(ii) First of all $P\{\widehat{\nu} = d(J-1)\} \to 1$ by Theorem 4. Since $\|\widehat{\Lambda}_{J,n}^k\| = O_P(\sigma_n^k)$ by (A.7), by (A.6), (4.23) and (4.7) with $(p,q)=(1,1)$,

$$\|\widehat{\Lambda}_{J,n}^k B_n\|_2^2 \le \|\widehat{\Lambda}_{J,n}^k\|^2\|B_n\|_2^2 = O_P(\sigma_n^{2k})O_P(J/\sigma_n^2)\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2 = o_P(\sigma_n^{2k-1}\sqrt{J}) \quad \text{(A.11)}$$

for all real k. In particular, we have $\|\widehat{\Lambda}_{J,n}^{1/2}B_n\|_2^2 = o_P(\sqrt{J})$, so that (4.3) holds as in the proof of (i) leading to (A.10).

For (4.4) we have, by (A.2),

$$\sum_{j=1}^J \Big\|\widehat{\Sigma}_{J,n}^{1/2}(\widehat{\beta}_j-\beta_j)\Big\|^2 = \Big\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\widehat{\Sigma}_{J,n}^{1/2}\Big\|_2^2 = \Big\|(\widehat{\Lambda}_{J,n}^{-1/2}Z_n+B_n)\widehat{\Sigma}_{J,n}^{1/2}\Big\|_2^2. \quad \text{(A.12)}$$

Since $\|\widehat{\Sigma}_{J,n}^{1/2}\| \sim \sigma_n^{1/2}$ as in the second paragraph of the proof of Theorem 4, by (A.6) and (A.11), $\|B_n\widehat{\Sigma}_{J,n}^{1/2}\|_2^2 \le \sigma_n\|B_n\|_2^2 = o_P(\sqrt{J})$. Since Z_n is a standard normal matrix, by (A.6) and (A.11),

$$\begin{aligned}E[\langle B_n\widehat{\Sigma}_{J,n}^{1/2}Z_n\widehat{\Sigma}_{J,n}^{1/2}\rangle_2^2|X] &= E[\langle\widehat{\Lambda}_{J,n}^{-1/2}B_n\widehat{\Sigma}_{J,n}, Z_n\rangle_2^2|X]\\ &= \|\widehat{\Lambda}_{J,n}^{-1/2}B_n\widehat{\Sigma}_{J,n}\|_2^2 \le \|\widehat{\Sigma}_{J,n}\|^2\|\widehat{\Lambda}_{J,n}^{-1/2}B_n\|_2^2 = o_P(\sqrt{J}).\end{aligned}$$

Inserting the above approximations into (A.12) yields

$$\sum_{j=1}^J \Big\|\widehat{\Sigma}_{J,n}^{1/2}(\widehat{\beta}_j-\beta_j)\Big\|^2 = \Big\|\widehat{\Lambda}_{J,n}^{-1/2}Z_n\widehat{\Sigma}_{J,n}^{1/2}\Big\|_2^2 + o_P(\sqrt{J}). \quad \text{(A.13)}$$

Since $\widehat{\Lambda}_{J,n}^{-1/2}Z_n\widehat{\Sigma}_{J,n}^{1/2} = (I_{J,0}\otimes\widehat{\Sigma}_{J,n}^{1/2})\widehat{\Lambda}_{J,n}^{-1/2}Z_n$ and $\|AZ_n\|_2^2 = \langle A'AZ_n, Z_n\rangle_2$, by (A.6) and (A.7)

$$E\Big[\Big(\|\widehat{\Lambda}_{J,n}^{-1/2}Z_n\widehat{\Sigma}_{J,n}^{1/2}\|_2^2 - \|Z_n\|_2^2\Big)^2 \Big| X\Big]$$

$$
\begin{aligned}
&= E\Big[\langle\{\widehat{\Lambda}_{J,n}^{-1/2}(I_{J,0}\otimes\widehat{\Sigma}_{J,n})\widehat{\Lambda}_{J,n}^{-1/2} - I_{J,0}\otimes I_d\}Z_n, Z_n\rangle_2^2\Big|X\Big] \\
&\le \Big\|\widehat{\Lambda}_{J,n}^{-1/2}(I_{J,0}\otimes\widehat{\Sigma}_{J,n})\widehat{\Lambda}_{J,n}^{-1/2} - I_{J,0}\otimes I_d\Big\|_2^2 3\widehat{\nu} \\
&\le \|\widehat{\Lambda}_{J,n}^{-1}\|^2\Big\|I_{J,0}\otimes\widehat{\Sigma}_{J,n} - \widehat{\Lambda}_{J,n}\Big\|_2^2\Big\{3d(J-1)\Big\} = o_P(J). \qquad \text{(A.14)}
\end{aligned}
$$

Note that for a standard normal matrix Z and self-adjoint operator A with eigenvalues $\lambda_k, k\le m_0$, $E\langle AZ,Z\rangle_2^2 = E(\sum_k \lambda_k\chi^2_{1,(k)})^2 \le \sum_k \lambda_k^2 m_0 E\chi_1^4 = \|A\|_2^2(3m_0)$, where $\chi^2_{1,(k)}$ are i.i.d. χ^2_1 variables. This fact is used in the derivation of the first inequality in (A.14). It follows from (A.13), (A.14) and (A.8) that $\sum_{j=1}^J \|\widehat{\Sigma}_{J,n}^{1/2}(\widehat{\beta}_j - \beta_j)\|^2/\widehat{\sigma}^2 = \|Z_n\|_2^2/\sigma^2 + o_P(\sqrt{J})$, which then implies (4.4) via (A.10).

The proof of (4.5) is simpler. We obtain from (A.2) and (A.11) that

$$
\sum_{j=1}^J \|\widehat{\beta}_j - \beta_j\|^2 = \|B_n + \widehat{\Lambda}_{J,n}^{-1/2}Z_n\|_2^2 = \|\widehat{\Lambda}_{J,n}^{-1/2}Z_n\|_2^2 + o_P(\sqrt{J}/\sigma_n),
$$

and then obtain from (A.7), as in (A.14),

$$
\begin{aligned}
&E\Big[\Big(\|\widehat{\Lambda}_{J,n}^{-1/2}Z_n\|_2^2 - \|(I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1/2}Z_n\|_2^2\Big)^2\Big|X\Big] \\
&= E\Big[\langle(\widehat{\Lambda}_{J,n}^{-1} - I_{J,0}\otimes\widehat{\Sigma}_{J,n}^{-1})Z_n, Z_n\rangle_2^2\Big|X\Big] \\
&\le \Big\|\widehat{\Lambda}_{J,n}^{-1} - I_{J,0}\otimes\widehat{\Sigma}_{J,n}^{-1}\Big\|_2^2\{3J(d-1)\} = o_P(J/\sigma_n^2).
\end{aligned}
$$

Since $\|(I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1/2}Z_n\|_2^2 = \|Z_n\widehat{\Sigma}_{J,n}^{-1/2}\|_2^2$, the above facts and (A.8) imply

$$
\frac{\sigma_n}{\widehat{\sigma}^2}\sum_{j=1}^J\|\widehat{\beta}_j - \beta_j\|^2 = \frac{\sigma_n\|Z_n\widehat{\Sigma}_{J,n}^{-1/2}\|_2^2}{\widehat{\sigma}^2} + o_P(\sqrt{J}) = \frac{\sigma_n\|Z_n\widehat{\Sigma}_{J,n}^{-1/2}\|_2^2}{\sigma^2} + o_P(\sqrt{J}). \qquad \text{(A.15)}
$$

Given X and $\widehat{\nu} = d(J-1)$, $\|Z_n\widehat{\Sigma}_{J,n}^{-1/2}\|_2^2/\sigma^2$ is a sum of $J-1$ i.i.d. $\|N(0,\widehat{\Sigma}_{J,n}^{-1})\|^2$ variables. In addition, $E[\|N(0,\widehat{\Sigma}_{J,n}^{-1})\|^2|X] = \mathrm{trace}(\widehat{\Sigma}_{J,n}^{-1})$ and $\mathrm{Var}[\|N(0,\widehat{\Sigma}_{J,n}^{-1})\|^2|X] = 2\,\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-2})$. Since $\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-k}) \sim \sigma_n^{-k}$ for integers k and $P\{\widehat{\nu} = d(J-1)\} \to 1$, by (A.15),

$$
\begin{aligned}
&\Big\{2J\,\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-2})\Big\}^{-1/2}\Big(\sum_{j=1}^J\|\widehat{\beta}_j - \beta_j\|^2/\widehat{\sigma}^2 - (J-1)\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-1})\Big) \\
&= \frac{\|Z_n\widehat{\Sigma}_{J,n}^{-1/2}\|_2^2/\sigma^2 - (J-1)\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-1})}{\{2J\,\mathrm{trace}(\widehat{\Sigma}_{J,n}^{-2})\}^{1/2}} + o_P(1) \xrightarrow{\mathrm{D}} N(0,1).
\end{aligned}
$$

Finally, let us prove (4.10). By (4.24), $\sigma_n^{1/2}\{\widehat{\beta}_j - \beta_j - (\widehat{\Lambda}_{J,n}^{-1/2}Z_n)'\mathbf{e}_j\} = o_P(1)$ uniformly in j. Since Z_n/σ is a standard normal matrix, $\langle A, BZ_n\rangle_2 = \langle B'A, Z_n\rangle_2$

is a normal variable with mean zero and variance $\sigma^2\|B'A\|_2^2 = \sigma^2\langle A, BB'A\rangle_2$. Thus, since $\mathbf{e}_j b'$ is a $J\times d$ matrix, conditionally on X and $\widehat{\nu} = J(d-1)$,

$$\sigma_n^{1/2} b'(\widehat{\Lambda}_{J,n}^{-1/2}Z_n)'\mathbf{e}_j/\sigma = \sigma_n^{1/2}\langle \mathbf{e}_j b', \widehat{\Lambda}_{J,n}^{-1/2}Z_n\rangle_2/\sigma \sim N\Big(0, \sigma_n\langle \mathbf{e}_j b', \widehat{\Lambda}_{J,n}^{-1}(\mathbf{e}_j b')\rangle_2\Big).$$

Let $\widetilde{\Sigma}_{J,n} = (1-1/J)^{-1}\widehat{\Sigma}_{J,n}$. Since

$$\langle \mathbf{e}_j b', (I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1}(\mathbf{e}_j b')\rangle_2 = \mathbf{e}_j' I_{J,0}\mathbf{e}_j\, b'\widehat{\Sigma}_{J,n}^{-1}b = (1-1/J)b'\widehat{\Sigma}_{J,n}^{-1}b = b'\widetilde{\Sigma}_{J,n}^{-1}b$$

and $\|\mathbf{e}_j b'\|_2^2 = \|\mathbf{e}_j\|^2\|b'\|^2 = 1$, we find by (A.7) that, uniformly in j,

$$\sigma_n\Big|\langle \mathbf{e}_j b', \widehat{\Lambda}_{J,n}^{-1}\mathbf{e}_j b'\rangle_2 - b'\widetilde{\Sigma}_{J,n}^{-1}b\Big| \le \sigma_n\|\widehat{\Lambda}_{J,n}^{-1} - (I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1}\| = o_P(1).$$

Thus, $\sigma_n^{1/2}(\widehat{\beta}_j-\beta_j)/\sigma$ are uniformly within $o_P(1)$ of some $N(0,\sigma_n\widehat{\Sigma}_{J,n}^{-1})$ vectors. Since $\|\widehat{\Sigma}_{J,n}^k\| = O(\sigma_n^k)$ for $k=\pm 1/2$, $\widehat{\Sigma}_{J,n}^{1/2}(\widehat{\beta}_j-\beta_j)$ are uniformly within $o_P(1)$ of some $N(0,\sigma^2 I_d)$ random vectors. Since $\widehat{\sigma} = \sigma + o_P(1)$ via the inequalities in (A.8) and the first part of condition (4.7) with $(p,q) = (1,0)$, (4.10) holds. Hence, the proof is complete.

Proof of Theorem 2. (i) By the definition of $\widetilde{f}_i$ and $\widehat{f}_i^*$ in (4.13) and (4.14), $\widehat{f}_i^*(\boldsymbol{x}_i) = Q_i^*(\boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}z_i) = \widetilde{f}_i(\boldsymbol{x}_i) - Q_i^*(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})z_i$. Thus, by (4.23)

$$\begin{aligned}\Big\|\widehat{f}_i^*(\boldsymbol{x}_i) - \widetilde{f}_i(\boldsymbol{x}_i)\Big\| &= \Big\|Q_i^*(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})z_i\Big\| \\ &\le \Big\|Q_i^*\widehat{\Lambda}_{J,n}^{-1/2}Z_n z_i\Big\| + O_P(\|z_i\|)\frac{\|Q_i^*\|}{\sigma_n}\Big(J\sum_{k=1}^n\|z_k\|^2\rho_{J,k}^2\Big)^{1/2}.\end{aligned}$$

Since $E[\langle Z_n, AZ_n\rangle_2|X] \le \sigma^2\mathrm{trace}(A)$ for all operators A, by (A.6) and (A.7),

$$\begin{aligned}&E\Big[\Big\|Q_i^*\widehat{\Lambda}_{J,n}^{-1/2}Z_n z_i\Big\|^2\Big|X\Big] = E\Big[\langle Z_n, (I_{J,0}\otimes z_i z_i')\widehat{\Lambda}_{J,n}^{-1/2}(Q_i^*)'Q_i^*\widehat{\Lambda}_{J,n}^{-1/2}Z_n\rangle_2\Big|X\Big]\\ &\le \sigma^2\mathrm{trace}\Big((I_{J,0}\otimes z_i z_i')\widehat{\Lambda}_{J,n}^{-1/2}(Q_i^*)'Q_i^*\widehat{\Lambda}_{J,n}^{-1/2}\Big) \le \sigma^2\|z_i\|^2\|\widehat{\Lambda}_{J,n}^{-1/2}\|^2\mathrm{trace}\Big((Q_i^*)'Q_i^*\Big)\\ &\le \|z_i\|^2 O_P(\sigma_n^{-1})\|Q_i^*\|_2^2.\end{aligned}$$

(ii) By (4.12), $\max_{k\le n}\rho_{J,k}^2 = O(J^{-2\gamma/(2\alpha+1)})$, so that (4.16) holds after inserting $\sigma_n\sim n$, $J = o(n^{2\alpha+1})$, and the rates of $\|z_i\|\|Q_i^*\|$ and $\|z_i\|\|Q_i^*\|_2$ into (4.15). If (4.12) holds with (α, Q_i) replaced by (γ, Q_i^*), then $\|\widetilde{f}_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_i)\| = O_P(J^{1-2\gamma/(2\gamma+1)})$ uniformly in i, so that (4.17) follows from (4.16), (4.18) and the condition $(2\gamma+1)^2 \ge 2\gamma(2\alpha+1)$.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Brown, P. O. and Botstein, D. (1999). Exploring the new world of the genome with microarrays. *Nat. Genet.* **21**, 33-37.

Chen, H. (1988). Convergence rates for a parametric component in a partially linear model. *Ann. Statist.* **16**, 136-146.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

Cleveland, W. S. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83**, 596-610

Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications*. Springer-Verlag, New York.

Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81**, 310-320.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Fan, J., Tam, P., Vande Woude, G. and Ren, Y. (2004). Normalization and analysis of cDNA micro-arrays using within-array replications applied to neuroblastoma cell response to a Cytokine. *Proc. Nat. Acad. Sci.* 1135-1140.

Fan, J., Peng, H. and Huang, T. (2004). Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency. Preprint. Dept. of Operations Research and Financial Engineering, Princeton University.

Härdle, H., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Springer, New York.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer, New York.

Heckman, P.(1986). Spline smoothing in partly linear model. *J. Roy. Statist. Soc. Ser. B* **48**, 244-248.

Huang, J., Kuo, H.-C., Koroleva, I., Zhang, C.-H. and Soares, M. B. (2003). A semi-linear model for normalization and analysis of cDNA microarray data. Tech Report 321, Department of Statistics, Univ. Iowa, Iowa City, U.S.A.

Huang, J., Wang, D. and Zhang, C. (2004). A two-way semi-linear model for normalization and significant analysis of cDNA microarray data. Preprint.

Huber, P. J. (1981). *Robust Statistics*. John Wiley, New York.

Kerr, K. and Churchill, G. (2000). Analysis of variance for gene expression microarray data. *J. Comput. Biology* **7**, 819-837.

Rice, J. (1986). Convergence rates for partially spline models. *Statist. Probab. Lett.* **4**, 203-208.

Schumaker, L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.

Tseng, G. C., Oh, M.-K., Rohlin, L., Liao, J. C. and Wong, W. H. (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variation and assessment of gene effects. *Nucleic Acids Research* **29**, 2549-2557.

Yang, Y. H., Dudoit, S., Luu, P. and Speed, T. P. (2001). Normalization for cDNA microarray data. In *Microarrays: Optical Technologies and Informatics* (Edited by M. L. Bittner, Y. Chen, A. N. Dorsel and E. R. Dougherty), Vol. 4266 of Proceedings of SPIE.

Department of Statistics and Actuarial Science, 241 SH, University of Iowa, Iowa City, IA 52242, U.S.A

E-mail: jian@stat.uiowa.edu

Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08855, U.S.A.

E-mail: czhang@stat.rutgers.edu

(Received September 2003; accepted May 2004)