Asymptotic Analysis of A Two-Way Semiparametric Regression Model for Microarray Data

[1]Jian Huang and [2]Cun-Hui Zhang

1. Jian Huang: Department of Statistics and Actuarial Science, 241 SH, University of Iowa, Iowa City, IA 52242

2. Cun-Hui Zhang: Department of Statistics, Hill Center, Busch Campus, Rutgers University, Piscataway, NJ 08855

Correspondence should be addressed to

JH (jian@stat.uiowa.edu) or CHZ (czhang@stat.rutgers.edu)

Key words: 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. The purpose of normalization is to remove bias in the observed gene expression levels. This is necessary because many experimental factors may 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. Many 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 et al. 2000). This method takes into account the variations due to normalization. However, it assumes that normalization is a linear factor in the overall ANOVA model. Another normalization method is to use local regression (*loess*, Cleveland 1979, 1986; Fan and Gijbels 1996) that first regress the log-intensity ratio on the log-intensity product using all the genes printed on a slide, and then use the residuals of the regression as the normalized data in the subsequent analysis (Yang et al. 2000). Thus this method takes into account nonlinear normalization effects. However, because this method uses all the genes, including those with differential expressions, the resulting normalization curves can be potentially biased. And variations due to normalization are not considered in the subsequent analysis. In addition, an underlying assumption of this normalization method is that the number of differentially expressed genes is relatively small, so that the *loess* normalization curves are not affected significantly by the differentially expressed genes. If it is expected that many genes have differential expressions, Yang et al (2000) suggested using dye-swap for normalization. This approach makes the assumption that the normalization curves in the two dye-swaped slides are symmetric. Because of the slide-to-slide variation, this assumption may not always be satisfied.

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 et al (2001) first used a rank based procedure to select a set of 'invariant genes' that are likely to be non-differentially expressed, and then use 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 that the final results depend upon the threshold value may need to be evaluated on a case by case basis.

We propose a two-way semiparametric regression model (TW-SRM) for microarray data. This model grows out of the idea of the *loess* and ANOVA normalization methods. In essence, the TW-SRM is a semiparametric analysis of covariance model that includes nonlinear normalization factors. In addition, as it can be seen below, the *loess* method can be considered as the first step in an iterative back-fitting algorithm in solving the proposed TW-SRM. There are two important features of the TW-SRM. First, when TW-SRM is used for normalization, it does not require the assumptions that are needed in the *loess* normalization method, such as the assumption that the percentage of differentially expressed genes is small, nor does it require pre-selection of constantly expressed genes. Second, when TW-SRM 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-SRM presents novel and challenging theoretical questions in the area of semiparametric statistics. As described below, in the TW-SRM, 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 the gene effects, J is the number of parameters, n is the sample size. But for estimating the normalization curves, n is the number of infinite-dimensional parameters, 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, because then 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 much intensive research in semiparametric statistics (Bickel, Klaassen, Ritov and Wellner 1993), we are not aware of any other semiparametric models in which both 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 provide theoretical justification that unbiased statistical inference is possible when self calibration is needed when $n/J \to 0$ as $(n, J) \to (\infty, \infty)$.

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-SRM 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 the 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 model (2.2), it is only made explicit that the normalization curve f_i is slide-dependent. It can also be made to be dependent upon regions of a slide to account for spatial effect. 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-SRM 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, and where f_i and ε_{ij} are the same as in (2.2).

The covariate vector 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 model (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.

Below, 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-SRM is an extension of the semiparametric regression model (SRM) proposed by Engle et al.(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-SRM 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), and 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-SRM.

3. Semiparametric least squares estimation and computation in the two-way SRM. 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 $\psi_{ik}, k \leq K_i$. For example, these basis functions can be splines, wavelets, trigonometry functions, or polynomials. We will 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-SRM 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 $f_i(x) + \sum_{k=1}^J \beta_k' z_i/J$ in (2.2). 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 $D^2(\boldsymbol{\beta}, \boldsymbol{f})$. That is,

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

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

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

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

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

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 between optimizations in $\boldsymbol{\beta}$ and $\boldsymbol{f}$, i.e., (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 othorgonalization of the design vectors in the TW-SRM. 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 orthoganolization 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 the TW-SRM (2.2) 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, i.e. the design matrices for the estimation of $\widehat{f}_i(\boldsymbol{x}_i)$, and $\widehat{K}_i$ be their ranks:

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

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

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

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

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

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

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

Here and in the sequel, A^{-1} denotes the generalized inverse of matrix A, defined by $A^{-1}\boldsymbol{x} \equiv \arg\min\big\{\|\boldsymbol{b}\| : A\boldsymbol{b} = \boldsymbol{x}\big\}$. 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 $A^{-1} = \sum_{\lambda_j \neq 0} \lambda_j^{-1}\mathbf{v}_j\mathbf{v}_j'$. For $d > 1$, (3.10) is still given by (3.11) with

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

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

$$s_i(x) = \lambda_{i0} + \sum_{k=1}^{K_i} b_k(x)\alpha_{ik} \equiv \boldsymbol{b}_i(x)'\alpha_i \in S_i$$

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 $B_{ij} = (b_1(x_{ij}), \ldots, b_{K_i}(x_{ij}))'$ 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

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

The projection matrix Q_i defined in (3.9) is

$$Q_i = B_i(B_i'B_i)^{-1}B_i', \quad i = 1, \ldots, n.$$

The iterative algorithm described earlier becomes the following. Set $\boldsymbol{\alpha}^{(0)} = 0$. For $k = 0, 1, 2, \ldots,$

Step 1: Compute $\boldsymbol{\beta}^{(k)}$ by minimizing $D_w(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. The explicit solution is given by (3.6).

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

$$\alpha_i^{(k+1)} = (B_i'B_i)^{-1}B_i'(\boldsymbol{y}_i - \boldsymbol{\beta}^{(k)}z_i), i = 1, \ldots, n.$$

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 meet 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

$$\widehat{f}_i(x) = \boldsymbol{b}(x)'\alpha_i^{(K)},\ i = 1, \ldots, n.$$

Local regression (loess) method: The loess method can also be used to estimate the TW-SRM. 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-SRM 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-SRM will be 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. 2000; 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 the SLSE (3.4) and the information operator (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 iid $N(0,\sigma^2)$ 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)

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

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). 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

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

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

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

and

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

Our inference procedures about individual β_j are based on

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

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 $\sigma^2\Sigma_n^{-1}$. Note that $\Sigma_n \min_{i\le n}(J-\widehat{K}_i)/(J-1) \le \widehat{\Sigma}_{J,n} \le \Sigma_n$.

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

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

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

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

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

Condition A. *The random vectors $\boldsymbol{x}_i, i\le n$, are pairwise independent, and for each i, the random variables $\{x_{ij}, j\le J\}$ 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), and $\lambda_{\max}(\Sigma_n)$ and $\lambda_{\min}(\Sigma_n)$ are respectively the largest and the smallest eigenvalues of Σ_n. 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

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

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

Corollary 1. *Suppose* $\limsup_{n\to\infty}\max_{i\le n}\|z_i\| < \infty$ *and* $\liminf_{n\to\infty}\lambda_{\min}(\Sigma_n)/n > 0$.

(i) *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).*

(ii) *Suppose Condition A holds and that f_i are uniformly continuous with a common compact support. Suppose spline or wavelet bases are used in (3.1) with* $\max_{i\le n} K_i = o(n \wedge J)$ *and* $\min_{i\le n} K_i \to \infty$. *Then,* $\max_i \rho^2_{J,i} = o(1)$ *and (4.10) holds.*

Approximate confidence regions and intervals can be easily constructed based on Theorem 1 and Corollary 1. By the central limit theorem for iid χ^2_1 variables, $\chi^2_{1-\alpha,\nu}$ could 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 $\sqrt{\widehat{\Sigma}_{J,n}}|\widehat{\beta}_j - \beta_j|/\widehat{\sigma} \le 1.96$.

4.2. Bounds on the normalization error. As indicated earlier, in addition to the detection of differentially expressed genes, the TW-SRM can also be used for normalization, which is an important step preceeding any down-stream analysis. Thus it is of interest to assess the quality of normalization.

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 *et al.* (2001), to generate ideal normalizing curves. Consider linear ideal normalizing curves of the form

$$\widetilde{f}_i(\boldsymbol{x}_i) \equiv Q^*_i(\boldsymbol{y}_i - \boldsymbol{\beta} z_i), \tag{4.11}$$

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

$$\widehat{f}^*_i(\boldsymbol{x}_i) \equiv Q^*_i(\boldsymbol{y}_i - \widehat{\boldsymbol{\beta}} z_i) \tag{4.12}$$

instead of (4.11). 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.

Theorem 2. *Suppose conditions of Theorem 1 hold and that for each i, Q_i^* depends on data only through $\{z_i, i \le n\}$ and the values of $\{x_{ij}, j \le J\}$. Suppose $\|Q_i^*\| \le M_0$ for all i and certain fixed constant $M_0 < \infty$. Then,*

$$\frac{1}{Jn}\sum_{i=1}^n \Big\|\widehat{f}_i^*(\boldsymbol{x}_i) - \widetilde{f}_i(\boldsymbol{x}_i)\Big\|^2 \le O_P(1)\Big\{\Big(\eta'_{J,n}\eta_{J,n} + \sqrt{J\eta'_{J,n}\eta^3_{J,n}}\Big)\sum_{i=1}^n \frac{\rho^2_{J,i}}{n} + \sigma^2\frac{\eta'_{J,n}}{n}\Big\}. \tag{4.13}$$

where $\eta_{J,n} \equiv \max_{i\le n}\|z_i\|^2/\sum_{i=1}^n\|z_i\|^2$, $\eta'_{J,n} \equiv \max_{i\le n}\big(E\|Q_i^*\|_2^2/J\big)$, and $\rho_{J,i}$ are as in (4.8).

Consider typical designs with $\eta_{J,n} = O(1/n)$ and $\max_{i\le n} E\|Q_i^*\|_2^2 = o(1)\sum_{i=1}^n E\|Q_i^*\|_2^2$. Theorem 2 implies that $\widehat{f}^*$ is essentially as good as the ideal 'estimator' $\widetilde{f}$ based on the knowledge of the unknown $\boldsymbol{\beta}$ in the sense that the right-hand side of (4.13) is of smaller order than

$$\frac{1}{Jn}\sum_{i=1}^n E\Big\|\widetilde{f}_i(\boldsymbol{x}_i) - f_i(\boldsymbol{x}_i)\Big\|^2 = \frac{1}{Jn}\sum_{i=1}^n E\Big\|(I_J - A_i^*)f_i(\boldsymbol{x}_i)\Big\|^2 + \frac{\sigma^2}{Jn}\sum_{i=1}^n E\|Q_i^*\|_2^2.$$

This is the case if the average squared approximation error $(Jn)^{-1}\sum_{i=1}^n E\big\|(I_J - Q_i^*)f_i(\boldsymbol{x}_i)\big\|^2$ using the linear estimators Q_i^* is a small fraction of $\big\{\max_{i\le n} E\|Q_i^*\|_2^2/n^3\big\}^{1/2}$ times the average squared approximation error $\sum_{i=1}^n \rho^2_{J,i}/n$ using Q_i.

4.3. Inequalities about the SLSE and the observed information operator. A crucial step in proving Theorem 1 is to understand the effects of approximation errors for the unknown functions f_i on the SLSE of $\boldsymbol{\beta}$ in (3.4) 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 $\big\langle A, B\big\rangle_2 \equiv \sum_{ij} a_{ij}b_{ij}$ be the inner product of matrices $A \equiv (a_{ij})$ and $B \equiv (b_{ij})$ of common finite-dimensions and $\|A\|_2 \equiv \big(\big\langle A, A\big\rangle_2\big)^{1/2}$ be the Hilbert-Schmidt norm of A. The inner-product $\big\langle \cdot,\cdot\big\rangle_2$ 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, $\big\langle A_1\otimes B_1, A_2\otimes B_2\big\rangle_2 = \big\langle A_1, A_2\big\rangle_2\big\langle B_1, B_2\big\rangle_2$. Let

$$\widehat{\Omega}_{J,n} \equiv \Big\{\widehat{\Lambda}_{J,n}\boldsymbol{\beta} : \boldsymbol{\beta} \in \Omega_0^{J\times d}\Big\} \subseteq \Omega_0^{J\times d} \tag{4.14}$$

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

$$\big\langle A, Z \big\rangle_2 \sim N\Big(0, \|A\|_2^2\Big), \quad \forall A \in \widehat{\Omega}_{J,n}. \tag{4.15}$$

Theorem 3. *Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_{J,n}$ be as in (3.4) and (3.13). 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*

$$E\Big\|\widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - Z_n\Big\|_2^2 \le J \sum_{i=1}^n \rho_{J,i}^2, \tag{4.16}$$

where $Z_n \equiv \widehat{\Lambda}_{J,n}^{-1/2} \sum_{i=1}^n (I_J - Q_i)\epsilon_i z_i'$. Moreover, if the errors $\{\epsilon_{ij}, i \le n, j \le J\}$ in (2.2) are iid $N(0,\sigma^2)$ given X, then Z_n/σ given X is a standard normal vector in $\widehat{\Omega}_{J,n}$ as in (4.15).

Theorem 3 is derived from 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). We state it below and give the proof in the Appendix.

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

$$\Big\|(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \widehat{\Lambda}_{J,n}^{-1/2} Z_n\Big\|_2^2 \le O_P(J/\sigma_n^2) \sum_{i=1}^n \|z_i\|^2 \rho_{J,i}^2 \tag{4.17}$$

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

$$\max_{1\le j\le J} E \min\left\{1, \sigma_n \Big\|(\widehat{\beta}_j - \beta_j) - \Big(\widehat{\Lambda}_{J,n}^{-1/2} Z_n\Big)' \mathbf{e}_j\Big\|^2\right\} \le o(1) + O(\sigma_n^{-1}) \sum_{i=1}^n \|z_i\|^2 \rho_{J,i}^2, \tag{4.18}$$

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

$$E\Big\|\widehat{\Lambda}_{J,n} - I_{J,0} \otimes \widehat{\Sigma}_{J,n}\Big\|_2^2 \le \sum_{i=1}^n \|z_i\|^4 E\widehat{K}_i = o(\sigma_n^2). \tag{4.19}$$

5. Concluding remarks. In this article, we studied the asymptotic properties of the TW-SRM, motivated from the problem of normalization and identification of differentially expressed genes using cDNA microarray data. We have shown that, under appropriate conditions, statistical inference about $\boldsymbol{\beta}$ in the TW-SRM can be carried out in the same order of precision (i.e., the rate of convergence is $n^{-1/2}$) as in a regular semiparametric model. This suggests that some important inference tools, such as the bootstrap, can be consistently applied to the TW-SRM. This is useful for assessing the variance of the estimator of $\boldsymbol{\beta}$ when numerical calculation involving high-dimensional matrices becomes difficult. We also shown that the normalization curves can be estimated accurately.

It was not intuitively clear to us at the outset whether many β_j could be unestimable due to the singularities of (3.12) if reasonably rich approximation spaces S_i are used to estimate f_i in the SLSE (3.4) in the TW-SRM. We were particularly intrigued by the presence of the large number of nonparametric components $f_i, i = 1, \ldots, n$, where as described earlier, n is the sample size for estimating $\boldsymbol{\beta}$. Thus from the asymptotic analysis point of view, the TW-SRM 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. Therefore, we may view such standard models *finite*-dimensional semiparametric models. This appears to be a key distinction between the TW-SRM and the standard semiparametric models. This distinction renders that the existing theory for semiparametric models (Bickel et al. 1983) cannot be directly applied to the TW-SRM.

There are several interesting and challenging questions that have not been addressed in the present report. For example, it is of interest to extend our results to the robust estimators (Huber 1981) of the TW-SRM. Our analysis makes use of the fact that the least squares estimators can be considered as orthogonal projections. This is no longer the case when a robust objective function is used in defining the estimators. The second extension is to allow heteroscedasticity in the TW-SRM. 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 computation and theoretical analysis of a heteroscedasticity TW-SRM will be more complicated. The third question is how to incorporate correlation into the TW-SRM. 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.

6. Appendix: Proofs

Proof of (3.10). Since Q_i in (3.9) are projections to V_i, $Q_i\boldsymbol{\beta}z_i + \mathbf{v} \in V_i$ for $\mathbf{v} \in V_i$. Thus,

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

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 $\mathbf{r}_i \equiv (I_J - Q_i)f_i(\boldsymbol{x}_i)$. Let

$$B_n \equiv \arg\min \sum_{i=1}^n \Big\|\mathbf{r}_i - (I_J - Q_i)\boldsymbol{\beta}z_i\Big\|^2 = \widehat{\Lambda}_{J,n}^{-1}\sum_{i=1}^n (I_J - Q_i)\mathbf{r}_i z_i'. \tag{6.20}$$

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 $\epsilon_{ij} = f_i(x_{ij}) = 0$, by (3.10) and (3.11)

$$\widehat{\Lambda}_{J,n}^{1/2}B_n = \widehat{\Lambda}_{J,n}^{-1/2}\sum_{i=1}^n (I_J - Q_i)f_i(\boldsymbol{x}_i)z_i' = \widehat{\Lambda}_{J,n}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - Z_n \tag{6.21}$$

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

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

This and (6.20) imply that $E\|\widehat{\Lambda}_{J,n}^{1/2}B_n\|_2^2 \le E\sum_{i=1}^n \|\mathbf{r}_i\|^2 = J\sum_{i=1}^n \rho_{J,i}^2$ by (4.8). Hence, in view of (6.21), the proof of Theorem 2 is complete.

We need the following proposition for the proof of Theorem 3.

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

$$\widehat{\tau}_{J,n} \equiv \Big\|\widehat{\Lambda}_{J,n} - I_{J,0} \otimes \widehat{\Sigma}_{J,n}\Big\|_2 = \min_C \Big\|\widehat{\Lambda}_{J,n} - I_{J,0} \otimes C\Big\|_2. \tag{6.22}$$

(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*

$$E\widehat{\Lambda}_{J,n} = I_{J,0} \otimes E\widehat{\Sigma}_{J,n} = I_{J,0} \otimes \sum_{i=1}^n \frac{J - E\widehat{K}_i}{J-1} z_i z_i'. \tag{6.23}$$

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

$$E\Big\|\widehat{\Lambda}_{J,n} - E\widehat{\Lambda}_{J,n}\Big\|_2^2 \le \sum_{i=1}^n \|z_i\|^4 E\widehat{K}_i. \tag{6.24}$$

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

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

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, $\big\langle I_J - Q_i, I_{J,0}\big\rangle_2 = \mathrm{trace}\big(I_J - Q_i\big) = J - \widehat{K}_i$, which implies $(J-1)\big\langle C, A\big\rangle_2 = \big\langle I_{J,0}, I_{J,0}\big\rangle_2 \big\langle C, A\big\rangle_2 = (J-1)\big\langle \widehat{\Sigma}_{J,n}, A\big\rangle_2$ by (4.2). This proves (6.22) 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 to $c_1 \equiv E\,\mathrm{trace}(I_J - Q_i)/J = (J - E\widehat{K}_i)/J$. 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 (6.23) 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)

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

This implies (6.24) since $\|z_iz_i'\|_2^2=\|z_i\|^4$ and $E\big\|(Q_i-EQ_i)\big\|_2^2\le E\big\|Q_i\big\|_2^2=E\widehat{K}_i$. The proof of Proposition 1 is complete.

Proof of Theorem 4. For linear mappings A in the parameter space $\Omega_0^{J\times d}=\{\boldsymbol{\beta}:\boldsymbol{\beta}'\mathbf{e}=0\}$, e.g. $A=\widehat{\Lambda}_{J,n}$, let $\|A\|=\max\big\{\|A\boldsymbol{\beta}\|:\|\boldsymbol{\beta}\|=1,\boldsymbol{\beta}\in\Omega_0^{J\times d}\big\}$ be the operator norm. Since linear mappings are matrices when $\boldsymbol{\beta}\in\Omega_0^{J\times d}$ are viewed as vectors with certain orthonormal basis,

$$\|A\|\le\|A\|_2,\quad \|AB\|_2\le\|A\|\|B\|_2,\quad \|(A+B)^{-1}\|\le\frac{\|A^{-1}\|}{1-\|A^{-1}\|\|B\|}. \tag{6.25}$$

The first two inequality follows easily from the Cauchy-Schwarz inequality, while the third follows from $\|A^{-1}\|^{-1}=\min\big\{\|A\boldsymbol{\beta}\|:\|\boldsymbol{\beta}\|=1,\boldsymbol{\beta}\in\Omega_0^{J\times d}\big\}$. 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 $\mathrm{trace}(\Sigma_n)=\sigma_n\equiv\sum_{i=1}^n\|z_i\|^2$. 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

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

so that by (6.25) and algebra, e.g. $A^{-1}-B^{-1}=A^{-1}(B-A)B^{-1}$, for all integers k

$$\|\widehat{\Lambda}_{J,n}^k\|=O_P(\sigma_n^k),\quad \|\widehat{\Lambda}_{J,n}^k-(I_{J,0}\otimes\widehat{\Sigma}_{J,n})^k\|_2^2=o_P(\sigma_n^{2k}). \tag{6.26}$$

It follows from (6.20), (6.21), (6.25) and (6.26) that $(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})-\widehat{\Lambda}_{J,n}^{-1/2}Z_n=B_n$ and

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

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 $\mathbf{r}_i$ are pairwise independent, $E\big\|\sum_{i=1}^n \mathbf{r}_i z_i'\big\|^2 = \sum_{i=1}^n E\|\mathbf{r}_i z_i'\|^2 = \sum_{i=1}^n \|z_i\|^2 E\|\mathbf{r}_i\|^2 = J\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2$. This and the bound on $\|B_n\|_2^2$ imply (4.17).

Finally, let us prove (4.18). As in the proof of (4.17) above, we find by (6.21) 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 (4.17), $\max_j E\min\Big(1, \sigma_n\|B_n'\mathbf{e}_j\|^2\Big) = E\sum_j \min\big(1,\sigma_n\|B_n'\mathbf{e}_j\|^2\big)/J$ is bounded by

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

which in turn is bounded by $o(1) + O(\sigma_n^{-1})\sum_{i=1}^n \|z_i\|^2\rho_{J,i}^2$. Therefore, (4.18) holds. Since (6.26) with $k=-1$ implies $P\{\widehat{\nu} = d(J-1)\} \to 1$ and Proposition 1 implies (4.19) directly, the proof of Theorem 3 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. Furthermore, $m \ge J\{n(1-\kappa^*) - d\}$ due to $\widehat{\nu} \le (J-1)d$, and conditionally on X the noncentrality parameter 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 (6.20). Thus, the first part of (4.7) with $q=1$ implies

$$E\left|\frac{\widehat{\sigma}^2}{\sigma^2} - 1\right| \le E\left\{\frac{2}{m} + \sum_{i=1}^n \frac{\|\mathbf{r}_i\|^2}{\sigma^2 m} + \left(\frac{2}{m}\right)^{1/2}\right\} \le O(1)\left\{\sum_{i=1}^n \frac{\rho_{J,i}^2}{\sigma^2 n} + \frac{1}{\sqrt{nJ}}\right\} = \frac{o(1)}{\sqrt{J}}. \qquad (6.27)$$

(i) Let $B_n \equiv \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} - \widehat{\Lambda}_{J,n}^{-1/2} Z_n$ as in (6.20) and (6.21). By Theorem 2 and (4.7) with $(p,q) = (0,1)$, $E\big[\big|\big\langle \widehat{\Lambda}_{J,n}^{1/2} B_n, Z_n\big\rangle_2\big|^2\big|X\big] = \|\widehat{\Lambda}_{J,n}^{1/2} B_n\|_2^2 = o_P(\sqrt{J})$, so that

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

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 (4.3) holds by (6.27) if $J = O(\widehat{\nu})$, by the central limit theorem for iid χ_1^2-variables. 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 $\sigma_n\widehat{\nu} \ge \mathrm{trace}\big(\widehat{\Lambda}_{J,n}\big) = \sum_{i=1}^n \|z_i\|^2(J - \widehat{K}_i)$. This implies $\widehat{\nu} \ge J(1-\kappa^*)$ and completes the proof of part (i).

(ii) First of all $P\{\widehat{\nu} = d(J-1)\} \to 1$ by Theorem 3. Since $\|\widehat{\Lambda}^p_{J,n}\| = O_P(\sigma_n^p)$ by (6.26), by (6.25), (4.17) and (4.7) with $(p,q) = (1,1)$

$$\|\widehat{\Lambda}^k_{J,n} B_n\|_2^2 \le \|\widehat{\Lambda}^k_{J,n}\|^2 \|B_n\|_2^2 = O_P(\sigma_n^{2k}) O_P(J/\sigma_n^2) \sum_{i=1}^n \|z_i\|^2 \rho_{J,i}^2 = o_P(\sigma_n^{2k-1}\sqrt{J}) \tag{6.28}$$

for all real k. The proof of (4.3) is as in part (i) since $\big\|\widehat{\Lambda}^{1/2}_{J,n} B_n\big\|_2^2 = o_P(\sqrt{J})$. For (4.4), we have

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

Since $E\big[\big|\big\langle B_n \widehat{\Sigma}^{1/2}_{J,n}, \widehat{\Lambda}^{-1/2}_{J,n} Z_n \widehat{\Sigma}^{1/2}_{J,n}\big\rangle_2\big|^2 \big| X\big] = \big\|\widehat{\Lambda}^{-1/2}_{J,n} B_n \widehat{\Sigma}_{J,n}\big\|_2^2 \le \|\widehat{\Sigma}_{J,n}\|^2 \|\widehat{\Lambda}^{-1/2}_{J,n} B_n\|_2^2 = o_P(\sqrt{J})$ and $\|B_n \widehat{\Sigma}^{1/2}_{J,n}\|_2^2 \le \sigma_n \|B_n\|_2^2 = o_P(\sqrt{J})$,

$$\sum_{j=1}^J \Big\|\widehat{\Sigma}^{1/2}_{J,n}(\widehat{\beta}_j - \beta_j)\Big\|^2 = \Big\|\widehat{\Lambda}^{-1/2}_{J,n} Z_n \widehat{\Sigma}^{1/2}_{J,n}\Big\|_2^2 + o_P(\sqrt{J}). \tag{6.29}$$

Since $\widehat{\Lambda}^{-1/2}_{J,n} Z_n \widehat{\Sigma}^{1/2}_{J,n} = (I_{J,0} \otimes \widehat{\Sigma}^{1/2}_{J,n})\widehat{\Lambda}^{-1/2}_{J,n} Z_n$, by Cauchy-Schwarz and (6.26)

$$\begin{aligned}
& E\Big[\Big|\|\widehat{\Lambda}^{-1/2}_{J,n} Z_n \widehat{\Sigma}^{1/2}_{J,n}\|_2^2 - \|Z_n\|_2^2\Big| \,\Big| X\Big] \\
= \;& E\Big[\Big|\big\langle\big\{\widehat{\Lambda}^{-1/2}_{J,n}(I_{J_0} \otimes \widehat{\Sigma}_{J,n})\widehat{\Lambda}^{-1/2}_{J,n} - I_{J,0} \otimes I_d\big\} Z_n, Z_n\big\rangle\Big| \,\Big| X\Big] \\
\le \;& \Big\|\widehat{\Lambda}^{-1/2}_{J,n}(I_{J_0} \otimes \widehat{\Sigma}_{J,n})\widehat{\Lambda}^{-1/2}_{J,n} - I_{J,0} \otimes I_d\Big\|_2 \sqrt{E[\|Z_n\|_2^2|X]} \\
\le \;& \|\widehat{\Lambda}^{-1}_{J,n}\| \Big\|I_{J_0} \otimes \widehat{\Sigma}_{J,n} - \widehat{\Lambda}_{J,n}\Big\|_2 \sqrt{d(J-1)} = o_P(\sqrt{J}).
\end{aligned} \tag{6.30}$$

This and (6.29) imply $\sum_{j=1}^J \big\|\widehat{\Sigma}^{1/2}_{J,n}(\widehat{\beta}_j - \beta_j)\big\|^2 = \big\|Z_n\big\|_2^2 + o_P(\sqrt{J})$, which then implies (4.4) via (6.27) and the proof of (4.3). The proof of (4.5) is simpler. We obtain from (6.28)

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

and then obtain $\|\widehat{\Lambda}^{-1/2}_{J,n} Z_n\|_2^2 = \|(I_{J,0} \otimes \widehat{\Sigma}_{J,n})^{-1/2} Z_n\|_2^2 + o_P(\sqrt{J}/\sigma_n)$ as in (6.30). These imply (4.5) since $\|(I_{J,0} \otimes \widehat{\Sigma}_{J,n})^{-1/2} Z_n\|_2^2/\sigma^2$ given X is a sum of $J-1$ iid $\|N(0, \widehat{\Sigma}^{-1}_{J,n})\|^2$ variables with common mean $\mathrm{trace}(\widehat{\Sigma}^{-1}_{J,n})$ and variance $2\,\mathrm{trace}(\widehat{\Sigma}^{-2}_{J,n})$ of the order σ_n^{-2}. We omit some details.

Finally, let us prove (4.10). By (4.18), $\sigma_n^{1/2}\big\{\widehat{\beta}_j - \beta_j - (\widehat{\Lambda}^{-1/2}_{J,n} Z_n)'\mathbf{e}_j\big\} = o_P(1)$ uniformly in j. Since Z_n/σ is a standard normal matrix and $\mathbf{e}_j \boldsymbol{b}'$ is a $J \times d$ matrix, conditionally on X

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

Since $\big\|\mathbf{e}_j\boldsymbol{b}'\big\|_2^2 = \|\mathbf{e}_j\|^2\|\boldsymbol{b}'\|^2$ and $\big\langle \mathbf{e}_j\boldsymbol{b}', (I_{J,0}\otimes\widehat{\Sigma}_{J,n})^{-1}(\mathbf{e}_j\boldsymbol{b}')\big\rangle_2 = (1-1/J)\boldsymbol{b}'\widehat{\Sigma}_{J,n}^{-1}\boldsymbol{b}$, by (6.26)

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

uniformly in j and $\|\boldsymbol{b}'\| = 1$. 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. Hence, since $\widehat{\sigma} = \sigma + o_P(1)$ via the inequalities in (6.27) and the first part of condition (4.7) with $q = 0$, (4.10) holds and the proof is complete.

Proof of Theorem 2. Since $I_{J,0}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, by the definition of $\widehat{f}_i^*$ and $\widetilde{f}_i$

$$\widehat{f}_i^*(\boldsymbol{x}_i) = Q_i^*(\boldsymbol{y}_i - \widehat{\boldsymbol{\beta}}z_i) = \widetilde{f}_i(\boldsymbol{x}_i) - Q_i^* I_{J,0}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})z_i. \tag{6.31}$$

Define $\zeta_n^2 \equiv \sum_{i=1}^n \|\widehat{f}_i^*(\boldsymbol{x}_i) - \widetilde{f}_i(\boldsymbol{x}_i)\|^2$, $A_i^* = \big(Q_i^* I_{J,0}\big)' Q_i^* I_{J,0}$ and $A^* \equiv \sum_{i=1}^n A_i^* \otimes z_i z_i'$. By (6.31) $\zeta_n^2 = \big\langle \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, A^*(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\big\rangle_2$. Since $\|Q_i^*\|_2^2 = \text{trace}(A_i^*)$, we find as in the proof of Proposition 1 that

$$EA^* = \sum_{i=1}^n I_{J,0} E A_i^* I_{J,0} \otimes z_i z_i' = I_{J,0} \otimes \sum_{i=1}^n \frac{E\|Q_i^*\|_2^2 - 1}{J-1} z_i z_i' \le \eta_{J,n}' I_{J,0} \otimes \Sigma_n$$

and $E\big\|A^* - EA^*\big\|_2^2 \le \sum_{i=1}^n \|z_i\|^4 E\|A_i^*\|_2^2 \le \sum_{i=1}^n M_0^2 E\|Q_i^*\|_2^2 \|z_i\|^4 \le M_0^2 \eta_{J,n}' J \eta_{J,n} \sigma_n^2$. Thus,

$$E\|A^*\| \le \|EA^*\| + E\|A^* - EA^*\|_2 \le \eta_{J,n}'\sigma_n + M_0\sqrt{J\eta_{J,n}'\eta_{J,n}}\sigma_n,$$

and $E\text{trace}(A^*) \le \eta_{J,n}' J\sigma_n$. By (6.21), Theorem 3, and (6.26)

$$\begin{aligned}
E\Big[\zeta_n^2\Big|X\Big] &= E\Big[\big\langle B_n + \widehat{\Lambda}_{J,n}^{-1/2}Z_n, A^*(B_n + \widehat{\Lambda}_{J,n}^{-1/2}Z_n)\big\rangle_2\Big|X\Big] \\
&= \big\langle B_n, A^* B_n\big\rangle_2 + \sigma^2\big\|(A^*)^{1/2}\widehat{\Lambda}_{J,n}^{-1/2}\big\|_2^2 \\
&\le \|A^*\|\|B_n\|_2^2 + \sigma^2 \text{trace}(A^*)/\sigma_n.
\end{aligned}$$

This implies (4.13) after applications of (4.17) of Theorem 4 and the above bounds for $\|A^*\|$ and trace(A^*).

References

1. Bickel PJ, Klaassen CAJ, Ritov Y, and Wellner JA (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, Baltimore.

2. Brown PO and Botstein D (1999). Exploring the new world of the genome with microarrays. Nat. Genet., 21 (suppl. 1), 33-37.

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

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

5. Cleveland, WS (1988). Locally-weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association, 83, 596-610

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

7. Engle RF, Granger CWJ, Rice J and Weiss A (1986). Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association, 81: 310-320.

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

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

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

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

12. Huber PJ (1981) *Robust Statistics*. John Wiley, New York.

13. Kerr K and Churchill G (2000). Analysis of variance for gene expression microarray data. Journal of Computational Biology 7, 819-837.

14. Rice J (1986). Convergence rates for partially spline models. Statistics and Probability Letter 4, 203-208.

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

16. Tseng GC, Oh M-K, Rohlin L, Liao JC and Wong WH (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variation and assessment of gene effects. Nucleic Acids Research, Vol 29, No. 12. 2549-2557.

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