Analysis of Variance and *F*-Tests for Partial Linear Models With Applications to Environmental Health Data

Li-Shan HUANG and Phillip W. DAVIDSON

Fish consumption during pregnancy exposes the fetus to both the neurotoxicant methylmercury and nutrients known to be beneficial for brain development. When nutrient status is not measured, maternal methylmercury levels may be a partial biomarker for both toxic and nutrient exposures. It is therefore necessary to employ a flexible model—such as the partial linear model—that will allow for possible nonlinear trends of methylmercury. To enhance interpretations of fitting a partial linear model, we propose analysis of variance (ANOVA) inference tools including ANOVA decomposition and significance tests. The ANOVA decomposition explicitly gives the proportion of variation explained by the model and separates the contributions from the parametric and nonparametric components. Semiparametric F-tests are constructed based on ANOVA decomposition with the normality assumption. The proposed F-tests are applicable to testing significance of the parametric, the nonparametric, and the combination of both. The ANOVA investigation also yields new byproduct estimators, which can be viewed as penalized least squares estimators. Simulation results demonstrate that the performance of the new estimators and ANOVA F-tests is comparable to alternative methods in practical applications. This methodology is applied to reanalyze the Seychelles Child Development Study Main Cohort data to explore nonlinear relationships of prenatal methylmercury exposure through maternal fish consumption with prenatal and postnatal child development.

KEY WORDS: Local linear regression; Methylmercury; Penalized least squares; Smoother matrix.

1. INTRODUCTION

Fish is an important source of nutrition worldwide. All fish consumption leads to some degree of methylmercury exposure, which is the primary exposure source in humans (Myers, Davidson, and Shamlaye 2006). The developing brain is especially sensitive to this neurotoxin, prompting concern that pregnant women should limit their fish consumption (ATSDR 1999), while it is unclear whether methylmercury exposure at the levels resulting from ocean fish consumption poses a significant threat to children's neurological development. Some studies have found subtle associations between exposure and endpoints while others have not (National Institute of Environmental Health Sciences 1998; National Research Council 2000). This is a critical issue because fish is an important source of nutrients, such as long chain polyunsaturated fatty acids (LCPUFA), that are important to the developing brain. Because both nutrients and methylmercury coexist in fish, the effects may be quite complex and whether pregnant mothers should limit fish consumption is presently unclear (Davidson et al. 2008; Strain et al. 2008).

To help clarify this important public health issue, this article attempts to develop statistical tools to better assess and interpret the nonlinear effects of prenatal methylmercury exposure from maternal fish consumption in the Seychelles Child Development Study (SCDS) (Myers et al. 2003). The SCDS data on the Main Cohort include the maternal methylmercury exposure but no nutrient measurements. It is therefore necessary to employ a flexible statistical model that allows for the possibility of nonlinear trends of methylmercury in order to explore the combined effects of both nutrients and methylmercury from fish consumption.

Extending from classical linear models, partial linear models (PLMs) (see Härdle, Liang, and Gao 2000 for a review) allow for nonparametric estimation of effects of one covariate while retaining linear relationships for the remaining explanatory variables. The form of a PLM is

$$Y = \mathbf{z}^T\boldsymbol{\alpha} + m(X) + \varepsilon, \tag{1}$$

where $\mathbf{z}$ is a p-dimensional covariate vector, X is a one-dimensional predictor, $m(\cdot)$ is an unknown smooth function, and ε has mean 0 and finite variance σ^2. For simplicity, we assume that the intercept term is incorporated in $m(X)$ but not in $\boldsymbol{\alpha}$. PLM is a special case of semiparametric additive models (SAMs) (Hastie and Tibshirani 1990), which allow for multiple nonparametric components. Though SAMs or PLMs are increasingly popular, their application tools such as Analysis of Variance (ANOVA) tables and significance tests are limited. The original motivation for the present paper came from the SCDS. After fitting SAMs in Huang et al. (2005) for the SCDS Main Cohort data at age nine years, we wanted to present ANOVA tables, to partition variation between different sources, and to assess the significance of $m(X)$ and other covariates using more standard tests.

In the case of univariate nonparametric regression, Huang and Chen (2008) studied ANOVA and related F-tests based on local linear regression (Fan and Gijbels 1996). Extending from Huang and Chen (2008), this paper presents a set of analysis tools for PLMs covering ANOVA decomposition and significance tests. The development also leads to new estimators. We are not aware of any ANOVA decomposition for PLMs in the

Li-Shan Huang is Associate Professor, Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Ave, Box 630, Rochester, NY 14642 (E-mail: *Lhuang@bst.rochester.edu*). Phillip W. Davidson is Professor of Pediatrics and Environmental Medicine, University of Rochester, 601 Elmwood Ave, Box 671, Rochester, NY 14642 (E-mail: *phil_davidson@urmc.rochester.edu*). The work was partially supported by National Institutes of Health grants R01-08442, R01-ES10219, R01-ES01247. The authors thank Drs. Christopher Cox and Gary J. Myers, the editor, the associate editor, and the referees for helpful comments and suggestions.

Journal of the American Statistical Association
Accepted for publication, Applications and Case Studies
DOI: 10.1198/jasa.2010.ap08274

literature, though the idea of ANOVA has been adopted to develop nonparametric tests (e.g., Dette and Neumeyer 2001). On hypothesis testing for PLMs, we found only a few papers, and most were focused on testing for the nonparametric term $m(\cdot)$ (e.g., Härdle, Muller, and Mammen 1998). In fact, most literature in the semiparametric context has been devoted to estimation rather than inference. Recently Fan and Huang (2005) proposed profile likelihood ratio (PLR) tests that accommodated for checking the significance of either parametric or nonparametric terms in semiparametric varying coefficient models (Hastie and Tibshirani 1993), which included PLMs as a special case.

The outline of this paper is as follows. Section 2 presents the SCDS nine-year data and introduces nonparametric ANOVA by Huang and Chen (2008). A global ANOVA decomposition for PLMs is discussed in Section 3.1 and an R-squared is defined to measure the proportion of variation explained by fitting a PLM. New estimators are developed in the ANOVA framework as well. Based on the ANOVA decomposition, it is straightforward to identify the sources of variation explained by the nonparametric or parametric components and then construct F-tests that parallel those in linear models. In Section 3.2, we discuss semiparametric F-tests and Theorem 1 shows that the proposed F-test statistics have asymptotic F-distributions when ε is normal. Compared to other F-tests in the literature (e.g., Hastie and Tibshirani 1990; Ruppert, Wand, and Carroll 2003), our F-tests come with an ANOVA decomposition and asymptotic F-distributions without further calibration. Section 3.3 discusses asymptotic properties of the ANOVA-based new estimators and Section 4 includes a simulation study to examine the performance of the proposed estimators and F-tests. We report the results of reanalyzing the SCDS nine-year data based on proposed methodologies in Section 5. The results on the child's IQ outcome are somewhat different from earlier analyses and highlight the needs to assess significance of nonlinear trends in applications. Technical details are postponed to the Appendix.

2. DATA AND BACKGROUND

2.1 Seychelles Study

Several outbreaks of poisoning following exposure to high concentrations of mercury took place during the 20th century. Prenatal poisoning from fish consumption in Japan and contaminated grain in Iraq were both reported to cause irreparable damage to the developing fetal central nervous system. Both outbreaks were high dose exposures. Following the Iraq exposure it was proposed that prenatal methylmercury exposure above 10 ppm in maternal hair may be associated with declines in neurodevelopment (Cox et al. 1989). Frequent fish consumers are known to reach such levels of exposure. Epidemiologic studies have examined this question and reported both presence and absence of an association between prenatal methylmercury exposure from fish consumption and child development (National Institute of Environmental Health Sciences 1998; National Research Council 2000). Early studies focused only on methylmercury exposure. However, given that fish contain nutrients that are important to fetal brain development, more recent studies have included nutrients. Still, the question remains unanswered. Therefore it is critical that statistical models adopted for evaluating complicated effects of nutrient and methylmercury coexposures are flexible enough to accommodate possible nonlinear effects of methylmercury when no nutrient variables are available.

The SCDS has examined the association between prenatal exposure to methylmercury from maternal fish consumption and child development. The Main Cohort was enrolled in 1989–1990 and consists of 779 mother–child pairs in the Republic of Seychelles. Prenatal methylmercury exposure was assessed in a segment of maternal hair corresponding to growth during pregnancy. Nutrients were not studied in the Main Cohort (but now studied in a new cohort). The children's neurological and developmental status was evaluated longitudinally. At 9 years of age, 643 children returned for testing. Analysis using conventional linear regression models (Myers et al. 2003) found that out of 21 endpoints, there was one association where a developmental score decreased with increasing prenatal exposure and one association where the endpoint score improved with increasing exposure. Huang et al. (2005), based on the SCDS nine-year data, suggested that the combined effects of prenatal methylmercury exposure and nutrients from fish, as a function of methylmercury, may be nonlinear. This was based on the application of SAMs with the effects of six continuous covariates, including methylmercury, modeled as smooth functions using smoothing splines. The significance of nonlinear methylmercury effects is yet to be investigated, since the approximate F-tests (Hastie and Tibshirani 1990) adopted by Huang et al. (2005) were ad hoc in nature.

In biomedical studies, research results are often reported focusing on p-values from statistical tests. From a public health perspective it is important to assess the significance of methylmercury effects from maternal fish consumption on child development. When the methylmercury effects may be nonlinear due to nutrient interactions, not only is a flexible procedure such as PLMs necessary, it is also important to extend traditional test statistics to enable the assessment of the significance of nonlinear fits. In addition, the SCDS investigators requested ANOVA tables and semiparametric R-squared to enhance the interpretations, which are standard in linear analysis. Based on classical ANOVA ideas and extending nonparametric ANOVA in Huang and Chen (2008), we have developed an array of analysis tools for PLMs in Section 3.

2.2 Nonparametric ANOVA

We briefly review the ANOVA results in Huang and Chen (2008) for univariate local linear regression (Fan and Gijbels 1996). Assume data pairs (X_i, Y_i), $i = 1, \ldots, n$, are independently drawn from the following model:

$$Y = m(X) + \varepsilon, \tag{2}$$

where X and ε are independent, and ε has a mean 0 and unknown variance σ^2. The local linear estimator $\hat{\beta}_0$ for estimating $m(x) = E(Y|X = x)$ and an estimate $\hat{\beta}_1$ for $m'(x)$ solve the following weighted least squares problem:

$$\min_{\beta_0,\beta_1} n^{-1} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(X_i - x))^2 K_h(X_i - x), \tag{3}$$

where $K_h(\cdot) = K(\cdot/h)/h$, and the dependence of β_0 and β_1 on x and h is suppressed. The function $K(\cdot)$ is a nonnegative weight function, typically a symmetric probability density function, and h is the smoothing parameter determining the neighborhood size for local fitting.

Huang and Chen (2008) proposed ANOVA quantities for (2) based on integrated sums of squares:

$$\begin{aligned} SSE_1(h) &= n^{-1} \int \sum_{i=1}^{n} (Y_i - \hat\beta_0(x) - \hat\beta_1(x)(X_i - x))^2 \\ &\qquad \times K_h(X_i - x)\, dx, \\ SSR_1(h) &= n^{-1} \int \sum_{i=1}^{n} (\hat\beta_0(x) + \hat\beta_1(x)(X_i - x) - \bar Y)^2 \\ &\qquad \times K_h(X_i - x)\, dx, \\ SST(h) &= n^{-1} \int \sum_{i=1}^{n} (Y_i - \bar Y)^2 K_h(X_i - x)\, dx, \end{aligned} \tag{4}$$

and showed that an ANOVA decomposition holds: $SSE_1(h) + SSR_1(h) = SST(h)$ and $SST(h) = SST \equiv n^{-1} \sum_{i=1}^{n} (Y_i - \bar Y)^2$ under some conditions. In a matrix expression, (4) is written as

$$\begin{aligned} SSE_1(h) &= n^{-1} \mathbf{y}^\top (\mathbf{I} - \mathbf{H}^*) \mathbf{y}, \\ SSR_1(h) &= n^{-1} \mathbf{y}^\top (\mathbf{H}^* - \mathbf{L}) \mathbf{y}, \end{aligned} \tag{5}$$

and $SST = n^{-1} \mathbf{y}^\top (\mathbf{I} - \mathbf{L}) \mathbf{y}$, where $\mathbf{y} = (Y_1, \ldots, Y_n)^\top$, $\mathbf{L}$ is an $n \times n$ matrix with entries $1/n$, and $\mathbf{H}^*$ is a symmetric matrix depending only on X_i's, bandwidth h, and $K(\cdot)$. Under conditions (A3)–(A5) in the Appendix, $\mathbf{H}^* \mathbf{1} = \mathbf{1}$ with $\mathbf{1}$ an n-vector of 1, and $\mathbf{H}^* \mathbf{x} = \mathbf{x}$ with $\mathbf{x} = (X_1, \ldots, X_n)^\top$. In other words, simple linear models are nested in the space spanned by $\mathbf{H}^*$. In addition, $\mathbf{H}^*$ is asymptotically idempotent based on the following definition:

Definition 1. Conditioned on $\mathbf{x}$, an $n \times n$ matrix $\mathbf{A}_n$ that depends only on X_i's, $K(\cdot)$, and h, is asymptotically idempotent if for any random n-vector $\mathbf{y}_n$ with entries of order $O(1)(1 + o_p(1))$, $E\{(\mathbf{A}_n - \mathbf{A}_n^2)\mathbf{y}_n | \mathbf{x}\}$ tends to a zero vector in probability as $n \to \infty$, $h \to 0$, and $nh \to \infty$.

Huang and Chen (2008) showed that elements of $E\{(\mathbf{H}^* - \mathbf{H}^{*2})\mathbf{y}|\mathbf{x}\}$ are of order h^2 and therefore $\mathbf{H}^*$ is an asymptotic projection matrix.

Under model (2), the projection estimate is of the form

$$\mathbf{H}^* \mathbf{y} = \mathbf{y}^* = \begin{pmatrix} \int (\hat\beta_0(x) + \hat\beta_1(x)(X_1 - x)) K_h(X_1 - x)\, dx \\ \vdots \\ \int (\hat\beta_0(x) + \hat\beta_1(x)(X_n - x)) K_h(X_n - x)\, dx \end{pmatrix}. \tag{6}$$

We see that $\mathbf{y}^*$ makes use of both $\hat\beta_0(\cdot)$ and $\hat\beta_1(\cdot)$, in contrast to conventional local linear regression using only $\hat\beta_0(\cdot)$. It was shown in Huang and Chen (2008) that the elements in $\mathbf{y}^* = (Y_1^*, \ldots, Y_n^*)^\top$ have an asymptotic bias of order h^4, while the asymptotic variance has the same rate of $n^{-1}h^{-1}$ as that of $\hat\beta_0(X_i)$:

$$\begin{aligned} E\{Y_i^* - m(X_i)|\mathbf{x}\} &= [1 + o_p(1)] h^4 b(X_i), \\ \operatorname{var}\{Y_i^* | \mathbf{x}\} &= [1 + o_p(1)] n^{-1} h^{-1} \kappa_0 \sigma^2 / f(X_i), \end{aligned} \tag{7}$$

where $b(X_i) = (\frac{\mu_2^2 - \mu_4}{4})\{m^{(4)}(X_i) + 2m^{(3)}(X_i)\frac{f'(X_i)}{f(X_i)} + m^{(2)}(X_i) \times \frac{f''(X_i)}{f(X_i)}\}$ and $\kappa_0 = \int (K_0^*(v) - \frac{1}{\mu_2} K_1^*(v))^2\, dv$, with μ_j the jth moment of $K(\cdot)$, $K_0^*(\cdot)$ being the convolution of the kernel function $K(\cdot)$ and itself, and $K_1^*(\cdot)$ the convolution of $uK(u)$ and itself. He and Huang (**?**HeHuang2008) explored further the estimator <ref:HeHuang with the form (6) for estimation at a grid point t (not necessarily at observed X_i's). Here

$$\int (\hat\beta_0(x) + \hat\beta_1(x)(t - x)) K_h(t - x)\, dx, \tag{8}$$

and they show that the asymptotic bias and variance of (8) are identical to (7) except that the argument X_i is replaced by t.

3. ANOVA FRAMEWORK

Analysis of variance for a regression model helps in understanding the quality of model fitting and the amount of variation explained by the fitted model. ANOVA is also a necessary component in classical F-tests for parametric models. In this section, we develop an ANOVA framework for PLMs, including an ANOVA decomposition, an empirical R^2, new estimators, and semiparametric F-tests.

3.1 ANOVA Decomposition and Estimators

Assume that data $(\mathbf{z}_i, X_i, Y_i)$, $i = 1, \ldots, n$, are drawn independently from model (1) with $\mathbf{z}_i = (Z_{i1}, \ldots, Z_{ip})^\top$. Let $\mathcal{Z}_{n \times p}$ be the centered data matrix for the $\mathbf{z}_i$'s, and $\mathbf{m} = (m(X_1), \ldots, m(X_n))^\top$. Model (1) in a matrix notation is

$$\mathbf{y} = \mathcal{Z}\boldsymbol\alpha + \mathbf{m} + \boldsymbol\varepsilon.$$

Intuitively, if an estimate of $\boldsymbol\alpha$, say $\hat{\boldsymbol\alpha}$ is available, then the SSE and SSR expressions in both (4) and (5) may be adapted for PLMs as

$$\begin{aligned} SSE_p(h) &= n^{-1} \int \sum_{i=1}^{n} (Y_i - \mathbf{z}_i \hat{\boldsymbol\alpha} - \hat\gamma_0(x) - \hat\gamma_1(x)(X_i - x))^2 \\ &\qquad \times K_h(X_i - x)\, dx \\ &= n^{-1} (\mathbf{y} - \mathcal{Z}\hat{\boldsymbol\alpha})^\top (\mathbf{I} - \mathbf{H}^*)(\mathbf{y} - \mathcal{Z}\hat{\boldsymbol\alpha}), \\ SSR_p(h) &= n^{-1} \int \sum_{i=1}^{n} (\mathbf{z}_i \hat{\boldsymbol\alpha} + \hat\gamma_0(x) + \hat\gamma_1(x)(X_i - x) - \bar Y)^2 \\ &\qquad \times K_h(X_i - x)\, dx, \end{aligned} \tag{9}$$

where $\hat\gamma_0(x)$ and $\hat\gamma_1(x)$ denote the resulting solution to (3) at a point x with the response $(\mathbf{y} - \mathcal{Z}\hat{\boldsymbol\alpha})$. Hence we will need an estimator for $\boldsymbol\alpha$. Mimicking the estimators in Speckman (1988) and Opsomer and Ruppert (1999) and replacing their smoother matrix $\mathbf{S}$ by $\mathbf{H}^*$ [$\mathbf{S}$ is defined as $\mathbf{S}\mathbf{y} = (\hat\beta_0(X_1), \ldots, \hat\beta_0(X_n))^\top$], a new estimator for $\boldsymbol\alpha$ is defined

$$\hat{\boldsymbol\alpha} = (\mathcal{Z}^\top (\mathbf{I} - \mathbf{H}^*) \mathcal{Z})^{-1} \mathcal{Z}^\top (\mathbf{I} - \mathbf{H}^*) \mathbf{y}. \tag{10}$$

Then under conditions (A3)–(A5) in the Appendix, we will show that the following ANOVA decomposition holds for PLMs with the new estimator $\hat{\boldsymbol\alpha}$:

$$SSE_p(h) + SSR_p(h) = SST, \tag{11}$$

where SST is as given in (4). After plugging (10) in (9),

$$(\mathcal{Z}\hat{\boldsymbol\alpha})^\top (\mathbf{I} - \mathbf{H}^*)(\mathcal{Z}\hat{\boldsymbol\alpha}) = (\mathcal{Z}\hat{\boldsymbol\alpha})^\top (\mathbf{I} - \mathbf{H}^*) \mathbf{y}.$$

In other words, $\mathcal{Z}\hat{\boldsymbol{\alpha}}$ and $(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})$ are orthogonal with respect to the $(\mathbf{I}-\mathbf{H}^*)$-weighted inner product, $(\mathcal{Z}\hat{\boldsymbol{\alpha}})^\top(\mathbf{I}-\mathbf{H}^*)(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})=\mathbf{0}$. Hence

$$SSE_p(h)=n^{-1}(\mathbf{y}^\top(\mathbf{I}-\mathbf{H}^*)\mathbf{y}-(\mathcal{Z}\hat{\boldsymbol{\alpha}})^\top(\mathbf{I}-\mathbf{H}^*)(\mathcal{Z}\hat{\boldsymbol{\alpha}})). \quad (12)$$

Compared to (4) under model (2), we clearly see in (12) that with the additional parametric term $\mathbf{z}^\top\hat{\boldsymbol{\alpha}}$, the SSE is reduced by the amount of $n^{-1}(\mathcal{Z}\hat{\boldsymbol{\alpha}})^\top(\mathbf{I}-\mathbf{H}^*)(\mathcal{Z}\hat{\boldsymbol{\alpha}})$. For $SSR_p(h)$, note that $\int\sum_i(\hat{\gamma}_0(x)+\hat{\gamma}_1(x)(X_i-x))^2K_h(X_i-x)\,dx=(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})^\top H^*(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})$ and $\int\sum_i(\hat{\gamma}_0(x)+\hat{\gamma}_1(x)(X_i-x))K_h(X_i-x)\,dx=\mathbf{H}^*(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})$. Then

$$SSR_p(h)=n^{-1}(\mathbf{y}^\top(\mathbf{H}^*-\mathbf{L})\mathbf{y}+(\mathcal{Z}\hat{\boldsymbol{\alpha}})^\top(\mathbf{I}-\mathbf{H}^*)(\mathcal{Z}\hat{\boldsymbol{\alpha}})). \quad (13)$$

Therefore the two terms in the RHS of (13) may be interpreted as nonparametric and parametric SSR respectively. From (12) and (13), it is clear that the ANOVA decomposition (11) holds, and a measure of variation explained by a fitted PLM can then be defined as

$$R^2(h)=\frac{SSR_p(h)}{SST}=1-\frac{SSE_p(h)}{SST}. \quad (14)$$

It is an extension of the nonparametric R-squared for model (2) defined in Huang and Chen (2008).

Similar to (6), we may define a new estimator for $\mathbf{m}$:

$$\hat{\mathbf{m}}=(\hat{m}(X_1),\ldots,\hat{m}(X_n))^\top=\mathbf{H}^*(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}}). \quad (15)$$

It is clear that $\hat{\mathbf{m}}$ is obtained by projecting $(\mathbf{y}-\mathcal{Z}\hat{\boldsymbol{\alpha}})$ to the space spanned by $\mathbf{H}^*$. In (10) and (15), estimators are expressed with a centered $\mathcal{Z}$ and an uncentered $\mathbf{H}^*$ to be consistent with (1). Alternatively, (10) and (15) can be expressed using a centered $\mathbf{H}^*$, which is $(\mathbf{H}^*-\mathbf{L})$, and including one more column $\mathbf{1}$ in the uncentered $\mathcal{Z}$ for the intercept. The new estimators are developed as a byproduct of seeking an ANOVA decomposition for PLMs. We delay the discussion on the properties of new estimators to Section 3.3 and focus next on developing ANOVA-based semiparametric F-tests.

3.2 Semiparametric F–Tests

Hypothesis testing to check the significance of terms in (1) is very useful in data analysis applications. For example, in the SCDS, it is important to assess whether the effect of prenatal methylmercury exposure is significant and whether a nonlinear curve is significantly better than a linear fit. Though some work has been done in this area as mentioned in the Introduction, it will be convenient to have some simple testing procedures analogous to those in the parametric setting. Motivated by the ANOVA decomposition in Section 3.1 and the partition of $SSR_p(h)$ into nonparametric and parametric SSR in (13), we develop ANOVA-based F-tests under the normality assumption for PLMs in this section. The model under the alternative hypothesis H_a is the PLM (1) and we will first focus the discussion on the following four null hypotheses:

- H_0(i): Testing overall model significance, $H_0: E\{Y|X,\mathbf{z}\}=c_0$ with c_0 a constant. This is based on the ANOVA decomposition (11) and the test will examine whether SSR is sufficiently large in comparison with SSE.
- H_0(ii): Testing overall significance of nonintercept parametric effects, $H_0:\boldsymbol{\alpha}=\mathbf{0}$. The parametric SSR term in (13) will become the difference in SSR under the null and alternative hypotheses.
- H_0(iii): Testing if the nonparametric term is significantly different from a constant, $H_0: m(x)=b_0$ with b_0 a constant. The nonparametric SSR in (13) will become the difference in SSR under the null and alternative hypotheses.
- H_0(iv): Testing if the nonparametric effect is linear, $H_0: m(x)=b_0+b_1x$. Since a linear model is nested in local linear regression, this will be analogous to classical F-tests between nested models.

It will be clear later that the proposed ANOVA F-tests are flexible and can accommodate checking the significance for parametric or nonparametric terms or their combinations.

We first rewrite both $SSE_p(h)$ and $SSR_p(h)$ in different quadratic forms from those in (12) and (13). Let $\mathbf{P}_{z|x}=\mathcal{Z}(\mathcal{Z}^\top(\mathbf{I}-\mathbf{H}^*)\mathcal{Z})^{-1}\mathcal{Z}^\top(\mathbf{I}-\mathbf{H}^*)$ so that $\mathcal{Z}\hat{\boldsymbol{\alpha}}=\mathbf{P}_{z|x}\mathbf{y}$. Note that $P_{z|x}$ is an orthogonal projector satisfying $\mathbf{P}_{z|x}^2=\mathbf{P}_{z|x}$ (idempotent) and $(\mathbf{I}-\mathbf{H}^*)\mathbf{P}_{z|x}=\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*)$ [symmetric with respect to the $(\mathbf{I}-\mathbf{H}^*)$-weighted inner product]. Therefore, $\mathbf{P}_{z|x}\mathcal{Z}=\mathcal{Z}$ and $(\mathbf{I}-\mathbf{P}_{z|x})^\top(\mathbf{I}-\mathbf{H}^*)\mathbf{P}_{z|x}=\mathbf{0}$. Thus $\mathcal{Z}\hat{\boldsymbol{\alpha}}$ is the projection of $\mathbf{y}$ into the spaced spanned by $\mathcal{Z}$ with respect to the $(\mathbf{I}-\mathbf{H}^*)$-weighted inner product that is asymptotically orthogonal to $\mathbf{H}^*$. Then

$$\begin{aligned}SSE_p(h)&=n^{-1}\mathbf{y}^\top\big(\mathbf{I}-\mathbf{P}_{z|x}^\top\big)(\mathbf{I}-\mathbf{H}^*)\mathbf{y},\\ SSR_p(h)&=n^{-1}\mathbf{y}^\top\big(\mathbf{H}^*-\mathbf{L}+\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*)\mathbf{P}_{z|x}\big)\mathbf{y}.\end{aligned} \quad (16)$$

Using the quadratic expressions, the proposed F-test statistics for H_0(i)–(iv) are developed by mimicking classical F-tests; that is, by forming a ratio with the numerator being the difference in SSR under H_0 and H_a divided by the difference in their degrees of freedom (DF), and the denominator being the SSE under H_a divided by the DF under H_a. To formalize the proposed F-tests, we also need criteria for "asymptotic orthogonality" and "asymptotic idempotency" for the matrices sandwiched in the quadratic forms. Following Huang and Chen (2008), we define "asymptotic orthogonality" between two matrices as follows:

Definition 2. Assume that $n\times n$ matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ depend only on $\{\mathbf{x},\mathcal{Z}\}$, $K(\cdot)$, and h. Conditioned on $\{\mathbf{x},\mathcal{Z}\}$, $\mathbf{A}_n$ and $\mathbf{B}_n$ are asymptotically orthogonal if for any random n-vector $\mathbf{y}_n$ with entries of order $O(1)(1+o_p(1))$, $E\{\mathbf{A}_n\mathbf{B}_n\mathbf{y}_n|\mathbf{x},\mathcal{Z}\}$ tends to a zero vector in probability as $n\to\infty$, $h\to 0$, and $nh\to\infty$.

Similarly, the asymptotic idempotency for matrices resulting from fitting PLMs is defined analogously as in Definition 1 except that the conditioned variables are $\mathbf{x}$ and $\mathcal{Z}$. We formally state the ANOVA F-tests for H_0(i)–(iv) in the following theorem.

Theorem 1. Assume that the bandwidth $h\to 0$ and $nh\to\infty$ as $n\to\infty$, and Conditions (A) in the Appendix hold with ε having a normal distribution. Conditioned on $\{\mathbf{x},\mathcal{Z}\}$, the F-statistics for testing H_0(i)–(iv) have a common denominator

$$\frac{nSSE_p(h)}{n-\mathrm{tr}(\mathbf{H}^*+\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*))}, \quad (17)$$

and the numerator is respectively

$$\begin{aligned} Q_1 &= \frac{nSSR_p(h)}{\mathrm{tr}(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*)) - 1}; \\ Q_2 &= \frac{\mathbf{y}^\top(\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*))\mathbf{y}}{\mathrm{tr}(\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*))}; \\ Q_3 &= \frac{\mathbf{y}^\top(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_z - \mathbf{L})\mathbf{y}}{\mathrm{tr}(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_z - \mathbf{L})}; \\ Q_4 &= \frac{\mathbf{y}^\top(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_{xz})\mathbf{y}}{\mathrm{tr}(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_{xz})}, \end{aligned} \tag{18}$$

where $\mathbf{H}_z = \mathcal{Z}(\mathcal{Z}^\top\mathcal{Z})^{-1}\mathcal{Z}^\top$, that is, $(\mathbf{H}_z + \mathbf{L})$ is the projection matrix under H_0(iii), and similarly $(\mathbf{H}_{xz} + \mathbf{L})$ is the projection matrix under H_0(iv). The matrices sandwiched in (17) and (18) are all symmetric and asymptotically idempotent, and the matrices sandwiched in (18) for the numerators are asymptotically orthogonal to the matrix in (17) for the denominator. Then the F-statistics for H_0(i)–(iv) have asymptotic F-distributions with DF indicated by the corresponding divisors in (18) and (17).

The proof of Theorem 1 is given in the Appendix. We remark that other analogous F-tests in the literature have not yet been shown to possess F-approximations (Hastie and Tibshirani 1990; Ruppert, Wand, and Carroll 2003). In Theorem 1, the DF employed by a PLM is $\mathrm{tr}(\mathbf{H}^*) + \mathrm{tr}(\mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*))$, which is analogous to Green and Silverman (1994), and the parametric DF is not equal to p. For the nonparametric DF, Huang and Chen (2008) showed that the asymptotic expression for $\mathrm{tr}(\mathbf{H}^*)$ based on local linear regression is $h^{-1}(\nu_0 + \nu_2/\mu_2)|\Omega|$ where ν_j is the jth moment of $K^2(\cdot)$ and $|\Omega|$ is the range of X. It is straightforward to show that the DF for the nonintercept parametric component is asymptotically p:

$$\begin{aligned} &\mathrm{tr}\big(P_{z|x}^\top(\mathbf{I}-\mathbf{H}^*)\big) \\ &\quad = \mathrm{tr}\big((\mathcal{Z}^\top(\mathbf{I}-\mathbf{H}^*)\mathcal{Z})^{-1}(\mathcal{Z}^\top(\mathbf{I}-\mathbf{H}^*)^2\mathcal{Z})\big) \\ &\quad = p - \mathrm{tr}\big((\mathcal{Z}^\top(\mathbf{I}-\mathbf{H}^*)\mathcal{Z})^{-1}\mathcal{Z}^\top\mathbf{H}^*(\mathbf{I}-\mathbf{H}^*)\mathcal{Z}\big). \end{aligned} \tag{19}$$

Since the elements in $\mathbf{H}^*$ are of order $O(n^{-1})$ and $O(n^{-1}h^{-1})$ for nondiagonal and diagonal elements, respectively (cf. Huang and Chen 2008), and since $\mathbf{H}^*$ and $(\mathbf{I}-\mathbf{H}^*)$ are asymptotically orthogonal, the last term in (19) tends to 0 as $n \to \infty$, $nh \to \infty$ and $h \to 0$. This result indicates that in a semiparametric setting, the DF for the nonintercept parametric term is no longer exactly equivalent to the number of parameters, and it needs to be adjusted by nonparametric modeling in finite-sample settings.

Using similar arguments, one may adapt the ANOVA F-tests for testing significance of a slope parameter, say α_1, in a PLM setting. Denote the reduced data matrix as $\mathcal{Z}'$ and the corresponding weighted projection matrix as $P_{z'|x}$. For $H_0 : \alpha_1 = 0$, it is easy to show that the numerator for the F-test statistic is

$$\frac{\mathbf{y}^\top(\mathbf{P}_{z|x} - \mathbf{P}_{z'|x})(\mathbf{I}-\mathbf{H}^*)\mathbf{y}}{\mathrm{tr}((\mathbf{P}_{z|x} - \mathbf{P}_{z'|x})(\mathbf{I}-\mathbf{H}^*))},$$

and $\mathrm{tr}((\mathbf{P}_{z|x} - \mathbf{P}_{z'|x})(\mathbf{I}-\mathbf{H}^*))$ is asymptotically equal to 1. Asymptotic idempotency of the matrix $(\mathbf{P}_{z|x} - \mathbf{P}_{z'|x})(\mathbf{I}-\mathbf{H}^*)$ can be checked easily and its asymptotic orthogonality to the matrix in (17) holds since $\mathbf{P}_{z'|x}^\top(\mathbf{I}-\mathbf{H}^*)(\mathbf{I}-\mathbf{P}_{z|x}) = \mathbf{0}$. Testing for significance of multivariate parameters works analogously.

The proposed F-tests also have the flexibility to be used for testing for joint significance of $m(X)$ and a slope parameter, say α_1, that is, $H_0 : \alpha_1 = 0$ and $m(X) = b_0$. Modifying the test statistic for H_0(iii) in (18), the numerator of this F-test is

$$\frac{\mathbf{y}^\top(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_{z'} - \mathbf{L})\mathbf{y}}{\mathrm{tr}(\mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I}-\mathbf{H}^*) - \mathbf{H}_{z'} - \mathbf{L})}, \tag{20}$$

where $\mathbf{H}_{z'}$ denotes the projection matrix for the reduced data matrix $\mathcal{Z}'$. It is easy to show that (20) divided by (17) is a valid F-statistic using arguments similar to those in the proof of Theorem 1. Therefore the proposed ANOVA F-tests are as general, flexible, simple, and easy as the classical F-tests; they can test significance of multiple terms in (1), either parametric, nonparametric, or a combination of both.

Another remark is that (17) may be used to estimate σ^2 for PLMs. In the univariate model (2), Huang and Chen (2008) showed that $SSE_1(h)$ in (5) is a $n^{1/2}$-consistent estimator for σ^2 if $nh^2 \to \infty$ and $nh^4 \to 0$, which implies that a smaller bandwidth is needed for estimating σ^2. We conjecture that a similar result holds for (17) for PLMs. Further work on estimating residual variance and testing for the constant-variance assumption is in progress.

3.3 Properties of New Estimators

The new estimators (10) and (15) arise naturally in developing the ANOVA framework. We investigate their properties in this section. The following theorem gives the rates of the bias and variance of $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{m}}$. The definitions of $\mathbf{V}$ and $\mathbf{g}(x)$ are given in the Appendix.

Theorem 2. Under Conditions (A), conditioned on $\mathcal{Z}$ and $\mathbf{x}$, as $n \to \infty$ and $h \to 0$,

(a) $E(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = h^4\mathbf{V}^{-1}\int \mathbf{g}(x)b(x)f(x)\,dx + o(h^4) + O(h^4/\sqrt{n})$.

If in addition $nh^2 \to \infty$, $\mathrm{var}(\hat{\boldsymbol{\alpha}}) = \sigma^2 n^{-1}\mathbf{V}^{-1} + o(n^{-1})$.

(b) If $nh^2 \to \infty$, then

$$\begin{aligned} \mathrm{bias}(\hat{m}(X_i)) &= \mathrm{bias}(\hat{m}_0(X_i)) + \mathbf{g}(X_i)^\top\,\mathrm{bias}(\hat{\boldsymbol{\alpha}})[1 + o(1)] \\ &= O(h^4), \\ \mathrm{var}(\hat{m}(X_i)) &= \mathrm{var}(\hat{m}_0(X_i))(1 + o(1)) \\ &= O(n^{-1}h^{-1}), \end{aligned} \tag{21}$$

where $(\hat{m}_0(X_1), \ldots, \hat{m}_0(X_n))^\top = \mathbf{H}^*(\mathbf{y} - \mathcal{Z}\boldsymbol{\alpha})$, the estimator of $\mathbf{m}$ if $\boldsymbol{\alpha}$ were known exactly.

A conclusion from Theorem 2 is that even if X and $\mathbf{z}$ are correlated, with the conventional optimal rate of bandwidth $n^{-1/5}$, then the bias of $\hat{\boldsymbol{\alpha}}$ is smaller than $n^{-1/2}$ and is asymptotically negligible compared with its variance. In contrast, estimators with $\mathbf{S}$ from a local linear estimator (Opsomer and Ruppert 1999) or Nadaraya–Watson (local constant) estimator (Speckman 1988) have a bias of rate h^2 for $\boldsymbol{\alpha}$, and the bias dominates the variance if the bandwidth is of order $n^{-1/5}$. The main advantage for employing $\mathbf{H}^*$ is that the projected response in the univariate case has a bias of order h^4 [see (7)] so that $\hat{\boldsymbol{\alpha}}$ does not suffer the bias h^2 problem and its asymptotic behavior is similar to the estimator based on partial residuals in Speckman (1988) and Hamilton and Truong (1997). While local cubic regression is also able to provide an estimator with a bias of order h^4, the proposed $\hat{m}(\cdot)$ remains in the local linear regression setting and achieves bias reduction by making use of the estimated slopes in fitting local lines [see (6) and (15)].

Note that $\hat{m}(X_i)$ offers an estimate for $m(X_i)$, but not for $m(t)$ at a general grid point t. To estimate $m(t)$, one simple solution is to apply local linear estimation using $(\mathbf{y} - \mathcal{Z}\hat{\alpha})$ as the response variable,

$$\min_{\gamma_0,\gamma_1} \sum_{i=1}^{n} (Y_i - \mathbf{z}_i\hat{\boldsymbol{\alpha}} - \gamma_0 - \gamma_1(X_i - t))^2 K_h(X_i - t). \tag{22}$$

The solution $\hat{\gamma}_0$ is an estimate of $m(t)$ and its asymptotic behavior will be similar to that of Opsomer and Ruppert (1999). Alternatively, one may extend (8) to form

$$\int (\hat{\gamma}_0(x) + \hat{\gamma}_1(x)(t - x)) K_h(t - x)\, dx,$$

which we expect to have the same asymptotic rate as $\hat{m}(X_i)$ in
ang2009?> Theorem 2(b), as He and Huang (?HeHuang2009) have shown for model (2).

To better understand the difference between estimators $\hat{\gamma}_0(X_i)$ in (22) and $\hat{m}(X_i)$ in (15) for PLMs, and between $\hat{\beta}_0(x)$ in (3) and Y_i^* in (6) for model (2), we derive the "equivalent kernel" of Y_i^* in the following theorem:

Theorem 3. For $i = 1, \ldots, n$, write Y_i^* as a weighted estimator, $Y_i^* = n^{-1} \sum_{k=1}^{n} W_h(X_k - X_i) Y_k$ with $W_h(\cdot) = W(\cdot/h)/h$. Under conditions (A3)–(A6), the "equivalent kernel" $W(\cdot)$ for Y_i^* is asymptotically

$$\frac{1}{f(X_i)} (K_0^*(\cdot) - \mu_2^{-1} K_1^*(\cdot)). \tag{23}$$

In other words, Y_i^* asymptotically uses a combination of two kernel functions: $K_0^*(\cdot)$ is an ordinary kernel function of order $(0, 2)$ (see Gasser, Muller, and Mammitzsch 1985 for definition); $\mu_2^{-1} K_1^*(\cdot)$ is of order $(2, 4)$ with second moment $2\mu_2$.

A sketch of the proof for Theorem 3 is given in the Appendix. Theorem 3 shows the effect of incorporating the estimated slopes in $\mathbf{H}^*$, which introduces another weight function of order $(2, 4)$. It is easy to see that the 3rd moments of both $K_0^*(\cdot)$ and $K_1^*(\cdot)$ are 0. Hence the combination of two kernel functions reduces the bias to the order of h^4 for Y_i^*.

The new estimators (10) and (15) also have an interesting property as penalized likelihood estimators. With a symmetric $\mathbf{H}^*$, they are the solutions to

$$\min_{\boldsymbol{\alpha},\mathbf{f}} (\mathbf{y} - \mathcal{Z}\boldsymbol{\alpha} - \mathbf{f})^\top (\mathbf{y} - \mathcal{Z}\boldsymbol{\alpha} - \mathbf{f}) + \mathbf{f}^\top (\mathbf{H}^{*^{-1}} - \mathbf{I})\mathbf{f}, \tag{24}$$

where $\mathbf{H}^{*^{-1}}$ denotes the inverse of $\mathbf{H}^*$ [since $\mathbf{H}^*$ is symmetric and positive definite (Huang and Chen 2008), its inverse exists] and $\mathbf{f}$ denotes some estimate of $(m(X_1), \ldots, m(X_n))^\top$ under the constraint that $\mathbf{f}$ lies in the space spanned by $\mathbf{H}^*$. Indeed, for model (2), $\mathbf{y}^*$ in (6) is also a penalized likelihood estimator in that it is the solution to the following penalized least squares problem:

$$\min_{\mathbf{f}} (\mathbf{y} - \mathbf{f})^\top (\mathbf{y} - \mathbf{f}) + \mathbf{f}^\top (\mathbf{H}^{*^{-1}} - \mathbf{I})\mathbf{f}. \tag{25}$$

The penalized least squares properties (24) and (25) are based on some arguments in Hastie and Tibshirani (1990). Further investigation on the penalty form $\mathbf{f}^\top (\mathbf{H}^{*^{-1}} - \mathbf{I})\mathbf{f}$ is in progress.

4. SIMULATION STUDY

In this section, we use simulated data to examine the performance of the new estimators (10) and (15) and the ANOVA F-tests in Theorem 1. The new estimators are compared with the estimators based on partial residuals of local linear regression (Hamilton and Truong 1997):

$$\begin{aligned} \tilde{\boldsymbol{\alpha}} &= (\mathcal{Z}^\top (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{S})\mathcal{Z})^{-1} \mathcal{Z}^\top (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{S})\mathbf{y}, \\ \tilde{\mathbf{m}} &= \mathbf{S}(\mathbf{y} - \mathcal{Z}\tilde{\boldsymbol{\alpha}}). \end{aligned} \tag{26}$$

The proposed F-tests for H_0(ii)–(iv) are compared with the PLR tests of Fan and Huang (2005), which did not discuss PLR tests for H_0(i). The number of simulations is 500 and the Epanechnikov kernel is used throughout the simulation study. For choosing the smoothing parameter, we try using both a fixed bandwidth at three values: $h = 0.4167$ (0.625/1.5), 0.625, and 0.9375 (0.625 × 1.5) which were used in Fan and Huang (2005), and the AICc (Hurvich, Simonoff, and Tsai 1998) criterion with the nonparametric $SSE_1(h)$ (5) and the nonparametric DF $\mathrm{tr}(\mathbf{H}^*)$:

$$AICc(h) = \log\left(\frac{SSE_1(h)}{n - \mathrm{tr}(\mathbf{H}^*)}\right) + 1 + \frac{2\,\mathrm{tr}(\mathbf{H}^*) + 1}{n - \mathrm{tr}(\mathbf{H}^*) - 2}.$$

We anticipate that the parametric SSE will have little influence on bandwidth selection and hence the AICc is based only on the nonparametric SSE. The AICc criterion is calculated on a grid of bandwidths that consist of 50 logarithmically equally-spaced points in the interval [0.3, 1.5], and the bandwidth that minimizes $AICc(h)$ is chosen. The projection matrix $\mathbf{H}^*$ is normalized so that $\mathbf{H}^*\mathbf{1} = \mathbf{1}$ to overcome boundary effects. For the F-tests, we use critical values from the F-distributions with the DF as given in Theorem 1. The critical values for the PLR tests are taken from the chi-squared distributions based on theorem 5.1 in Fan and Huang (2005). The significance level is 5%.

Example 1. Take

$$Y = 2 + 2\theta_1 X + 4\theta_1 \exp(-16X^2) + 4\theta_2 Z_1 + 2\theta_2 Z_2 + \theta_2 Z_3 + \varepsilon,$$

where $\varepsilon \sim \mathrm{N}(0, 1)$, Z_3 is binary, 0 or 1, with $P(Z_3 = 1) = 0.4$, and X, Z_1, and Z_2 are initially simulated as jointly normal with mean 0, variance 1, and pairwise correlation 0.6. Then the dataset of size $n = 100$ is truncated so that $|X| \le 1.645$, which eliminates about 10% of data to prevent sparsity at boundaries. The pairwise correlation 0.6 is selected so that the results reflect relatively difficult cases for statistical estimation. For convenience we denote the parameters for Z_j as α_j, $j = 1, 2, 3$, that is, $\alpha_1 = 4\theta_2$, $\alpha_2 = 2\theta_2$, and $\alpha_3 = \theta_2$.

We compare the performance of estimators when $\theta_1 = \theta_2 = 0.5$. The average values of $\hat{\boldsymbol{\alpha}}$ are compared with $\tilde{\boldsymbol{\alpha}}$, together with their sample standard deviations (SD), and their sample standard errors are also calculated as the square-root of

$$\begin{aligned} \mathrm{var}(\hat{\boldsymbol{\alpha}}) &= \sigma^2 (\mathcal{Z}^\top (\mathbf{I} - \mathbf{H}^*)\mathcal{Z})^{-1} \mathcal{Z}^\top (\mathbf{I} - \mathbf{H}^*)^2 \\ &\quad \times \mathcal{Z}(\mathcal{Z}^\top (\mathbf{I} - \mathbf{H}^*)\mathcal{Z})^{-1}, \\ \mathrm{var}(\tilde{\boldsymbol{\alpha}}) &= \sigma^2 (\mathcal{Z}^\top (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{S})\mathcal{Z})^{-1} \mathcal{Z}^\top [(\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{S})]^2 \\ &\quad \times \mathcal{Z}(\mathcal{Z}^\top (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{S})\mathcal{Z})^{-1}, \end{aligned}$$

Table 1. Average squared bias, variance, and mean squared errors (MSE) of nonparametric estimates for Example 1 when $\theta_1 = \theta_2 = 0.5$

	Bias^2		Var		MSE	
	$\hat{m}(X_i)$ (SD)	$\tilde{m}(X_i)$ (SD)	$\hat{m}(X_i)$ (SD)	$\tilde{m}(X_i)$ (SD)	$\hat{m}(X_i)$ (SD)	$\tilde{m}(X_i)$ (SD)
$h = 0.4167$	0.1926 (0.1565)	0.2096 (0.1549)	0.0645 (0.0044)	0.0497 (0.0036)	0.2571 (0.1567)	0.2593 (0.1552)
$h = 0.625$	0.2475 (0.1572)	0.2850 (0.1499)	0.0449 (0.0031)	0.0347 (0.0024)	0.2923 (0.1574)	0.3197 (0.1501)
$h = 0.9375$	0.3500 (0.1630)	0.3819 (0.1400)	0.0323 (0.0023)	0.0258 (0.0018)	0.3823 (0.1631)	0.4076 (0.1402)
AICc	0.2865 (0.1968)	0.2938 (0.1754)	0.0446 (0.0162)	0.0352 (0.0115)	0.3311 (0.1881)	0.3290 (0.1690)

where σ^2 is assumed to be known for ease of comparison of their relative variation. It turns out that the performance of the two parametric estimates is quite close and hence the details are omitted. When $\theta_1 = \theta_2 = 0.5$, Table 1 compares the average squared bias, variance, and MSE of the estimated smooth functions at those interior points with $|X_i| \le 1$, where the variances of $\hat{\mathbf{m}}$ and $\tilde{\mathbf{m}}$ are calculated empirically based on (15) and (26) [e.g., $\mathrm{Var}(\hat{\mathbf{m}}) = \sigma^2 \operatorname{tr}(\mathbf{A}\mathbf{A}^T)$ with $\mathbf{A} = \mathbf{H}^*(\mathbf{I} - \mathbf{P}_{z|x})$]. As expected, the proposed estimator $\hat{m}(\cdot)$ generally has a smaller average squared bias than $\tilde{m}(\cdot)$ at the price of a larger variance. The average MSE of $\hat{m}(\cdot)$ is comparable with $\tilde{m}(\cdot)$ in the cases of $h = 0.4167$ and when using the AICc bandwidth, and is smaller when $h = 0.625$ and 0.9375. When using the AICc-bandwidth, there is more variation as the SDs in the row of AICc are generally larger than those with a fixed bandwidth.

When $\theta_1 = \theta_2 = 0.5$, the AICc-selected bandwidth has a mean of 0.8273 and the first-quantile and third-quartile are 0.4751 and 1.5. On the DF, the values of the mean (SD) of $\operatorname{tr}(\mathbf{H}^*)$ are 8.70 (0.26), 6.06 (0.17), and 4.29 (0.11) respective to the three fixed values of the bandwidth, and 5.94 (2.27) for AICc. In comparison, the values of $\operatorname{tr}(\mathbf{S})$ are smaller, 7.40 (0.24), 5.45 (0.17), and 4.14 (0.12) respectively, and for AICc 5.34 (1.69). The values of the mean (SD) parametric DF $\operatorname{tr}(\mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*))$ are 2.95 (0.016), 2.97 (0.014), and 2.98 (0.011) respectively for fixed bandwidths and 2.97 (0.024) for AICc, which are close to the asymptotic value 3.

For testing hypotheses, $\theta_1 = 0, 0.25, 0.5$, and $\theta_2 = 0, 0.25$. For the overall model significance H_0(i), the empirical Type I error of the ANOVA F-test when $\theta_1 = \theta_2 = 0$ are close to 5%, 0.038, 0.036, and 0.042 for $h = 0.4167$, 0.625, and 0.9375, respectively, and 0.054 for AICc. When $(\theta_1, \theta_2) = (0.25, 0)$, the proportion of rejection is 0.86, 0.874, 0.822, and 0.876, respectively. The rejection rate is 100% for the remaining combinations of (θ_1, θ_2). Therefore, the proposed F-test for H_0(i) performs reasonably well both under the null and alternative hypotheses with departures either in the parametric or nonparametric terms.

For testing H_0(ii), H_0(iii), and H_0(iv), the proportions of rejection of the F-tests and PLR tests are given in Table 2. For H_0(ii), both tests are sensitive in detecting the parametric effect when $\theta_2 = 0.25$ and the PLR test has a level slightly larger than 5%. For testing whether the nonparametric effect is significant, H_0(iii), when using a fixed bandwidth, the PLR test has a level larger than 5%, which possibly leads to a higher rejection rate than the F-test when $\theta_1 \ne 0$. When using the AICc bandwidth in testing H_0(iii), both tests reject more than 5% of the time when $\theta_1 = 0$. When testing H_0(iv), whether the nonparametric function is linear, the two tests have similar performance in most cases except that when $\theta_1 = 0.25$, the PLR test is more powerful than the F-test with a fixed bandwidth. But the F-test is more powerful based on the AICc bandwidth. When $m(\cdot)$ is nonlinear ($\theta_1 \ne 0$), using a smaller bandwidth for testing H_0(iv) is more powerful for both the F- and PLR tests. The F-tests with AICc have greater power than with a fixed bandwidth in testing H_0(iii) and H_0(iv); this phenomenon is not observed for the PLR test.

Example 2. Take

$$Y = 2 + 2\theta_1 \sin(2X) + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + \varepsilon,$$

where X, Z_1, Z_2, and Z_3 have identical distributions as in Example 1 and α_1, α_2, and α_3 are the same as Example 1. We modify Example 1 a little bit to compare the estimators when the nonparametric curve $2\theta_1 \sin(2X)$ is easier to estimate than that

Table 2. Proportions of rejection for Example 1. For combinations of (θ_1, θ_2) not shown for H_0(ii), the proportions of rejections are all 100%

	F	PLR	F	PLR	F	PLR
			Test H_0(ii)			
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0.5, 0)	
$h = 0.4167$	0.038	0.068	0.038	0.062	0.050	0.072
$h = 0.625$	0.042	0.064	0.046	0.056	0.040	0.058
$h = 0.9375$	0.052	0.060	0.038	0.052	0.036	0.060
AICc	0.048	0.066	0.042	0.052	0.050	0.072
			Test H_0(iii)			
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0.5, 0)	
$h = 0.4167$	0.032	0.090	0.80	0.906	1	1
$h = 0.625$	0.036	0.084	0.820	0.924	1	1
$h = 0.9375$	0.030	0.078	0.798	0.910	1	1
AICc	0.078	0.114	0.876	0.898	1	1
(θ_1, θ_2)	(0, 0.25)		(0.25, 0.25)		(0.5, 0.25)	
$h = 0.4167$	0.038	0.106	0.776	0.882	1	1
$h = 0.625$	0.046	0.096	0.802	0.870	1	1
$h = 0.9375$	0.062	0.084	0.770	0.874	1	1
AICc	0.104	0.098	0.872	0.894	1	1
			Test H_0(iv)			
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0.5, 0)	
$h = 0.4167$	0.034	0.058	0.416	0.488	0.972	0.978
$h = 0.625$	0.036	0.044	0.414	0.450	0.966	0.958
$h = 0.9375$	0.048	0.030	0.276	0.334	0.880	0.874
AICc	0.088	0.060	0.496	0.292	0.958	0.882
(θ_1, θ_2)	(0, 0.25)		(0.25, 0.25)		(0.5, 0.25)	
$h = 0.4167$	0.032	0.050	0.430	0.530	0.970	0.976
$h = 0.625$	0.046	0.058	0.436	0.474	0.960	0.950
$h = 0.9375$	0.058	0.044	0.324	0.358	0.852	0.894
AICc	0.152	0.040	0.546	0.304	0.970	0.902

Table 3. Proportions of rejection for Example 3, H_0(i). For combinations of (θ_1, θ_2) not shown here, the proportions of rejections are all 100%

	$a = 0.2$		$a = 0.5$		$a = 1.0$	
(θ_1, θ_2)	(0, 0)	(0.25, 0)	(0, 0)	(0.25, 0)	(0, 0)	(0.25, 0)
$h = 0.4167$	0.074	0.938	0.128	0.966	0.268	0.994
$h = 0.625$	0.066	0.944	0.116	0.960	0.254	0.986
$h = 0.9375$	0.066	0.900	0.104	0.926	0.204	0.964
AICc	0.088	0.948	0.180	0.968	0.278	0.976

of Example 1. For $\theta_1 = \theta_2 = 0.5$, the two parametric estimators' performance is similar to that of Example 1 and hence, for brevity, is omitted here. This seems to imply that nonparametric estimation may not have much effect on the parametric estimate when the model is correctly specified. Given that $2\theta_1 \sin(2X)$ is easier to estimate, we do not see a clear MSE improvement of $\hat{\mathbf{m}}$ when $h = 0.4167$, 0.625, and AICc. With $h = 0.9375$, $\tilde{\mathbf{m}}$ in (26) has a larger bias, a smaller variance, and a larger MSE than $\hat{\mathbf{m}}$. The degrees of freedom and comparison between the F- and PLR tests in this example are similar to those of Example 1 and are omitted.

Example 3. Take

$$Y = 2 + 2\theta_1 X + 4\theta_1 \exp(-16X^2) + \alpha_1 Z_1 + \alpha_2 Z_2 + \alpha_3 Z_3 + |X|^a \varepsilon.$$

This example is identical to Example 1 except that the error variance depends on X with a =0.2, 0.5, 1.0. We want to examine the robustness issue with respect to departure from the constant error variance assumption. To our surprise, the behavior of $\hat{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\alpha}}$, and $\hat{m}(\cdot)$ and $\tilde{m}(\cdot)$ is similar to Example 1 and is omitted. This informally demonstrates moderate robustness of all PLM estimators.

The proportions of rejection under H_0(i) of the F-tests are given in Table 3, indicating that the F-tests are affected when the constant variance assumption is violated. As the degree of heteroscedasticity increases, the rejection rate when $\theta_1 = \theta_2 = 0$ deviates further from 0.05. Note that H_0(i) involves testing significance of both the parametric and nonparametric terms. The results for testing overall parametric effects H_0(ii) are shown in Table 4. The case when $a = 0.5$ is similar to $a = 0.2$ and 1.0 and hence is omitted. The proportions of rejection are all 100% when $\theta_2 = 0.25$. We observe that both the F- and PLR tests have excellent robustness properties against heteroscedasticity for testing parametric effects.

Table 5 shows the results for testing overall significance of the nonparametric term H_0(iii). The rejection rates when $\theta_1 = 0.5$ are all 100% and hence, omitted. We see that when $\theta_1 = 0$, the proportion of rejections increases as the degree of heteroscasdasticity a increases. The deviation from 0.05 is larger for the PLR tests than the F-tests. The power of both tests when $\theta_1 \neq 0$ is generally comparable. This implies that the F-tests have better performance than the PLR tests in this example for H_0(iii) when data are heteroscedastic: a smaller false rejection rate and comparable power. We also note that the overall results of $\theta_2 = 0.25$ are similar to those of $\theta_2 = 0$, which may imply that testing nonparametric H_0(iii) is not affected by the parametric part in a PLM. In most cases, the F-tests are more powerful with AICc than with a fixed bandwidth.

The rejection rates for testing whether $m(x)$ is linear H_0(iv) are given in Table 6. The PLR tests behave better under the null as compared to the case of testing H_0(iii). When $\theta_1 = 0$ and with a fixed bandwidth, the deviation from 0.05 is generally larger for the PLR tests than the F-tests; with the AICc bandwidth, the F-tests have a larger Type I error. Some observations are similar to the case of testing H_0(iii): with a fixed bandwidth, the F-tests have better performance under H_0 and comparable power under H_a than the PLR tests. With AICc, the F-tests have greater power than the PLR tests but with a price of a higher false rejection rate under H_0.

Based on the limited scales of this simulation study, we find that the ANOVA F-tests perform reasonably well. It retains the advantages of parametric F-tests when testing significance of parametric terms, in terms of robustness against nonparametric heteroscedasticity. On testing the nonparametric term, when errors are homoscedastic the F-tests generally have a Type I error close to 0.05 and reasonable power performance under alternative hypotheses. Both the F- and PLR tests are not robust for testing nonparametric terms when heteroscasdasticity is a function of the nonparametric variable. This may suggest that one needs to estimate and correct for the variance function for heteroscasdasticity before applying testing procedures based on homoscasdastic errors. Further investigation on estimating and testing the variance function will be an interesting problem for future research.

5. DATA ANALYSIS

In this section, we apply the methodology in Section 3 to the SCDS data and compare the results with the analysis using SAMs reported by Huang et al. (2005) and the multiple linear regression analysis reported by Myers et al. (2003). We reanalyze the relationships between prenatal methylmercury exposure and five developmental measures determined at nine years of age. The five endpoints are WISC-III Full-Scale IQ (Full IQ) (Wechsler 1992), California Verbal Learning Test—Short Delay (CVLTSD) (Delis et al. 1994), the Grooved Pegboard (GP) (Knights and Moule 1968) using the dominant hand,

Table 4. Proportions of rejection for Example 3, H_0(ii). The proportions of rejection when $\theta_2 = 0.25$ are all 100%

	$a = 0.2$						$a = 1.0$					
	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0.5, 0)		(0, 0)		(0.25, 0)		(0.5, 0)	
$h = 0.4167$	0.044	0.072	0.038	0.060	0.052	0.082	0.048	0.066	0.046	0.078	0.032	0.042
$h = 0.625$	0.052	0.072	0.040	0.060	0.044	0.066	0.050	0.064	0.046	0.072	0.024	0.042
$h = 0.9375$	0.050	0.068	0.038	0.060	0.044	0.052	0.048	0.062	0.044	0.060	0.040	0.054
AICc	0.054	0.062	0.056	0.062	0.046	0.060	0.050	0.060	0.036	0.056	0.054	0.076

JASA jasa v.2009/02/18 Prn:2010/04/02; 14:57 F:jasaap08274.tex; (Milda) p. 9

Table 5. Proportions of rejection for Example 3, H_0(iii)

	$a=0.2$				$a=0.5$				$a=1.0$			
	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0, 0)		(0.25, 0)		(0, 0)		(0.25, 0)	
$h=0.4167$	0.070	0.18	0.920	0.970	0.132	0.290	0.956	0.988	0.306	0.514	0.986	0.996
$h=0.625$	0.072	0.168	0.924	0.968	0.128	0.260	0.946	0.986	0.290	0.470	0.970	0.994
$h=0.9375$	0.078	0.172	0.878	0.952	0.124	0.246	0.888	0.970	0.250	0.428	0.934	0.978
AICc	0.128	0.196	0.942	0.958	0.224	0.302	0.986	0.974	0.334	0.412	0.972	0.980
(θ_1, θ_2)	(0, 0.25)		(0.25, 0.25)		(0, 0.25)		(0.25, 0.25)		(0, 0.25)		(0.25, 0.25)	
$h=0.4167$	0.090	0.194	0.912	0.960	0.184	0.330	0.972	0.990	0.304	0.478	0.980	0.996
$h=0.625$	0.096	0.188	0.914	0.960	0.166	0.298	0.962	0.988	0.290	0.480	0.962	0.992
$h=0.9375$	0.102	0.170	0.878	0.952	0.162	0.266	0.906	0.972	0.280	0.442	0.920	0.980
AICc	0.164	0.154	0.918	0.942	0.278	0.288	0.974	0.972	0.384	0.400	0.956	0.978

Bruininks–Osetetsky Test of Motor Proficiency (BO) (Bruininks 1978), and the Connors Teacher Rating Scale (CTRS) (Connors 1985). These tests measure global intelligence (Full IQ), verbal memory (CVLTSD), motor function (GP and BO), and the child's activity level (CTRS). For GP and CTRS, a higher score indicates poorer performance, while, for the remaining tests, an increase in the score is associated with improved performance. The CTRS was log-transformed and GP was transformed as the negative reciprocal in the linear analysis (Myers et al. 2003) to better satisfy the normal error assumption and we use the same transformed version here. The GP using the nondominant hand is not included in the present analysis since Myers et al. (2003) adopted a model with gender by prenatal exposure interactions for this endpoint, which is different from PLMs and will be investigated in future studies.

In this reanalysis, the PLM includes the nonparametric term for the prenatal methylmercury exposure (range 0.54 to 23.0 ppm) and linear trends for 13 covariates that were chosen a priori (see Myers et al. 2003 for details). The bandwidth is first fixed as 7.28 with the Epanechnikov kernel so that $\mathrm{tr}(\mathbf{H}^*)$ is 4 (including the intercept) in the case of Full IQ, which is comparable with 3 (without the intercept) in Huang et al. (2005). Then we also try AICc as the bandwidth selector with 50 bandwidths equally spaced in the logarithm scale ranging from 3.5 to 10.5.

5.1 IQ Significance

Let us first focus on the Full IQ results, since there is some attention to IQ effects raised by Axelrad et al. (2007). In addition to a fixed bandwidth $h=7.28$, the AICc selection results in a bandwidth of 8.21 with DF $\mathrm{tr}(\mathbf{H}^*)=3.64$. The methodology in Section 3 offers an ANOVA table (Table 7 for $h=7.28$), which explicitly summarizes the quality of fitting by various SS quantities and DF. When $h=8.21$, the SSR quantities are: 2028.27 for nonparametric SSR and 12,858.94 for parametric SSR, which are close to those of $h=7.28$ in Table 7. The R-squareds are approximately the same, 0.24 for both bandwidths, which is better than 0.20 obtained by linear analysis (Myers et al. 2003). The estimate of the residual standard error by the square root of (17) is 9.76 for either of the bandwidths while it was 9.95 in the linear analysis. The parametric estimates as well as their standard errors with either bandwidth are close to those in Myers et al. (2003) and, hence, are omitted. The estimated curves for prenatal methylmercury exposure are given in Figure 1(a) with pointwise ± 1.96 standard errors and the standardized residual plots are shown in Figures 2(a) and (b). The standard errors of $\hat{\mathbf{m}}$ are calculated empirically based on (15) with the estimated residual standard error. There are two outliers (standardized residual with an absolute value > 3) with prenatal methylmercury 4.97 and 6.41 ppm and IQ 139 and 118, respectively, which are also outliers in the linear analysis.

Table 6. Proportions of rejection for Example 3, H_0(iv). The proportions of rejection are either 100% or close to 100% when $\theta_1 = 0.5$ and hence are omitted

	$a=0.2$				$a=0.5$				$a=1.0$			
	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR	F	PLR
(θ_1, θ_2)	(0, 0)		(0.25, 0)		(0, 0)		(0.25, 0)		(0, 0)		(0.25, 0)	
$h=0.4167$	0.062	0.108	0.612	0.686	0.128	0.212	0.748	0.838	0.286	0.422	0.856	0.918
$h=0.625$	0.064	0.090	0.560	0.636	0.104	0.162	0.714	0.774	0.252	0.352	0.786	0.860
$h=0.9375$	0.060	0.072	0.396	0.490	0.108	0.130	0.474	0.580	0.192	0.290	0.530	0.722
AICc	0.160	0.084	0.638	0.456	0.246	0.148	0.748	0.566	0.312	0.240	0.792	0.650
(θ_1, θ_2)	(0, 0.25)		(0.25, 0.25)		(0, 0.25)		(0.25, 0.25)		(0, 0.25)		(0.25, 0.25)	
$h=0.4167$	0.068	0.118	0.622	0.690	0.160	0.240	0.770	0.852	0.264	0.380	0.850	0.922
$h=0.625$	0.082	0.090	0.578	0.608	0.138	0.196	0.728	0.768	0.240	0.316	0.784	0.870
$h=0.9375$	0.092	0.084	0.400	0.450	0.134	0.144	0.494	0.598	0.234	0.276	0.566	0.736
AICc	0.174	0.068	0.668	0.422	0.280	0.150	0.764	0.588	0.388	0.264	0.810	0.698

Table 7. Analysis of Variance table for Full IQ with $h = 7.28$: the numbers in the SS column are corrected sums of squares, corrected by the sample mean

	DF	SS	MS	$F = \frac{MSR}{MSE}$	p-value
SSR nonparametric	3.00	2069.61			
SSR parametric	19.97	12,847.22			
SSR	22.98	14,916.83	649.14	6.82	0
SSE	484.02	46,126.90	95.30		
SST	507	61,043.72			

The residual plots show moderate compatibility with the normal error assumption with homoscedastic variance.

Since the results on testing H_0(i)–(iv) for the Full IQ are similar between $h = 7.28$ and 8.21, we describe mainly the results with $h = 7.28$. The F-tests for H_0(i) on model significance and H_0(ii) on overall parametric effects are all highly significant with p-values < 0.0001. The nonparametric SSR for prenatal methylmercury exposure accounts for about 13.9% of SSR (Table 7) and is significant based on the ANOVA F-test (see Table 8). Applying the F-test for H_0(iv) to examine whether a nonlinear trend is significantly better than a linear trend, the nonlinear trend is significant at a 5% level for Full IQ. Table 8 also includes a comparison to results from the approximate F-tests in Huang et al. (2005), which are not significant in the case of Full IQ. Since the proposed semiparametric F-tests have asymptotic F-distributions, they are more appealing than the approximate F-tests. To check whether our F-test results for Full IQ are sensitive to the choice of bandwidth, we try different values of the bandwidth ranging from 5 to 20 and the rejection of H_0(iii) and H_0(iv) continues to hold. We also examine the significance results from the PLR tests. With $h = 7.28$, the PLR p-value is 0.56 for H_0(iv) and is 0.24 for H_0(iii), which result in different conclusions from the proposed F-tests. In examining a step further, it may be due to difference in defining SSR or SSE. The SSE expression in Fan and Huang (2005) is $\sum_i (Y_i - \tilde{m}(X_i))^2$ and equals 47,924.69 when $h = 7.28$, which is substantially different from 46,126.90 in Table 7. If we plug in 46,126.90 for the SSE under H_a, the PLR tests for H_0(iii) and H_0(iv) will become significant. We recall that the proposed SSE quantity has a valid ANOVA decomposition, while SSE defined by $\sum_i (Y_i - \tilde{m}(X_i))^2$ does not have an ANOVA decomposition in the case of local linear regression.

The significant F-test results in this IQ reanalysis further emphasize the nonlinear behavior of methylmercury effects. Let us examine the way the nonlinear curve behaves [Figure 1(a)]. From the estimated values $\hat{\mathbf{m}}$ with $h = 7.28$, the curve is nearly flat from 1.02 to 5 ppm, decreasing by 0.85 (95% CI $[-2.26, 0.56]$) IQ points from 5 to 10.02 ppm, de-

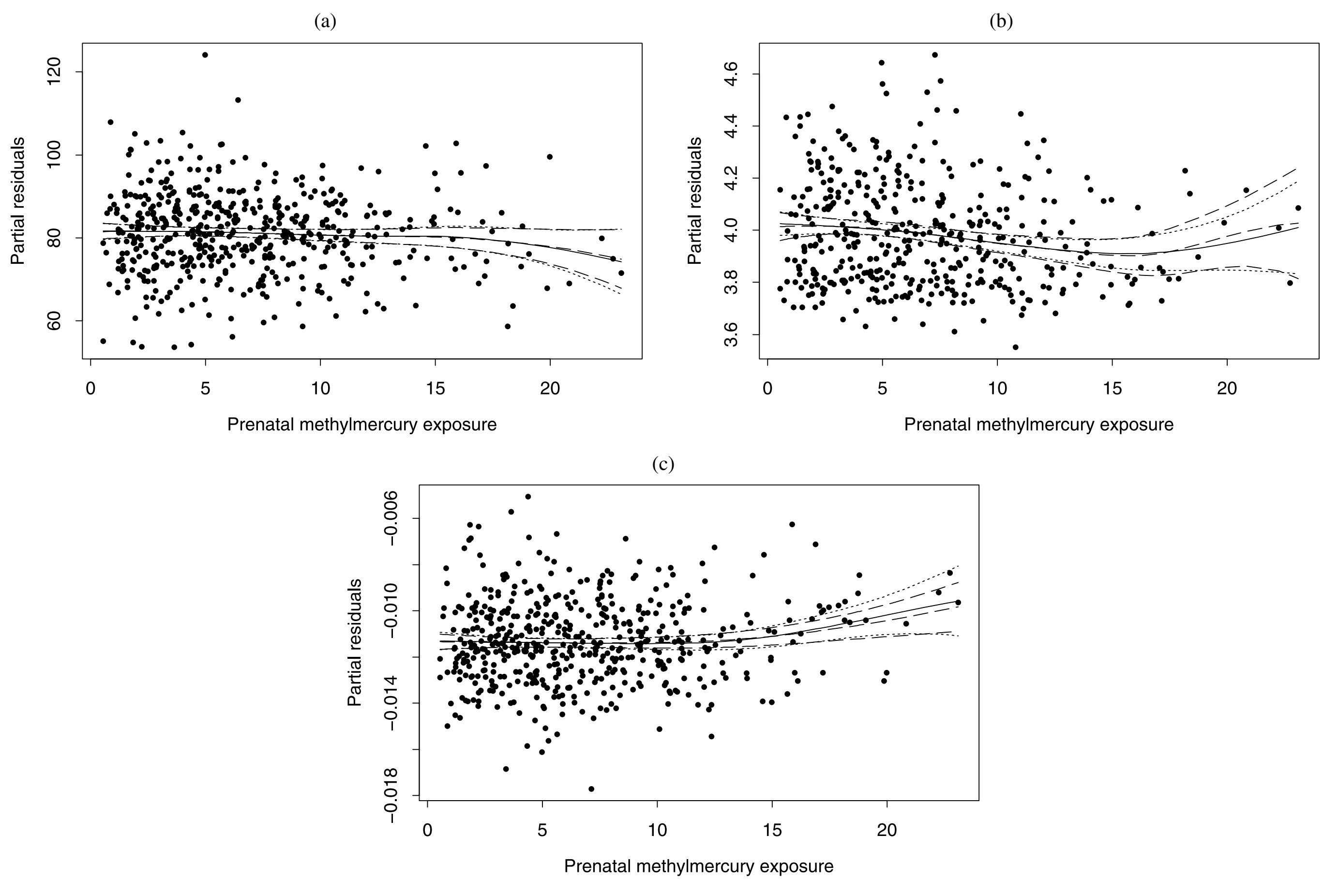

Figure 1. Partial residuals plots from fitting the PLMs for (a) Full IQ, (b) CTRS, and (c) GP. The solid line shows the nonparametric fit for the effects of prenatal methylmercury exposure with $h = 7.28$ and the dotted lines denote its pointwise 95% confidence intervals. The dashed lines show the results when using the AICc bandwidth, 8.21 (Full IQ), 5.24 (CTRS), and 10.50 (GP).

JASA jasa v.2009/02/18 Prn:2010/04/02; 14:57 F:jasaap08274.tex; (Milda) p. 11

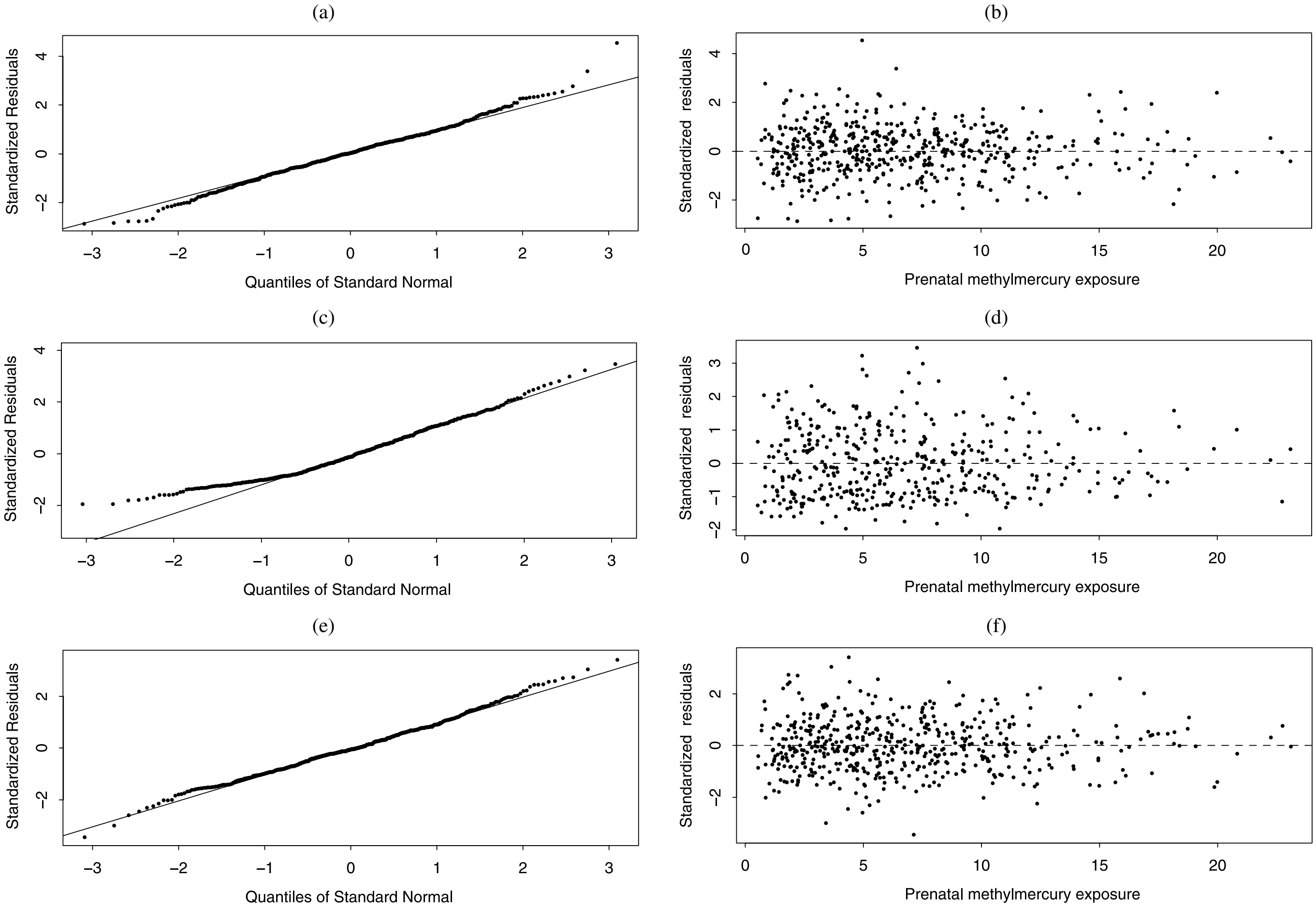

Figure 2. Standardized residual plots when $h = 7.28$: (a) qq-plot and (b) standardized residuals against prenatal methylmercury exposure for Full IQ; (c) and (d) are the same plots for CTRS; and (e) and (f) are the same plots for GP.

creasing by 0.28 (95% CI $[-2.46, 1.89]$) points from 10.02 ppm to 15.09 ppm, and decreasing by 2.70 (95% CI $[-6.51, 1.11]$) points from 15.09 to 19.98 ppm. However, all of these pointwise comparisons are not significant. If the comparison is made between 1.02 ppm and 19.98 ppm, the IQ decreases by 3.88 points (95% CI $[-8.58, 0.83]$, not significant). Though the point estimate -3.88 is surprisingly close to the estimate by Axelrad et al. (2007) of -3.42 IQ points from 1 ppm to 20 ppm (-0.18 IQ points for every 1 ppm, 95% CI $[-0.378, -0.009]$), their significance results and interpretations differ. Our reanalysis focuses on the overall significant nonlinear effects with unsignificant point estimates, while Axelrad et al. (2007) examined a linear dose-response relationship with a significant slope. When both methylmercury and nutrients coexist in fish, we argue that a nonlinear effect is more realistic and that a linear dose-response assessment may result in a misleading picture at the low levels of exposure. In addition, Axelrad et al. (2007) employed a questionable approach of calculating an estimated “IQ” score from only two subtests of the WISC-III Full-Scale IQ. The Seychelles study (Myers et al. 2003) administered the entire WISC-III and used all its subtest scores to compute the Full IQ based upon the test manual instructions (Wechsler 1992).

Table 8. Comparison of the significance tests for prenatal methylmercury exposure in SCDS nine-year data based on the PLMs with ANOVA F-tests ($h = 7.28$), SAMs with approximate F-tests (Hastie and Tibshirani 1990) taken from Huang et al. (2005), and linear model (LM) analysis from Myers et al. (2003)

	PLM ANOVA F-test p-value		SAM approximate F-test p-value		LM p-value
Outcome	Nonlinearity H_0(iv)	Overall H_0(iii)	Nonlinearity H_0(iv)	Overall H_0(iii)	Overall H_0(iii)
Full IQ	<0.0001	<0.0001	0.284	0.181	0.196
CVLTSD	0.837	0.618	0.811	0.643	0.194
BO	0.230	0.100	0.364	0.271	0.096
GP (dominant hand)	0.180	0.138	0.050	0.044	0.084
CTRS	0.033	0.003	0.246	0.021	0.004

We remark that the change in IQ with increasing prenatal methylmercury exposure levels in SCDS is very small, even when the exposure is above 10 ppm. A change in IQ of five points—one-third of a SD in the normative sample—is well within the normal random variation associated with IQ tests such as the WISC-III and such a change is usually clinically interpreted as within normal limits.

5.2 Other Outcomes

For the remaining four outcomes, we focus the discussion on testing nonparametric hypotheses H_0(iii) and H_0(iv). As in the case of Full IQ, we fit both $h = 7.28$ and the AICc bandwidth. Though the p-values change a little, the conclusion mostly agrees between $h = 7.28$ and the AICc bandwidth. The ANOVA F-test results with $h = 7.28$ are given in Table 8, with a comparison to Huang et al. (2005) and Myers et al. (2003).

For the CTRS outcome, the AICc bandwidth is 5.24 (DF 5.27). Both F-test p-values for testing H_0(iii) and H_0(iv) are significant. The p-values are smaller when the AICc bandwidth is used, 0.0026 and 0.0002 for H_0(iii) and H_0(iv), respectively. In comparison, the approximate F-test results in Huang et al. (2005) did not show significant nonlinear effects, while the overall significance remains consistent. Similar results are seen for the PLR tests: p-value for nonlinearity H_0(iv) is 0.376 and 0.408 when $h = 7.28$ and 5.24, respectively, and highly significant for testing H_0(iii) with the two values of the bandwidth. In the case of CTRS, the difference in significance between the ANOVA F- and PLR tests is similar to that for the Full IQ. The proposed F-test uses a SSE 16.73 when $h = 7.28$ under H_a, which is a bit smaller than $\sum_i (Y_i - \tilde{m}(X_i))^2$, 16.90. If we plug in the the proposed SSE in the PLR test statistic, then the PLR test has a p-value for H_0(iv) of 0.052.

The nonlinear prenatal methylmercury effects for CTRS is shown in Figure 1(b), with pointwise ± 1.96 standard errors. Note that a decreased CTRS means improved score. From $\hat{\mathbf{m}}$, which estimates the CTRS in the log scale, the curve decreases by 0.022 (95% CI $[-0.059, 0.015]$) from 1.02 to 5.04 ppm, decreases by 0.050 (95% CI $[-0.083, -0.018]$, significant) from 5.04 to 10.05 ppm, decreases by 0.040 (95% CI $[-0.089, 0.009]$) from 10.05 to 14.94 ppm, and increases by 0.036 (95% CI $[-0.049, 0.122]$) from 14.94 to 19.86 ppm. The standardized residual plots are shown in Figures 2(c) and (d). There are two outliers with prenatal methylmercury 4.97 and 7.29 ppm and the original CTRS score 95 and 100. The residual plots show moderate compatibility with the homoscedastic normal error assumption. Similar to the Full IQ, CTRS demonstrates a significant nonlinear effect when coexposed by methylmercury and nutrients.

Neither F- nor PLR tests for H_0(iii) and H_0(iv) with $h = 7.28$ are significant for the remaining three endpoints. In Huang et al. (2005), the GP was significant based on the approximate F-tests. For the GP, the AICc selection chooses the largest bandwidth 10.5 (DF 3.00) and it improves the power of the ANOVA F-tests, p-values 0.007 and 0.005 for H_0(iii) and H_0(iv), respectively. It also improves the power of the PLR tests with p-values 0 and 0.088, respectively. Due to the negative reciprocal transform, the transformed GP has a range of $(-0.017, -0.0051)$. To examine the nonlinear prenatal effect with $h = 10.5$, we arbitrarily magnify it by 10^3. Using the same pointwise comparison from 1.02 to 5 ppm, ..., from 15.09 to 19.98 ppm, there is a significant increase (adverse effect) by 0.32 (95% CI [0.04, 0.60]) from 10.02 to 15.09 ppm, and a significant increase by 0.67 (95% CI [0.19, 1.14]) from 15.09 to 19.98 ppm. From Figure 1(c), the curve is nearly flat below 12 ppm and slowly increasing above 12 ppm, which confirms the results in Huang et al. (2005).

In this reanalysis, we formally evaluate the nonlinear trends of prenatal methylmercury using the new methodology developed in Section 3. The proposed ANOVA F-tests enjoy more theoretical support than the approximate F-tests. In attempting to interpret the nonlinear effects, we have found that it is not as simple as the linear effect. When a nonlinear curve is significant based on the proposed F-tests, it means that its contribution to reduction of SSE (or increment of SSR) is significant when compared to the model under H_0. It does not have a simple interpretation of the amount of change per 1 ppm increase of exposure as in the linear model. In other words, the confidence intervals in the parametric setting directly reflect significance tests of the slope parameters; however, in the nonparametric setting, such a correspondence is not warranted. In the case of Full IQ, the pointwise results may reflect the nonsignificant p-value in Myers et al. (2003) by linear regression analysis.

Nevertheless, the nonlinear curves in Figure 1 show interesting relationships: When no nutrient measurements are available, the methylmercury effects are confounded with nutrient effects, which may lead to a nonlinear trend in statistical modeling. Further investigation of the combined effects from both nutrients and methylmercury in fish will be important. The SCDS team has examined nutrient and methylmercury relationships in another cohort by linear regression analysis (Davidson et al. 2008) and further nonlinear analysis is in progress to explore nutrient effects modified by methylmercury.

6. DISCUSSION

Though smoothing techniques such as local linear regression and spline methods have been applied widely in scientific investigations, we note that there are not yet standard ANOVA tools to enhance interpretations of curve estimates. This paper provides a complete framework for estimation, ANOVA, and F-tests for PLMs. The proposed methodology enjoys being a compelling extension of classical methods. As in the classical settings, the semiparametric F-tests depend on the assumptions of normality and homoscedasticity. We are interested in further studying the robustness properties of the ANOVA F-tests and investigating an extension for heteroscasdastic cases. In the simulation study, the AICc bandwidth selector surprisingly improves the power of the ANOVA F-tests. It will be interesting to explore the performance of different selection criteria such as BIC, C_p, or GCV, based on the definitions of SSE and DF in the paper. With the proposed methodology, future work also includes nonparametric analysis of covariance (Dette and Neumeyer 2001), for example, separately fitting female and male methylmercury curves and testing for significance of methylmercury-by-gender interactions. Extensions of the ANOVA methods to additive models, SAMs, and semiparametric varying coefficient models, and to analysis of deviance for generalized SAMs are in progress. Another direction is to extend the semiparametric methods to models for multiple outcomes to allow investigation of dependencies across different response variables.

Using the proposed methodology, we reanalyzed the SCDS nine-year Main Cohort data. The result of the Full IQ is particularly important, as some authors have suggested there is a reduction in intelligence due to methylmercury exposure (Axelrad et al. 2007). However, fish contain nutrients, especially LCPUFA, that appear to modify the impact of methylmercury on child development (Davidson et al. 2008; Strain et al. 2008). The significance of nonlinear methylmercury effects on IQ may imply that the nonlinear assumption is more reasonable to model methylmercury effects when no nutrient status is available. Linear dose–response assessment may result in a misleading picture at the low levels of exposure. The lack of significance in pointwise comparison of methylmercury effects on IQ may be because we did not adjust for nutrients. More research is necessary to confirm this prediction.

While most studies in the literature assume a linear methylmercury exposure effect, Huang et al. (2005) and the present paper employ a more flexible nonlinear trend. This reanalysis agrees with Huang et al. (2005), Cox et al. (1989), and WHO (1990) that the methylmercury effects may be nonlinear. Our results for the Full IQ, CTRS, and GP outcomes suggest a possible adverse effect in the uppermost range of prenatal exposure included in the SCDS Main Cohort.

APPENDIX

Proof of Theorem 1

For F-type test statistics, $F = (\mathbf{y}^\top \mathbf{A}\mathbf{y}/\operatorname{tr}(\mathbf{A}))/(\mathbf{y}^\top \mathbf{B}\mathbf{y}/\operatorname{tr}(\mathbf{B}))$, the F-distribution is warranted if $\mathbf{A}$ and $\mathbf{B}$ are both projection matrices (symmetric and idempotent) and they are orthogonal to each other. For all four F-statistics in Theorem 3, it is easy to see that the symmetric property is satisfied. We will verify that the matrices are asymptotically idempotent and the numerator and denominator are asymptotically orthogonal. All the expectations in this section are conditioned on $\mathbf{x}$ and $\mathcal{Z}$ and for simplicity, abbreviated as $E_c\{\cdot\} = E\{\cdot|\mathbf{x}, \mathcal{Z}\}$.

The matrices in the quadratic forms of $SSR_p(h)$ and $SSE_p(h)$ are $\mathbf{A}_1 = (\mathbf{I} - \mathbf{P}_{z|x}^\top)(\mathbf{I} - \mathbf{H}^*)$ and $\mathbf{B}_1 = (\mathbf{H}^* - \mathbf{L} + \mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*))$, respectively. For $\mathbf{A}_1$, note that $\mathbf{A}_1(\mathbf{I} - \mathbf{P}_{z|x}) = \mathbf{A}_1$ since $\mathbf{A}_1\mathbf{P}_{z|x} = \mathbf{0}$. Then

$$\begin{aligned} E_c\{(\mathbf{A}_1 - \mathbf{A}_1^2)\mathbf{y}\} &= E_c\{\mathbf{A}_1\mathbf{H}^*(\mathbf{I} - \mathbf{P}_{z|x})\mathbf{y}\} \\ &= E_c\{(\mathbf{I} - \mathbf{P}_{z|x}^\top)(\mathbf{I} - \mathbf{H}^*)\mathbf{H}^*(\mathbf{I} - \mathbf{P}_{z|x})\mathbf{y}\}, \end{aligned}$$

which tends to a zero vector since $\mathbf{H}^*$ and $(\mathbf{I} - \mathbf{H}^*)$ are asymptotically orthogonal and elements of $(\mathbf{I} - \mathbf{P}_{z|x}^\top)$ are conditionally of order $O(1)$. Hence $\mathbf{A}_1$ is asymptotically idempotent. For $\mathbf{B}_1$, note that $\mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*)$ is asymptotically idempotent since $E_c\{\mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*)(\mathbf{I} - \mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*))\mathbf{y}\} = E_c\{\mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*)\mathbf{H}^*\mathbf{P}_{z|x}\mathbf{y}\}$, which tends to a zero vector as $n \to \infty$. Clearly, $(\mathbf{H}^* - \mathbf{L})$ is asymptotically idempotent and asymptotically orthogonal to $\mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*)$. Therefore $\mathbf{B}_1$ is asymptotically idempotent. The orthogonality between $\mathbf{A}_1$ and $\mathbf{B}_1$ follows analogously from the asymptotic orthogonality between $(\mathbf{I} - \mathbf{H}^*)$ and $(\mathbf{H}^* - \mathbf{L})$ and exact orthogonality between $\mathbf{A}_1$ and $\mathbf{P}_{z|x}$. Hence the test statistic for H_0(i) has an asymptotic F distribution under the normality assumption.

For Q_2, the matrix $\mathbf{A}_2 = \mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*)$ is a component in $\mathbf{B}_1$ and hence the arguments discussed for $\mathbf{B}_1$ is applicable for $\mathbf{A}_2$. The matrix in Q_3 is $\mathbf{A}_3 = \mathbf{H}^* + \mathbf{P}_{z|x}^\top(\mathbf{I} - \mathbf{H}^*) - \mathbf{H}_z - \mathbf{L} = \mathbf{B}_1 - \mathbf{H}_z$. Since $\mathbf{H}_z$ is idempotent and $\mathbf{B}_1$ is asymptotically idempotent, it is easy to see that $\mathbf{A}_3$ is asymptotically idempotent given that $\mathbf{B}_1\mathbf{H}_z = \mathbf{H}_z$ due to $\mathbf{P}_{z|x}\mathbf{H}_z = \mathbf{H}_z$. For asymptotic orthogonality between $\mathbf{B}_1 - \mathbf{H}_z$ and $\mathbf{A}_1$, it follows from the fact that $\mathbf{A}_1$ and $\mathbf{B}_1$ are asymptotically orthogonal as shown above and $\mathbf{H}_z\mathbf{A}_1 = \mathbf{0}$.

The matrix in Q_4 is $\mathbf{A}_4 = \mathbf{I} - \mathbf{A}_1 - \mathbf{H}_{xz}$. We first show $\mathbf{A}_1\mathbf{H}_{xz} = \mathbf{0}$. Denote the local projection matrix for fitting local linear regression (3) as $\mathbf{H}_x = \mathbf{X}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{W}$. Then $\mathbf{H}_x\mathbf{x} = \mathbf{x}$, where $\mathbf{x} = (X_1, \ldots, X_n)$, and $\mathbf{H}_x\mathbf{L} = L$. Hence the asymptotic projection matrix $\mathbf{H}^*\mathbf{x} = (\int \mathbf{W}\mathbf{H}_x\hat{f}(x)\,dx)\mathbf{x} = \mathbf{x}$ and $\mathbf{H}^*\mathbf{L} = \mathbf{L}$. It follows that $\mathbf{A}_1\mathbf{H}_z = \mathbf{0}$, $\mathbf{A}_1\mathbf{L} = \mathbf{0}$, and $\mathbf{A}_1\mathbf{x} = \mathbf{0}$, which together imply $\mathbf{A}_1\mathbf{H}_{xz} = \mathbf{0}$. The asymptotic idempotency of $\mathbf{A}_4$ follows from $E_c\{(\mathbf{A}_4 - \mathbf{A}_4^2)\mathbf{y}\} = E_c\{(\mathbf{I} - \mathbf{A}_1)\mathbf{A}_1\mathbf{y}\}$. Then the asymptotic orthogonality between $\mathbf{A}_4$ and $\mathbf{A}_1$ follows from the properties of $\mathbf{A}_1$ as shown above and $\mathbf{A}_1\mathbf{H}_{xz} = \mathbf{0}$.

Conditions for Theorem 2

The following conditions, similar to Speckman (1988), are imposed for Theorem 2.

Condition (A).

(A1) For the jth covariate Z_j in $\mathbf{z}$, $j = 1, \ldots, p$, assume Z_j and X are related via a regression model

$$Z_j = g_j(X) + \eta_j, \tag{A.1}$$

where the η_j's are independent random variables with mean 0 and are independent of ε, and $g_j(\cdot)$ are smooth functions with continuous fourth derivatives. The covariance matrix of $(\eta_1, \ldots, \eta_p)$ denoted by $\mathbf{V} = (V_{ij})$ is finite and positive definite. Let $\mathbf{g}(x) = (g_1(x), \ldots, g_p(x))^\top$. Denote the data realization from (A.1) as $\boldsymbol{\eta}_j = (\eta_{1j}, \ldots, \eta_{nj})^\top$, $\mathbf{g}_j = (g_j(X_1), \ldots, g_j(X_n))^\top$, $\mathbf{g} = (\mathbf{g}_1, \ldots, \mathbf{g}_p)$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p)$. Then in matrix notation, data from (A.1) is

$$\mathcal{Z} = \mathbf{g} + \boldsymbol{\eta}.$$

(A2) The asymptotic projection matrix $\mathbf{H}^*$ has the following properties (Huang and Chen 2008): $\{(\mathbf{I} - \mathbf{H}^*)\mathbf{m}|X_1, \ldots, X_n\} = h^4(b(X_1), \ldots, b(X_n))^\top$; $\|(\mathbf{I} - \mathbf{H}^*)\mathbf{m}\|^2 = O(nh^8)$, where $\|\cdot\|$ denotes the Euclidean norm; for $1 \le j \le p$, $\|\mathbf{H}^*\boldsymbol{\eta}_j\|^2 = O(h^{-1})$; $\operatorname{tr}(\mathbf{H}^{*2}) = O(h^{-1})$; and for $1 \le i \le n$, from (A.1), the ith element of $(\mathbf{H}^* - I)\mathbf{g}_j$ is asymptotically $h^4 b_j(X_i) + o(h^4)$ where $b_j(X_i)$ is $b(X_i)$ with the function $m(\cdot)$ replaced by $g_j(\cdot)$.

(A3) The marginal density function $f(x)$ for X is bounded away from 0 and ∞, and $f''(x)$ exists. X has a bounded support.

(A4) The kernel $K(\cdot)$ is a Lipschitz continuous, bounded and symmetric probability density function, having a support on a compact interval, say $[-1, 1]$.

(A5) The error ε is from a symmetric distribution with mean 0, variance σ^2, and a finite fourth moment.

(A6) The fourth derivative of $m(\cdot)$ exists.

Proof of Theorem 2

Condition (A1) justifies $E(\boldsymbol{\eta}_j^\top(\mathbf{I} - \mathbf{H}^*)\mathbf{m}) = \mathbf{0}$ and hence $n^{-1}\boldsymbol{\eta}_j^\top(\mathbf{I} - \mathbf{H}^*)\mathbf{m} = O(n^{-1/2}h^4)$ since $\operatorname{Var}(\boldsymbol{\eta}_j^\top(\mathbf{I} - \mathbf{H}^*)\mathbf{m}) = V_{jj}\|(\mathbf{I} - \mathbf{H}^*)\mathbf{m}\|^2 = O(nh^8)$ from condition (A2). It is easy to check that all the assumptions in Speckman (1988) are satisfied for the setting here with $\mathbf{H}^*$, and theorems 1 and 3 in Speckman (1988) can be modified for the new estimators (10) and (15) as in Theorem 2.

Proof of Theorem 3

It is known that $\hat{\beta}_0(x)$ and $\hat{\beta}_1(x)$ are linear estimators of $\mathbf{y}$ (Fan and Gijbels 1996). Denote $\hat{\beta}_0(x) = \sum_{j=1}^n b_j(x)Y_j$ and $\hat{\beta}_1(x) = \sum_{j=1}^n c_j(x)Y_j$. Then

$$Y_i^* = \int \left\{ \left(\sum_{j=1}^n b_j(x)Y_j \right) + (X_I - x) \left(\sum_{j=1}^n c_j(x)Y_j \right) \right\} K_h(X_i - x)\,dx.$$

Using the asymptotic expression of $b_j(x)$ and $c_j(x)$ from Fan and Gijbels (1996),

$$\begin{aligned}
&\int b_j(x) K_h(X_i - x)\, dx \\
&\quad = n^{-1} h^{-1} (1 + o_p(1)) \int K_h(X_j - x) K_h(X_i - x)/f(x)\, dx \\
&\quad = n^{-1} h^{-2} (1 + o_p(1)) \int K(u) K\left(\frac{X_j - X_i}{h} + u\right) \Big/ f(X_i - hu)\, du \\
&\quad = n^{-1} h^{-2} (1 + o_p(1)) f^{-1}(X_i) K_0^*\left(\frac{X_j - X_i}{h}\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\int c_j(x)(X_i - x) K_h(X_i - x)\, dx \\
&\quad = n^{-1} h^{-2} (1 + o_p(1)) \\
&\qquad \times \int (X_j - x) K_h(X_j - x)(X_i - x) K_h(X_i - x)/(f(x)\mu_2)\, dx \\
&\quad = -\mu_2^{-1} n^{-1} h^{-2} (1 + o_p(1)) \\
&\qquad \times \int (X_j - X_i - hu) K\left(\frac{X_j - X_i}{h} - u\right) u K(u)/f(X_i + hu)\, du \\
&\quad = -\mu_2^{-1} n^{-1} h^{-3} (1 + o_p(1)) f^{-1}(X_i) K_1^*\left(\frac{X_j - X_i}{h}\right).
\end{aligned}$$

Thus, Theorem 3 is proved.

[Received May 2008. Revised December 2009.]

REFERENCES

ATSDR (1999), *Toxicological Profile for Mercury: Update*, Atlanta, GA: Agency for Toxic Substances and Disease Registry. [1]

Axelrad, D. A., Bellinger, D. C., Ryan, L. M., and Woodruff, T. J. (2007), "Dose–Response Relationship of Prenatal Mercury Exposure and IQ: An Integrative Analysis of Epidemiologic Data," *Environmental Health Perspectives*, 115, 609–615. [9,11,13]

Bruininks, R. H. (1978), *Bruinink–Oseretsky Test of Motor Proficiency*, Circle Pines, MN: American Guidance Service. [9]

Connors, K. A. (1985), *The Conners Rating Scales: Instruments for the Assessment of Childhood Psychopathology*, Odessa, FL: Psychological Assessment Resources. [9]

Cox, C., Clarkson, T. W., Marsh, D. O., Amin-Zaki, L., Tikriti, S., and Myers, G. (1989), "Dose–Response Analysis of Infants Prenatally Exposed to Methylmercury: An Application of a Single Compartment Model to Single-Strand Hair Analysis," *Environmental Research*, 49, 318–332. [2,13]

Davidson, P. W., Myers, G. J., Strain, J. J., Thurston, S. W., Bonham, M. P., Shamlaye, C. F., Stokes-Riner, A., Wallace, J. M. W., Robson, P. J., Duffy, E. M., Georger, L., Sloane-Reeves, J., Cernuchiarti, E., Canfield, R. L., Cox, C., Huang, L.-S., Janciuras, J., and Clarkson, T. W. (2008), "Neurodevelopmental Effects of Maternal Nutritional Status and Exposure to Methylmercury From Eating Fish During Pregnancy," *NeuroToxicology*, 29, 767–775. [1,12,13]

Delis, D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. (1994), *California Verbal Learning Test—Children's Version*, San Antonio, TX: The Psychological Corporation. [8]

Dette, H., and Neumeyer, N. (2001), "Nonparametric Analysis of Covariance," *The Annals of Statistics*, 29, 1361–1400. [2,12]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall. [1,2,13,14]

Fan, J., and Huang, T. (2005), "Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, 11, 1031–1058. [2,6,10]

Gasser, T., Müller, H.-G., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society, Ser. B*, 47, 238–252. [6]

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman & Hall. [5]

Hamilton, S. A., and Truong, Y. K. (1997), "Local Linear Estimation in Partly Linear Models," *Journal of Multivariate Analysis*, 60, 1–19. [6]

Härdle, W., Liang, H., and Gao, J. (2000), *Partially Linear Models*, Heidelberg: Physica-Verlag. [1]

Härdle, W., Muller, M., and Mammen, E. (1998), "Testing Parametric Versus Semiparametric Modeling in Generalized Linear Models," *Journal of the American Statistical Association*, 93, 1461–1474. [2]

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman & Hall. [1,2,5,6,11]

——— (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society, Ser. B*, 55, 757–796. [2]

He, H., and Huang, L.-S. (2006), "Double-Smoothing for Bias Reduction in Local Linear Regression," *Journal of Statistical Planning and Inference*, 139, 1056–1072. <uncited>

Huang, L.-S., and Chen, J. (2008), "Analysis of Variance, Coefficient of Determination, and *F*-Test for Local Polynomial Regression," *The Annals of Statistics*, 36, 2085–2109. [1-6,13]

Huang, L.-S., Cox, C., Myers, G. J., Davidson, P. W., Cernuchiarti, E., Sloane-Reeves, J., Shamlaye, C. F., and Clarkson, T. W. (2005), "Exploring Nonlinear Association Between Prenatal Methylmercury Exposure From Fish Consumption and Child Development: Evaluation of the Seychelles Child Development Study Nine-Year Data Using Semiparametric Additive Models," *Environmental Research*, 97, 100–108. [1,2,8-13]

Hurvich, C. M., Simonoff, J. S., and Tsai, C.-L. (1998), "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society, Ser. B*, 60, 271–293. [6]

Knights, R. M., and Moule, P. D. (1968), "Normative Data on the Motor Steadiness Battery for Children," *Perceptual Motor Skills*, 26, 643–650. [8]

Myers, G. J., Davidson, P. W., Cox, C., Shamlaye, C. F., Palumbo, D., Cernichiari, E., Sloane-Reeves, J., Wilding, G., Kost, J., Huang, L.-S., and Clarkson, T. W. (2003), "Prenatal Methylmercury Exposure From Ocean Fish Consumption in the Seychelles Child Development Study," *The Lancet*, 361, 1686–1692. [1,2,8,9,11,12]

Myers, G. J., Davidson, P. W., and Shamlaye, C. F. (2006), "Developmental Disabilities Following Prenatal Exposure to Methyl Mercury From Maternal Fish Consumption: A Review of the Evidence," in *International Review of Mental Retardation Research*, Vol. 30, San Diego, CA: Elsevier Academic Press, pp. 141–170. [1]

National Institute of Environmental Health Sciences (1998), *Workshop Report on Scientific Issue Relevant to Assessment of Health Effects From Exposure to Methylmercury*, Raleigh, North Carolina. [1,2]

National Research Council (2000), *Toxicological Effects of Methylmercury*, Washington, DC: National Academy Press. [1,2]

Opsomer, J. D., and Ruppert, D. (1999), "A Root-*n* Consistent Backfitting Estimator for Semiparametric Additive Modeling," *Journal of Computational and Graphical Statistics*, 8, 715–732. [3,5,6]

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge, U.K.: Cambridge University Press. [2,5]

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society, Ser. B*, 50, 413–436. [3,5,6,13]

Strain, J. J., Davidson, P. W., Bonham, M. P., Duffy, E. M., Stokes-Riner, A., Thurston, S. W., Wallace, J. M. W., Robson, P. J., Shamlaye, C. F., Georger, L., Sloane-Reeves, J., Cernuchiarti, E., Canfield, R. L., Cox, C., Huang, L.-S., Janciuras, J., Myers, G. J., and Clarkson, T. W. (2008), "Associations of Maternal Long Chain Polyunsaturated Fatty Acids, Methyl Mercury, and Infant Development in the Seychelles Child Development and Nutrition Study," *NeuroToxicology*, 29, 776–782. [1,13]

Wechsler, D. (1991), *Wechsler Intelligence Scale for Children—Third Revision (WISC-III)*, San Antonio, TX: The Psychological Corporation. [8,11]

WHO (1990), *Methylmercury. Environmental Health Criteria*, Vol. 101, Geneva: World Health Organization. [13]

META DATA IN THE PDF FILE

Following information will be included as pdf file Document Properties:

Title : Analysis of Variance and F-Tests for Partial Linear Models With Applications to Environmental Health Data
Author : Li-Shan Huang, Phillip W. Davidson
Subject : Journal of the American Statistical Association, Vol.0, No.0, 2010, 1-15
Keywords: Local linear regression, Methylmercury, Penalized least squares, Smoother matrix

THE LIST OF URI ADRESSES

Listed below are all uri addresses found in your paper. The non-active uri addresses, if any, are indicated as ERROR. Please check and update the list where necessary. The e-mail addresses are not checked – they are listed just for your information. More information can be found in the support page: *http://www.e-publications.org/ims/support/urihelp.html*.

--- mailto:Lhuang@bst.rochester.edu [2:pp.1,1] Check skip
--- mailto:phil_davidson@urmc.rochester.edu [2:pp.1,1] Check skip
200 http://www.amstat.org [2:pp.1,1] OK
302 http://pubs.amstat.org/loi/jasa [2:pp.1,1] Found
200 http://dx.doi.org/10.1198/jasa.2010.ap08274 [2:pp.1,1] OK