Published in final edited form as:

J Stat Plan Inference. 2013 December 1; 143(12): 2160–2170. doi:10.1016/j.jspi.2013.05.002.

Model Averaging Procedure for Partially Linear Single-index Models^a

Yao Yu¹, Sally W. Thurston¹, Russ Hauser², and Hua Liang¹

¹ Department of Biostatistics and Computational Biology, University of Rochester, NY 14642, U.S.A.

² Department of Environmental Health, Harvard School of Public Health Boston, Massachusetts 02115, U.S.A.

Abstract

This paper is concerned with model selection and model averaging procedures for partially linear single-index models. The profile least squares procedure is employed to estimate regression coefficients for the full model and submodels. We show that the estimators for submodels are asymptotically normal. Based on the asymptotic distribution of the estimators, we derive the focused information criterion (FIC), formulate the frequentist model average (FMA) estimators and construct proper confidence intervals for FMA estimators and FIC estimator, a special case of FMA estimators. Monte Carlo studies are performed to demonstrate the superiority of the proposed method over the full model, and over models chosen by AIC or BIC in terms of coverage probability and mean squared error. Our approach is further applied to real data from a male fertility study to explore potential factors related to sperm concentration and estimate the relationship between sperm concentration and monobutyl phthalate.

Keywords

Profile least squares procedure; AIC; BIC; Focused information criterion (FIC); Frequentist model averaging (FMA)

1 Introduction

Partially linear single-index models (PLSIM) are an extension of single-index models, $E(Y|Z) = \eta(Z^T \mathbf{a})$, to also adjust for additional covariates X which are assumed to be linearly related to the outcome. Various methods have been proposed in the literature for parameter estimators in the PLSIM. For example, Carroll et al. (1997) studied a more general case and proposed one-step and fully iterated estimation procedures; Yu and Ruppert (2002) developed the penalized spline estimation procedure; Xia and Häardle (2006) borrowed some ideas from dimension reduction (Xia et al., 1999), and then proposed the minimum

^aThe authors gratefully thank the Editor, an Associate Editor and two referees for their helpful comments. Dr. Liang's research was partially supported by NSF grants DMS-1007167 and DMS-1207444 and by Award Number 11228103, made by National Natural Science Foundation of China. Dr. Hauser's research was supported by the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH) grant ES009718. The contents of this publication are the sole responsibility of the authors and do not necessarily represent the official views of the NIEHS, NIH, or NSF.

average variance estimation (MAVE) method; Liang et al. (2010) proposed a profile least squares approach and obtained efficient estimators reaching the efficient bound.

Although recent research can provide us with a variety of estimation methods, the methods mentioned above are based on the assumption that the true model is known which may not be realistic. Moreover, including extraneous variables in a model will cause the parameter estimates to be more variable than those given in simpler models. As a consequence, a variety of model selection criteria have been proposed to select the “best model” and then make inference based on the selected model, such as the Akaike information criterion (AIC, Akaike, 1973), the Bayesian information criterion (BIC, Schwarz, 1978) and the deviance information criterion (DIC, Spiegelhalter et al., 2002). However, as observed by Yang (2001) and Leung and Barron (2006), model selection criteria are often unstable; that is, a small perturbation of the data may lead to a significant change of the “best” model chosen according to a certain model selection criterion. Consequently, a small change in the data may cause different results. Similar conclusions can also be found in Danilov and Magnus (2004) and Leeb and Pötscher (2006) which show that parameter estimates resulting from models selected by some model selection information criteria such as AIC or by a hypothesis testing procedure may not be reasonably accurate. Furthermore, in real data analysis, it isn't uncommon for two or more models to be close in terms of a certain information criterion which makes it difficult to conclude that the model chosen is better than the other models.

To make full use of the information from candidate models and overcome the problem of model structure uncertainty commonly encountered in information criteria, strategies weighting the candidate models were adopted and studied in the literature to avoid inaccurate scientific summaries and overconfident decisions. Draper (1995) suggested a Bayesian approach to solve this problem by considering a weighted average of posterior means over possible models. However, prior probabilities of all possible models need to be specified. Buckland et al. (1997) not only emphasized the necessity of incorporating model selection uncertainty into statistical inference, but also proposed a method to weight candidate models, where the weights can be obtained from information criteria or from the bootstrap. Yang (2001) proposed an adaptive regression by mixing algorithm, assigning weights to the candidate models by assessing the performance of estimators. Leung and Barron (2006) proposed a convex mixture of the component estimators with weights that may depend on the data. Overall, the above procedures aim to combine candidate models via smoothing estimators, instead of relying entirely upon a single model selected by a particular model selection criterion. The confidence intervals constructed based on the asymptotic distribution of model average estimators can improve coverage probabilities while reducing the mean squared error of estimators. A summary of recent progress in the area of model averaging has been provided in Claeskens and Hjort (2008).

Model averaging procedures have been well studied for parametric and nonparametric models in the above research. Recently, pioneering research has investigated model averaging procedures for semiparametric models; see, for example, Claeskens and Carroll (2007) who developed FIC and FMA in partial linear models, and established asymptotic properties for FMA estimators; and Zhang and Liang (2011) who studied FIC and FMA

procedures in generalized additive partially linear models. However, so far model averaging procedures haven't been studied in PLSIM. In this paper, we propose a model averaging procedure based on FIC for PLSIM. We first apply the profile least squares approach to estimate parameters for submodels and establish asymptotic properties for the resulting estimators. We then derive FIC based on the unbiased estimator of the mean squared error of focus parameters, which are prespecified functions of model parameters (Hjort and Claeskens, 2003), to establish a weight function for candidate models. Based on the weight function constructed in previous steps, we develop a frequentist model averaging (FMA) procedure for PLSIM, and construct confidence intervals for focus parameters with improved coverage probability and mean squared error compared to models chosen by information criteria such as AIC or BIC. The rest of the paper is organized as follows. Section 2 gives the model framework, provides profile least squares estimators and shows asymptotic normality of the estimators. Section 3 introduces FIC and FMA procedures, and constructs confidence intervals for the focus parameters on the basis of FMA estimators. Results from simulation studies and a real data analysis are presented in Sections 4 and 5 respectively. Regularity conditions and technical proofs are presented in the Appendix.

2 Model Framework and Estimation

Consider a partially linear single-index model

$$Y = \eta(Z^T \alpha) + X^T \beta + \varepsilon, \quad (2.1)$$

where $\eta(\cdot)$ is an unknown differentiable function; α is an unknown $p \times 1$ single-index parameter vector, β is an unknown $q \times 1$ regression parameter vector; $\|\alpha\| = 1$ and the first element of α is positive (for identifiability), where $\|\cdot\|$ denotes the Euclidean norm; ε are independent of $\{Z, X\}$ with mean zero and finite variance σ^2 .

When model selection criteria are applied, there is a list of models to be considered as possible candidates for a final model. Some variables, namely “certain” variables, are included in all candidate models, while the remaining variables may be either included or

excluded. Define $\alpha = (\alpha_c^T, \alpha_u^T)^T$ and $\beta = (\beta_c^T, \beta_u^T)^T$ where α_c and β_c are parameters corresponding to certain variables; α_u and β_u are parameters corresponding to other variables for which there is uncertainty as to whether or not they will be included in the selected model. In a general setting, a model is selected from the $2^{d_{\alpha,u} + d_{\beta,u}}$ submodels of the full model, where $d_{\alpha,u}$ and $d_{\beta,u}$ are the number of elements in α_u and β_u respectively. Let

$\alpha_{u,s}^T$ and $\beta_{u,s}^T$ be the subset of α_u^T and β_u^T which correspond to the uncertain parameters included in the s^{th} submodel and π_α and π_β be the projection matrix of 0's and 1's mapping α_u^T and β_u^T to $\alpha_{u,s}^T$ and $\beta_{u,s}^T$. Two extreme cases of the submodels are the full model where $\alpha_{u,s}^T = \alpha_u^T$ and $\beta_{u,s}^T = \beta_u^T$, and the reduced model where $\alpha_{u,s}^T = 0_{d_{\alpha,u}}^T$ and $\beta_{u,s}^T = 0_{d_{\beta,u}}^T$.

Suppose $\{(Y_i, Z_i, X_i), i = 1, \dots, n\}$ are a sequence of i.i.d. samples from model (2.1). In the s^{th} submodel, we consider the random sample $\{(Y_i, \Pi_\alpha Z_i, \Pi_\beta X_i), i = 1, \dots, n\}$, where

$\Pi_\alpha = \text{diag}(I_{d_{\alpha,c}}, \pi_\alpha)$ and $\Pi_\beta = \text{diag}(I_{d_{\beta,c}}, \pi_\beta)$. For simplicity, we denote the covariates in

the s^{th} submodel by $\Pi_{\alpha}Z_i = Z_{i,s}$ and $\Pi_{\beta}X_i = X_{i,s}$ respectively. Now we employ the profile least squares procedure proposed by Liang et al. (2010) to estimate the unknown parameter vectors $\alpha_s = (\alpha_c^T, \alpha_{u,s}^T)^T$ and $\beta_s = (\beta_c^T, \beta_{u,s}^T)^T$.

Let $\Lambda_{i,s} = Z_{i,s}^T \alpha_s$ and $\zeta_s = (\alpha_s^T, \beta_s^T)^T$. Suppose we have α_s, β_s . Then in the neighborhood of $u = z^T \alpha$, $\eta(\cdot)$ can be estimated utilizing the local linear regression method (Fan and Gijbels, 1996), i.e., minimizing

$$\sum_{i=1}^n \left\{ a + b(\Lambda_{i,s} - u) + X_{i,s}^T \beta_s - Y_i \right\}^2 K_h(\Lambda_{i,s} - u), \quad (2.2)$$

with respect to a and b , where $K_h(\cdot) = 1/hK(\cdot/h)$, $K(\cdot)$ is a kernel function and h is a bandwidth. Then, we can obtain

$$\hat{\eta}(u, \zeta_s) = \hat{a} = \frac{K_{20}(u, \zeta_s) K_{01}(u, \zeta_s) - K_{10}(u, \zeta_s) K_{11}(u, \zeta_s)}{K_{00}(u, \zeta_s) K_{20}(u, \zeta_s) - K_{10}^2(u, \zeta_s)}, \quad (2.3)$$

where $K_{jl}(u, \zeta_s) = \sum_{i=1}^n K_h(Z_{i,s}^T \alpha_s - u) (Z_{i,s}^T \alpha_s - u)^j (X_{i,s}^T \beta_s - y_i)^l$ for $j = 0, 1, 2$ and $l = 0, 1$. Consequently, the profile least squares function can be represented as a function of ζ_s

$$Q(\zeta_s) = \sum_{i=1}^n \left\{ Y_i - \hat{\eta}(Z_{i,s}^T \alpha_s, \zeta_s) - X_{i,s}^T \beta_s \right\}^2. \quad (2.4)$$

Then the Newton-Raphson iterative method can be applied to find the profile least squares

estimator $\hat{\zeta}_s = (\hat{\alpha}_c^T, \hat{\alpha}_{u,s}^T, \hat{\beta}_c^T, \hat{\beta}_{u,s}^T)^T$ in the s^{th} submodel, which minimizes the profile least squares function (2.4). It is worth mentioning that the Newton-Raphson iterative method is distinct from the more commonly used iterative (backfitting) procedure in partially single-index models which iteratively updates estimations of the nonparametric component and the parametric component obtained from their corresponding objective functions.

Now we derive the asymptotic distribution of $\hat{\zeta}_s$ under a local misspecification framework (Hjort and Claeskens, 2003) that assumes the true model has an unknown parameter vector

$\zeta_0 = (\alpha_{c,0}^T, \delta_{\alpha}^T / \sqrt{n}, \beta_{c,0}^T, \delta_{\beta}^T / \sqrt{n})^T$ where $\delta_{\alpha}^T / \sqrt{n}$ and $\delta_{\beta}^T / \sqrt{n}$ are the parameter-specific distances between the true model and the simplest submodel for α_u and β_u . This implies that the true model is $O(1/\sqrt{n})$ away from the reduced submodel. For simplicity, we introduce the following notation: $A^{\otimes 2} = AA^T$ for any matrix A , $\delta_{\alpha\beta} = (0, \delta_{\alpha}^T, 0, \delta_{\beta}^T)^T$, $\hat{\delta}_{\alpha\beta} = (0, \hat{\delta}_{\alpha}^T, 0, \hat{\delta}_{\beta}^T)^T$ where $\hat{\delta}_{\alpha}$ and $\hat{\delta}_{\beta}$ are the estimators of δ_{α} and δ_{β} from the full model, $\Lambda = Z^T \alpha$, $\tilde{Z} = Z - E(Z|\Lambda)$, $\tilde{X} = X - E(X|\Lambda)$, $\Gamma_{\tilde{X}\tilde{X}} = E(\tilde{X}^{\otimes 2})$, $\Gamma_{\tilde{X}\tilde{Z}} = E\{\tilde{Z}\tilde{X}^T \eta'(\Lambda)\}$ and $\Gamma_{\tilde{Z}\tilde{Z}} = E\{\tilde{Z}\tilde{Z}^T \eta'(\Lambda)\}^{\otimes 2}$ where η' is the first derivative of η . Zero vectors or matrices, and

identity matrices are presented by 0 and I respectively. The following theorem shows that for any s^{th} submodel, $\hat{\zeta}_s$ is asymptotically normal.

Theorem 1 Under a local misspecification framework and assuming conditions (i)-(vi) in the Appendix hold, then

$$n^{1/2} \begin{pmatrix} \hat{\alpha}_s - (\alpha_{c,0}^T, 0)^T \\ \hat{\beta}_s - (\beta_{c,0}^T, 0)^T \end{pmatrix} = (\Pi \Omega \Pi^T)^{-} \Pi G_n + (\Pi \Omega \Pi^T)^{-} \Pi \Omega \delta_{\alpha\beta} + o_p(1) \\ \xrightarrow{d} (\Pi \Omega \Pi^T)^{-} \Pi G + (\Pi \Omega \Pi^T)^{-} \Pi \Omega \delta_{\alpha\beta},$$

with $G_n \xrightarrow{d} G \sim N(0, \Sigma)$ and $\Sigma = \Omega \sigma^2$, where $\Omega = \begin{pmatrix} \Gamma_{\tilde{Z}\tilde{Z}} & \Gamma_{\tilde{X}\tilde{Z}} \\ \Gamma_{\tilde{X}\tilde{Z}}^T & \Gamma_{\tilde{X}\tilde{X}} \end{pmatrix}$, $\alpha_{c,0}$ and $\beta_{c,0}$ are parameter vectors for the true model, Π is a block diagonal matrix with blocks defined by $I_{d_{\alpha,c}}$, π_{α} , $I_{d_{\beta,c}}$ and π_{β} , and “ $-$ ” denotes the Moore-Penrose inverse.

3 Focused Information Criterion and Frequentist Model Averaging

Many model selection criteria such as AIC or BIC only select one final model regardless of the intended use of the selected model, for example, predicting outcomes or estimating coefficients and so on. Following Hjort and Claeskens (2003), we derive FIC for PLSIM to find the best submodel for a given parameter of interest, namely the focus parameter. The focus parameter can be the quantity of interest such as coefficients of certain variables or predicted outcomes under certain conditions, and can be presented as a function of parameters in PLSIM. Recalling that the FIC strategy is to find the submodel with the smallest mean squared error of $\sqrt{n}(\hat{\mu}_s - \mu_0)$, where μ_0 is the true value of the focus parameter and $\hat{\mu}_s$ is the focus parameter estimator obtained from the s^{th} submodel, we derive modelling bias, variance under large sample approximation and then obtain the FIC in Section 3.1. In other words, based on the asymptotic distribution of $\hat{\zeta}_s$ and a Taylor expansion, we first derive the asymptotic distribution for the estimators of the focus parameters obtained from submodels. We then develop FIC, the unbiased estimator of the mean squared error of $\sqrt{n}(\hat{\mu}_s - \mu_0)$.

3.1 Focused Information Criterion

Recalling that the focus parameter is a function of parameters in PLSIM and a local

misspecification framework, we let $\mu_0 = \mu(\alpha_{c,0}, \delta_{\alpha}/\sqrt{n}, \beta_{c,0}, \delta_{\beta}/\sqrt{n})$,

$\hat{\mu}_s = \mu(\Pi_{\alpha}^T \hat{\alpha}_s, \Pi_{\beta}^T \hat{\beta}_s)$. Denote $R = \Pi^T (\Pi \Omega \Pi^T)^{-} \Pi$ and the partial derivative of μ with respect

$\mu_{\zeta} = \left(\frac{\partial \mu}{\partial \alpha_c}^T, \frac{\partial \mu}{\partial \alpha_u}^T, \frac{\partial \mu}{\partial \beta_c}^T, \frac{\partial \mu}{\partial \beta_u}^T \right) \Big|_{\alpha_c = \alpha_{c,0}, \alpha_u = 0, \beta_c = \beta_{c,0}, \beta_u = 0}$. Assuming that μ_{ζ} is continuous in the neighborhood of $\alpha_{c,0}$ and $\beta_{c,0}$, we show the asymptotic normality of $\hat{\mu}_s$ in Theorem 2.

Theorem 2 Under a local misspecification framework and assuming conditions (i)-(vi) hold,

$$\sqrt{n}(\hat{\mu}_s - \mu_0) = \mu_\zeta^T \mathbf{R} \mathbf{G}_n + \mu_\zeta^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I}) \boldsymbol{\delta}_{\alpha\beta} + o_p(1) \xrightarrow{d} \mu_\zeta^T \mathbf{R} \mathbf{G} + \mu_\zeta^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I}) \boldsymbol{\delta}_{\alpha\beta}.$$

A direct calculation yields the MSE of $\sqrt{n}(\hat{\mu}_s - \mu_0)$ as follows:

$$MSE \{ \sqrt{n}(\hat{\mu}_s - \mu_0) \} = \mu_\zeta^T \mathbf{R} \boldsymbol{\Sigma} \mathbf{R} \mu_\zeta + \mu_\zeta^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I}) \boldsymbol{\delta}_{\alpha\beta} \boldsymbol{\delta}_{\alpha\beta}^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I})^T \mu_\zeta.$$

By Theorem 1, we obtain

$$\hat{\boldsymbol{\delta}}_{\alpha\beta} = [0, I, 0, I] \boldsymbol{\Omega}^{-1} \mathbf{G}_n + \boldsymbol{\delta}_{\alpha\beta} + o_p(1) \xrightarrow{d} \boldsymbol{\Delta}_{\alpha\beta} \equiv [0, I, 0, I] \boldsymbol{\Omega}^{-1} \mathbf{G} + \boldsymbol{\delta}_{\alpha\beta} \sim N \left(\boldsymbol{\delta}_{\alpha\beta}, [0, I, 0, I] \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1} [0, I, 0, I]^T \right).$$

The FIC for the s^{th} submodel can be written as

$$FIC_s = \mu_\zeta^T \left\{ \mathbf{R} \boldsymbol{\Sigma} \mathbf{R} + (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I}) \hat{\boldsymbol{\delta}}_{\alpha\beta} \hat{\boldsymbol{\delta}}_{\alpha\beta}^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I})^T - (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I}) [0, I, 0, I] \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1} [0, I, 0, I]^T (\mathbf{R} \boldsymbol{\Omega} - \mathbf{I})^T \right\} \mu_\zeta.$$

3.2 Frequentist Model Average

As an alternative to model selection, the model averaging procedure has many merits. For example, the model averaging procedure is stable due to smoothing estimators across candidate models; the model average estimators can avoid overly optimistical confidence intervals. Therefore, in this section, we obtain the model average estimators by using the weight function consisting of AIC, BIC or FIC to smooth the focus parameter estimators obtained from the candidate models, and derive the asymptotic distribution for model average estimators. Based on the distribution of model average estimators, we construct a confidence interval having correct coverage probability for the weighted estimators.

Model averaging procedures require weights to combine focus parameter estimators from candidate models. Denote $\mathcal{A}$ as the set of all candidate models and let s be the index for the s^{th} submodel. The model average estimators can be described as

$$\hat{\mu} = \sum_{s \in \mathcal{A}} w(s) \hat{\mu}_s, \quad (3.1)$$

where $w(s)$ is the weight for the s^{th} submodel which is a function of AIC, BIC or FIC. Following the suggestions of Buckland et al. (1997), the weights of smoothed AIC (S-AIC)

and smoothed BIC (S-BIC) estimators are given as $w_{AIC}(s) = \frac{\exp(-AIC_{s/2})}{\sum_{s \in \mathcal{A}} \exp(-AIC_{s/2})}$ and

$w_{BIC}(s) = \frac{\exp(-BIC_{s/2})}{\sum_{s \in \mathcal{A}} \exp(-BIC_{s/2})}$ respectively. For smoothed FIC (S-FIC) (Claeskens and Hjort, 2008),

$$w_{FIC}(s) = \frac{\exp\left(\frac{-FIC_s^*}{\mu_\zeta^T \Omega^{-1} \Sigma \Omega^{-1} \mu_\zeta}\right)}{\sum_{s \in \mathcal{A}} \exp\left(\frac{-FIC_s^*}{\mu_\zeta^T \Omega^{-1} \Sigma \Omega^{-1} \mu_\zeta}\right)},$$

where FIC_s^* is defined as

$$FIC_s^* = \begin{cases} FIC(s) & \text{if the estimated squared bias minus variance is positive,} \\ \mu_\zeta^T \mathbf{R} \Sigma \mathbf{R} \mu_\zeta & \text{otherwise.} \end{cases}$$

The FIC estimator can be treated as a special case of the S-FIC model average estimator, which is a weighted sum over the individual candidate estimates with weights of a special form: 1 for the selected model and 0 for the remaining models. Now we give the asymptotic distribution and explain how to construct confidence intervals for the model average estimators $\hat{\mu}$ in equation (3.1).

Theorem 3 Assume a local misspecification framework and conditions (i)-(vi) in the Appendix hold. If the weight functions are not continuous on at most a countable set, then

$$\sqrt{n}(\hat{\mu} - \mu_0) = \mu_\zeta^T \Omega^{-1} G_n + \mu_\zeta^T \left\{ Q(s) \hat{\delta}_{\alpha\beta} - \hat{\delta}_{\alpha\beta} \right\} + o_p(1),$$

where $Q(s) = \sum_s w(s) \mathbf{R} \Omega$.

Based on Theorem 3, we can easily obtain

$$\left[\sqrt{n}(\hat{\mu} - \mu_0) - \mu_\zeta^T \left\{ Q(s) \hat{\delta}_{\alpha\beta} - \hat{\delta}_{\alpha\beta} \right\} \right] / \hat{\kappa} \xrightarrow{d} N(0, 1),$$

where $\hat{\kappa}$ is an unbiased estimator of $\mu_\zeta^T \Omega^{-1} \Sigma \Omega^{-1} \mu_\zeta$. As a result, the confidence intervals for μ_0 with the lower bound(low_n) and upper bound(up_n) can be constructed as following:

$$\hat{\mu} - \mu_\zeta^T \left\{ Q(s) \hat{\delta}_{\alpha\beta} - \hat{\delta}_{\alpha\beta} \right\} / \sqrt{n} \mp z_{j/2} \hat{\kappa} / \sqrt{n}, \quad (3.2)$$

where $z_{j/2}$ is the $j/2$ quantile of a standard normal distribution. Therefore, we obtain the confidence intervals (low_n , up_n) of μ_0 with the probability $Pr\{\mu_0 \in (low_n, up_n)\} \rightarrow 1 - 2\Phi(z_{j/2})$, where $\Phi(\cdot)$ is the standard normal distribution function.

4 Simulation study

We now perform Monte Carlo simulation studies to examine the performance of our proposed methods. We assess the accuracy of the focus parameter estimators in terms of

mean squared error $\sum_{j=1}^{1000} (\hat{\mu}_j - \mu_0)^2 / 1000$, where j denotes the j^{th} replication. Smaller mean squared errors imply better estimators of the focus parameters. Empirical coverage

probabilities (CP) are applied to identify possible occurrences of overly-optimistic coverage probabilities for a claimed 95% confidence interval, from the proposed method or from models selected by model selection criteria.

The data were generated from the model:

$$y_i = \sin(Z_i^T \alpha) + X_i^T \beta + 0.1 \varepsilon_i, \quad (4.1)$$

where $\alpha = (\alpha_c, \alpha_u)$ with $\alpha_c = (2, 3, 3) / \sqrt{22+0.055/n}$, $\alpha_u = (0.25, 0.1, 0.15) / \sqrt{22n+0.055}$; $\beta = (\beta_c, \beta_u)$, with $\beta_c = (3, 2)$, $\beta_u = (0.75, 0.75) / \sqrt{n}$, where α_c and β_c are certain parameters and α_u and β_u are uncertain parameters. The number of potential submodels is 2^5 .

For each of the 1000 independent simulation iterations, X_i , Z_i , ε_i are mutually independent. Simulation studies each with $n = 200$ or 300 were conducted to examine the robustness, stability and the performance of our proposed method when X includes some discrete variables: (i), X_i and Z_i follow normal(0,1), $|\varepsilon_i| \sim \text{lognormal}(0,1.5)$; (ii), $(X_{1i}, X_{2i}) \sim \text{bernoulli}(0.5)$, $(X_{3i}, X_{4i}) \sim \text{normal}(0,1)$, $Z_i \sim \text{normal}(0,1)$, $|\varepsilon_i| \sim \text{lognormal}(0,1.5)$. We defined the following five focus parameters for each scenario: (a) $\mu_1 = \alpha_1 + \alpha_2 + \alpha_3 + \beta_1 + \beta_2$; (b) $\mu_2 = \alpha_1 + \alpha_2 + \alpha_3 + \beta_1 + \beta_2 + \beta_3 + \beta_4$; (c) $\mu_3 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \beta_1 + \beta_2$; (d) $\mu_4 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \beta_1 + \beta_2 + \beta_3 + \beta_4$; (e) $\mu_5 = \eta(1.04\alpha_1 + 0.7\alpha_2 - 1.49\alpha_3 + 2.33\alpha_4 + 0.21\alpha_5 - 0.19\alpha_6) + 2.05\beta_1 + 0.06\beta_2 - 1.59\beta_3 + 0.17\beta_4$; or $\mu_5^* = \eta(1.04\alpha_1 + 0.7\alpha_2 - 1.49\alpha_3 + 2.33\alpha_4 + 0.21\alpha_5 - 0.19\alpha_6) + 1\beta_1 + 0\beta_2 - 1.59\beta_3 + 0.17\beta_4$. μ_1 is the sum of all the certain parameters α_c and β_c ; μ_2 is the sum of α_c and β ; μ_3 is the sum of α and β_c ; μ_4 is the sum of α and β ; μ_5 and μ_5^* are the point estimates of y_i with randomly chosen x_i and z_i and are used to examine whether the proposed method works well for complex focus parameters, which may even depend on the nonparametric function. μ_5^* is the focus parameter for the bernoulli setting in (ii); this setting requires only that μ_5 be altered to μ_5^* since the other focus parameters have a 0 or 1 coefficient for $X_{1,i}$ and $X_{2,i}$.

We implement the FIC model selection and model averaging as follows: **step 1**. Apply the profile least squares method to estimate α and β for the full model. We apply multiple linear regression to obtain initial values of β and use $(1/\sqrt{d_\alpha}, \dots, 1/\sqrt{d_\alpha})^T$ as initial values of α in the Newton-Naphason method. We use the Epanechnikov kernel function $K(u) = 3/4(1-u^2)I(|u| < 1)$. For bandwidth selection, we generate a sequence of bandwidths based on the ranges of $Z_i^T \alpha$ in model (4.1). We then compute the generalized cross-validation (GCV) score for every bandwidth. The bandwidth with a minimum GCV score is selected for the full model. **step 2**. Estimate $\tilde{Z}$ by $\hat{Z}$, where $\hat{Z} = Z - \hat{E}(Z|\Lambda)$, and $\tilde{X}$ by $\hat{X}$, where $\hat{X} = X - \hat{E}(X|\Lambda)$ by using $\hat{E}(Z|\Lambda)$ and $\hat{E}(X|\Lambda)$ as the local linear estimators of $E(Z|\Lambda)$ and $E(X|\Lambda)$. Also estimate σ^2 as the variance of the residuals, and then compute the covariance matrix of the estimated $\tilde{Z}$ and $\tilde{X}$ to obtain the estimate of Ω . **step 3**. Denote

$$\alpha = \left\{ \alpha_1 = \left(1 - \sum_{r=2}^p \alpha_r^2\right)^{1/2}, \alpha_2, \dots, \alpha_p \right\} \quad \text{with} \quad \sum_{r=2}^p \alpha_r^2 < 1 \quad \text{and}$$

$\hat{\alpha}^{(1)} = (\hat{\alpha}_2, \hat{\alpha}_3, \dots, \hat{\alpha}^p)$. Following Cui et al. (2011), we estimate the Moore-Penrose inverse Ω^- by $\mathbf{J}(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T$, where $\mathbf{J}$ is a $p \times (p-1)$ Jacobian matrix

$$\mathbf{J} = \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\alpha}^{(1)}} = \begin{pmatrix} -\hat{\boldsymbol{\alpha}}^{(1)T} / \sqrt{1 - \|\hat{\boldsymbol{\alpha}}^{(1)}\|^2} \\ I_{p-1} \end{pmatrix}.$$

step 4. Estimate $\boldsymbol{\alpha}_s$ and $\boldsymbol{\beta}_s$ for submodels using the selected bandwidth obtained in step 1.

step 5. Calculate the bias-modified FIC_s^* given in Section 3.2 for submodels. **step 6.**

Formulate the smoothed AIC (S-AIC), smoothed BIC (S-BIC) and smoothed FIC (S-FIC) estimators based on equation (3.1), and construct confidence intervals for smoothed estimators and FIC estimator according to function (3.2). **step 7.** Repeat step 1 to step 6 for 1000 generated datasets, calculate the mean squared errors for the estimated focus parameters and empirical coverage probabilities.

The mean squared errors and coverage probabilities for the full model, the models chosen by AIC, BIC, FIC and the smoothed estimators: S-AIC, S-BIC and S-FIC are summarized in Table 1. It can be seen from Table 1 that the coverage property and accuracy of model average estimators are good because the confidence intervals based on model average estimators may be shifted to the correct position. Meanwhile FIC shows competitive performance in our simulation studies. In all settings, for both $n = 200$ and $n = 300$, model average estimators outperform the competitive information criteria in terms of accuracy and coverage. The S-AIC, S-BIC and S-FIC have similar performance, and all perform better than the full model and models chosen by AIC, BIC and FIC. Regarding model selection criteria, FIC shows better performance than the full model and than the models chosen by AIC and BIC, not only in terms of mean squared errors, but also for coverage probabilities. Compared with setting (i), setting (ii) has a larger mean squared error. However, the coverage probabilities for S-AIC, S-BIC and S-FIC are between 89.9% and 93.8% when the sample size is $n = 300$. In addition, in both settings (i) and (ii), the model selection criteria for AIC and BIC gave overly optimistic standard errors resulting in small coverage probabilities. The coverage probabilities of the nominal 95% intervals range from 74.6 % to 89.6% for AIC while they range from 64.3 % to 89.8% for BIC. As expected, MSE and coverage probabilities for all methods improve with larger sample size.

5 Real Data Example-Semen Quality Data

Phthalates are esters of phthalic acid which are mainly used as additives to increase the plasticity or fluidity of the material to which they are added such as plastics. Since phthalates are reproductive and developmental toxicants in laboratory animals and are commonly used in consumer products leading to widespread human exposure, there is increased scientific and public attention regarding the potential human health risks associated with phthalate exposure. Studies on laboratory animals demonstrated that some phthalate esters can adversely affect semen quality (Borch et al., 2006; Ge et al., 2007; Jurewicz et al., 2009). In humans, limited research on subfertile males found associations between urinary phthalate levels commonly experienced by the public and impaired semen quality such as sperm concentration, morphology or motility (Duty et al., 2005, 2003;

Hauser et al., 2006; Jonsson et al., 2005; Pant et al., 2008; Wirth et al., 2008; Zhang et al., 2006). Research on the relationship between phthalates and reproductive hormones in fertile men have shown suggestive associations (Meeker et al., 2009; Mendiola et al., 2011, 2012). Often, statistical conclusions are obtained via linear regression models, which assume a linear relationship between sperm concentration and all confounding variables. As a consequence, some nonlinear relationships were ignored unconsciously. When nonlinear relationships are present, models which capture the nonlinearity and lead to more accurate estimates of exposure-response relationships become necessary.

Here we examine data from $n = 450$ men who are part of a semen quality study conducted at the Massachusetts General Hospital Fertility Center. Our focus is on the relationship between sperm concentration and urinary monobutyl phthalate concentrations, MBP, a metabolite of di-n-butyl phthalate. Our outcome variable, Y , is the logarithm of sperm concentration (millions of sperm per millilitre), and we adjust for four continuous covariates, namely, Z_1 : BMI (kg/m^2), Z_2 : Age (years), X_1 : Abstinence time (days since last ejaculation, truncated to 10 days), X_2 : Log(MBP) (logarithm of urinary MBP, in nanogram monoester per millilitre), and four discrete covariates including X_3 : Race (1=white, 0=otherwise), X_4 : HiFever (1=had high fever in last 3 months, 0=otherwise), X_5 : FSmoker (1=former smoker, 0=otherwise), and X_6 : CSmoker (1=current smoker, 0=otherwise). The outcome and all explanatory variables were standardized (divided by the sample standard deviation after subtracting the sample mean) before analysis.

An earlier study indicated that the logarithm of sperm concentration is nonlinearly related to age and to BMI (Liang et al., 2008). We use PLSIM to capture the curvature while exploring significant predictors for the outcome variable. To satisfy the identifiability requirement of PLSIM, we restrict α_1 to be > 0 , and choose BMI as the first covariate for the single-index. Hence the full model can be presented as:

$$Y = \eta(\alpha_1 Z_1 + \alpha_2 Z_2) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6. \quad (5.1)$$

The profile least squares procedure isn't sensitive to initial values when covariates $\mathbf{X}$ are continuous or $\mathbf{X}$ only includes a few discrete covariates. However when covariates $\mathbf{X}$ include many discrete variables, convergence problem may occur if the initial values are far from true values. We first applied the MAVE method (Xia et al., 1999) to obtain good initial values for $\hat{\alpha}$ and $\hat{\beta}$. We then employed the profile least squares method given in Section 2 to estimate α and β for the full model (5.1). We used bandwidths ranging from $n^{-1/3}$ to $n^{-1/6}$, and based on GCV we chose $h = n^{-1/6}$ for the full model. We applied the selected bandwidth to submodels. The kernel function used is the Epanechnikov kernel function

$K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$. The estimated results are presented in Table 2. Covariates are considered significant when the P -value is less than 0.05.

Since the meaning of α is very difficult to explain for the PLSIM, in this paper, we only report the estimated β , their corresponding standard errors, Z -values and P -values for the PLSIM in Table 2. The linear coefficient estimates provided by the PLSIM show that abstinence time is a statistically significant predictor with an estimated coefficient of 0.1497,

which indicates that greater abstinence time is associated with higher sperm concentration. Although $\log(\text{MBP})$ is not a significant predictor in the PLSIM and the linear regression model, we see that the MBP effect is estimated with a smaller coefficient and P -value in PLSIM than in the linear regression model (see Table 2). Moreover the plot $\eta(Z^T \mathbf{a})$ versus the single-index $Z^T \mathbf{a}$ from the full model presented in Figure 1 shows the nonlinear relationship between the outcome variable and the single-index. Thus, the need to accommodate curvature is important when we analyze this dataset.

Based on the estimated results of the full model, we now conduct model selection and model averaging. We choose the significant covariates: BMI, age and abstinence time as certain variables and the remaining 5 variables as uncertain variables. We then compute FIC and perform model averaging on the 2^5 submodels. The model selection criteria applied are AIC, BIC and three versions of FICs. $\text{FIC}(\mu_1)$ corresponds to the coefficient of abstinence time $\mu_1 = \beta_1$. To employ the model average estimator to examine the influence of MBP on sperm concentration, we consider μ_2 and μ_3 , the logarithm of sperm concentrations corresponding to two different conditions: μ_2 is for a white male with median values of standardized age, BMI, abstinence time and $\log(\text{MBP})$; μ_3 is for a white male with median values of age, BMI and abstinence time and the third quantile of $\log(\text{MBP})$. The μ_2 and μ_3 can be described as follows

$$\begin{aligned}\mu_2 &= \eta(-0.1479\alpha_1 - 0.1790\alpha_2) - 0.4255\beta_1 + 0.0275\beta_2 + 0.4287\beta_3, \\ \mu_3 &= \eta(-0.1479\alpha_1 - 0.1790\alpha_2) - 0.4255\beta_1 + 0.5655\beta_2 + 0.4287\beta_3.\end{aligned}$$

The “best” submodels chosen according to each criteria and the corresponding estimated focus parameters are presented in Table 3. For simplicity, we label the submodels with “1”, “12”, . . . , “123456” indicating the inclusion of variables in the linear term of the PLSIM, namely, abstinence time, $\log(\text{MBP})$, Race, HiFever, FSmoker and CSmoker. For example, the submodel “1234” contains $\{Z_1, Z_2, X_1, X_2, X_3, X_4\}$. It turns out that AIC yields the submodel with $\{Z_1, Z_2, X_1, X_3\}$. On the other hand, BIC selects the simplest model $\{Z_1, Z_2, X_1\}$ excluding all uncertain variables. As expected, the submodels selected by FIC depend on the focus parameters: $\text{FIC}(\mu_1)$ chooses the submodel with $\{Z_1, Z_2, X_1, X_2, X_3\}$; $\text{FIC}(\mu_2)$ selects $\{Z_1, Z_2, X_1, X_3, X_4\}$; $\text{FIC}(\mu_3)$ selects $\{Z_1, Z_2, X_1\}$.

As to the “best” models, except for $\text{FIC}(\mu_1)$, information criteria, including AIC, BIC, $\text{FIC}(\mu_2)$ and $\text{FIC}(\mu_3)$, prefer the submodels that exclude $\log(\text{MBP})$. On the other hand, we observed that the information criteria values are very close for some submodels. For example, $\text{FIC}(\mu_3)$ values for submodels “1” and “14” differ by 0.001. If we focused on the “best” model chosen by information criteria, we may ignore the information from the remaining submodels. Since one purpose of this study is to identify the relationship between sperm concentration and urinary MBP concentrations, observing the change of sperm concentration under different MBP concentrations is one way to explore the influence of MBP. Therefore we applied the model averaging procedure to μ_2 and μ_3 , the predicted sperm concentration when $\log(\text{MBP})$ takes the median and third quantile respectively. The model average estimates of μ_2 and μ_3 are 0.0139 and -0.0068 respectively; their corresponding confidence intervals are $(-0.0292, 0.0800)$ and $(-0.903, 0.0683)$. For the

models chosen from variable selection, $\log(\text{MBP})$ isn't included in the models chosen by AIC, BIC, $\text{FIC}(\mu_2)$ and $\text{FIC}(\mu_3)$. This conclusion is consistent with results from the full model of the linear regression model and the PLSIM.

6 Discussion

Despite the fact that a small change in a data set may result in the selection of very different final models, the same data set is usually used for model selection and parameter estimation in real data analysis. As a result, estimators and explanations of models' coefficients based on model selection procedure often have large variability. This paper proposed the focus information criterion and model averaging procedures over candidate models for PLSIM. One merit of model average estimators is stability. Consequently, the risk of selecting a poor model is reduced. The method has been shown to have good theoretical properties, and our simulation studies also show the superiority of the methods in various settings. In addition, increasing sample size can improve the accuracy of the estimators with decreased mean squared errors.

In this paper, we studied only the case of a relatively small number of covariates. The procedure can be applied to PLSIM with a very large number of covariates. However, such an application is not trivial due to the computational burden with an exponential growth in the number of submodels. When we apply PLSIM to data with a very large number of uncertain covariates, variable screening is desirable to remove clearly irrelevant variables before the model averaging procedure is applied. Other potential extensions may be to the generalized partially linear single index models and robust regression. Further effort is needed to develop similar procedures and theoretical results under these settings.

Appendix: Proofs of Theorems

A.1 Assumptions

- (i) The function $\eta(\cdot)$ is differentiable and not constant on the support $\mathcal{A}$ of $Z^T \mathbf{a}$. (ii) The function $\eta(z^T \mathbf{a})$ and the density function of $Z^T \mathbf{a}, f_{\mathbf{a}}(z)$, are three times continuously differentiable with respect to z . The third derivatives are uniformly Lipschitz continuous over $\mathcal{U} \subset \mathbb{R}^p$ for all $u \in \{u = z^T \mathbf{a} : \mathbf{a} \in \mathcal{A}, z \in \mathcal{Z} \subset \mathbb{R}^p\}$. (iii) $E(|Y|^{m_1}) < \infty$ for some $m_1 \geq 3$. The conditional variance of Y given (X, Z) is bounded and bounded away from 0. (iv) The kernel function $K(\cdot)$ is twice continuously differentiable on the support $(-1, 1)$. In addition, its second derivative is Lipschitz continuous. Moreover, $\int u^j K(u) du = 1$ if $j = 0$ and 0 if $j = 1, 2$. (v) The bandwidth h satisfies $\left\{nh^{3+3/(m_1-1)}\right\} \frac{1}{\log n} \rightarrow \infty$ and $nh^8 \rightarrow 0$ as $n \rightarrow \infty$. (vi) The Moore-Penrose inverse exists for $E\left(\tilde{X}^{\otimes 2}\right)$ and $E\left\{\tilde{Z}\eta'(\Lambda)\right\}^{\otimes 2}$, where η' is the first derivative of η .

A.2 Proof of Theorem 1

Recall the definition of profile least squares. It follows that

$$0 = \left\{ \frac{\partial Q(\hat{\alpha}_s, \hat{\beta}_s)}{\partial \alpha_s}, \frac{\partial Q(\hat{\alpha}_s, \hat{\beta}_s)}{\partial \beta_s} \right\} = l(\hat{\alpha}_s, \hat{\beta}_s), \quad (\text{A.1})$$

where

$$l(\hat{\alpha}_s, \hat{\beta}_s) = \frac{1}{n} \sum_{i=1}^n \left\{ Y_i - \hat{\eta}(\Lambda_{i,s}, \hat{\zeta}_s) - X_{i,s}^T \hat{\beta} \right\} \left\{ \frac{\frac{\partial \hat{\eta}(\Lambda_{i,s}, \hat{\zeta}_s)}{\partial \alpha_s}}{\frac{\partial \hat{\eta}(\Lambda_{i,s}, \hat{\zeta}_s)}{\partial \beta_s} + X_{i,s}} \right\}$$

with $\hat{\Lambda}_{i,s} = Z_{i,s}^T \hat{\alpha}_s$, and $\hat{\zeta}_s = (\hat{\alpha}_s^T, \hat{\beta}_s^T)^T$ being the profile least squares estimator of $\zeta_s = (\alpha_s^T, \beta_s^T)^T$. Applying the Taylor expansion, we obtain

$$\begin{aligned} l(\hat{\alpha}_s, \hat{\beta}_s) &= \frac{1}{n} \sum_{i=1}^n \left\{ Y_i - \hat{\eta}(\Lambda_{i,s}, \Pi \zeta_0) - X_{i,s}^T \Pi \beta_0 \right\} \left\{ \frac{\frac{\partial \hat{\eta}(\Lambda_{i,s}, \Pi \zeta_0)}{\partial \alpha_s}}{\frac{\partial \hat{\eta}(\Lambda_{i,s}, \Pi \zeta_0)}{\partial \beta_s} + X_{i,s}} \right\} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\frac{\partial \hat{\eta}(\bar{\Lambda}_{s,i}, \bar{\zeta}_s)}{\partial \alpha_s}}{\frac{\partial \hat{\eta}(\bar{\Lambda}_{s,i}, \bar{\zeta}_s)}{\partial \beta_s} + X_{i,s}} \right\}^{\otimes 2} (\hat{\zeta}_s - \Pi \zeta_0) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \left\{ Y_i - \hat{\eta}(\bar{\Lambda}_{s,i}, \bar{\zeta}_s) - X_{i,s}^T \bar{\beta}_s \right\} \frac{\partial^2 \hat{\eta}(\bar{\Lambda}_{s,i}, \bar{\zeta}_s)}{\partial \zeta_s \partial \zeta_s^T} (\hat{\zeta}_s - \Pi \zeta_0), \end{aligned}$$

where $\bar{\beta}_s$, $\bar{\Lambda}_{s,i}$ and $\bar{\zeta}_s$ are the interior points between β_0 and $\hat{\beta}_s$, $\Lambda_{0,i}$ and $\hat{\Lambda}_{s,i}$, $\Pi \zeta_0$ and $\hat{\zeta}_s$, respectively. Following a similar argument to those used by Liang et al. (2010), we have that

$$l(\hat{\alpha}_s, \hat{\beta}_s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{\Pi_\alpha \tilde{Z}_i \eta'(\Lambda_i)}{\Pi_\beta \tilde{X}_i} \right\} \varepsilon_i - \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\Pi_\alpha \tilde{Z}_i \eta'(\Lambda_i)}{\Pi_\beta \tilde{X}_i} \right\}^{\otimes 2} \sqrt{n} \begin{pmatrix} \hat{\alpha}_s - \Pi_\alpha \alpha_0 \\ \hat{\beta}_s - \Pi_\beta \beta_0 \end{pmatrix} + o_p(1).$$

Moreover, the summand of the matrix over n in the second term of above equation

converges to $\Pi \begin{pmatrix} \Gamma_{\tilde{Z}\tilde{Z}} & \Gamma_{\tilde{X}\tilde{Z}} \\ \Gamma_{\tilde{Z}\tilde{X}} & \Gamma_{\tilde{X}\tilde{X}} \end{pmatrix} \Pi^T$. These results together with (A.1) lead to

$$\begin{aligned} &\Pi \begin{pmatrix} \Gamma_{\tilde{Z}\tilde{Z}} & \Gamma_{\tilde{X}\tilde{Z}} \\ \Gamma_{\tilde{Z}\tilde{X}} & \Gamma_{\tilde{X}\tilde{X}} \end{pmatrix} \Pi^T n^{1/2} \begin{pmatrix} \hat{\alpha}_s - \Pi_\alpha \alpha_0 \\ \hat{\beta}_s - \Pi_\beta \beta_0 \end{pmatrix} \\ &= \Pi \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{\tilde{Z}_i \eta'(\Lambda_i)}{\tilde{X}_i} \varepsilon_i \right\} + o_p(1). \end{aligned} \quad (\text{A.2})$$

Meanwhile, we have

$$n^{1/2} \begin{pmatrix} \hat{\alpha}_s - \Pi_{\alpha} \alpha_0 \\ \hat{\beta}_s - \Pi_{\beta} \beta_0 \end{pmatrix} = n^{1/2} \begin{pmatrix} \hat{\alpha}_s - (\alpha_{c,0}^T, 0)^T \\ \hat{\beta}_s - (\beta_{c,0}^T, 0)^T \end{pmatrix} + \Pi \delta_{\alpha\beta}. \quad (\text{A.3})$$

Equations (A.2) and (A.3), together with the central limit theorem yield Theorem 1, which completes the proof.

A.3 Proof of Theorem 2

Write $\mu(\alpha_c, \alpha_u, \beta_c, \beta_u) = \mu$, $\hat{\mu}_s = \mu \left([I, 0] \Pi_{\alpha}^T \hat{\alpha}_s, [0, I] \Pi_{\alpha}^T \hat{\alpha}_s, [I, 0] \Pi_{\beta}^T \hat{\beta}_s, [0, I] \Pi_{\beta}^T \hat{\beta}_s \right)$

$\mu_{\zeta} = \left(\frac{\partial \mu}{\partial \alpha_c}^T, \frac{\partial \mu}{\partial \alpha_u}^T, \frac{\partial \mu}{\partial \beta_c}^T, \frac{\partial \mu}{\partial \beta_u}^T \right) \Big|_{\alpha_c = \alpha_{c,0}, \alpha_u = 0, \beta_c = \beta_{c,0}, \beta_u = 0}$. Applying the Taylor expansion, we obtain

$$\mu_0 = \mu(\alpha_{c,0}, \delta_{\alpha}/\sqrt{n}, \beta_{c,0}, \delta_{\beta}/\sqrt{n}) = \mu(\alpha_{c,0}, 0, \beta_{c,0}, 0) + \mu_{\zeta} \frac{1}{\sqrt{n}} \delta_{\alpha\beta} + o_p(n^{-1/2}),$$

$$\hat{\mu}_s = \mu(\alpha_{c,0}, 0, \beta_{c,0}, 0) + \mu_{\zeta}^T \Pi^T \begin{pmatrix} \hat{\alpha}_s - (\alpha_{c,0}^T, 0)^T \\ \hat{\beta}_s - (\beta_{c,0}^T, 0)^T \end{pmatrix} + o_p(n^{-1/2}),$$

$$\hat{\mu}_s - \mu_0 = \mu_{\zeta}^T \Pi^T \begin{pmatrix} \hat{\alpha}_s - (\alpha_{c,0}^T, 0)^T \\ \hat{\beta}_s - (\beta_{c,0}^T, 0)^T \end{pmatrix} - \mu_{\zeta}^T \frac{1}{\sqrt{n}} \delta_{\alpha\beta} + o_p(n^{-1/2}).$$

Recalling the definition of $\mathbf{R}$, by Theorem 1, we can obtain

$$\sqrt{n}(\hat{\mu} - \mu_0) = \mu_{\zeta}^T \mathbf{R} \mathbf{G}_n + \mu_{\zeta}^T \mathbf{R} \mathbf{\Omega} \delta_{\alpha\beta} - \mu_{\zeta}^T \delta_{\alpha\beta} + o_p(1) \xrightarrow{d} \mu_{\zeta}^T \mathbf{R} \mathbf{G} + \mu_{\zeta}^T \mathbf{R} \mathbf{\Omega} \delta_{\alpha\beta} - \mu_{\zeta}^T \delta_{\alpha\beta}.$$

The proof is complete.

A.4 Proof of Theorem 3

Let $\mathbf{v}_1$ be the block diagonal matrix with blocks defined by $(I_{d_{\alpha,c} \times d_{\alpha,c}}, 0, I_{d_{\beta,c} \times d_{\beta,c}}, 0)$ and $\mathbf{v}_2$ be the block diagonal matrix with blocks defined by $(0, I_{d_{\alpha,u} \times d_{\alpha,u}}, 0, I_{d_{\beta,u} \times d_{\beta,u}})$

$$\begin{aligned}
\sqrt{n}(\hat{\mu} - \mu_0) &= \sum_s w(s) \sqrt{n}(\hat{\mu}_s - \mu_0) \\
&= \sum_s w(s) \left\{ \mu_\zeta^T \mathbf{R} \mathbf{G}_n + \mu_\zeta^T \mathbf{R} \mathbf{\Omega} \boldsymbol{\delta}_{\alpha\beta} - \mu_\zeta^T \boldsymbol{\delta}_{\alpha\beta} + \mathbf{o}_p(1) \right\} \\
&= \mu_\zeta^T \sum_s w(s) \mathbf{R} \mathbf{\Omega} \left\{ \mathbf{\Omega}^- \mathbf{G}_n + \boldsymbol{\delta}_{\alpha\beta} \right\} - \mu_\zeta^T \boldsymbol{\delta}_{\alpha\beta} + \mathbf{o}_p(1) \\
&= \mu_\zeta^T \sum_s w(s) \mathbf{R} \mathbf{\Omega} \begin{pmatrix} [\mathbf{I}, 0, 0, 0] \mathbf{\Omega}^- \mathbf{G}_n \\ [0, \mathbf{I}, 0, 0] \mathbf{\Omega}^- \mathbf{G}_n + \boldsymbol{\delta}_\alpha \\ [0, 0, \mathbf{I}, 0] \mathbf{\Omega}^- \mathbf{G}_n \\ [0, 0, 0, \mathbf{I}] \mathbf{\Omega}^- \mathbf{G}_n + \boldsymbol{\delta}_\beta \end{pmatrix} - \mu_\zeta^T \boldsymbol{\delta}_{\alpha\beta} + \mathbf{o}_p(1).
\end{aligned}$$

By Theorem 1, we obtain

$$\begin{aligned}
\sqrt{n}(\hat{\mu} - \mu_0) &= \mu_\zeta^T \sum_s w(s) \mathbf{v}_1 \mathbf{\Omega}^- \mathbf{G}_n - \mu_\zeta^T \sum_s w(s) \mathbf{R} \mathbf{\Omega} \hat{\boldsymbol{\delta}}_{\alpha\beta} - \mu_\zeta^T \hat{\boldsymbol{\delta}}_{\alpha\beta} + \mu_\zeta^T \mathbf{v}_2 \mathbf{\Omega}^- \mathbf{G}_n + \mathbf{o}_p(1) \\
&= \mu_\zeta^T \mathbf{\Omega}^- \mathbf{G}_n - \mu_\zeta^T \mathbf{Q}(s) \hat{\boldsymbol{\delta}}_{\alpha\beta} - \mu_\zeta^T \hat{\boldsymbol{\delta}}_{\alpha\beta} + \mathbf{o}_p(1) \\
&\xrightarrow{d} \mu_\zeta^T \mathbf{\Omega}^- \mathbf{G} - \mu_\zeta^T \{ \mathbf{Q}(s) \boldsymbol{\Delta}_{\alpha\beta} + \boldsymbol{\Delta}_{\alpha\beta} \},
\end{aligned}$$

where $\boldsymbol{\delta}_{\alpha\beta} = [0, \mathbf{I}, 0, \mathbf{I}] \mathbf{\Omega}^- \mathbf{G} + \boldsymbol{\delta}_{\alpha\beta}$.

References

- Akaike, H. Second International Symposium on Information Theory (Tsahkadsor, 1971). Akadémiai Kiadó; Budapest: 1973. Information theory and an extension of the maximum likelihood principle; p. 267-281.
- Borch J, Axelstad M, Vinggaard A, Dalgaard M. Diisobutyl phthalate has comparable anti-androgenic effects to di-n-butyl phthalate in fetal rat testis. *Toxicology Letters*. 2006; 163:183–190. [PubMed: 16458459]
- Buckland S, Burnham K, Augustin N. Model selection: An integral part of inference. *Biometrics*. 1997; 53:603–618.
- Carroll R, Fan J, Gijbels I, Wand MP. Generalized partially linear single-index models. *Journal of the American Statistical Association*. 1997; 92:477–489.
- Claeskens G, Carroll RJ. An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika*. 2007; 94:249–265.
- Claeskens, G.; Hjort, NL. *Model Selection and Model Averaging*. Cambridge University Press; Cambridge: 2008.
- Cui X, Häardle W, Zhu LX. The EFM approach for single-index models. *The Annals of Statistics*. 2011; 39:1658–1688.
- Danilov D, Magnus JR. On the harm that ignoring pretesting can cause. *Journal of Econometrics*. 2004; 122:27–46.
- Draper D. Assessment and propagation of model uncertainty, *Journal of the Royal Statistical Society. Series B*. 1995; 57:45–70.
- Duty S, Calafat A, Silva M, Ryan L, Hauser R. Phthalate exposure and reproductive hormones in adult men. *Human Reproduction*. 2005; 20:604–610. [PubMed: 15591081]
- Duty SM, Silva MJ, Barr DB, Brock JW, Ryan L, Chen ZY, Herrick RF, Christiani DC, Hauser R. Phthalate exposure and human semen parameters. *Epidemiology*. 2003; 14:269–277. [PubMed: 12859026]
- Fan, J.; Gijbels, I. Vol. 66. Chapman & Hall; London: 1996. *Local Polynomial Modelling and Its Applications*.
- Ge R, Chen G, Tanrikut C, Hardy M. Phthalate ester toxicity in Leydig cells: Developmental timing and dosage considerations. *Reproductive Toxicology*. 2007; 23:366–373. [PubMed: 17258888]

- Hauser R, Meeker JD, Duty S, Silva MJ, Calafat AM. Altered semen quality in relation to urinary concentrations of phthalate monoester and oxidative metabolites. *Epidemiology*. 2006; 17:682–691. [PubMed: 17003688]
- Hjort N, Claeskens G. Frequentist model average estimators. *Journal of the American Statistical Association*. 2003; 98:879–899.
- Jonsson B, Richthoff J, Rylander L, Giwercman A, Hagmar L. Urinary phthalate metabolites and biomarkers of reproductive function in young men. *Epidemiology*. 2005; 16:487–493. [PubMed: 15951666]
- Jurewicz J, Hanke W, Radwan M, Bonde J. Environmental Factors and Semen Quality. *International Journal of Occupational Medicine and Environmental Health*. 2009; 22:305–329. [PubMed: 20053623]
- Leeb H, Pötscher BM. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*. 2006; 34:2554–2591.
- Leung G, Barron AR. Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory*. 2006; 52:3396–3410.
- Liang H, Liu X, Li R, Tsai C-L. Estimation and testing for partially linear single-index models. *The Annals of Statistics*. 2010; 38:3811–3836.
- Liang H, Thurston SW, Ruppert D, Apanasovich T, Hauser R. Additive partial linear models with measurement errors. *Biometrika*. 2008; 95:667–678.
- Meeker J, Calafat A, Hauser R. Urinary metabolites of di (2-ethylhexyl) phthalate are associated with decreased steroid hormone levels in adult men. *Journal of andrology*. 2009; 30:287. [PubMed: 19059903]
- Mendiola J, Jørgensen N, Andersson A, Calafat A, Silva M, Redmon J, Sparks A, Drobnis E, Wang C, Liu F, et al. Associations between urinary metabolites of di (2-ethylhexyl) phthalate and reproductive hormones in fertile men. *International Journal of Andrology*. 2011; 34:369–378. [PubMed: 20633195]
- Mendiola J, Meeker J, Jorgensen N, Andersson A, Liu F, Calafat A, Redmon J, Drobnis E, Sparks A, Wang C, et al. Urinary concentrations of di (2-ethylhexyl) phthalate metabolites and serum reproductive hormones: Pooled analysis of fertile and infertile men. *Journal of Andrology*. 2012; 33:488–498. [PubMed: 21597090]
- Pant N, Shukla M, Patel DK, Shukla Y, Mathur N, Gupta YK, Saxena DK. Correlation of phthalate exposures with semen quality. *Toxicology and Applied Pharmacology*. 2008; 231:112–116. [PubMed: 18513777]
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6:461–464.
- Spiegelhalter SD, Best NG, Carlin BP, Linde AVD. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*. 2002; 64:583–639.
- Wirth JJ, Rossano MG, Potter R, Puscheck E, Daly DC, Paneth N, Krawetz SA, Protas BM, Diamond MP. A pilot study associating urinary concentrations of phthalate metabolites and semen quality. *Systems Biology in Reproductive Medicine*. 2008; 54:143–154. [PubMed: 18570050]
- Xia YC, Härdle W. Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*. 2006; 97:1162–1184.
- Xia Y, Tong H, Li WK. On extended partially linear single-index models. *Biometrika*. 1999; 86:831–842.
- Yang Y. Adaptive regression by mixing. *Journal of the American Statistical Association*. 2001; 96:574–588.
- Yu Y, Ruppert D. Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*. 2002; 97:1042–1054.
- Zhang X, Liang H. Focused information criterion and model averaging for generalized additive partial linear models. *The Annals of Statistics*. 2011; 39:174–200.
- Zhang Y, Zheng L, Chen B. Phthalate exposure and human semen quality in Shanghai: A cross-sectional study. *Biomedical and Environmental Sciences*. 2006; 19:205–209. [PubMed: 16944777]

Figure 1.

The estimated curve of $\hat{\eta}(\cdot)$ against estimated single-index $0.4357\text{BMI} + 0.9001\text{Age}$ and the associated 95% pointwise confidence intervals (dashed lines).

Table 1

The coverage probability (CP) corresponding to a claimed 95 % confidence interval and mean squared errors (MSE) ($\times 10^{-2}$) for focus parameter estimates (μ_5 is for setting (i); μ_5^* is for setting (ii)) from the full model and from models selected by AIC, BIC, FIC, S-FIC, S-AIC and S-BIC using 1000 replications.

Setting	n	Method	μ_1			μ_2			μ_3			μ_4			$\mu_5(\mu_5^*)$		
			CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	
i	200	FULL	89.1	1.684	90.5	2.695	88.6	3.797	90.2	4.722	89.5	7.696					
		AIC	86.4	10.632	77.5	12.033	79.4	12.415	74.6	13.601	82.9	7.259					
		BIC	87.4	9.625	64.9	11.311	83.9	10.437	64.3	11.811	81.4	6.465					
		FIC	95.9	2.845	94.7	3.462	90.7	5.024	91.5	5.266	88.8	7.040					
		S-FIC	95.4	1.271	94.1	2.196	91.7	2.012	91.3	2.716	91.1	5.032					
		S-AIC	95.2	1.504	93.5	2.476	90.7	1.745	91.2	2.457	91.1	4.465					
	300	S-BIC	95.2	1.502	93.5	2.475	90.7	1.741	91.2	2.454	91.1	4.460					
		FULL	91.6	1.033	91.9	1.611	90.3	2.593	90.9	3.095	88.9	6.208					
		AIC	89.6	3.104	78.5	3.937	82.6	5.774	76.8	6.548	82.5	5.841					
		BIC	89.8	3.079	69.0	4.061	87.0	4.980	69.0	5.924	84.1	4.642					
		FIC	95.3	0.913	94.7	1.176	92.1	2.535	92.8	2.338	88.7	5.677					
		S-FIC	95.4	0.684	95.0	1.206	93.1	1.308	93.5	1.651	92.9	3.824					
ii	200	S-AIC	95.1	0.756	95.0	1.269	93.1	1.092	93.5	1.519	92.4	3.306					
		S-BIC	95.1	0.755	95.0	1.269	93.1	1.089	93.5	1.517	92.4	3.302					
		FULL	85.5	5.443	86.0	6.147	81.2	8.816	81.7	9.511	82.9	10.372					
		AIC	85.2	10.419	80.6	11.798	82.7	13.308	79.2	14.305	75.8	9.619					
		BIC	85.0	10.219	76.6	12.102	83.3	12.408	76.3	13.802	74.5	8.844					
		FIC	91.6	3.534	91.2	4.448	87.3	7.502	85.5	9.804	82.0	10.697					
	300	S-FIC	93.7	3.071	93.8	3.932	89.5	3.838	90.0	4.039	86.2	6.638					
		S-AIC	93.2	3.301	93.6	4.279	89.3	3.395	89.1	3.529	85.6	5.592					
		S-BIC	93.2	3.297	93.6	4.277	89.3	3.388	89.1	3.524	85.5	5.582					
		FULL	89.0	3.314	88.7	3.918	86.2	5.244	88.1	5.667	84.0	7.578					
		AIC	87.6	5.592	83.1	6.502	85.1	7.466	82.5	8.177	78.7	6.805					
		BIC	88.0	4.794	80.8	5.804	85.5	5.751	80.9	6.534	78.9	6.044					

Setting	n	Method	μ_1		μ_2		μ_3		μ_4		$\mu_5(\ast)$	
			CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE	CP(%)	MSE
		FIC	92.2	2.566	92.0	3.157	89.0	3.800	89.7	4.467	84.3	6.545
		S-FIC	93.6	2.335	92.4	2.858	92.1	2.751	91.9	2.949	90.3	4.449
		S-AIC	93.5	2.442	92.0	3.033	92.3	2.620	92.0	2.808	89.9	3.871
		S-BIC	93.5	2.441	92.0	3.032	92.3	2.617	92.0	2.806	89.9	3.866

Table 2

Results for the semen quality data analysis. Left panel: Estimated coefficient (Esti), standard error (Std), t -value and p -value using the linear model. Right panel: Estimated coefficient, standard error, z -value and p -value using the PLSIM.

	Linear model				PLSIM			
	Esti	Std	z value	Pr(> z)	Esti	Std	z value	Pr(> z)
(Intercept)	-0.0000	0.0461	-0.00	1.0000				
BMI	-0.0925	0.0465	-1.99	0.0474				
Age	-0.0610	0.0471	-1.29	0.1964				
Abstinence time	0.1514	0.0464	3.26	0.0012	0.1497	0.0448	3.3395	0.0008
Log(MBP)	-0.0562	0.0474	-1.19	0.2365	-0.0662	0.0459	-1.4440	0.1487
Race	0.0978	0.0480	2.04	0.0421	0.0732	0.0463	1.5801	0.1141
HiFever	-0.0351	0.0470	-0.75	0.4550	-0.0273	0.0453	-0.6037	0.5460
FSmoker	-0.0611	0.0476	-1.28	0.1996	-0.0300	0.0461	-0.6512	0.5149
Csmoker	-0.0288	0.0471	-0.61	0.5413	-0.0333	0.0458	-0.7266	0.4675

Table 3

Selected submodels via AIC, BIC, FIC and corresponding estimated focus parameters for the semen quality data analysis.

Submodel	1	12	13	14	15	16	123	124
AIC	-0.0922	-0.0940	-0.0984	-0.0900	-0.0882	-0.0888	-0.0976	-0.0917
BIC	-0.0648	-0.0575	-0.0618	-0.0535	-0.0516	-0.0523	-0.0519	-0.0461
FIC(μ_1)	0.9080	0.9031	0.9020	0.9045	0.9069	0.9052	0.8994	0.9008
μ_1	0.1572	0.1531	0.1536	0.1540	0.1569	0.1558	0.1533	0.1542
FIC(μ_2)	0.4361	0.4889	0.3363	0.4254	0.4217	0.4202	0.3445	0.4780
μ_2	-0.0070	-0.0054	0.0360	-0.0061	-0.0079	-0.0089	0.0302	-0.0062
FIC(μ_3)	0.1627	0.9577	0.7651	0.1627	0.1631	0.1630	0.7281	0.9496
μ_3	-0.0070	-0.0503	0.0360	-0.0061	-0.0079	-0.0089	-0.0041	-0.0526

Submodel	125	126	134	135	136	145	146	156
AIC	-0.0904	-0.0903	-0.0947	-0.0945	-0.0947	-0.0861	-0.0865	-0.0851
BIC	-0.0447	-0.0446	-0.0490	-0.0489	-0.0490	-0.0404	-0.0409	-0.0394
FIC(μ_1)	0.9030	0.9021	0.9012	0.9027	0.9013	0.9042	0.9027	0.9059
μ_1	0.1531	0.1515	0.1543	0.1538	0.1519	0.1537	0.1525	0.1550
FIC(μ_2)	0.4735	0.4732	0.3352	0.3361	0.3374	0.4107	0.4089	0.4003
μ_2	-0.0068	-0.0071	0.0339	0.0345	0.0337	-0.0071	-0.0079	-0.0102
FIC(μ_3)	0.9409	0.9161	0.7179	0.7639	0.7714	0.1631	0.1631	0.1636
μ_3	-0.0511	-0.0510	0.0339	0.0345	0.0337	-0.0071	-0.0079	-0.0102

Submodel	1234	1235	1236	1245	1246	1256	1345	1346
AIC	-0.0941	-0.0937	-0.0938	-0.0879	-0.0883	-0.0870	-0.0908	-0.0910
BIC	-0.0393	-0.0389	-0.0391	-0.0331	-0.0335	-0.0322	-0.0360	-0.0362
FIC(μ_1)	0.8998	0.9012	0.9009	0.9018	0.9010	0.9041	0.9025	0.9015
μ_1	0.1519	0.1540	0.1519	0.1536	0.1535	0.1507	0.1514	0.1536
FIC(μ_2)	0.3486	0.3445	0.3447	0.4623	0.4618	0.4525	0.3353	0.3366
μ_2	0.0291	0.0283	0.0283	-0.0070	-0.0085	-0.0089	0.0342	0.0312
FIC(μ_3)	0.7358	0.7282	0.7290	0.9345	0.9067	0.8910	0.7184	0.7268
μ_3	-0.0073	-0.0064	-0.0055	-0.0531	-0.0547	-0.0526	0.0342	0.0312

Submodel	1356	1456	12345	12346	12456	12356	13456	123456
AIC	-0.0912	-0.0831	-0.0904	-0.0903	-0.0843	-0.0902	-0.0874	-0.0868
BIC	-0.0364	-0.0283	-0.0265	-0.0264	-0.0204	-0.0262	-0.0235	-0.0138
FIC(μ_1)	0.9039	0.9046	0.9020	0.9016	0.9047	0.9039	0.9047	0.9042
μ_1	0.1515	0.1520	0.1515	0.1509	0.1518	0.1523	0.1498	0.1497
FIC(μ_2)	0.3384	0.3890	0.3486	0.3488	0.4393	0.3447	0.3376	0.3487
μ_2	0.0312	-0.0095	0.0267	0.0266	-0.0087	0.0256	0.0310	0.0239
FIC(μ_3)	0.7762	0.1636	0.7359	0.7368	0.8802	0.7292	0.7326	0.7367

Submodel	1356	1456	12345	12346	12456	12356	13456	123456
μ_3	0.0312	-0.0095	-0.0098	-0.0095	-0.0539	-0.0083	0.0310	-0.0117