Corrected Local Polynomial Estimation in Varying-Coefficient Models with Measurement Errors

Author(s): Jinhong You, Yong Zhou and Gemai Chen

Source: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 34, No. 3 (Sep., 2006), pp. 391-410
Published by: Statistical Society of Canada
Stable URL: http://www.jstor.org/stable/20445211
Accessed: 05/12/2014 15:54

The Canadian Journal of Statistics
Vol. 34, No. 3, 2006, Pages 391–410
La revue canadienne de statistique

Corrected local polynomial estimation in varying-coefficient models with measurement errors

Jinhong YOU, Yong ZHOU and Gemai CHEN

Key words and phrases: Asymptotic normality; attenuation; local linear estimator; measurement error; regression model; varying coefficient; wild bootstrap.

MSC 2000: Primary 62H12; secondary 62A10.

Abstract: The authors study a varying-coefficient regression model in which some of the covariates are measured with additive errors. They find that the usual local linear estimator (LLE) of the coefficient functions is biased and that the usual correction for attenuation fails to work. They propose a corrected LLE and show that it is consistent and asymptotically normal, and they also construct a consistent estimator for the model error variance. They then extend the generalized likelihood technique to develop a goodness of fit test for the model. They evaluate these various procedures through simulation studies and use them to analyze data from the Framingham Heart Study.

Estimation polynomiale locale corrigée dans les modèles à coefficients variables comportant des erreurs de mesure

Résumé : Les auteurs s'intéressent à un modèle de régression à coefficients variables dont certaines covariables sont entachées d'erreurs additives. Ils montrent que l'estimateur localement linéaire (ELL) usuel des coefficients fonctionnels est biaisé et que le facteur de correction habituel du phénomène d'atténuation est inefficace. Ils proposent une version corrigée de l'ELL qui s'avère convergente et asymptotiquement normale ; ils suggèrent aussi une estimation convergente de la variance du terme d'erreur du modèle. Une adaptation de la technique de vraisemblance généralisée leur permet en outre d'élaborer un test d'adéquation du modèle. Ils évaluent ces diverses procédures par voie de simulation et s'en servent pour analyser des données issues de l'étude Framingham sur les risques cardiométaboliques.

1. INTRODUCTION

Parametric regression models are very useful for providing a parsimonious description of the relationship between a response variable and its covariates. However, they are used at the risk of introducing modelling biases. Nonparametric regression models, on the other hand, do not assume prior information on the model structure, but an entire nonparametric approach is hampered by some serious drawbacks, such as the curse of dimensionality, difficulty of interpretation, and lack of extrapolation capability. To avoid these drawbacks several modelling approaches have been proposed, such as additive modelling (Hastie & Tibshirani 1990), low-dimensional interaction modelling (Friedman 1991, Gu & Wahba 1993), multiple-index modelling (Härdle & Stoker 1989), partially linear modelling (Härdle, Liang & Gao 2000) and some hybrids (Carroll, Fan, Gijbels & Wand 1997). An important addition to these approaches is varying-coefficient modelling (Cleveland, Gross & Shyu 1991 and Hastie & Tibshirani 1993). If one replaces the coefficients of a linear regression model with smooth nonparametric functions and allows these functions to take other covariates as arguments, one gets a varying-coefficient regression model formally written as

$$Y_i = X_{i1}\alpha_1(U_i) + \cdots + X_{ip}\alpha_p(U_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where Y_i are the responses, $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^\top$ and U_i are covariates, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_p(\cdot))^\top$ is a p-dimensional vector of unknown functions, ε_i are independent errors with mean zero and variance σ^2 and the superscript "$^\top$" denotes the transpose of a vector or matrix. The varying-coefficient $\alpha(\cdot)$ has been estimated using penalized least squares (Wahba 1984 and Chiang, Rice & Wu 2001), kernel smoothing (Wu, Chiang & Hoover 1998 and Wu, Yu & Chiang 2000), local polynomial (Fan & Zhang 1999 and Fan & Huang 2005), series approximation (Huang, Wu & Zhou 2002 and Huang, Wu & Zhou 2004), and so on.

In some applications the covariates are measured with errors. For example, serum cholesterol level (Carroll, Ruppert & Stefanski 1995), urinary sodium chloride level (Wang, Carroll & Liang 1996), and exposure to pollutants (Tosteson, Stefanski & Schafer 1989) are often subject to measurement errors. It is well known that measurement errors can cause difficulties and complexities in statistical analysis. For example, measurement errors result in the usual regression estimators to be biased in the direction of 0. Bias of this nature is called *attenuation* or *attenuation to the null* (Carroll, Ruppert & Stefanski 1995, p. 21). A detailed study of linear regression models with measurement errors was considered in Fuller (1987) and Chen & Van Ness (1999). Carroll, Ruppert & Stefanski (1995) summarized much of the recent work for nonlinear regression models with measurement errors. Fan (1993), Fan & Truong (1993) used the deconvolution method to estimate the nonparametric regression functions when the covariates are measured with errors.

In this paper we extend the study of measurement errors to the varying-coefficient regression model (1). We consider the case where the covariate U_i is free of error and the covariate $\mathbf{X}_i$ is measured with additive errors, with observed covariate $\mathbf{W}_i$ given through

$$\mathbf{W}_i = \mathbf{X}_i + \boldsymbol{\zeta}_i, \quad i = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\zeta}_i$ are the measurement errors which are independent of $(\mathbf{X}_i^\top, U_i, \varepsilon_i)^\top$ and have the same covariance matrix $\operatorname{cov}(\boldsymbol{\zeta}) = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$. Our study is motivated by the Framingham Heart Study which consists of a series of exams on 1,615 men aged from 31 to 65. Of the various variables measured, we are interested in the dependence of blood pressure on serum cholesterol level, smoking status, and age. There are measurement errors in serum cholesterol levels, measured with two replicates. We will analyze this data set in Section 6.

For the model (1)-(2), the usual local linear estimator (LLE) of the coefficient functions is no longer consistent. By taking the measurement errors into account we propose a corrected local linear estimator and show that it is consistent and asymptotically normal.

With the varying-coefficient functions estimated consistently, we extend the generalized likelihood technique of Fan, Zhang & Zhang (2001) to our model (1)-(2) and develop a goodness-of-fit test to check whether there exists a parametric structure for the varying-coefficient functions.

The rest of this paper is organized as follows. The corrected local linear estimator is presented in Section 2. Its large sample properties are investigated in Section 3. A consistent estimator for the error variance is constructed in Section 4. The goodness-of-fit test is developed in Section 5. The proposed estimators and test are evaluated and applied to the Framingham Heart Study in Section 6. Some remarks are given in Section 7, and proofs of the main results are relegated to the Appendix.

2. LOCAL LINEAR ESTIMATION AND ITS CORRECTION

In this paper we use the local polynomial method to estimate the coefficient functions $\alpha_j(\cdot)$, although any other method mentioned in the introduction will lead to the same results. In our discussion we assume that the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$ is known. The case of unknown $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$ is discussed in Section 7.

Suppose that $(Y_i, \mathbf{X}_i^\top, U_i)_{i=1}^n$ is a sample from model (1). For U in a small neighbourhood of u, one can approximate $\alpha_j(U)$ locally by a linear function

$$\alpha_j(U) \approx \alpha_j(u) + \alpha_j'(u)(U-u) \equiv a_j + b_j(U-u), \quad j = 1, \ldots, p,$$

where $\alpha_j'(u) = \partial\alpha_j(u)/\partial u$. This leads to the following weighted local least-squares problem:

find $\{(a_j, b_j),\ j = 1, \ldots, p\}$ to minimize

$$\sum_{i=1}^{n}\left[Y_i - \sum_{j=1}^{p}\{a_j + b_j(U_i - u)\}X_{ij}\right]^2 K_h(U_i - u), \tag{3}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$. The solution to problem (3) is given by

$$\left\{\tilde{a}_1(u), \ldots, \tilde{a}_p(u), h\tilde{b}_1(u), \ldots, h\tilde{b}_p(u)\right\}^\top = \left\{(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}}\right\}^{-1}(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \mathbf{Y}, \tag{4}$$

$$\begin{aligned}
\mathbf{D}_u^{\mathbf{X}} &= \begin{pmatrix} \mathbf{X}_1^\top & \frac{U_1 - u}{h}\mathbf{X}_1^\top \\ \vdots & \vdots \\ \mathbf{X}_n^\top & \frac{U_n - u}{h}\mathbf{X}_n^\top \end{pmatrix}, \\
\boldsymbol{\omega}_u &= \operatorname{diag}\left(K_h(U_1 - u), \ldots, K_h(U_n - u)\right) \quad \text{and} \\
\mathbf{Y} &= (Y_1, \ldots, Y_n)^\top.
\end{aligned}$$

The first p components in (4) together is called the local linear estimator.

For our model (1)-(2), we observe $\mathbf{W}_i$ instead of $\mathbf{X}_i$. If we naively replace $\mathbf{X}_i$ with $\mathbf{W}_i$ in (4) directly, we will get an inconsistent estimator due to attenuation. Since $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$ is assumed to be known, we may straightforwardly apply a correction for attenuation as in the classical linear model situation, namely, we replace $\left\{(\mathbf{D}_u^{\mathbf{X}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}}\right\}^{-1}$ with $\left\{(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - n\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} \otimes \mathbf{I}_2\right\}^{-1}$, where $\mathbf{I}_2$ is the 2×2 identity matrix and the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is $(a_{ij}\mathbf{B})$. But it turns out that the corrected estimator is still inconsistent (cannot obtain (A1) in the Appendix). This means that if a correction is going to work at all, we need a new correction for our model.

By Lemma 8.1 in the Appendix we note that

$$\begin{aligned}
&\frac{1}{n}(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T \otimes \begin{pmatrix} 1 & (U_i - u)/h \\ (U_i - u)/h & \{(U_i - u)/h\}^2 \end{pmatrix} K_h(U_i - u) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\Sigma}_{\boldsymbol{\zeta}} \otimes \begin{pmatrix} 1 & (U_i - u)/h \\ (U_i - u)/h & \{(U_i - u)/h\}^2 \end{pmatrix} K_h(U_i - u) + O\left\{\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \quad \text{a.s.},
\end{aligned}$$

where $\mathbf{D}_u^{\mathbf{W}}$ has the same form as $\mathbf{D}_u^{\mathbf{X}}$ except that $\mathbf{X}_i$ are replaced by $\mathbf{W}_i$. Denote

$$\boldsymbol{\Omega} = \sum_{i=1}^{n} \boldsymbol{\Sigma}_{\boldsymbol{\zeta}} \otimes \begin{pmatrix} 1 & (U_i - u)/h \\ (U_i - u)/h & \{(U_i - u)/h\}^2 \end{pmatrix} K_h(U_i - u).$$

We modify (4) to define the following estimator for

$$\boldsymbol{\Psi}(u) = (\alpha_1(u), \ldots, \alpha_p(u), \alpha_1'(u), \ldots, \alpha_p'(u))^\top$$

as

$$\hat{\boldsymbol{\Psi}}(u) = \mathbf{H}^{-1}\left\{(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}\right\}^{-1}(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{Y},$$

where $\mathbf{H} = \operatorname{diag}(1, h) \otimes \mathbf{I}_p$ and $\mathbf{I}_p$ is the $p \times p$ identity matrix. Let $\mathbf{e}_{j,2p}$ be the $2p$-column vector with 1 in the jth position and zeros elsewhere. We can write the estimator of $\alpha_j(u)$ as

$$\hat{\alpha}_j^{\mathbf{W}}(u) = \mathbf{e}_{j,2p}^T \mathbf{H}^{-1}\left\{(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}\right\}^{-1}(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{Y}.$$

We call $\widehat{\boldsymbol{\alpha}}^{\mathbf{W}}(u) = (\hat{\alpha}_1^{\mathbf{W}}(u), \ldots, \hat{\alpha}_p^{\mathbf{W}}(u))^\top$ the corrected local linear estimator.

3. ASYMPTOTIC PROPERTIES OF THE CORRECTED LOCAL LINEAR ESTIMATOR

We make the following standard assumptions.

ASSUMPTION 1. *$(X_i^\top, U_i, \varepsilon_i, \zeta_i^\top)^\top$ are independent and identically distributed random vectors. $(\mathbf{X}_i^\top, U_i)^\top$ and $(\varepsilon_i, \boldsymbol{\zeta}_i^\top)^\top$ are independent, and ε_i and $\boldsymbol{\zeta}_i$ are independent. U_1 has a bounded support $\mathcal{U}$, its density function $f_U(\,\cdot\,)$ is Lipschitz continuous and bounded away from 0 on $\mathcal{U}$.*

ASSUMPTION 2. *The $p \times p$ matrix $\mathbf{\Gamma}(u) = E(\mathbf{X}_1\mathbf{X}_1^\top \mid U_1 = u)$ is nonsingular for each $u \in \mathcal{U}$.*

ASSUMPTION 3. *There is an $s > 2$ such that $E\varepsilon_1^{2s} < \infty$, $E\|\mathbf{X}_1\|^{2s} < \infty$ and $E\|\boldsymbol{\zeta}_1\|^{2s} < \infty$, where $E\|\cdot\|$ is the Euclidean norm, and for some $\delta < 2 - s^{-1}$ there is $n^{2\delta-1}h \to \infty$ as $n \to \infty$, where h is the bandwidth.*

ASSUMPTION 4. *$\{\alpha_j(\,\cdot\,),\ j = 1, \ldots, p\}$ have continuous second derivatives on $\mathcal{U}$.*

ASSUMPTION 5. *The function $K(\,\cdot\,)$ is a density function with compact support and the bandwidth h satisfies $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$ as $n \to \infty$.*

Let
$$\mu_j = \int_{-\infty}^{\infty} u^j K(u)\,du, \quad \nu_j = \int_{-\infty}^{\infty} u^j K^2(u)\,du, \quad j = 1, 2, 3.$$

The following theorem shows that $\hat{\mathbf{\Psi}}(\,\cdot\,)$ is asymptotically normal.

THEOREM 3.1. *Suppose that Assumptions 1 to 5 hold. Then it holds that*
$$\sqrt{nh}\left[\mathbf{H}\left\{\hat{\mathbf{\Psi}}(u) - \mathbf{\Psi}(u)\right\} - \frac{h^2}{2}\frac{1}{\mu_2 - \mu_1^2}\begin{pmatrix}(\mu_2 - \mu_1\mu_3)\boldsymbol{\alpha}''(u)\\(\mu_3 - \mu_1\mu_2)\boldsymbol{\alpha}''(u)\end{pmatrix} + o(h^2)\right] \xrightarrow{D} \mathsf{N}(0, \mathbf{\Sigma})$$
as $n \to \infty$, where $\boldsymbol{\alpha}''(u) = (\alpha_1''(u), \ldots, \alpha_p''(u))^\top$, $\alpha_j''(u) = \partial^2\alpha_j(u)/\partial u^2$, "$\xrightarrow{D}$" denotes convergence in distribution, and
$$\mathbf{\Sigma} = \frac{1}{f_U(u)}\mathbf{\Sigma}^* \otimes \frac{1}{(\mu_2 - \mu_1^2)^2}\begin{pmatrix}\mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2 & (\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2\\(\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 & \nu_2 - \mu_1(2\nu_1 + \mu_1\nu_0)\end{pmatrix}$$
with
$$\mathbf{\Sigma}^* = \mathbf{\Gamma}^{-1}(u)\left[\sigma^2\mathbf{\Gamma}(u) + \sigma^2\mathbf{\Sigma}_{\boldsymbol{\zeta}} + E\{\boldsymbol{\xi}_1\boldsymbol{\alpha}(u)\boldsymbol{\alpha}^\top(u)\boldsymbol{\xi}_1^\top \mid U_1 = u\}\right]\mathbf{\Gamma}^{-1}(u)$$
and
$$\boldsymbol{\xi}_1 = \mathbf{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_1\boldsymbol{\zeta}_1^\top - \mathbf{X}_1\boldsymbol{\zeta}_1^\top.$$

COROLLARY 3.1. *Under the conditions of Theorem 3.1, we have*
$$\sqrt{nh}\left\{\widehat{\boldsymbol{\alpha}}^W(u) - \boldsymbol{\alpha}(u) - \frac{h^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}''(u) + o(h^2)\right\} \xrightarrow{D} \mathsf{N}(0, \mathbf{\Sigma}^W) \quad \text{as } n \to \infty,$$
where $\widehat{\boldsymbol{\alpha}}^W(u) = (\hat{\alpha}_1^W(u), \ldots, \hat{\alpha}_p^W(u))^\top$, $\boldsymbol{\alpha}(u) = (\alpha_1(u), \ldots, \alpha_p(u))^\top$ and
$$\mathbf{\Sigma}^W = \frac{(c_0^2\nu_0 + 2c_0c_1\nu_1 + c_1^2\nu_2)}{f_U(u)}\mathbf{\Sigma}^*,$$
with $c_0 = \mu_2/(\mu_2 - \mu_1^2)$ and $c_1 = -\mu_1/(\mu_2 - \mu_1^2)$.

Remark 3.1. Compared with the results in Cai, Fan & Yao (2000), our asymptotic covariance matrix of the corrected local linear estimator has two additional terms $\sigma^2\mathbf{\Sigma}_{\boldsymbol{\zeta}} + E\{\boldsymbol{\xi}_1\boldsymbol{\alpha}(u)\boldsymbol{\alpha}^\top(u)\boldsymbol{\xi}_1^\top \mid U_1 = u\}$, due to the measurement errors.

To apply Corollary 3.1 to make statistical inference using $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(\cdot)$, a consistent estimator of $\boldsymbol{\Sigma}^{\mathrm{W}}$ is needed. Since μ_j and ν_j are known constants, we need only a consistent estimator for $\boldsymbol{\Sigma}^*/f_U(u)$. Define

$$\begin{aligned}
\widehat{\boldsymbol{\Gamma}}(u) &= \frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u)\mathbf{X}_i\mathbf{X}_i^{T},\\
\mathbf{G}_i &= \mathbf{W}_i(Y_i-\mathbf{W}_i^{T}\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))+\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i),\\
\hat{\boldsymbol{\Phi}}(u) &= \frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u)\mathbf{G}_i\mathbf{G}_i^{\mathrm{T}}.
\end{aligned}$$

The next theorem shows that $\{\widehat{\boldsymbol{\Gamma}}(u)\}^{-1}\hat{\boldsymbol{\Phi}}(u)\{\widehat{\boldsymbol{\Gamma}}(u)\}^{-1}$ is a consistent estimator of $\boldsymbol{\Sigma}^*/f_U(u)$.

THEOREM 3.2. *Suppose that Assumptions 1 to 5 hold. Then* $\{\widehat{\boldsymbol{\Gamma}}(u)\}^{-1}\hat{\boldsymbol{\Phi}}(u)\{\widehat{\boldsymbol{\Gamma}}(u)\}^{-1}\to\boldsymbol{\Sigma}^*/f_U(u)$ *in probability as* $n\to\infty$.

From Corollary 3.1 the asymptotic bias of $\hat{\alpha}_j^{\mathrm{W}}(u)$ is $\{\alpha_j''(u)h^2(\mu_2^2-\mu_1\mu_3)\}/\{2(\mu_2-\mu_1^2)\}$ and the asymptotic variance is $(nh)^{-1}e_{j,2p}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{W}}e_{j,2p}$. If we define the optimal bandwidth $h_{j,\mathrm{opt}}$ for estimating $\alpha_j(\cdot)$ to be the one which minimizes the squared bias plus variance, we have

$$h_{j,\mathrm{opt}}=\left\{\frac{\mu_2^2\nu_0-2\mu_1\mu_2\nu_1+\mu_1^2\nu_2}{f_U(u)(\mu_2^2-\mu_1\mu_3)^3}\times\frac{e_{j,2p}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{W}}e_{j,2p}}{\{\alpha_j''(u)\}^2}\right\}.$$

The data-driven bandwidth selection schemes by Ruppert, Sheather & Wand (1995) can be applied to our case with minor changes.

The following theorem shows the uniform strong convergence rate of $\hat{\alpha}_j^{\mathrm{W}}(u)$.

THEOREM 3.3. *Suppose that Assumptions 1 to 5 hold. Then*

$$\max_{1\le j\le p}\sup_{u\in\mathcal{U}}|\hat{\alpha}_j^{W}(u)-\alpha_j(u)|=O\{h^2+(\log n/nh)^{1/2}\}\quad \textit{almost surely.}$$

Remark 3.2. If h takes the optimal bandwidth, that is $h=cn^{-1/5}$ where c is a constant, according to Theorem 3.3 we have

$$\max_{1\le j\le p}\sup_{u\in\mathcal{U}}|\hat{\alpha}_j^{\mathrm{W}}(u)-\alpha_j(u)|=O\{n^{-\frac{2}{5}}(\log n)^{\frac{1}{2}}\}\quad\text{a.s.}$$

This means that the corrected local linear estimator achieves the optimal uniform strong convergence rate of the nonparametric estimation in nonparametric regression (Stone 1986).

4. ESTIMATION OF THE ERROR VARIANCE

The error variance $\sigma^2=E(\varepsilon_1^2)$ is a quantity that describes the noise level. Apart from its intrinsic interest as a model parameter, an estimator of σ^2 is essential in such tasks as the construction of confidence regions, model-based tests, model selection procedures, signal-to-noise ratio determination, bandwidth selection, and so on. Since $E(Y_i-\mathbf{X}_i^{T}\boldsymbol{\alpha}(U_i))^2=\sigma^2$ and $E(Y_i-\mathbf{W}_i^{T}\boldsymbol{\alpha}(U_i))^2=\sigma^2+E\big(\boldsymbol{\alpha}^{T}(U_1)\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\boldsymbol{\alpha}(U_1)\big)$, we define our estimator of σ^2 as

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mathbf{W}_i^{T}\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^2-\frac{1}{n}\sum_{i=1}^{n}(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^{T}\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i),$$

where $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(\cdot) = (\hat{\alpha}_1^{\mathrm{W}}(\cdot), \ldots, \hat{\alpha}_p^{\mathrm{W}}(\cdot))^\top$. The next two theorems give the asymptotic properties of $\hat{\sigma}_n^2$.

THEOREM 4.1. *Suppose that Assumptions 1 to 5 hold. Then*

$$\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{D} \mathrm{N}(0, \Delta) \quad \text{as } n \to \infty,$$

where $\Delta = \operatorname{var}\{\varepsilon_1^2 - \boldsymbol{\alpha}^T(U_1)(\boldsymbol{\zeta}_1\boldsymbol{\zeta}_1^T - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}})\boldsymbol{\alpha}(U_1) - 2\varepsilon_1\boldsymbol{\zeta}_1^T\boldsymbol{\alpha}(U_1)\}$.

THEOREM 4.2. *Suppose that Assumptions 1 to 5 hold. Then*

$$\lim_{n\to\infty} \sqrt{\frac{n}{2\log\log n}}\,|\hat{\sigma}_n^2 - \sigma^2| \le \sqrt{\Delta} \quad a.s.$$

Remark 4.1. Theorems 4.1 and 4.2 imply that our estimator of the error variance is root-n consistent and satisfies the law of the iterated logarithm. This means that the estimator is not infinitely inefficient compared with the conventional parametric approaches even though the model is not restricted within a finite-dimensional space.

We also need a consistent estimator of the asymptotic variance of $\hat{\sigma}_n^2$ to do inference. Let

$$\varpi_i = (Y_i - \mathbf{W}_i^T\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^2 - (\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^T\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) \text{ and } \hat{\Delta}_n = \frac{1}{n}\sum_{i=1}^n\left(\varpi_i - \frac{1}{n}\sum_{i=1}^n \varpi_i\right)^2.$$

THEOREM 4.3. *Suppose that Assumptions 1 to 5 hold. Then* $\hat{\Delta}_n \to \Delta$ *in probability as* $n \to \infty$.

5. A GOODNESS-OF-FIT TEST

To test whether model (1)-(2) holds with a specified parametric form for the varying-coefficient functions such as a linear regression model, we extend the generalized likelihood technique in Fan, Zhang & Zhang (2001) to the current setting involving measurement errors. The null hypothesis is

$$\mathcal{H}_0 : \alpha_j(u) = a_j(u, \boldsymbol{\theta}_j), \quad \text{for all } u \in \mathcal{U},\ j = 1, \ldots, n, \tag{5}$$

where $a_j(\cdot, \boldsymbol{\theta}_j)$ is a given family of functions indexed by an unknown parameter vector $\boldsymbol{\theta}_j$. Let $\hat{\boldsymbol{\theta}}_j$ be a consistent estimator of $\boldsymbol{\theta}_j$ and denote $\hat{\mathrm{a}}(U_i) = (\hat{a}_1(U_i, \hat{\boldsymbol{\theta}}_1), \ldots, \hat{a}_p(U_i, \hat{\boldsymbol{\theta}}_p))^\top$. Since

$$\begin{aligned} n^{-1}\sum_{i=1}^n (Y_i - \mathbf{W}_i^\top\hat{\mathrm{a}}(U_i))^2 &= n^{-1}\sum_{i=1}^n\{\mathbf{W}_i^\top(\boldsymbol{\alpha}(U_i) - \hat{\mathrm{a}}(U_i)\}^2 + n^{-1}\sum_{i=1}^n(\varepsilon_i - \boldsymbol{\zeta}_i^\top\boldsymbol{\alpha}(U_i))^2 \\ &\quad - 2n^{-1}\sum_{i=1}^n \boldsymbol{\alpha}^\top(U_i)\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\top(\boldsymbol{\alpha}(U_i) - \hat{\mathrm{a}}(U_i)) + o_p(1), \end{aligned}$$

the pseudo-residual sum of squares under the null hypothesis is

$$\mathrm{PRSS}_0 = n^{-1}\sum_{i=1}^n(Y_i - \mathbf{W}_i^\top\hat{\mathrm{a}}(U_i))^2 + 2n^{-1}\sum_{i=1}^n\widehat{\boldsymbol{\alpha}}^{\mathrm{W}T}(U_i)\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \hat{\mathrm{a}}(U_i)),$$

and the pseudo-residual sum of squares corresponding to model (1)-(2) is

$$\mathrm{PRSS}_1 = n^{-1}\sum_{i=1}^n(Y_i - \mathbf{W}_i^\top\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^2.$$

Our test statistic is defined as

$$Q_n = (\mathrm{PRSS}_0 - \mathrm{PRSS}_1)/\mathrm{PRSS}_1 = \mathrm{PRSS}_0/\mathrm{PRSS}_1 - 1,$$

for which we have the following theorem.

THEOREM 5.1. *Suppose that Assumptions 1 to 5 hold. Then under $\mathcal{H}_0$, $Q_n \to 0$ in probability as $n \to \infty$. Otherwise, if $\inf_{\boldsymbol{\theta}_j} E(\alpha_j(U_1) - a_j(U_1, \boldsymbol{\theta}_j))^2 > 0$ for some $j = 1, \ldots, p$, then there exists a constant $d > 0$ such that $Q_n > d$ with probability approaching one as $n \to \infty$.*

Remark 5.1. We note that PRSS_0 has a corrected term $2n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{\alpha}}^{\mathrm{W}T}(U_i)\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \hat{\mathbf{a}}(U_i))$. This term is necessary to guarantee that $Q_n > \delta$ with probability approaching one under alternative hypotheses when measurement errors are present as in our case.

Theorem 5.1 suggests that we should reject the null hypothesis (5) for large values of Q_n. However, the distribution of Q_n is hard to obtain. Inspired by Stute, González & Presedo (1998), Godfrey & Tremayne (2005) and Ioannidis & Peel (2005), among others who have successfully used the wild bootstrap method to calculate the p-values of their tests in similar situations, we also develop a wild bootstrap procedure to compute the p-value of Q_n. The wild bootstrap was proposed by C. F. J. Wu (1986), which samples from an external distribution to generate bootstrap distributions. Liu (1988) showed that it has some nice properties such as robustness to heteroscedasticity. We choose the wild bootstrap because we do not want to make a specific assumption on the form of our error distribution and sampling from model based distributions can be complicated when measurement errors are involved.

Following the steps below we can implement our goodness-of-fit test:

1. Fit the varying-coefficient regression model (1)-(2) as described in Section 2 and calculate Q_n and the estimated pseudo-residuals $\{\hat{\varepsilon}_i\}_{i=1}^n$, where

$$\hat{\varepsilon}_i = Y_i - \mathbf{W}_i^T \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i), \quad i = 1, \ldots, n.$$

2. Generate a sequence of independent and identically distributed random variables $\{\tau_i\}_{i=1}^n$ from a symmetric distribution function $F(\cdot)$ such that $E\tau_1 = 0$, $E\tau_1^2 = 1$ and $E|\tau_1|^3 < \infty$. We stress that $F(\cdot)$ is chosen independently of the given regression model.
3. Set $Y_i^* = \mathbf{W}_i^T \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) + \hat{\varepsilon}_i\tau_i$, $i = 1, \ldots, n$ and calculate the bootstrap test statistic Q_n^* based on the sample $\{Y_i^*, \mathbf{W}_i^T, U_i,\}_{i=1}^n$.
4. Repeat Step 2 and Step 3 a large number of times to generate a bootstrap distribution of Q_n^*.
5. Reject the null hypothesis $\mathcal{H}_0$ at level α if Q_n is greater than the $100(1-\alpha)\%$ quantile of the bootstrap distribution of Q_n^*.

The p-value of our test is simply the relative frequency of the event $\{Q_n^* \geq Q_n\}$ in the replications of the bootstrap sampling. For simplicity, we use the same bandwidth to calculate Q_n^* as was used to calculate Q_n. Note that we bootstrap the estimated pseudo-residuals from the nonparametric fit instead of the parametric fit, because the nonparametric estimate of the pseudo-residuals is always consistent under the null or the alternative hypothesis.

6. PERFORMANCES STUDIES

In this section we evaluate the performance of our proposed estimators and the goodness-of-fit test.

6.1. Performance of the error variance estimator and the corrected local linear estimator.

The data are generated from the following varying-coefficient regression model

$$Y_i = X_{1i}\alpha_1(U_i) + X_{2i}\alpha_2(U_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{6}$$

where $X_{1i} \sim \mathrm{N}(0,1)$, $X_{2i} \sim \mathrm{N}(0,1)$, $U_i \sim U(0,1)$, $\alpha_1(u_i) = 2\sin(2\pi u_i)$, $\alpha_2(u_i) = -1.5\cos(1.5\pi u_i)$ and $\varepsilon_i \sim \mathrm{N}(0,1)$. The measurement errors satisfy $W_{1i} = X_{1i} + \zeta_{1i}$ and $W_{2i} = X_{2i} + \zeta_{2i}$ with $(\zeta_{1i}, \zeta_{2i})^T \sim \mathrm{N}(0, \mathbf{\Sigma}_{\zeta})$, where we take $\mathbf{\Sigma}_{\zeta} = \mathrm{diag}\,(0.8^2, 0.8^2)$ and $\mathrm{diag}\,(0.4^2, 0.4^2)$.

For a fixed sample size n ($n = 300$ or $n = 500$) and a fixed $\mathbf{\Sigma}_{\zeta}$, we generate 1,000 random samples of $\{Y_i, \mathbf{W}_i^T, U_i\}_{i=1}^n$ according to model (6). For each sample, we estimate $\alpha_j(u)$ at $u = 1/n, 2/n, \dots, (n-1)/n$, $j = 1, 2$, and $\sigma^2 = 1$ as described in Sections 2 and 4 using the Gaussian kernel. The bandwidths are selected by cross-validation.

To make comparisons, we consider three types of estimators: our estimators $\hat{\sigma}_n^2$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$, the benchmark estimators $\tilde{\sigma}_n^2$ and $\tilde{\boldsymbol{\alpha}}^{\mathrm{X}}(u)$ in which the $\mathbf{X}_i$ are assumed to be observed exactly, namely,

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \mathbf{X}_i^{\mathrm{T}}\tilde{\boldsymbol{\alpha}}(U_i))^2, \quad \tilde{\boldsymbol{\alpha}}^{\mathrm{X}}(u) = (\mathbf{I}_2, \mathbf{0}_2)\mathbf{H}^{-1}\left[(\mathbf{D}_u^{\mathrm{X}})^T\boldsymbol{\omega}_u\mathbf{D}_u^{\mathrm{X}}\right]^{-1}(\mathbf{D}_u^{\mathrm{X}})^T\boldsymbol{\omega}_u\mathbf{Y},$$

and the naive estimators $\bar{\sigma}_n^2$ and $\bar{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$ which ignore the measurement errors and are given by

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \mathbf{X}_i^{\mathrm{T}}\bar{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^2, \quad \bar{\boldsymbol{\alpha}}^{\mathrm{W}}(u) = (\mathbf{I}_2, \mathbf{0}_2)\mathbf{H}^{-1}\left[(\mathbf{D}_u^{\mathrm{W}})^T\boldsymbol{\omega}_u\mathbf{D}_u^{\mathrm{W}}\right]^{-1}(\mathbf{D}_u^{\mathrm{W}})^T\boldsymbol{\omega}_u\mathbf{Y}.$$

For $\hat{\sigma}_n^2$, $\tilde{\sigma}_n^2$ and $\bar{\sigma}_n^2$, we calculate the sample means and the sample standard deviations (sd) over the 1,000 runs. For $\hat{\sigma}_n^2$ alone, we also compute the averages of the estimated asymptotic standard deviations (se) $\sqrt{\hat{\Delta}_n/n}$. Table 1 contains these performance measures.

TABLE 1: The sample means, sample standard deviations (sd) and averages of the estimated asymptotic standard deviations (for $\hat{\sigma}_n^2$ only, denoted by se) of three estimators of $\sigma^2 = 1$. $\mathbf{\Sigma}_{\zeta}$ is indexed by its diagonal entries.

$(n, \mathbf{\Sigma}_{\zeta})$		$(300, 0.8^2)$	$(300, 0.4^2)$	$(500, 0.8^2)$	$(500, 0.4^2)$
$\tilde{\sigma}_n^2$	mean	0.946	0.948	0.966	0.964
	sd	0.078	0.087	0.059	0.059
$\hat{\sigma}_n^2$	mean	0.673	0.905	0.824	0.941
	sd	0.781	0.121	0.289	0.092
	se	0.821	0.117	0.318	0.101
$\bar{\sigma}_n^2$	mean	2.071	1.357	2.154	1.373
	sd	0.174	0.114	0.146	0.875

For the three estimators of $\alpha_j(u)$, $j = 1, 2$, we plot the average of the 1,000 estimated curves for each estimator at $u = 1/n, 2/n, \dots, (n-1)/n$ in Figure 1 ($n = 300$) and Figure 2 ($n = 500$). For estimator $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$ alone and $n = 300$, we plot the the averages as in Figure 1, and plot the point-wise 95% confidence bands based on the asymptotic results in Section 3 as well as the simulated point-wise 95% confidence bands (the 2.5% and the 97.5% quantiles over 1,000 runs) in Figure 3.

From Table 1 and Figures 1 to 3 we make the following observations:

1. The naive estimator $\bar{\sigma}_n^2$ of the error variance is biased. The bias increases as the measurement variability increases and does not decrease with the increase of the sample size.

2. The bias and the standard deviation of our estimator $\hat{\sigma}_n^2$ decrease as the sample size increases. The estimator $\sqrt{\hat{\Delta}_n/n}$ gives results that are close to the simulated ones, showing the consistency predicted by the theory.

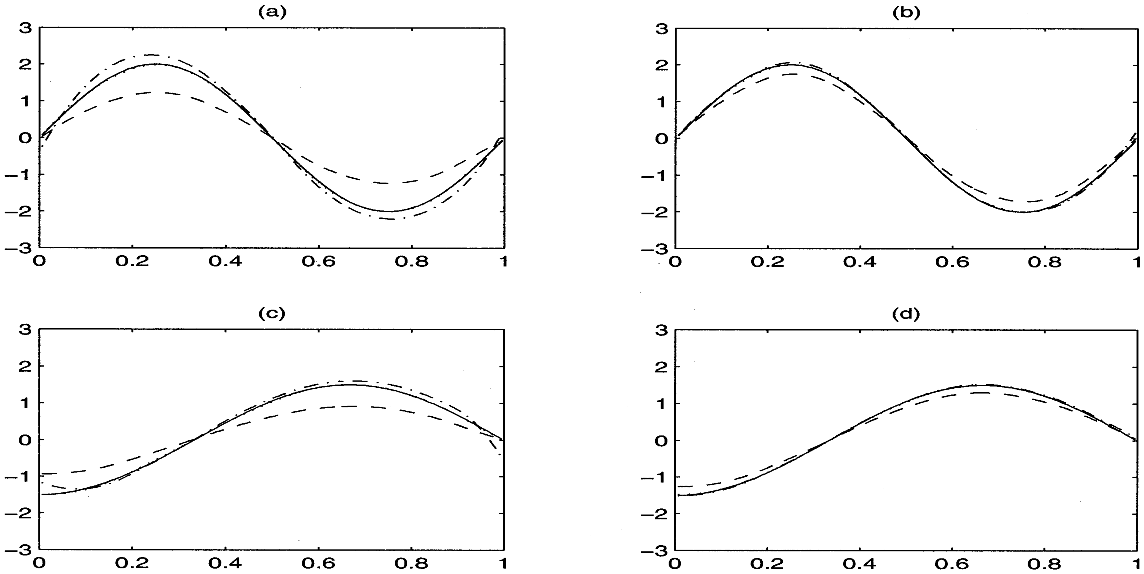

Figure 1: Estimates of the varying-coefficient functions $\alpha_1(\cdot)$ ((a) and (b)) and $\alpha_2(\cdot)$ ((c) and (d)) when $n = 300$. The left panel is based on $\mathbf{\Sigma}_\zeta = \operatorname{diag}(0.8^2, 0.8^2)$ and the right panel is based on $\mathbf{\Sigma}_\zeta = \operatorname{diag}(0.4^2, 0.4^2)$. In each plot, the dashed curve is for $\bar{\alpha}^{\mathrm{W}}(\cdot)$, the dash-dotted curve for $\hat{\alpha}^{\mathrm{W}}(\cdot)$, the dotted curve for $\tilde{\alpha}^{\mathrm{W}}(\cdot)$, and the solid curve for the true varying-coefficient function.

3. The performance of our estimator $\hat{\sigma}_n^2$ gets closer to that of the benchmark estimator $\tilde{\sigma}_n^2$ as the sample size increases.

4. Neglecting the measurement errors, the naive LLE $\bar{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$ is biased toward zero. The bias increases as the measurement variability increases and does not decrease as the sample size increases.

5. As expected, the benchmark LLE $\tilde{\boldsymbol{\alpha}}^{\mathrm{X}}(u)$ nearly coincides with the true varying-coefficient functions, and the performance of our LLE $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$ is very close to that of the benchmark LLE $\tilde{\boldsymbol{\alpha}}^{\mathrm{X}}(u)$ and improves as the sample size increases.

6. The asymptotic confidence bands are very close to the simulated quantile confidence bands, so our estimator of the asymptotic covariance matrix of $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(u)$ is consistent with the theory.

6.2. The power of the goodness-of-fit test.

For model (6), we consider the null hypothesis $\mathcal{H}_0$: $\alpha_j(u) = \theta_j u$, $u \in n(0,1)$, that is, we have linear varying-coefficient functions under $\mathcal{H}_0$, against the alternative $\mathcal{H}_1$: $\alpha_j(u) = \theta_j u + \sin(c\pi u)$, $u \in (0,1)$ and $c \in [0, 1.4]$, where $j = 1, 2$, $\theta_1 = 1$ and $\theta_2 = 1.5$.

For sample size $n = 500$, $\mathbf{\Sigma}_\zeta = \operatorname{diag}(0.8^2, 0.8^2)$ and for each c from the list 0(0.2)1.4, we generate 1,000 samples under $\mathcal{H}_1$. For each sample we estimate $\theta_1 = 1$ and $\theta_2 = 1.5$ using corrected least square estimators under $\mathcal{H}_0$. Then we generate $\{\tau_i : i = 1, \ldots, 500\}$ according to $\tau_i = -(\sqrt{5} - 1)/2$ with probability $(\sqrt{5} + 1)/(2\sqrt{5})$ and $\tau_i = (\sqrt{5} + 1)/2$ with probability $1 - (\sqrt{5} + 1)/(2\sqrt{5})$, and bootstrap 500 times to carry out our goodness-of-fit test at 5% level. Figure 4 plots the simulated powers against c. We see from Figure 4 that when the null hypothesis holds ($c = 0$), the size of our test is close to the nominal 5%. This demonstrates that the bootstrap estimate of the null distribution is approximately correct. When the alternative hypothesis is true ($c > 0$), the power of our test increases to 1 quickly as c increases.

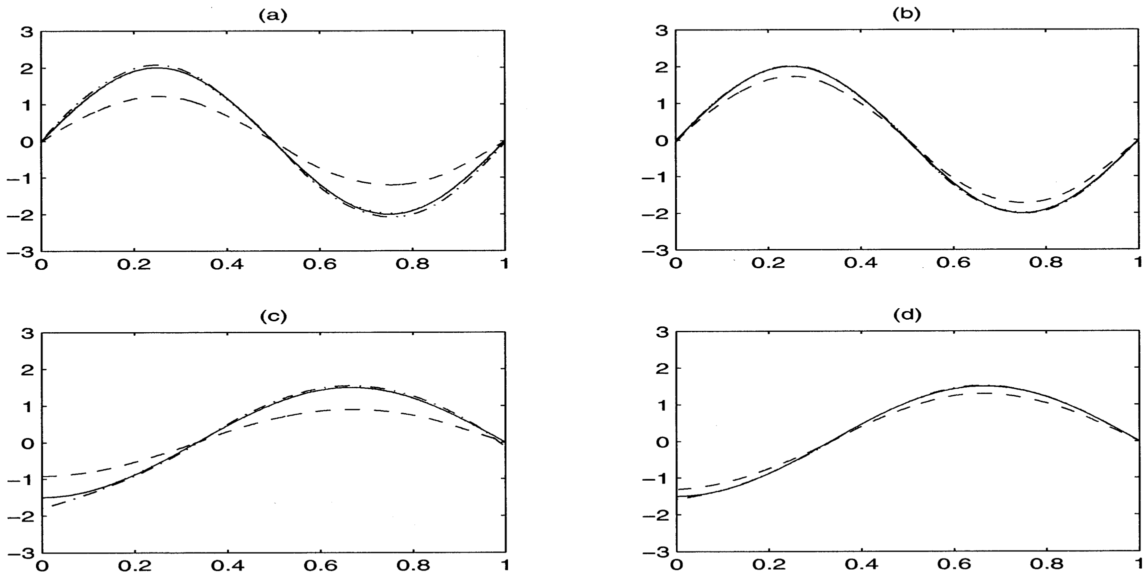

Figure 2: Estimates of the varying-coefficient functions $\alpha_1(\cdot)$ ((a) and (b)) and $\alpha_2(\cdot)$ ((c) and (d)) when $n = 500$. The left panel is based on $\mathbf{\Sigma}_\zeta = \text{diag}\,(0.8^2, 0.8^2)$ and the right panel is based on $\mathbf{\Sigma}_\zeta = \text{diag}\,(0.4^2, 0.4^2)$. In each plot, the dashed curve is for $\bar{\alpha}^{\mathrm{W}}(\cdot)$, the dash-dotted curve for $\hat{\alpha}^{\mathrm{W}}(\cdot)$, the dotted curve for $\tilde{\alpha}^{\mathrm{W}}(\cdot)$, and the solid curve for the true varying-coefficient function.

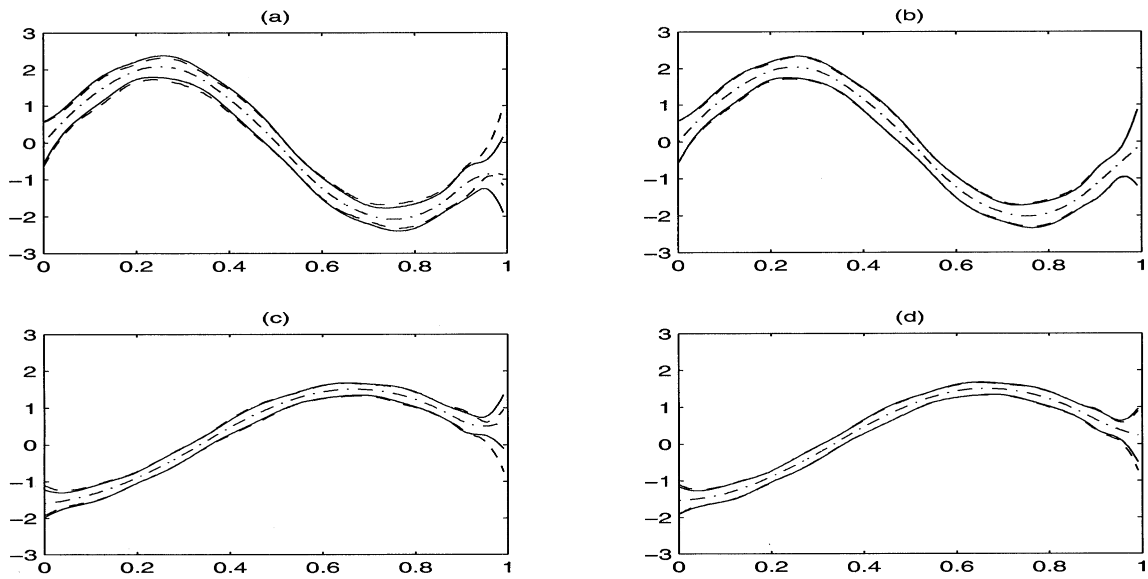

Figure 3: Estimates of the varying-coefficient functions $\alpha_1(\cdot)$ ((a) and (b), dash-dotted curves) and $\alpha_2(\cdot)$ ((c) and (d), dash-dotted curves) when $n = 300$, together with the 95% simulation confidence bands (dashed curves) and the 95% asymptotic confidence bands (solid curves). The left panel is based on $\mathbf{\Sigma}_\zeta = \text{diag}\,(0.8^2, 0.8^2)$ and the right panel is based on $\mathbf{\Sigma}_\zeta = \text{diag}\,(0.4^2, 0.4^2)$.

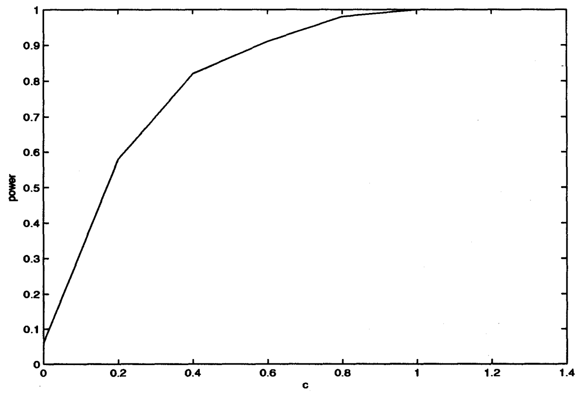

Figure 4: The simulated powers of our goodness-of-fit test versus the alternative hypothesis index c when sample size $n = 500$ and $\mathbf{\Sigma}_{\zeta} = \text{diag}\,(0.8^2, 0.8^2)$.

6.3. An application.

We analyze the data set from the Framingham Heart Study to illustrate the methodology developed in this paper. The Framingham Heart Study consists of a series of exams taken two years apart. There are 1,615 men aged from 31 to 65 in the study. For each man, Y is his average blood pressure in a fixed two-year period, U is his age at the beginning of the study, X_1 is his true cholesterol level and X_2 is his smoking status (0 = nonsmoking and 1 = smoking). We do not have his true cholesterol level but just the observed cholesterol level W, for which there are two replicates. All of the variables are scaled into $[0, 1]$ in the following discussion.

We want to study the effects of the subjects' serum cholesterol level, smoking status and age on blood pressure by fitting the following varying-coefficient regression model

$$Y_i = X_{1i}\alpha_1(U_i) + X_{2i}\alpha_2(U_i) + \varepsilon_i,\ i = 1, \ldots, 1615, \quad W_{ik} = X_{1i} + \zeta_{ik},\ k = 1, 2,$$

where only X_{1i} have measurement errors ζ_{ik} with $E(\zeta_{ik}) = 0$ and $V(\zeta_{ik}) = \eta^2$. In other words, we have $\mathbf{\Sigma}_{\zeta} = \text{diag}\,(\eta^2, 0)$. This model allows us to see the effects of the serum cholesterol level and smoking status on blood pressure at different ages.

We perform two analyses. In the first, we use only the first cholesterol measurement to find the naive LLE of $\alpha_j(\cdot)$, $j = 1, 2$. In this analysis, the measurement errors are ignored. In the second analysis, we first estimate η^2 to be 0.0026 (the average of 1,615 sample variances) and set $\widehat{\mathbf{\Sigma}}_{\zeta} = \text{diag}\,(0.0026, 0)$. We then find our corrected LLE of $\alpha_j(\cdot)$, $j = 1, 2$, using $\widehat{\mathbf{\Sigma}}_{\zeta}$ to replace $\mathbf{\Sigma}_{\zeta}$. The results are plotted in (a) and (c) of Figure 5. Also plotted in Figure 5 are the 95% asymptotic confidence bands and the corrected LLE ((b) and (d)).

We see from Figure 5(a) and Figure 5(c) that the naive LLE underestimates the effect of the serum cholesterol level due to attenuation associated with the measurement errors, and in turn overestimates the effect of smoking. We also see from Figure 5(b) and Figure 5(d) that there exist effects from serum cholesterol level and smoking on blood pressure (zero is not contained in the 95% confidence bands except for smoking effect at ages 31 to 32) and these effects are different at different ages, which is confirmed by our goodness-of-fit tests of $\mathcal{H}_0$: $\alpha_1(u) = c_1$ and $\alpha_2(u) = c_2$, c_1 and c_2 are two constants, with a p-value of 0.003.

7. CONCLUDING REMARKS

We have studied the estimation and goodness-of-fit of a varying-coefficient regression model when there are measurement errors. Because our measurement errors show up in the model in a way similar to that in a linear regression model with measurement errors, at first it seems that the correction that works for the linear regression case will work for our case. This turns out not to

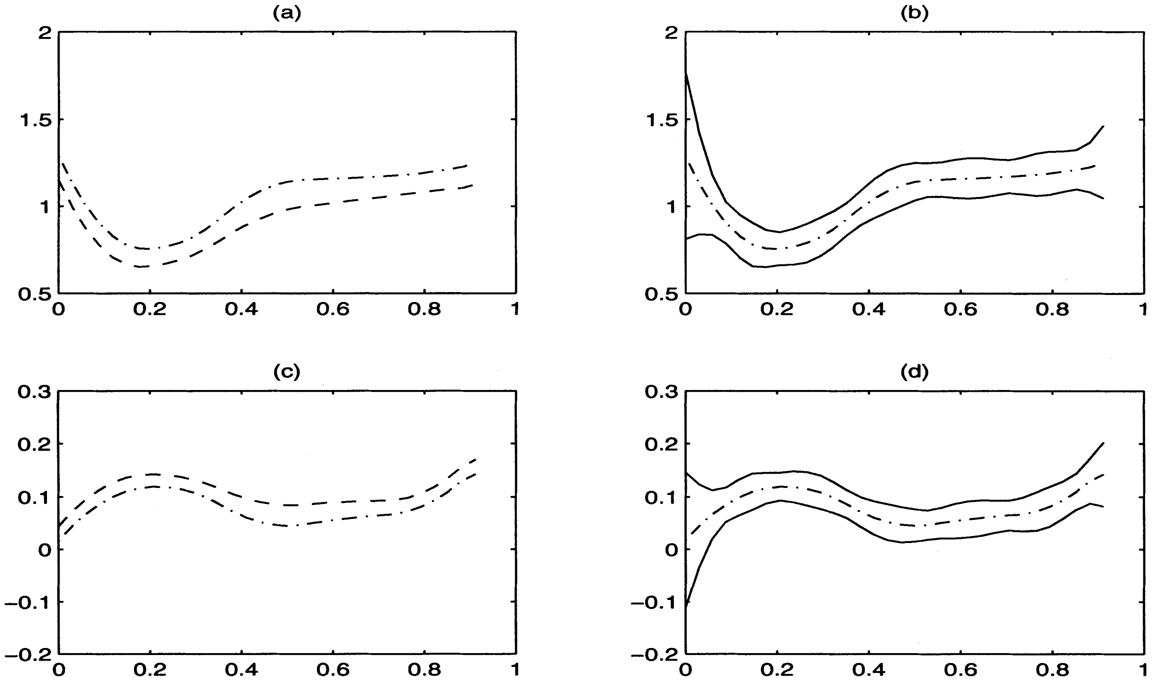

Figure 5: The naive local linear estimates of the serum cholesterol effect ((a), dashed curve) and the smoking effect ((c), dashed curve) over different scaled ages together with the corrected local linear estimates of the serum cholesterol effect ((a) and (b), dash-dotted curves) and the smoking effect ((c) and (d), dash-dotted curves). The solid curves in (b) and (d) are the 95% asymptotic confidence bands.

be true: applying the usual correction to our case cannot get rid of the attenuation issue because (we believe) the coefficients now are nonparametric and varying instead of being constant as in the linear regression case. We have however found a correction that works for our model.

We assume a known Σ_ζ in our discussion in Section 2. This is not really a tight restriction. As long as we have replicates, we can obtain a root-n consistent estimator $\widehat{\Sigma}_\zeta$ (Carroll, Ruppert & Stefanski 1995 and Liang, Härdle & Carroll 1999) and replacing the Σ_ζ in Section 2 with the estimator $\widehat{\Sigma}_\zeta$ gives us the same asymptotic theory. It is interesting to note that the same is not true for the linear regression case: using the true Σ_ζ to perform the correction and using the estimator $\widehat{\Sigma}_\zeta$ to perform the correction will both solve the attenuation problem but will lead to two different asymptotic variance structures.

There are a few related open issues that need further research. We have considered additive measurement errors in the X covariates. More needs to be studied when additivity does not hold (Hwang 1986, Iturria, Carroll & Firth 1999). If U is measured with additive normally distributed errors and X is exact, we conjecture that the nonparametric varying-coefficient functions can only be estimated at a logarithmic rate. We also conjecture that the same will happen when both X and U are measured with additive errors.

8. APPENDIX: PROOFS OF THE MAIN RESULTS

To prove the main results in the paper we first cite a lemma from Xia & Li (1999, Lemma A2).

LEMMA 8.1. *Suppose that Assumptions 1 to 5 hold. Then, as $n \to \infty$,*

$$\sup_{u\in\mathcal{U}}\left|\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{U_i-u}{h}\right)\left(\frac{U_i-u}{h}\right)^k X_{ij_1}X_{ij_2}-f_U(u)\Gamma_{j_1j_2}(u)\mu_k\right| = O\left\{h^2+\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \text{ a.s.,}$$

$$\sup_{u\in\mathcal{U}} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{U_i-u}{h}\right)\left(\frac{U_i-u}{h}\right)^k X_{ij}\varepsilon_i = O\left\{\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \quad a.s.$$

and

$$\sup_{u\in\mathcal{U}} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{U_i-u}{h}\right)\left(\frac{U_i-u}{h}\right)^k X_{ij}\zeta_{ij} = O\left\{\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \quad a.s.$$

where $j, j_1, j_2 = 1, \ldots, p$, $k = 0, 1, 2, 4$ and $\Gamma_{j_1,j_2}(u)$ is the (j_1, j_2)th element of $\mathbf{\Gamma}(u)$.

Proof of Theorem 2.1. Since the coefficient functions $\alpha_j(u)$ $(j = 1, 2, \ldots, p)$ are smooth in the neighbourhood of $|U_i - u| < h$, by a Taylor expansion there is

$$\mathbf{X}_i^\top \boldsymbol{\alpha}(U_i) = \mathbf{X}_i^\top \boldsymbol{\alpha}(u) + (U_i - u)\mathbf{X}_i^\top \boldsymbol{\alpha}'(u) + \frac{h^2}{2}\left(\frac{U_i-u}{h}\right)^2 \mathbf{X}_i^\top \boldsymbol{\alpha}''(u) + o(h^2) \quad \text{a.s.,}$$

where $\boldsymbol{\alpha}'(u)$ and $\boldsymbol{\alpha}''(u)$ are the vectors consisting of the first and the second derivatives of the functions $\alpha_j(\cdot)$. Since $\mathbf{D}_u^{\mathbf{W}} = \mathbf{D}_u^{\mathbf{X}} + \mathbf{D}_u^{\boldsymbol{\zeta}}$, we have

$$\begin{aligned}
\mathbf{H}\hat{\mathbf{\Psi}}(u) &= (\hat{\alpha}_1^{\mathbf{W}}(u), \ldots, \hat{\alpha}_p^{\mathbf{W}}(u), h\hat{\alpha}_1^{\mathbf{W}'}(u), \ldots, h\hat{\alpha}_p^{\mathbf{W}'}(u)) \\
&= \left\{(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \mathbf{\Omega}\right\}^{-1} (\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{Y} \\
&= \left\{(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \mathbf{\Omega}\right\}^{-1} \\
&\quad \times \Bigg[(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \boldsymbol{\varepsilon} + (\mathbf{D}_u^{\boldsymbol{\zeta}})^\top \boldsymbol{\omega}_u \boldsymbol{\varepsilon} + (\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}} \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix} \\
&\qquad + \frac{h^2}{2} (\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{A}_u \mathbf{X}^\top \boldsymbol{\alpha}''(u) + o(h^2) \cdot (\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{1}_n\Bigg],
\end{aligned}$$

where

$$\mathbf{A}_u = \text{diag}\left\{\left(\frac{U_1-u}{h}\right)^2, \ldots, \left(\frac{U_n-u}{h}\right)^2\right\},$$

$\mathbf{1}_n$ is an n-vector with each element equal to 1. It follows from Lemma 8.1 that

$$\frac{1}{n}(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \frac{1}{n}\mathbf{\Omega} = \frac{1}{n}(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}} + O\left\{\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \quad \text{a.s.} \tag{A1}$$

$$\frac{1}{n}(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}} = f_U(u)\mathbf{\Gamma}(u) \otimes \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} (1 + o(1)) \quad \text{a.s.} \tag{A2}$$

and $n^{-1}(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{1}_n = O(1)$ a.s. Therefore, it holds that

$$o(h^2) \cdot \left\{(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \mathbf{\Omega}\right\}^{-1} (\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{1}_n = o(h^2) \quad \text{a.s.}$$

Further,

$$\begin{aligned}
(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{X}} \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix} &= \left\{(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \mathbf{\Omega}\right\} \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix} \\
&\quad + \left\{-(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\zeta} + \mathbf{\Omega}\right\} \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix}.
\end{aligned}$$

So

$$\begin{aligned}\mathbf{H}\hat{\mathbf{\Psi}}(u) &= o(h^2) + \begin{pmatrix}\boldsymbol{\alpha}(u)\\ h\boldsymbol{\alpha}'(u)\end{pmatrix} + \{(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}\}^{-1}\\ &\quad\cdot\Bigg[(\mathbf{D}_u^{\mathbf{X}})^\top \boldsymbol{\omega}_u \boldsymbol{\varepsilon} + (\mathbf{D}_u^{\boldsymbol{\zeta}})^\top \boldsymbol{\omega}_u \boldsymbol{\varepsilon} + \{-(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{D}_u^{\boldsymbol{\zeta}} + \boldsymbol{\Omega}\}\begin{pmatrix}\boldsymbol{\alpha}(u)\\ h\boldsymbol{\alpha}'(u)\end{pmatrix}\\ &\qquad + \frac{h^2}{2}(\mathbf{D}_u^{\mathbf{W}})^\top \boldsymbol{\omega}_u \mathbf{A}_u \mathbf{X}^\top \boldsymbol{\alpha}''(u)\Bigg]\end{aligned}$$

almost everywhere.

For simplicity, write

$$J_1 = \begin{pmatrix}\sum_{i=1}^n \mathbf{X}_i \varepsilon_i K_h(U_i-u)\\ \sum_{i=1}^n \mathbf{X}_i \varepsilon_i K_h(U_i-u)(U_i-u)/h\end{pmatrix}, \quad J_2 = \begin{pmatrix}\sum_{i=1}^n \boldsymbol{\zeta}_i \varepsilon_i K_h(U_i-u)\\ \sum_{i=1}^n \boldsymbol{\zeta}_i \varepsilon_i K_h(U_i-u)(U_i-u)/h\end{pmatrix}$$

and

$$J_3 = \begin{pmatrix}\sum_{i=1}^n \boldsymbol{\xi}_i K_h(U_i-u) & \sum_{i=1}^n \boldsymbol{\xi}_i\left(\frac{U_i-u}{h}\right)K_h(U_i-u)\\ \sum_{i=1}^n \boldsymbol{\xi}_i\left(\frac{U_i-u}{h}\right)K_h(U_i-u) & \sum_{i=1}^n \boldsymbol{\xi}_i\left(\frac{U_i-u}{h}\right)^2 K_h(U_i-u)\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}(u)\\ h\boldsymbol{\alpha}'(u)\end{pmatrix},$$

where $\boldsymbol{\xi}_i = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T - \mathbf{X}_i\boldsymbol{\zeta}_i^T$. It is easy to see that J_1, J_2 and J_3 are uncorrelated random vectors made up with sums of independent and identically distributed random variables. For any $2p \times 1$ nonzero vector $\boldsymbol{\iota} = (\iota_1, \ldots, \iota_{2p})^T$ we have

$$\frac{\sqrt{h}}{\sqrt{n}}\boldsymbol{\iota}^T J_1 = \frac{\sqrt{h}}{\sqrt{n}}\sum_{i=1}^n\left\{\sum_{j=1}^p \iota_j X_{ij} K_h(U_i-u)\varepsilon_i + \sum_{j=1}^p \iota_{j+p} X_{ij} K_h(U_i-u)\left(\frac{U_i-u}{h}\right)\varepsilon_i\right\}.$$

It is easy to see that

$$\sqrt{h}\left\{\sum_{j=1}^p \iota_j X_{ij} K_h(U_i-u)\varepsilon_i + \sum_{j=1}^p \iota_{j+p} X_{ij} K_h(U_i-u)\left(\frac{U_i-u}{h}\right)\varepsilon_i\right\}$$

are independent and identically distributed random variables with mean zero and variance $\boldsymbol{\iota}^T\boldsymbol{\Sigma}_1\boldsymbol{\iota}$, where

$$\boldsymbol{\Sigma}_1 = \sigma^2 f_U(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}.$$

Therefore, by the multivariate central limit theorem we have that

$$\frac{\sqrt{nh}}{n} J_1 \xrightarrow{D} \mathsf{N}(0, \boldsymbol{\Sigma}_1) \quad \text{as } n \to \infty. \tag{A3}$$

By the same argument, we can show that

$$\frac{\sqrt{nh}}{n} J_2 \xrightarrow{D} \mathsf{N}(0, \boldsymbol{\Sigma}_2) \quad \text{as } n \to \infty, \tag{A4}$$

where

$$\boldsymbol{\Sigma}_2 = \sigma^2 f_U(u) E\{\boldsymbol{\zeta}_1\boldsymbol{\zeta}_1^T \mid U_1 = u\} \otimes \begin{pmatrix}\nu_0 & \nu_1\\ \nu_1 & \nu_2\end{pmatrix}.$$

For J_3 we have

$$\begin{aligned} J_3 &= \begin{pmatrix} \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\alpha}(u) K_h(U_i-u) \\ \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\alpha}(u)\left(\dfrac{U_i-u}{h}\right) K_h(U_i-u) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^n \boldsymbol{\xi}_i \left(\dfrac{U_i-u}{h}\right) K_h(U_i-u) h\boldsymbol{\alpha}'(u) \\ \sum_{i=1}^n \boldsymbol{\xi}_i \left(\dfrac{U_i-u}{h}\right)^2 K_h(U_i-u) h\boldsymbol{\alpha}'(u) \end{pmatrix} \\ &= J_{31} + J_{32}, \quad \text{say.} \end{aligned}$$

Following the same line as for J_1 and J_2. we have

$$\frac{\sqrt{nh}}{n} J_{31} \xrightarrow{D} \mathsf{N}(0, \boldsymbol{\Sigma}_3) \quad \text{as } n \to \infty, \tag{A5}$$

where

$$\boldsymbol{\Sigma}_3 = f_U(u) E\{\boldsymbol{\xi}_1 \boldsymbol{\alpha}(u)\boldsymbol{\alpha}^T(u)\boldsymbol{\xi}_1^T \mid U_1 = u\} \otimes \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}.$$

By Lemma 8.1 we have

$$J_{32} = O(nh^2) \cdot O\left\{ \left(\frac{\log n}{nh}\right)^{\frac{1}{2}} \right\} \quad \text{a.s.} \tag{A6}$$

On the other hand,

$$\begin{aligned} \frac{1}{n}(\mathbf{D}_u^{\mathrm{W}})^T \boldsymbol{\omega}_u \mathbf{A}_u \mathbf{X}^T \boldsymbol{\alpha}''(u) &= \frac{1}{n}(\mathbf{D}_u^{\mathrm{X}})^T \boldsymbol{\omega}_u \mathbf{A}_u \mathbf{X}^T \boldsymbol{\alpha}''(u) + O\left\{ \left(\frac{\log n}{nh}\right)^{\frac{1}{2}} \right\} \\ &= f_U(u)\boldsymbol{\Gamma}(u)\boldsymbol{\alpha}''(u) \otimes (\mu_2, \mu_3)^{\top}\{1 + o(1)\} \quad \text{a.s.} \end{aligned}$$

Therefore, by (A1) and (A2), we see that

$$\begin{aligned} &\{(\mathbf{D}_u^{\mathrm{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathrm{W}} - \boldsymbol{\Omega}\}^{-1} (\mathbf{D}_u^{\mathrm{X}})^T \boldsymbol{\omega}_u \mathbf{A}_u \mathbf{X}^T \boldsymbol{\alpha}''(u) \\ &\qquad = \frac{1}{\mu_2 - \mu_1^2} \begin{pmatrix} (\mu_2 - \mu_1\mu_3)\boldsymbol{\alpha}''(u) \\ (\mu_3 - \mu_1\mu_2)\boldsymbol{\alpha}''(u) \end{pmatrix} \{1 + o(1)\} \quad \text{a.s.} \end{aligned} \tag{A7}$$

Combining (A1)–(A7), we have

$$\sqrt{nh}\left[\mathbf{H}\{\hat{\boldsymbol{\Psi}}(u) - \boldsymbol{\Psi}(u)\} - \frac{h^2}{2}\frac{1}{\mu_2 - \mu_1^2} \begin{pmatrix} (\mu_2 - \mu_1\mu_3)\boldsymbol{\alpha}''(u) \\ (\mu_3 - \mu_1\mu_2)\boldsymbol{\alpha}''(u) \end{pmatrix} + o(h^2) \right] \xrightarrow{D} \mathsf{N}(0, \boldsymbol{\Sigma})$$

as $n \to \infty$, where

$$\begin{aligned} \boldsymbol{\Sigma} &= \left\{ f_U(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \right\}^{-1} \{ f_U(u)\big[\sigma^2\boldsymbol{\Gamma}(u) + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} + E\{\boldsymbol{\xi}_1\boldsymbol{\alpha}(u)\boldsymbol{\alpha}^T(u)\boldsymbol{\xi}_1^T \mid U_1 = u\} \\ &\quad \otimes \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \Big\} \left\{ f_U(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \right\}^{-1} \\ &= \frac{1}{f_U(u)} \boldsymbol{\Gamma}^{-1}(u)\big[\sigma^2\boldsymbol{\Gamma}(u) + \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} + E\{\boldsymbol{\xi}_1\boldsymbol{\alpha}(u)\boldsymbol{\alpha}^T(u)\boldsymbol{\xi}_1^T \mid U_1 = u\}\big]\boldsymbol{\Gamma}^{-1}(u) \\ &\quad \otimes \frac{1}{(\mu_2 - \mu_1^2)^2} \begin{pmatrix} \mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2 & (\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 \\ (\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\nu_2 & \nu_2 - \mu_1(2\nu_1 + \mu_1\nu_0) \end{pmatrix}. \end{aligned}$$

This implies that Theorem 2.1 holds.

Proof of Corollary 3.1. Corollary 3.1 comes directly from Theorem 3.1.

Proof of Theorem 3.2. According to Lemma 8.1 we have

$$\widehat{\mathbf{\Gamma}}(u) \rightarrow_p \mathbf{\Gamma}(u) f_U(u) \quad \text{as } n \rightarrow \infty.$$

On the other hand, since $\hat{\mathbf{\Phi}}(u)$ is a sandwich-type estimator, combining Lemma 8.1 and the argument in Section 3 of Liang, Härdle & Carroll (1999), we have

$$\hat{\mathbf{\Phi}}(u) \rightarrow_p \left[\sigma^2 \mathbf{\Gamma}(u) + \sigma^2 \mathbf{\Sigma}_{\boldsymbol{\zeta}} + E\{\boldsymbol{\xi}_1 \boldsymbol{\alpha}(u) \boldsymbol{\alpha}^\top(u) \boldsymbol{\xi}_1^\top \mid U_1 = u\}\right] f_U(u) \quad \text{as } n \rightarrow \infty.$$

This shows that Theorem 3.2 is true.

Proof of Theorem 3.3. In our proof of Theorem 3.1 we deliberately argue in terms of almost surely wherever we can rather than in terms of convergence in probability. This paves a way for Theorem 3.3 to come out naturally.

Proof of Theorem 4.1. According to the definition of $\hat{\sigma}_n^2$ we have

$$\begin{aligned} \hat{\sigma}_n^2 - \sigma^2 &= \left\{\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbf{B}f X_i^\top \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^2 - \sigma^2\right\} \\ &\quad - \frac{1}{n}\sum_{i=1}^n (\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^\top (\mathbf{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top) \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) \\ &\quad - \frac{2}{n}\sum_{i=1}^n (Y_i - \mathbf{X}_i^\top \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i)) \boldsymbol{\zeta}_i^\top \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) = J_1 - J_2 - J_3, \quad \text{say.} \end{aligned}$$

Further, J_1 can be decomposed as

$$\begin{aligned} J_1 &= n^{-\frac{1}{2}} \sum_{i=1}^n (\varepsilon_i^2 - \sigma^2) + \frac{1}{n}\sum_{i=1}^n \left\{\mathbf{X}_i^\top (\boldsymbol{\alpha}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))\right\}^2 \\ &\quad + \frac{2}{n}\sum_{i=1}^n (\boldsymbol{\alpha}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^\top \mathbf{X}_i \varepsilon_i = J_{11} + J_{12} + J_{13}, \quad \text{say.} \end{aligned}$$

By Theorem 3.3 it is easy to show that

$$J_{12} \leq \left\{\max_{1 \leq i \leq n} \|\boldsymbol{\alpha}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))\|\right\}^2 \cdot \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i = o(n^{-\frac{1}{2}}) \quad \text{a.s.}$$

Moreover, since

$$\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i)) = (\mathbf{I}_p, \mathbf{0}_{p \times p}) \{(\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \mathbf{D}_{u_i}^{\mathrm{W}} - \mathbf{\Omega}\}^{-1} (\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \{\boldsymbol{\varepsilon} + (\mathbf{X}_1^\top \boldsymbol{\alpha}(U_1), \ldots, \mathbf{X}_n^\top \boldsymbol{\alpha}(U_n))^\top\},$$

J_{13} can be written as

$$\begin{aligned} J_{13} &= \frac{2}{n}\sum_{i=1}^n \varepsilon_i \mathbf{X}_i (\mathbf{I}_q, \mathbf{0}_{q \times q}) \{(\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \mathbf{D}_{u_i}^{\mathrm{W}} - \mathbf{\Omega}\}^{-1} (\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \boldsymbol{\varepsilon} \\ &\quad + \frac{2}{n}\sum_{i=1}^n \varepsilon_i \mathbf{X}_i \Big[\boldsymbol{\alpha}(U_i) - (\mathbf{I}_p, \mathbf{0}_{p \times p}) \{(\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \mathbf{D}_{u_i}^{\mathrm{W}} - \mathbf{\Omega}\}^{-1} \\ &\quad \cdot (\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} (\mathbf{X}_1^\top \boldsymbol{\alpha}(U_1), \ldots, \mathbf{X}_n^\top \boldsymbol{\alpha}(U_n))^\top\Big] = J_{131} + J_{132}, \quad \text{say.} \end{aligned}$$

By the same argument leading to (A2) of Lemma A.4 in You & Chen (2006), we can show that $J_{131} = o(n^{-1/2})$ a.s. Let $\tilde{m}_j(U_i)$ be the jth element of

$$\boldsymbol{\alpha}(U_i) - (\mathbf{I}_p, \mathbf{0}_{p \times p}) \{(\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} \mathbf{D}_{u_i}^{\mathrm{W}} - \mathbf{\Omega}\}^{-1} (\mathbf{D}_{u_i}^{\mathrm{W}})^\top \boldsymbol{\omega}_{u_i} (\mathbf{X}_1^\top \boldsymbol{\alpha}(U_1), \ldots, \mathbf{X}_n^\top \boldsymbol{\alpha}(U_n))^\top.$$

To show $J_{132} = o(n^{1/2})$ a.s., we just need to show that $n^{-1}\sum_{i=1}^{n}\varepsilon_i X_{ij}\tilde{m}_j(U_i) = o(n^{-1/2})$ a.s. for $j = 1, \ldots, p$. Following the proof of Theorem 3.1, we have

$$\tilde{m} = \max_{1\le i\le n} |\tilde{m}_j(U_i)| = O\left\{h^2 + \left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right\} \quad \text{a.s.}$$

Put $\tau_{ij} = \varepsilon_i X_{ij}\tilde{m}_j(U_i)$. For any $\delta_1 > 0$, set

$$\Pi'_{ij} = \varepsilon_i X_{ij} I_{\{|\varepsilon_i X_{ij}|\le \delta_1^2 i^{1/2}\}} \quad \text{and} \quad \Pi''_{ij} = \varepsilon_i X_{ij} I_{\{|\varepsilon_i X_{ij}|> \delta_1^2 i^{1/2}\}}$$

such that $\tau_{ij} = \tilde{m}_j(U_i)\Pi'_{ij} + \tilde{m}_j(U_i)\Pi''_{ij}$. Since $E\|\varepsilon_1 X_{1j}\|^2 \le E|\varepsilon_1^4| + E\|X_{1j}\|^4 < \infty$ by the three-series theorem we obtain $\sum_{i=1}^{\infty}|\Pi''_{ij}| < \infty$. This implies that $\sum_{i=1}^{n} 1/\sqrt{n}\,\Pi''_{ij}\tilde{m}_j(U_i) = o(1)$ a.s. For $i = 1, \ldots, n$, let $\tau'_{ij} = 1/\sqrt{n}\,\Pi'_{ij}\tilde{m}_j(U_i)$. Given $\tilde{\boldsymbol{\Delta}}_n = \{U_1, \ldots, U_n, \mathbf{X}_1, \ldots, \mathbf{X}_n\}$, $\tau'_{1j}, \ldots, \tau'_{nj}$ are independent and

$$E(\tau'_{ij} \mid \tilde{\boldsymbol{\Delta}}_n) = 0, \quad \max_{1\le i\le n} |\tau'_{ij}| \le \tilde{m}\delta_1^2 \quad \text{and} \quad E(\tau'^2_{ij} \mid \tilde{\boldsymbol{\Delta}}_n) \le n^{-1}X_{ij}^2\tilde{m}^2\sigma^2.$$

Set $\theta_n = I_{\{n^{-1}\sum_{i=1}^{n} X_{ij}^2 \le EX_{1j}^2+1\}}$. By Bernstein's inequality we have

$$\begin{aligned} p_k &= Pr\left[\bigcup_{n\ge k}\left\{\left|\sum_{i=1}^{n}\tau'_i\right| \ge \delta_1, n^{-1}\sum_{i=1}^{n}X_{ij}^2 \le EX_{1j}^2 + 1\right\}\right] \\ &\le 2\sum_{n\ge k}\sum_{i=1}^{n} E\left[\theta_n \exp\left\{-\frac{n(\delta_1/n)^2}{(2/n)\sum_{i=1}^{n}E[(\tau'_i)^2|\tilde{\boldsymbol{\Delta}}_n] + \delta_1^2\tilde{m}(\delta_1/n)}\right\}\right] \\ &\le 2\sum_{n\ge k}\sum_{i=1}^{n} E\left[\theta_n \exp\left\{-\frac{\delta_1^2}{2\delta_1^3\tilde{m}}\right\}\right] \le 2\sum_{n\ge k} n^{-2} \to 0, \quad \text{as } k \to \infty. \end{aligned}$$

It follows from the above equation and the strong law of large numbers that

$$Pr\left\{\bigcup_{n\ge k}\left|\sum_{i=1}^{n}\tau'_i\right| \ge \delta_1\right\} \xrightarrow{p} 0 \quad \text{as } k \to \infty.$$

Therefore, $J_{132} = o(n^{1/2})$ a.s.

Combining the above we have

$$J_1 = \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma^2) + o\big(n^{-\frac{1}{2}}\big) \quad \text{a.s.}$$

On the other hand, J_2 can be written as

$$\begin{aligned} J_2 &= \frac{1}{n}\sum_{i=1}^{n}(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))^T(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\{\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)\} + \frac{1}{n}\sum_{i=1}^{n}\{\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)\}^T \\ &\quad \cdot(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\boldsymbol{\alpha}(U_i) + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^T(U_i)(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\boldsymbol{\alpha}(U_i) = J_{21} + J_{22} + J_{23}, \quad \text{say.} \end{aligned}$$

By Theorem 3.3 we have

$$\begin{aligned} |J_{21}| &\le \max_{1\le i\le n}\|\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)\|^2 \cdot \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_p^T(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\mathbf{1}_p \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^T(U_i)(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\{\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)\} \\ &= O(n^{-\frac{1}{2}}) + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^T(U_i)(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T)\{\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)\}, \quad \text{a.s.} \end{aligned}$$

By the same argument used in handling J_{13}, we can show that the second term in the above equation is $o(n^{-1/2})$ a.s. Therefore, $J_{21} = o(n^{-1/2})$ a.s. follows. Similarly, we have $J_{22} = o(n^{-1/2})$.

For J_3 we have

$$\begin{aligned} J_3 &= \frac{2}{n}\sum_{i=1}^{n} \mathbf{X}_i^T(\boldsymbol{\alpha}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))\boldsymbol{\zeta}_i^T(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)) \\ &\quad + \frac{2}{n}\sum_{i=1}^{n} \mathbf{X}_i^T(\boldsymbol{\alpha}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))\boldsymbol{\zeta}_i^T\boldsymbol{\alpha}(U_i) + \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i\boldsymbol{\zeta}_i^T(\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i) - \boldsymbol{\alpha}(U_i)) \\ &\quad + \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i\boldsymbol{\zeta}_i^T\boldsymbol{\alpha}(U_i) = J_{31} + J_{32} + J_{33} + J_{34}, \quad \text{say.} \end{aligned}$$

By the way we treat J_{21} we can show that $J_{3i} = o(n^{-1/2})$ a.s. for $i = 1, 2$ and 3. Therefore, it holds that

$$\sqrt{n}\,(\hat{\sigma}_n^2 - \sigma^2) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\{(\varepsilon_i^2 - \sigma^2) - \boldsymbol{\alpha}^T(U_i)(\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}})\boldsymbol{\alpha}(U_i) - 2\varepsilon_i\boldsymbol{\zeta}_i^T\boldsymbol{\alpha}(U_i)\} + o(1) \quad \text{a.s.}$$

Since $(\varepsilon_i^2 - \sigma^2) - \boldsymbol{\alpha}^T(U_i)(\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^T - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}})\boldsymbol{\alpha}(U_i) - 2\varepsilon_i\boldsymbol{\zeta}_i^T\boldsymbol{\alpha}(U_i)$ are independent and identically distributed random variables with mean zero and variance Δ, we see that Theorem 4.1 holds after applying the central limit theorem.

Proof of Theorem 4.2. From the proof of Theorem 4.1 and the Hartman–Wintner law of the iterated logarithm, it is easy to complete the proof.

Proof of Theorem 4.3. By Theorems 3.3, 4.2 and the Cauchy–Schwarz inequality, we have

$$\frac{1}{n}\sum_{i=1}^{n}\varpi_i^2 \xrightarrow{p} E\{\varepsilon_1^2 - 2\varepsilon_1\boldsymbol{\zeta}_1^T\boldsymbol{\alpha}(U_1) - \boldsymbol{\alpha}^T(U_1)(\boldsymbol{\zeta}_1\boldsymbol{\zeta}_1^T - \boldsymbol{\Sigma}_{\boldsymbol{\zeta}})\boldsymbol{\alpha}(U_1)\}^2 \quad \text{as} \quad n \to \infty.$$

Therefore, by Theorem 4.2 the proof is complete. □

Proof of Theorem 5.1. Under the null hypothesis, there is

$$\begin{aligned} \mathrm{PRSS}_0 - \mathrm{PRSS}_1 &= n^{-1}\sum_{i=1}^{n}\{\mathbf{W}_i^T(\mathbf{a}(U_i) - \hat{\mathbf{a}}(U_i))\}^2 \\ &\quad - n^{-1}\sum_{i=1}^{n}\{\mathbf{W}_i^T(\mathbf{a}(U_i) - \widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(U_i))\}^2 + o_p(1) \\ &= J_1 + o_p(1), \quad \text{say.} \end{aligned}$$

By the consistency of $\hat{\boldsymbol{\theta}}_j$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{W}}(\cdot)$ we have $J_1 \xrightarrow{p} 0$ as $n \to \infty$. This implies that $\mathrm{PRSS}_0 - \mathrm{PRSS}_1 \xrightarrow{p} 0$ under the null hypothesis. From the definition of PRSS_1 and Theorem 3.3, there is

$$\mathrm{PRSS}_1 \xrightarrow{p} E(\varepsilon_1 - \boldsymbol{\zeta}_i^T\boldsymbol{\alpha}(U_i))^2 > 0 \quad \text{as } n \to \infty.$$

This proves the first result of Theorem 5.1. The second result can be proved in a similar way.

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, the Associate Editor and the referees for their constructive comments and suggestions which have helped to improve the paper substantially. The third author's research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Z. Cai, J. Fan & Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95, 941–956.

R. J. Carroll, J. Fan, I. Gijbels & M. P. Wand (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92, 477–489.

R. J. Carroll, D. Ruppert & L. A. Stefanski (1995). *Measurement Error in Nonlinear Models.* Chapman & Hall, London.

C. L. Chen & J. Van Ness (1999). *Statistical Regression With Measurement Error.* Arnold, London.

C. T. Chiang, J. A. Rice & C. O. Wu (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96, 605–619.

W. S. Cleveland, E. Gross & W. M. Shyu (1991). Local regression models. In *Statistical Models in S* (J. M. Chambers. & T. J. Hastie, eds.), Wadsworth & Brooks, Pacific Grove, pp. 309–376.

J. Fan (1993). Adaptively local one-dimensional subproblems with applications to a deconvolution problem. *The Annals of Statistics*, 21, 600–610.

J. Fan & T. Huang (2005). Profile Likelihood Inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11, 1037–1051.

J. Fan & Y. K. Truong (1993). Nonparametric regression with errors-in-variables. *The Annals of Statistics*, 21, 1900–1925.

J. Fan, C. Zhang & J. Zhang (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29, 153–193.

J. Fan & W. Zhang (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27, 1491–1518.

J. H. Friedman (1991). Multivariate adaptive regression splines (with discussion). *The Annals of Statistics*, 19, 1–141.

W. A. Fuller (1987). *Measurement Error Models.* Wiley, New York.

L. G. Godfrey & A. R. Tremayne (2005). The wild bootstrap and heteroskedasticity-robust tests for serial correlation in dynamic regression models. *Computational Statistics and Data Analysis*, 49, 377–395.

C. Gu & G. Wahba (1993). Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2, 97–117.

W. Härdle, H. Liang & J. T. Gao (2000). *Partially Linear Models.* Physica-Verlag, Heidelberg.

W. Härdle & T. M. Stoker (1989). Investigating smooth multiple regression by the method of average derivative. *Journal of the American Statistical Association*, 84, 986–995.

T. J. Hastie & R. Tibshirani (1990). *Generalized Additive Models.* Chapman & Hall, London.

T. J. Hastie & R. Tibshirani (1993). Varying-coefficient models. *Journal of Royal Statistical Society Series B*, 55, 757–796.

J. Huang, C. O. Wu & L. Zhou (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89, 111–128.

J. Huang, C. O. Wu & L. Zhou (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistic Sinica*, 14, 763–788.

J. Hwang (1986). Multiplicative errors-in-variables models with applications to the recent data released by the U.S. department of energy. *Journal of the American Statistical Association*, 81, 680–688.

C. Ioannidis & D. A. Peel (2005). Testing for market efficiency in gambling markets when the errors are non-normal and heteroskedastic an application of the wild bootstrap. *Economical Letters*, 87, 221–226.

S. J. Iturria, R. J. Carroll & D. Firth (1999). Polynomial regression and estimating functions in the presence of multiplicative measurement error. *Journal of Royal Statistical Society Series B*, 61, 547–561.

H. Liang, W. Härdle & R. J. Carroll (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics*, 27, 1519–1535.

Y. Liu (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16, 1696–1708.

D. Ruppert, S. J. Sheather & M. P. Wand (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90, 1257–1270.

C. J. Stone (1986). Additive regression and other nonparametric models. *The Annals of Statistics*, 13, 689–705.

W. Stute, M. W. Gonzâlez & Q. M. Presedo (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93, 141–149.

T. Tosteson, L. A. Stefanski & D. W. Schafer (1989). A measurement error model for binary and ordinal regression. *Statistical Medicine*, 8, 1139–1147.

G. Wahba (1984). Partial spline models for semiparametric estimation of functions of several variables. In *Statistical Analysis of Time Series*, Proceedings of the Japan U.S. Joint Seminar, Tokyo, 319–329. Institute of Statistical Mathematics, Tokyo.

N. Wang, R. J. Carroll & K. Y. Liang (1996). Quasi-likelihood and variance functions in measurement error models with replicates. *Biometrics*, 52, 423–432.

C. F. J. Wu (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14, 1261–1295.

C. O. Wu, C. T. Chiang & D. R. Hoover (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93, 1388–1402.

C. O. Wu, K. F. Yu & C. T. Chiang (2000). A two-step smoothing method for varying-coefficient models with repeated measurements. *Annals of Institute of Statistics and Mathematics*, 52, 519–543.

Y. Xia & W. K. Li (1999). On the estimation and testing of functional-coëfficient linear models. *Statistics Sinica*, 9, 737–757.

J. You & G. Chen (2006). Estimation in a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis*, 97, 324–341.

Received 20 May 2004
Accepted 27 February 2006

Jinhong You: jyou@bios.unc.edu
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7400, USA

Yong Zhou: zhouyong8@hotmail.com
Institute of Applied Mathematics
Academy of Mathematics and System Science
Chinese Academy of Science, Beijing 100080, P. R. China

Gemai Chen: gchen@math.ucalgary.ca
Department of Mathematics & Statistics, University of Calgary
Calgary, Alberta, T2N 1N4 Canada