

Nonlinear Varying-Coefficient Models with Applications to a Photosynthesis Study

Esra Kürüm, Runze Li, Yang Wang & Damla Şentürk

**Journal of Agricultural, Biological,
and Environmental Statistics**

ISSN 1085-7117

Volume 19

Number 1

JABES (2014) 19:57-81

DOI 10.1007/s13253-013-0157-7

Volume 19 • Number 1

March 2014

Journal of Agricultural, Biological and Environmental Statistics



An Applied Journal Published by
The International Biometric Society
and The American Statistical Association



Springer

13253 • ISSN 1085-7117
19(1) 1–160 (2014)

Available
online
www.springerlink.com

Your article is protected by copyright and all rights are held exclusively by International Biometric Society. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Nonlinear Varying-Coefficient Models with Applications to a Photosynthesis Study

Esra KÜRÜM, Runze LI, Yang WANG, and Damla ŞENTÜRK

Motivated by a study on factors affecting the level of photosynthetic activity in a natural ecosystem, we propose nonlinear varying-coefficient models, in which the relationship between the predictors and the response variable is allowed to be nonlinear. One-step local linear estimators are developed for the nonlinear varying-coefficient models and their asymptotic normality is established leading to point-wise asymptotic confidence bands for the coefficient functions. Two-step local linear estimators are also proposed for cases where the varying-coefficient functions admit different degrees of smoothness; bootstrap confidence intervals are utilized for inference based on the two-step estimators. We further propose a generalized F -test to study whether the coefficient functions vary over a covariate. We illustrate the proposed methodology via an application to an ecology data set and study the finite sample performance by Monte Carlo simulation studies.

Key Words: Generalized F -test; Local linear regression; Nonlinear regression model; Varying coefficient models.

1. INTRODUCTION

This work was motivated by an empirical study in the field of ecology. It is known that the rate of photosynthesis in an ecosystem is affected by sunlight intensity. The level of photosynthetic activity in a natural ecosystem is measured by the *Net Ecosystem Exchange of CO₂* (NEE), since leaves absorb carbon dioxide (CO₂) during photosynthesis. NEE depends on the amount of *Photosynthetically Active Radiation* (PAR) available to an ecosys-

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA, NCI, NIDDK or the NIH.

Esra Kürüm (✉) is Assistant Professor, Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey (E-mail: esra.kurum@medeniyet.edu.tr). Runze Li is Distinguished Professor, Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802-2111, USA (E-mail: rzli@psu.edu). Yang Wang is Head, Division of Strategic Investment, Marketing and Treasury, Quantitative Marketing Research, China Vanke, Shenzhen 518093, China (E-mail: wangy39@vanke.com). Damla Şentürk is Associate Professor, Department of Biostatistics, University of California, Los Angeles, CA 90095, USA (E-mail: dsenturk@ucla.edu).

© 2013 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Volume 19, Number 1, Pages 57–81

DOI: [10.1007/s13253-013-0157-7](https://doi.org/10.1007/s13253-013-0157-7)

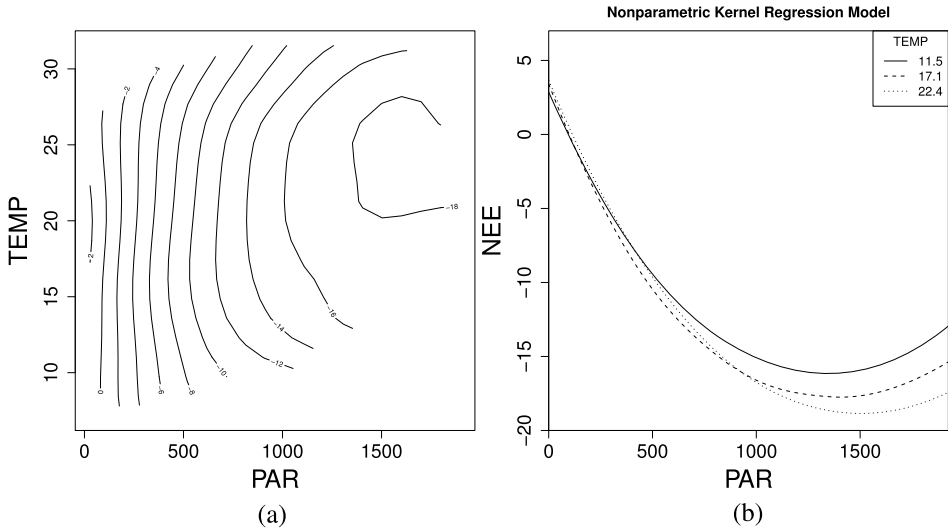


Figure 1. (a) Contour plot of the two-dimensional kernel regression fit of NEE on PAR and TEMP, (b) Slices from the kernel regression fit at three quartiles of the observed temperature values.

tem. It is believed that the relationship between NEE and PAR is nonlinear (Monteith 1972) and follows the model,

$$NEE = \beta_1 - \frac{\beta_2 PAR}{PAR + \beta_3} + \varepsilon, \quad (1.1)$$

where ε is the random error with mean zero, and β_1 , β_2 , β_3 are the unknown parameters. The coefficient β_1 represents the dark respiration rate, that is, the rate of respiration in plants where carbon dioxide is released without the aid of sun light, and β_2 represents the light saturated net photosynthetic rate. The light saturation point refers to the amount of light that the chloroplast cannot absorb. At this point, photosynthesis still occurs, but the amount of light is more than the number of pigments in chlorophyll cells that are available for absorption. Finally, β_2/β_3 is the apparent quantum yield. The quantum yield of photosynthesis is a measure of efficiency which refers to the amount of the products of photosynthesis, such as O_2 , per unit light input (Garrett and Grisham 2005). When the quantum yield measured is based only on the incident light, that is, the light falling on a surface, and does not take into account the scattered light, it is referred to as the apparent quantum yield or lower limits of true quantum yield.

Model (1.1) was obtained in a laboratory where climate variables such as temperature and moisture variability can be controlled. However, when the temperature (T) cannot be controlled, the parameters β_1 , β_2 , and β_3 are likely to depend on temperature. To demonstrate that the relationship between NEE and PAR depends on temperature, we fit a two-dimensional kernel regression of NEE ($\mu\text{mol m}^{-2} \text{s}^{-1}$) on PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$) and temperature ($^{\circ}\text{C}$). Figure 1(a) gives the contour plot of the regression fit, where three slices of the estimated regression fit are given at three quartiles of the observed temperature values in Figure 1(b). In Figure 1(a), there seems to be a weak temperature effect on the association of NEE and PAR for low values of PAR; however, as PAR increases, the non-parallel

pattern demonstrates that the effect of temperature on this relation is getting stronger. The profiles in Figure 1(b) strengthen this observation: with large PAR values, the estimates seem to differ for different temperature values. Hence, we consider the following model allowing the nonlinear regression relation between NEE and PAR to change with temperature:

$$NEE = \beta_1(T) - \frac{\beta_2(T)PAR}{PAR + \beta_3(T)} + \varepsilon. \quad (1.2)$$

A detailed analysis of the ecological data set will be given in Section 4.

Powerful modeling techniques have been proposed for dealing with ‘curse of dimensionality’ while modeling high dimensional data. Examples include additive models (Breiman and Friedman 1985; Hastie and Tibshirani 1990), low-dimensional interaction models (Friedman 1991; Gu and Wahba 1993; Stone et al. 1997), multiple-index models (Härdle and Stoker 1989; Li 1991), and partially linear models (Wahba 1984; Green and Silverman 1994). With the aim of increasing the flexibility of linear regression models and reducing the modeling bias, varying-coefficient models

$$Y = \mathbf{X}^T \boldsymbol{\beta}(U) + \varepsilon, \quad (1.3)$$

was first introduced by Cleveland, Grosse, and Shyu (1992) and became popular in the statistical literature due to the work by Hastie and Tibshirani (1993). In (1.3), Y denotes the response variable, X the vector of predictors, and U a scalar predictor, such as temperature. The varying-coefficient models explore the dynamic features of the data by allowing the model coefficients to vary over a covariate U , such as temperature. At a fixed $U = u$ value, the regression coefficients in a varying-coefficient model can be interpreted in a similar way to a linear regression model.

Model (1.2) motivates us to consider a new class of models, a natural extension of varying-coefficient models, namely, nonlinear varying-coefficient models. We define the nonlinear varying-coefficient model as

$$Y = f\{\mathbf{X}, \boldsymbol{\beta}(U)\} + \varepsilon, \quad (1.4)$$

where $f(\cdot, \cdot)$ is a pre-specified function and $\boldsymbol{\beta}(U)$ consists of unknown coefficient functions. The error term satisfies $E(\varepsilon|\mathbf{X}, U) = 0$. Similar to the interpretation of a varying-coefficient model, at a fixed value $U = u$, the nonlinear varying-coefficient model can be interpreted as a nonlinear regression model.

Estimation in varying-coefficient models has been studied extensively. Earlier work on estimation of the varying-coefficient models includes Wu, Chiang, and Hoover (1998), Hoover et al. (1998), Fan and Zhang (1999) and Kauermann and Tutz (1999). Fan and Zhang (2008) provide an excellent review of the literature on estimation and inference procedures for varying-coefficient models. In this paper, we propose one-step and two-step local linear estimators for nonlinear varying-coefficient models using local linear regression techniques (Fan and Gijbels 1996). While the one-step estimators estimate the varying-coefficient functions using a single bandwidth, the two-step estimators adopt to cases where the varying-coefficient functions admit different degrees of smoothness. Since there is no closed form solution for the local parameters, we propose an iterative linear

regression algorithm to search for the local solutions. Thus, the proposed estimation procedures can be easily carried out. We establish the asymptotic normality of the proposed one-step estimators and derive consistent estimators for their asymptotic standard error, leading to asymptotic point-wise confidence intervals. Bootstrap percentile confidence intervals are utilized for inference based on the two-step estimators.

It is of interest to test whether the proposed nonlinear varying-coefficient model reduces to a regular nonlinear regression model, which amounts to testing whether the varying-coefficient functions in (1.4) are in fact varying over the covariate U . In this case the null hypothesis is parametric, while the alternative is nonparametric. For this hypothesis testing problem, we develop a generalized likelihood ratio test where the proposed test statistic follows a χ^2 distribution and it inherits Wilks phenomenon (Fan, Zhang, and Zhang 2001), i.e. the asymptotic null distribution of the test statistic is independent of the nuisance parameters $\beta(U)$ and the density function of U . Based on the Wilks phenomenon, we utilize a bootstrap procedure to estimate the degrees of freedom of the χ^2 distribution of the proposed test statistic (Cai, Fan, and Li 2000). We apply the proposed methodology for a detailed empirical analysis of the aforementioned ecological data set. The efficacy of the proposed algorithms is studied via Monte Carlo simulations.

The paper is organized as follows. In Section 2, we develop one-step and two-step estimation procedures and propose a generalized F -test for the nonlinear varying-coefficient models. Asymptotic properties of the proposed one-step estimation procedure are also given in this section. Monte Carlo simulation studies are presented in Section 3. A detailed empirical analysis of the ecology data set is given in Section 4. Regularity conditions and technical proofs are given in the supplementary material.

2. STATISTICAL INFERENCE PROCEDURES

Suppose that $\{U_i, \mathbf{X}_i, Y_i\}$, $i = 1, 2, \dots, n$ is an independent and identically distributed sample from the nonlinear varying-coefficient model

$$Y_i = f\{\mathbf{X}_i, \beta(U_i)\} + \varepsilon_i,$$

where $E(\varepsilon_i|\mathbf{X}_i, U_i) = 0$ and $\beta(U_i) = \{\beta_1(U_i), \beta_2(U_i), \dots, \beta_p(U_i)\}^T$ consists of p unknown nonparametric coefficient functions. At every fixed $U = u$ value, a nonlinear varying-coefficient model implies a nonlinear regression model. Hence, its identifiability requires the identifiability of the nonlinear models holding at each fixed $U = u$, i.e. that there exists no nonzero parameter $\beta^0(u) \neq \beta(u)$ such that $f(x; \beta^0(u)) = f(x; \beta(u))$ for all x (Seber and Wild 2003).

2.1. ONE-STEP ESTIMATION

One approach for estimating the coefficient functions $\beta_j(\cdot)$, $j = 1, \dots, p$, is to employ local linear fitting techniques. In what follows, we will use h_1 to denote the bandwidth choice used in obtaining the one-step estimator and use h_0 and h_2 to denote the initial and

final bandwidths used in obtaining the two-step estimators. In order to estimate the non-parametric coefficient functions at a fixed point u_0 , we locally approximate the functions in the neighborhood of a fixed point u_0 by the Taylor expansion:

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j(u_0)(u - u_0) \equiv a_j + b_j(u - u_0),$$

for $j = 1, \dots, p$. Let $\mathbf{a} = (a_1, \dots, a_p)^T$ and $\mathbf{b} = (b_1, \dots, b_p)^T$. We minimize the local least squares,

$$\ell(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n [Y_i - f\{\mathbf{X}_i, \mathbf{a} + \mathbf{b}(U_i - u_0)\}]^2 K_{h_1}(U_i - u_0), \quad (2.1)$$

with respect to the local parameters (\mathbf{a}, \mathbf{b}) , where $K_{h_1}(t) = h_1^{-1}K(t/h_1)$ is the scaled kernel function $K(\cdot)$ with bandwidth h_1 . The estimators of the nonlinear regression coefficients can be given as $\hat{\beta}(u_0) = \hat{\mathbf{a}}$.

Note that $f(\mathbf{X}, \beta)$ is pre-specified, but may be a nonlinear function of \mathbf{X} and β . Thus, it is typical that there is no closed form for $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. The Newton–Raphson algorithm may be employed to search the solution. However, based on our limited experience, the Hessian matrix of $\ell(\mathbf{a}, \mathbf{b})$ may not be positive definite during the iteration of the Newton–Raphson algorithm. This poses a challenge in finding the minimizer of the objective function. Hence, we propose the following iterative linear regression algorithm to obtain the values that minimize the local least squares function (2.1). During the course of the iteration, suppose we have the value $(\mathbf{a}_0, \mathbf{b}_0)$ of (\mathbf{a}, \mathbf{b}) at the current step. We approximate $f\{\mathbf{X}_i, \mathbf{a} + \mathbf{b}(U - u_0)\}$ in the neighborhood of $(\mathbf{a}_0, \mathbf{b}_0)$ using the Taylor expansion,

$$f\{\mathbf{X}_i, \mathbf{a} + \mathbf{b}(U_i - u_0)\} \approx f\{\mathbf{X}_i, \mathbf{a}_0 + \mathbf{b}_0(U_i - u_0)\} + \{(\mathbf{a} - \mathbf{a}_0) + (\mathbf{b} - \mathbf{b}_0)(U_i - u_0)\}^T f'\{\mathbf{X}_i, \mathbf{a}_0 + \mathbf{b}_0(U_i - u_0)\},$$

where $f'(\mathbf{X}_i, \beta)$ is the $p \times 1$ vector equal to $\partial f(\mathbf{X}_i, \beta)/\partial \beta$. Denote the values of \mathbf{a} and \mathbf{b} at the k th iteration as $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Let

$$\mathbf{F}_k = \begin{bmatrix} f'^T\{\mathbf{X}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_1 - u_0)\} & (U_1 - u_0)f'^T\{\mathbf{X}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_1 - u_0)\} \\ \vdots & \vdots \\ f'^T\{\mathbf{X}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_n - u_0)\} & (U_n - u_0)f'^T\{\mathbf{X}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_n - u_0)\} \end{bmatrix}_{n \times 2p},$$

$$\begin{aligned} \mathbf{Y}_{i,k} &= Y_i - f\{\mathbf{X}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_i - u_0)\} \\ &\quad + \{\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_i - u_0)\}^T f'\{\mathbf{X}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(U_i - u_0)\}, \end{aligned}$$

and $\mathbf{Y}_k = (Y_{1,k}, \dots, Y_{n,k})^T$. Then, we update (\mathbf{a}, \mathbf{b}) according to

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (\mathbf{F}_k^T \mathbf{W} \mathbf{F}_k)^{-1} \mathbf{F}_k^T \mathbf{W} \mathbf{Y}_k,$$

where $\mathbf{W} = \text{diag}(K_{h_1}(U_1 - u_0), \dots, K_{h_1}(U_n - u_0))$. The solution of this iterative linear regression algorithm will satisfy $\ell(\mathbf{a}, \mathbf{b}) = 0$, and the estimators will be given by $\hat{\beta}(u_0) =$

$\hat{\mathbf{a}}$ and $\hat{\boldsymbol{\beta}}'(u_0) = \hat{\mathbf{b}}$. The proposed algorithm is indeed the Fisher scoring algorithm if we further assume that the error follows a normal distribution. Hence, our algorithm shares the convergence properties of the Fisher scoring algorithm. The asymptotic rate of convergence of the scoring algorithm should approach that of the Newton method as the sample size goes to infinity (Osborne 1992). In other words, the scoring algorithm approaches a second order convergence rate as the sample size increases. We demonstrate that the proposed algorithm converges quickly in simulation studies.

A question that arises in the practical implementation of the proposed procedure is the bandwidth selection. We suggest using a multi-fold cross-validation technique (Geisser 1975). Specifically, we propose minimizing the following cross-validation score:

$$CV(h) = \sum_j \|Y_j - \hat{Y}_{-j}\|^2, \quad (2.2)$$

where \hat{Y}_{-j} denotes the fitted value with j th group of subjects excluded. The bandwidth that minimizes (2.2) is chosen.

2.2. TWO-STEP ESTIMATION

As a second estimation proposal, we extend the two-step local linear estimation procedure of Fan and Zhang (1999) proposed for varying-coefficient models to nonlinear varying-coefficient models. While the proposed one-step estimation is easier to implement, the two-step estimators have been reported to lead to efficiency gains when the varying-coefficient functions admit different degrees of smoothness and hence a single bandwidth is not optimal in estimating all varying-coefficient functions. The proposed two-step estimation procedure is sketched as follows. In the first step, we obtain initial estimators $\tilde{\beta}_j(U_i)$, $j = 1, \dots, p$, by minimizing (2.1) with a small bandwidth h_0 . These estimators would have a smaller bias, but a larger variance. As Fan and Zhang (1999) pointed out the two-step estimator is not sensitive to the initial bandwidth h_0 choice, since the two-step estimator is optimal for a wide range of bandwidths. We suggest using a multi-fold cross-validation (2.2) or generalized cross-validation to obtain the bandwidth \hat{h} and then use $h_0 = 0.5\hat{h}$ as the initial bandwidth (Fan and Zhang 1999). In the second step, we obtain the two-step estimator of $\beta_j(\cdot)$, $j = 1, \dots, p$, by replacing all the varying-coefficient functions except $\beta_j(\cdot)$ with their corresponding initial estimators. In this step, we reduce the variance by using further smoothing where we recommend using a multi-fold cross-validation (2.2) in choosing the optimal bandwidth. To obtain the two-step estimator of $\beta_j(\cdot)$ at a fixed point u_0 , we start by locally approximating the function $\beta_j(\cdot)$ in the neighborhood of a fixed point u_0 as follows:

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j(u_0)(u - u_0) \equiv a_j^* + b_j^*(u - u_0).$$

This leads to minimizing the following local least squares in order to obtain the two-step estimator of $\beta_j(u_0)$:

$$l^*(a^*, b^*) = \sum_{i=1}^n [Y_i - f^*\{\mathbf{X}_i, a_j^* + b_j^*(U_i - u_0)\}]^2 K_{h_2}(U_i - u_0),$$

with respect to (a_j^*, b_j^*) , where $f^*(\cdot, \cdot)$ is the function $f(\cdot, \cdot)$ with all coefficients except $\beta_j(u_0)$ substituted with their corresponding initial estimators, and $K_{h_2}(t) = h_2^{-1} K(t/h_2)$ is the scaled kernel function $K(\cdot)$ with bandwidth h_2 . The final estimator of $\beta_j(u_0)$ is $\hat{a}^*(u_0)$.

In the first step of the proposed estimation, the iterative algorithm described in Section 2.1 is employed and the initial estimators $\hat{\beta}_j(U_i)$ for $i = 1, \dots, n$ and $j = 1, \dots, p$, are obtained. For minimization of the local likelihood in the second step, we propose an iterative linear regression algorithm similar to the one given in Section 2.1. We denote the values of a^* and b^* at the k th iteration as $a^{*(k)}$ and $b^{*(k)}$, respectively, and $f^{*'}(\mathbf{X}_i, \beta_j)$ is the derivative of $f^*(\mathbf{X}_i, \beta_j)$ with respect to β_j , that is, $\partial f^*(\mathbf{X}_i, \beta_j)/\partial \beta_j$. Let

$$\mathbf{F}_k^* = \begin{bmatrix} f^{*'}(\mathbf{X}_1, a_j^{*(k)} + b_j^{*(k)}(U_1 - u_0)) & (U_1 - u_0) f^{*'}(\mathbf{X}_1, a_j^{*(k)} + b_j^{*(k)}(U_1 - u_0)) \\ \vdots & \vdots \\ f^{*'}(\mathbf{X}_n, a_j^{*(k)} + b_j^{*(k)}(U_n - u_0)) & (U_n - u_0) f^{*'}(\mathbf{X}_n, a_j^{*(k)} + b_j^{*(k)}(U_n - u_0)) \end{bmatrix}_{n \times 2},$$

$$\mathbf{Y}_{i,k}^* = Y_i - f^*\{\mathbf{X}_i, a_j^{*(k)} + b_j^{*(k)}(U_i - u_0)\} \\ + \{a_j^{*(k)} + b_j^{*(k)}(U_i - u_0)\}^T f^{*'}\{\mathbf{X}_i, a_j^{*(k)} + b_j^{*(k)}(U_i - u_0)\},$$

and $\mathbf{Y}_k = (Y_{1,k}, \dots, Y_{n,k})^T$. Then, we update $\mathbf{a}_j^* = (a_j^*, b_j^*)^T$ according to

$$\mathbf{a}_j^{*(k+1)} = (\mathbf{F}_k^{*T} \mathbf{W} \mathbf{F}_k^*)^{-1} \mathbf{F}_k^{*T} \mathbf{W} \mathbf{Y}_k^*,$$

where $\mathbf{W} = \text{diag}(K_{h_2}(U_1 - u_0), \dots, K_{h_2}(U_n - u_0))$. The solution of this iterative linear regression algorithm will satisfy $\ell^*(a^*, b^*) = 0$, and the two-step estimator of $\beta_j(u_0)$ is given by \hat{a}_j^* .

2.3. ASYMPTOTIC PROPERTIES OF THE ONE-STEP ESTIMATOR

We now derive the asymptotic distributions of our one-step estimators $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. Define $\boldsymbol{\theta}(u_0) = (a_1, \dots, a_p, b_1, \dots, b_p)^T$ and $\hat{\boldsymbol{\theta}}(u_0) = (\hat{\mathbf{a}}^T, \hat{\mathbf{b}}^T)^T$. Further define $\mu_k = \int t^k K(t) dt$, $v_k = \int t^k K^2(t) dt$ and $\mathbf{H} = \text{diag}(1, h_1) \otimes \mathbf{I}_p$ with \otimes denoting the Kronecker product and \mathbf{I}_p the $p \times p$ identity matrix. Let $c(u)$ denote the marginal density of U , and let

$$\boldsymbol{\Gamma}_1(u_0) = E(f'\{\mathbf{X}; \boldsymbol{\beta}(u_0)\} [f'\{\mathbf{X}; \boldsymbol{\beta}(u_0)\}]^T | U = u_0)_{p \times p},$$

$$\boldsymbol{\Gamma}_2(u_0) = E(\sigma^2(u_0, \mathbf{X}) f'\{\mathbf{X}; \boldsymbol{\beta}(u_0)\} [f'\{\mathbf{X}; \boldsymbol{\beta}(u_0)\}]^T | U = u_0)_{p \times p}.$$

The following theorem establishes the asymptotic normality of the one-step estimators of the regression coefficient functions $\boldsymbol{\beta}(u)$.

Theorem 1. Under the regularity conditions (A)–(G) in the supplementary material, we have the following result for $\hat{\theta}(u_0)$:

$$\sqrt{nh_1} \left[\mathbf{H} \{ \hat{\theta}(u_0) - \theta(u_0) \} - \frac{h_1^2}{2(\mu_2 - \mu_1^2)} \begin{Bmatrix} (\mu_2^2 - \mu_1\mu_3)\boldsymbol{\beta}''(u_0) \\ (\mu_3 - \mu_1\mu_2)\boldsymbol{\beta}''(u_0) \end{Bmatrix} + o_p(h_1^2) \right] \\ \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Delta}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Delta}^{-1})$$

as $n \rightarrow \infty$, where

$$\boldsymbol{\Delta} = c(u_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \boldsymbol{\Gamma}_1(u_0) \quad \text{and} \quad \boldsymbol{\Lambda} = c(u_0) \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes \boldsymbol{\Gamma}_2(u_0).$$

If $K(\cdot)$ is symmetric, we obtain the following simplification:

$$\sqrt{nh_1} \left\{ \hat{\mathbf{a}}(u_0) - \boldsymbol{\beta}(u_0) - \frac{h_1^2 \mu_2}{2} \boldsymbol{\beta}''(u_0) + o_p(h_1^2) \right\} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma}(u_0))$$

as $n \rightarrow \infty$, where

$$\boldsymbol{\Sigma}(u_0) = v_0 \boldsymbol{\Gamma}_1^{-1}(u_0) \boldsymbol{\Gamma}_2(u_0) \boldsymbol{\Gamma}_1^{-1}(u_0) / c(u_0).$$

Theorem 2. Under the regularity conditions given in the supplementary material, we have

$$\mathbf{H}(\mathbf{F}^T \mathbf{W} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{W} \boldsymbol{\Omega} \mathbf{W} \mathbf{F} (\mathbf{F}^T \mathbf{W} \mathbf{F})^{-1} \mathbf{H} \xrightarrow{\mathcal{P}} \boldsymbol{\Delta}^{-1} \boldsymbol{\Lambda} \boldsymbol{\Delta}^{-1}$$

as $n \rightarrow \infty$, where $\mathbf{F} = \mathbf{F}_k$, $\boldsymbol{\Omega} = \text{diag}(e_1^2, \dots, e_n^2)$ with $e_i = Y_i - f\{\mathbf{X}_i, \hat{\boldsymbol{\beta}}(U_i)\}$.

The proofs of Theorems 1 and 2 are given in the supplementary material. Theorem 1 shows that when $K(\cdot)$ is symmetric, the asymptotic bias of $\hat{\mathbf{a}}(u)$ is the same as that in standard varying-coefficient models (Fan and Zhang 2008). Theorem 1 also establishes the asymptotic variance of $\hat{\mathbf{a}}(u)$; however, we need a consistent estimator of this asymptotic variance to obtain confidence intervals for $\boldsymbol{\beta}(u)$. To this end, we use Theorem 2 to derive a consistent estimator of the asymptotic covariance matrix of $\hat{\theta}(u_0)$ given in Theorem 1,

$$\widehat{\text{cov}}\{\hat{\theta}(u_0)\} = (\mathbf{F}^T \mathbf{W} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{W} \boldsymbol{\Omega} \mathbf{W} \mathbf{F} (\mathbf{F}^T \mathbf{W} \mathbf{F})^{-1}. \quad (2.3)$$

The accuracy of this estimator will be tested in our simulation studies. Thus, the $(1 - \alpha)100\%$ asymptotic pointwise confidence intervals for $\theta(u_0)$ can be given as $\hat{\beta}_j(u_0) \pm \mathcal{C}_{\alpha/2} [\widehat{\text{var}}\{\hat{\beta}_j(u_0)\}]^{1/2}$, where $\mathcal{C}_{\alpha/2}$ is the $100 \times (1 - \alpha/2)$ -percentile of $N(0, 1)$.

2.4. GENERALIZED F-TEST

It is of scientific interest to test whether the coefficient functions in (1.2) really change over the temperature T . In general, we may formulate this problem as a hypothesis test to study whether the coefficients of the nonlinear varying-coefficient model (1.4) actually vary over the covariate U as follows:

$$H_0 : \beta_j(U) \equiv \beta_{j0} \quad \text{vs} \quad H_1 : \beta_j(U) \neq \beta_{j0} \quad \text{for } j = 1, \dots, p,$$

where the β_{j0} 's are unknown constants. This hypothesis testing is challenging because the corresponding null hypothesis is parametric, whereas the alternative is nonparametric. Thus, the parameter space under the null hypothesis is finite dimensional, while it is infinite dimensional under the alternative hypothesis. Intuitively, we would compare the residual sum of squares under H_0 and under H_1 . Let $\hat{\beta}(\cdot)$ and $\hat{\beta}_0$ be the estimators of β under H_1 and H_0 , respectively. Denote by $RSS(H_1) = \sum_{i=1}^n [Y_i - f\{\mathbf{X}_i, \hat{\beta}(U_i)\}]^2$ and $RSS(H_0) = \sum_{i=1}^n [Y_i - f(\mathbf{X}_i, \hat{\beta}_0)]^2$, the residual sum of squares under H_0 and H_1 , respectively. Define

$$F_0 = \frac{RSS(H_0) - RSS(H_1)}{n^{-1}RSS(H_1)}.$$

Note that F_0 is similar to the F -test for linear regression models. However, the parameter space under H_1 is infinite dimensional. Recall that an F -test for linear hypothesis in linear regression models is equivalent to the likelihood ratio test under the normality assumption. Specifically, assume that ε_i are iid $N(0, \sigma^2)$; then, the local log-likelihood function conditional on $\{U_i, \mathbf{X}_i\}$ equals

$$\ell\{\beta(U_i)\} = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [Y_i - f\{\mathbf{X}_i, \beta(U_i)\}]^2.$$

Then the likelihood ratio test is defined by

$$T_0 = \ell_n(H_1) - \ell_n(H_0) = \frac{n}{2} \log \left\{ \frac{RSS(H_0)}{RSS(H_1)} \right\}, \quad (2.4)$$

where $\ell_n(H_0)$ and $\ell_n(H_1)$ are the maximum likelihood functions under the null and alternative hypotheses, respectively, and (2.4) is a monotone function of the ratio $RSS(H_0)/RSS(H_1)$. Noting that $\log(1+x) \approx x$ for small x by the Taylor expansion, we obtain

$$T_0 \approx \frac{1}{2} \frac{RSS(H_0) - RSS(H_1)}{n^{-1}RSS(H_1)} = \frac{1}{2} F_0$$

under H_0 . Therefore, motivated by the generalized likelihood ratio test proposed in Fan, Zhang, and Zhang (2001), we define the generalized F -test statistic to be

$$F = \frac{r_K}{2} F_0,$$

where $r_K = \{K(0) - 0.5 \int K^2(u) du\} / [\int \{K(u) - 0.5K * K(u)\} du]$, $K * K$ denotes the convolution of K , and $K(\cdot)$ is the kernel function used to estimate the regression coefficients. The value of r_K is 1.2000, 2.1153, 2.3061, 2.3797, and 2.5375 for uniform, Epanechnikov, biweight, triweight, and Gaussian kernel, respectively. Note here that the generalized F -test does not rely on the normality assumption of the error distribution in general.

Intuitively, under the alternative hypothesis, $RSS(H_0)$ would be larger than $RSS(H_1)$ which would make the test statistic F large and we would reject the null hypothesis. However, under H_0 , $RSS(H_0)$ will be close to $RSS(H_1)$; hence, F would be small and we would be in favor of H_0 . Fan, Zhang, and Zhang (2001) showed that the generalized likelihood ratio test statistic F has a χ^2 limiting distribution for a variety of models including Gaussian

white noise models, nonparametric regression models, varying-coefficient models and generalized varying-coefficient models. However, the null limiting distribution given in Fan, Zhang, and Zhang (2001) does not keep the type I error rate well. As demonstrated by Cai, Fan, and Li (2000), a bootstrap procedure may provide a better estimate of the null distribution with moderate sample size than the asymptotic distribution. Thus, we propose to use a nonparametric bootstrap method to estimate the null distribution. In the proposed nonparametric bootstrap procedure, we first estimate the model coefficients under the null and alternative hypotheses using the original data. Let $\tilde{\beta}(U_i)$ and $\hat{\beta}(U_i)$ denote the estimates of the model coefficients under the null and the alternative hypotheses, respectively, from which the test statistic F is computed. We then obtain residuals $e_i = Y_i - f\{X_i, \tilde{\beta}(U_i)\}$. Next, we obtain each bootstrap sample by generating data (Y_i^*, X_i, U_i) using the model

$$Y_i^* = f\{X_i, \tilde{\beta}(U_i)\} + e_i^*,$$

where e_i^* denotes the residuals sampled from $\{e_1, \dots, e_n\}$ with replacement. The distribution of the test statistics F^* computed from the bootstrap samples estimate the null distribution of F . We study the properties of the proposed generalized F -test by Monte Carlo simulations in Section 3. Note that in simulation studies and data application, we employ the one-step estimator to estimate the null distribution of the test statistics F , but the two-step estimator can also be used and the procedure would be the same.

3. SIMULATION STUDIES

In this section, we study the performance of the proposed one-step and two-step estimators via Monte Carlo simulation studies. All simulations were conducted using R. The Epanechnikov kernel, $K(t) = 0.75(1 - t^2)_+$, is used in our simulations. We evaluate the performance of the proposed procedures using root average squared error (RASE):

$$RASE_j = \left(\frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} \left[\frac{\beta_j(u_k) - \hat{\beta}_j(u_k)}{\max\{\beta_j(u)\} - \min\{\beta_j(u)\}} \right]^2 \right)^{1/2},$$

where $\{u_k, k = 1, \dots, n_{\text{grid}}\}$ is an equidistant set of grid points over the range of U with $n_{\text{grid}} = 200$ and $j = 1, \dots, p$. We further study the accuracy of the proposed estimator of the covariance matrix from Section 2.3.

Example 1. In this example, data are generated from the following nonlinear varying-coefficient model:

$$Y_i = \exp\{\beta_1(U_i) + X_i\beta_2(U_i)\} + \varepsilon_i,$$

where $\beta_1(U_i) = \sin(\pi U_i)$ and $\beta_2(U_i) = \sin\{4\pi(U_i - 1/8)\}$. The predictor and error variables, X_i and ε_i , are generated from two independent $N(0, 1)$, while U_i are generated from $U[0, 1]$, the uniform distribution on $[0, 1]$. We consider sample sizes $n = 250, 500$ and 1000 and 500 Monte Carlo runs for each sample size.

In our simulation, we generate several pilot simulation data sets, and use a cross-validation bandwidth selector to get an overall picture about the optimal bandwidth. To

Table 1. RASE for one-step estimates in Example 1.

h	$\beta_1(u)$			$\beta_2(u)$		
	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
$\frac{1}{2}h_1$	0.224(0.112)	0.160(0.048)	0.116(0.033)	0.117(0.058)	0.076(0.029)	0.049(0.018)
h_1	0.154(0.054)	0.114(0.038)	0.081(0.022)	0.083(0.030)	0.053(0.019)	0.035(0.009)
$2h_1$	0.198(0.125)	0.132(0.053)	0.095(0.028)	0.147(0.030)	0.088(0.017)	0.059(0.009)

Table 2. RASE for two-step estimates in Example 1.

h	$\beta_1(u)$			$\beta_2(u)$		
	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
$\frac{1}{2}h_2$	0.194(0.162)	0.128(0.084)	0.086(0.032)	0.106(0.061)	0.066(0.026)	0.043(0.013)
h_2	0.125(0.057)	0.095(0.038)	0.064(0.021)	0.081(0.038)	0.050(0.020)	0.033(0.010)
$2h_2$	0.191(0.128)	0.121(0.089)	0.075(0.035)	0.143(0.033)	0.082(0.019)	0.052(0.009)

save computing time, we fix the bandwidth to be close to the optimal ones from the pilot simulation data sets. Specifically, we set the bandwidth to be $h_1 = 0.10, 0.075$ and 0.06 for $n = 250, 500$ and 1000 , respectively, for the one-step procedure. We use $h_0 = h_1/2$ for the initial bandwidth of the two-step estimator and $h_2 = (0.2, 0.1), (0.15, 0.075), (0.12, 0.06)$ (corresponding to that used for $(\beta_1$ and $\beta_2)$) for the final bandwidth of the two-step estimator for $n = 250, 500$ and 1000 , respectively. It is of interest to examine the performance of the proposed estimators with a wide range of the bandwidth. Thus, for each sample size, we also set the bandwidth to be $h_1/2, 2h_1, h_2/2$ and $2h_2$, corresponding to undersmoothing and oversmoothing simulations.

The sample mean and sample standard deviation of the RASE values for the one-step and two-step procedures from the 500 replications are summarized in Tables 1 and 2, respectively. Figure 2 depicts the one-step estimates of the parameter functions and 95 % asymptotic point-wise confidence intervals based on a typical sample which has the median RASE value over the 500 Monte Carlo runs. The estimates target the underlying true coefficient functions in which the proposed confidence intervals get narrower with increasing sample size. In this simulation, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ admit different degrees of smoothness and the one-step estimate for $\beta_1(\cdot)$ is undersmoothed (Figure 2). Hence, we employ the two-step procedure. Figure 3 displays the two-step estimates of $\beta_1(\cdot)$ for each sample size based on a sample with the median RASE value over 500 Monte Carlo runs along with 95 % percentile bootstrap confidence intervals based on 500 bootstrap samples. According to Figure 3, the two-step estimates are smoother and closer to the underlying true coefficient function than the one-step estimates. The two-step estimates of $\beta_2(\cdot)$ are close to the one-step estimates and hence are omitted. Since the two varying-coefficient functions admit different degrees of smoothness in this simulation, the two-step estimator leads to efficiency gains across all sample sizes, as demonstrated with smaller RASE values in Table 2. Even though the two-step procedure outperforms the one-step procedure in this sim-

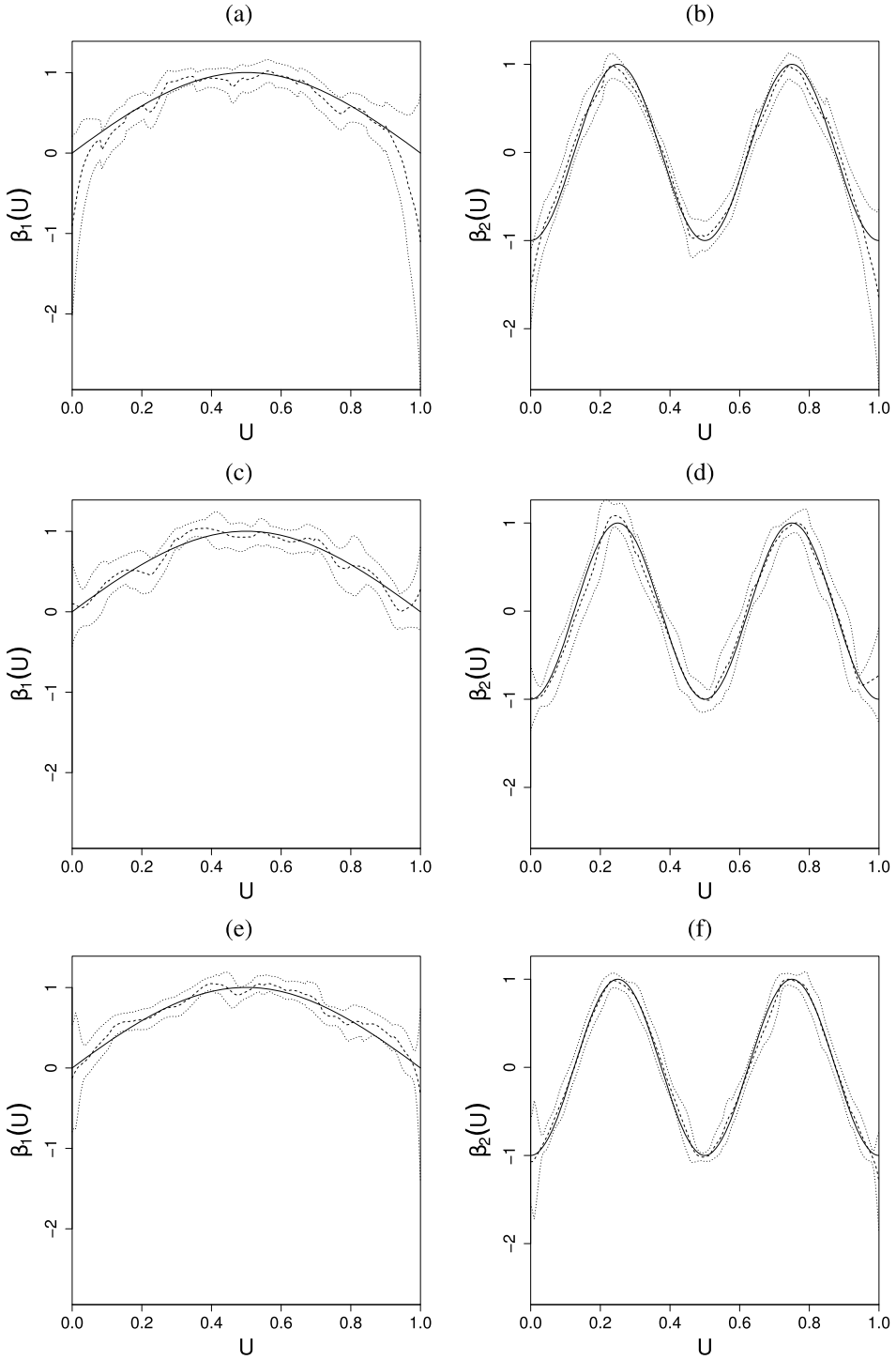


Figure 2. One-step estimates for the varying-coefficient functions (dashed) overlaying the true functions (solid) along with 95 % point-wise asymptotic confidence intervals (dotted) in (a) and (b) for $n = 250$, (c) and (d) for $n = 500$, (e) and (f) for $n = 1000$.

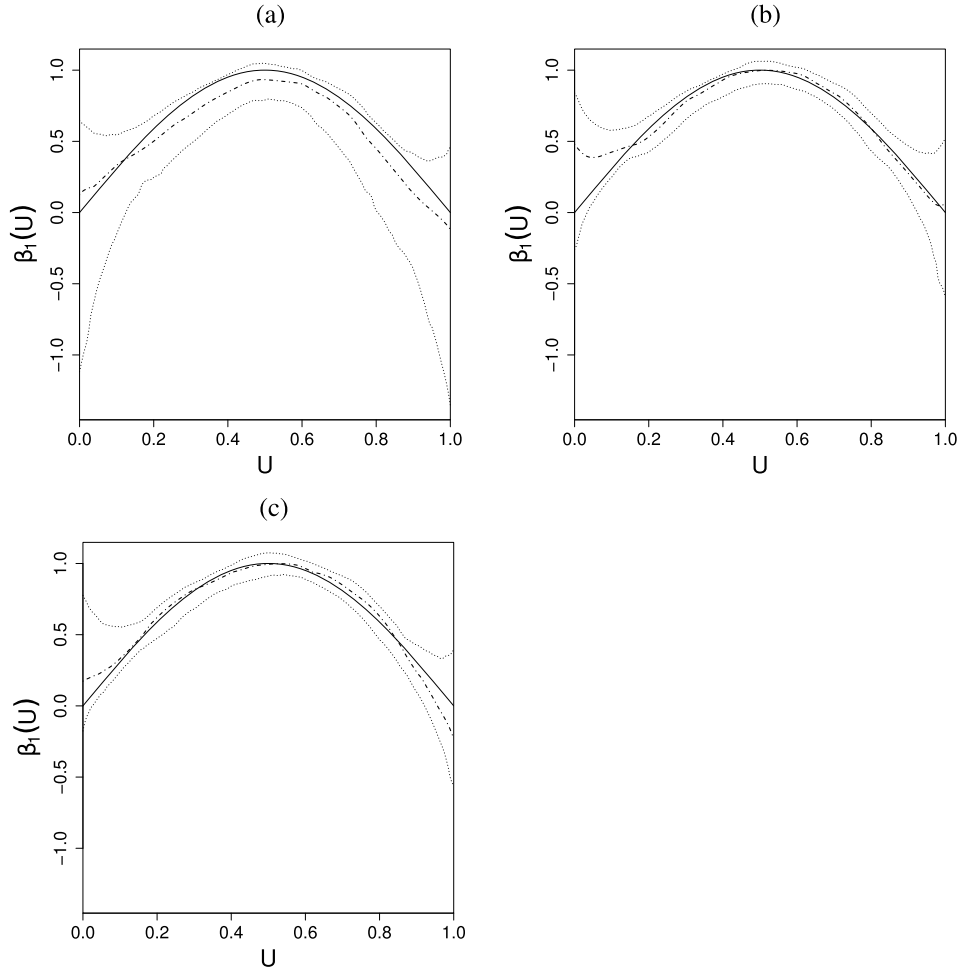


Figure 3. Two-step estimates of the varying-coefficient function $\beta_1(\cdot)$ (dash-dotted) overlaying the true functions (solid) and 95 % percentile bootstrap confidence intervals based on 500 bootstrap samples (dotted) in (a) for $n = 250$, (b) for $n = 500$, and (c) for $n = 1000$.

Table 3. Bias and standard errors for one-step estimates in Example 1 ($h = h_1$).

n	$\beta_1(u)$ with $u = 0.5$			$\beta_2(u)$ with $u = 0.5$		
	Bias	SD	SE(SD _{se})	Bias	SD	SE(SD _{se})
250	0.081	0.097	0.074(0.015)	0.125	0.112	0.064(0.024)
500	0.067	0.082	0.065(0.020)	0.081	0.081	0.053(0.021)
1000	0.046	0.056	0.048(0.014)	0.052	0.050	0.036(0.013)

ulation, it requires 3 times more computation time than the one-step procedure for $n = 500$ implemented with CPU 2.3 GHz Intel Core i5.

To examine the bias of the proposed one-step estimator, we also summarize the empirical bias of $\hat{\beta}_j(u)$. Table 3 reports the empirical bias at $u = 0.5$. The empirical bias of

$\hat{\beta}_j(u)$ at other locations is similar, and is not reported here to save space. We also test the accuracy of the proposed standard error formula in (2.3). The standard deviation, denoted by SD in Table 3, of 500 $\hat{\beta}_j(u)$, based on 500 simulations, can be viewed as the true standard error. The sample average and the sample standard deviation of the 500 estimated standard errors of $\hat{\beta}_j(u)$, denoted by SE and SD_{se} in Table 3, respectively, summarize the overall performance of the standard error formula (2.3). Table 3 presents the results at $u = 0.5$. Table 3 shows that the proposed covariance estimator slightly underestimates the true one, but the difference between SD and SE is less than twice of the SD_{se} . We conclude that the proposed covariance estimator works reasonably well (Cai, Fan, and Li 2000; Fan and Li 2004). Results for other locations are similar, and are not presented here to save space.

We next investigate the behavior of the generalized F -test and its power. It is expected that the null distribution approximately follows a χ^2 -distribution. Furthermore, the limiting χ^2 distribution does not depend on the specific null values considered. This property has been called the Wilks phenomenon in Fan, Zhang, and Zhang (2001). To illustrate the Wilks phenomenon, we consider

$$H_0 : \beta_j(U) = \gamma_{j0} \quad \text{vs} \quad H_1 : \beta_j(U) \neq \gamma_{j0}, \quad j = 1, 2, \quad (3.1)$$

with five different sets for $(\gamma_{10}, \gamma_{20})$: $(\mu_{\beta_1}, \mu_{\beta_2})$ and four combinations of $(\mu_{\beta_1} \pm 2\text{std}(\beta_1), \mu_{\beta_2} \pm 2\text{std}(\beta_2))$, where $\mu_{\beta_j} = E\{\beta_j(U)\}$ and $\text{std}(\beta_j) = \sqrt{\text{var}\{\beta_j(U)\}}$ for $j = 1, 2$.

A nonparametric bootstrap procedure with resampling from residuals is applied for $n = 500$ to estimate the null distribution of the test statistic F at different null values. After one sample is generated from the above simulation set-up, the residuals $e_i = Y_i - f(\mathbf{X}_i, \boldsymbol{\gamma}_0)$ are obtained where $\boldsymbol{\gamma}_0$ contains the null values. Next, 500 bootstrap samples $(Y_i^*, \mathbf{X}_i, U_i)$ of size $n = 500$ are generated under the null according to

$$Y_i^* = f(\mathbf{X}_i, \boldsymbol{\gamma}_0) + e_i^*,$$

where e_i^* is randomly drawn from $\{e_1, \dots, e_n\}$ with replacement. The test statistic is computed for each bootstrap sample. The estimated density of the test statistic F for five different null values is plotted in Figure 4(a) along with the density of the chi-square distribution. The degrees of freedom for the null distribution of the test statistic F is chosen to be close to the sample mean of bootstrap test statistic values across different null values, which is approximately 31.2. The plotted densities in Figure 4(a) are very close to each other and to the chi-square density, which confirms that the null distribution of the test statistic follows a χ^2 distribution with degrees of freedom 31.2 and that it is independent of the γ_{j0} null values. Hence, the Wilks phenomenon holds for the nonlinear varying-coefficient models. We also study the power of the proposed hypothesis testing procedure. Consider the hypotheses given in (3.1). The power is evaluated at a sequence of alternatives indexed by δ ,

$$H_1 : \beta_j(U, \delta) = (1 - \delta)\gamma_{j0} + \delta\beta_j^0(U),$$

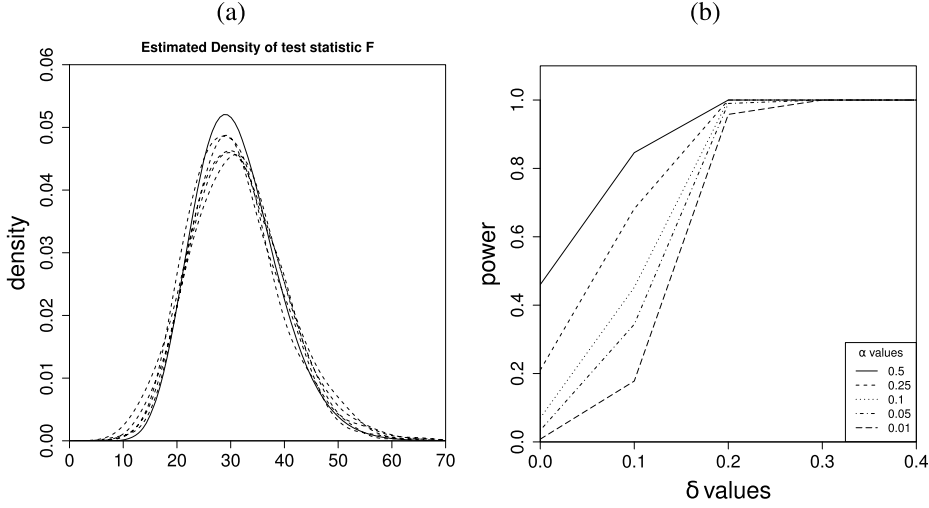


Figure 4. (a) Plots for estimated densities of the test statistic F from the proposed generalized F -test using five different null values. The solid line in (a) is the density function of a χ^2 distribution with 31.2 degrees of freedom; (b) is the estimated power functions.

where $j = 1, 2$ and $\delta \in [0, 0.4]$. Hence, larger values of δ correspond to further deviation from the null hypothesis. Figure 4(b) depicts the five power functions at five different significance levels: 0.5, 0.25, 0.10, 0.05 and 0.01. The estimated power at $\delta = 0$ for the five significance levels are 0.498, 0.242, 0.074, 0.048 and 0.012 based on 500 replications at $n = 500$. This implies that the test statistic keeps type I error rate very well. As expected, the power functions increase rapidly for increasing δ values.

Example 2. In this example, we generate data from

$$Y_i = \beta_1(U_i) - \frac{\beta_2(U_i)X_i}{X_i + \beta_3(U_i)} + \varepsilon_i,$$

which is the same model as that was fit to the ecological data. In our simulation, $\varepsilon_i \sim N(0, 0.1)$; the X_i follows a truncated normal distribution with mean and variance equal to the corresponding values of PAR variable in the ecological data and lies between the minimum value of PAR variable and the maximum value of PAR variable; and the U_i follows the uniform distribution over $[11.5, 24.7]$, which is the range for temperature in the ecological data. Furthermore, we set $\beta_1(U_i) = 0.5 \exp(0.5U_i)$, $\beta_2(U_i) = 23 + \sin(0.25U_i)$ and $\beta_3(U_i) = 300 + 2(U_i - 10)^2$ so that their ranges are similar to those of estimated parameters in Section 4. We consider three sample sizes, $n = 500, 1000$ and 2000 . We take the bandwidth to be $h_1 = 5.3, 4.6$ and 4.0 for $n = 500, 1000$ and 2000 , respectively, for the one-step procedure. We use $h_0 = h_1/2$ for the initial bandwidth of the two-step estimator and $h_2 = (5.3, 5.3, 5.3), (4.6, 4.6, 4.6), (4.0, 4.0, 4.0)$ (corresponding to that used for $(\beta_1, \beta_2, \beta_3)$) for the final bandwidth of the two-step estimator for $n = 250, 500$ and 1000 , respectively. For each sample size, we also set the bandwidth to be $h_1/2, 2h_1, h_2/2$ and $2h_2$, corresponding to undersmoothing and oversmoothing simulations for each sample size. We conduct 500 simulations for each case.

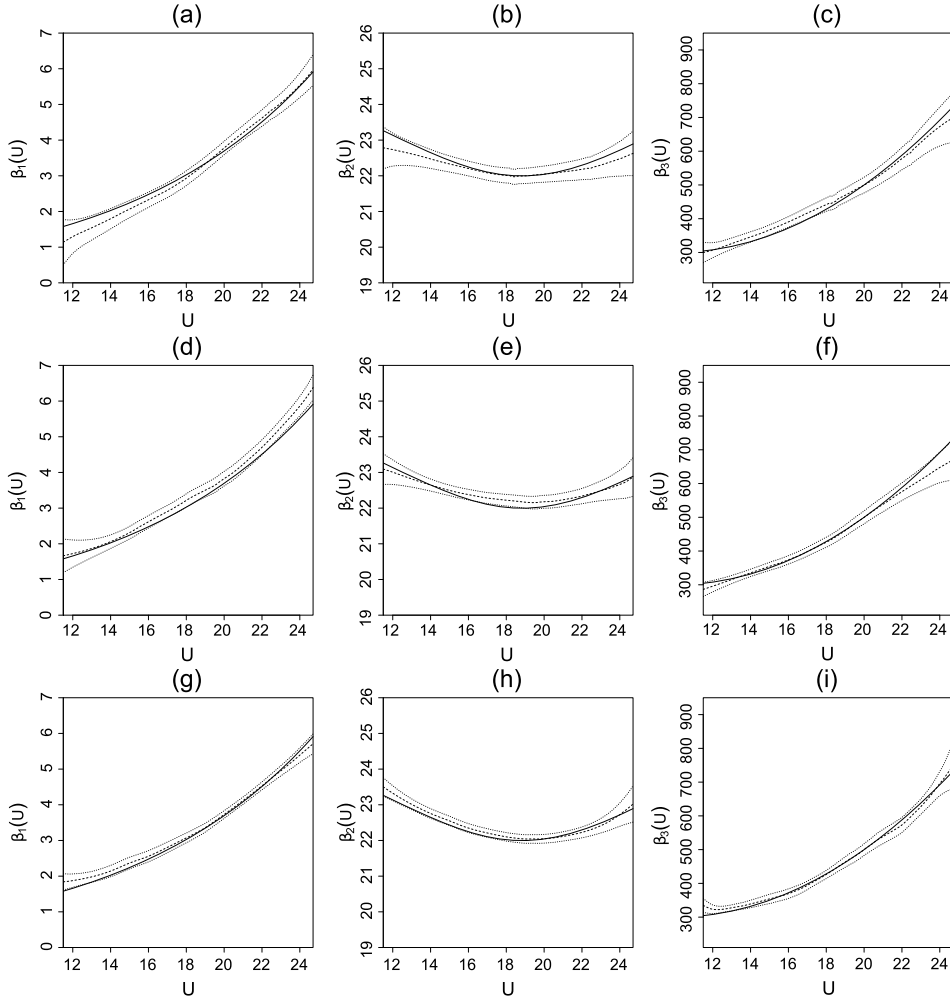


Figure 5. One-step estimates for the parameter functions and their confidence intervals. The dashed lines are the estimated parameter functions, the dotted lines are the corresponding 95 % confidence intervals, and the solid lines are the true parameter functions. (a), (b), and (c) are for $n = 500$, (d), (e), and (f) for $n = 1000$, (g), (h), and (i) for $n = 2000$.

The sample mean and standard deviation of the 500 RASE values for the one-step and two-step procedures are reported in Tables 4 and 5, respectively. Since $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\beta_3(\cdot)$ admit similar degrees of smoothness, the one-step and two-step procedures perform similarly as expected and the RASEs decrease as the sample size increases. Figure 5 depicts the one-step estimates of the parameter functions and their 95 % point-wise asymptotic confidence intervals based on a typical sample which has the median RASE value over the 500 Monte Carlo runs. The two-step estimates are similar and are not displayed here. The estimated coefficient functions are close to the underlying true coefficient functions. In the second simulation, the two-step procedure requires 6 times more computation time than the one-step procedure for $n = 1000$ implemented with CPU 2.3 GHz Intel Core i5.

Table 4. RASE for one-step estimates in Example 2.

h	β_1			β_2			β_3		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
$\frac{1}{2}h_1$	0.051(0.019)	0.035(0.012)	0.026(0.007)	0.172(0.060)	0.119(0.038)	0.091(0.025)	0.046(0.019)	0.034(0.012)	0.025(0.008)
h_1	0.046(0.019)	0.031(0.012)	0.023(0.008)	0.143(0.056)	0.102(0.038)	0.077(0.026)	0.041(0.019)	0.031(0.013)	0.022(0.009)
$2h_1$	0.105(0.027)	0.078(0.016)	0.056(0.011)	0.211(0.059)	0.159(0.041)	0.119(0.028)	0.110(0.020)	0.083(0.015)	0.061(0.011)

Table 5. RASE for two-step estimates in Example 2.

h	β_1			β_2			β_3		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
$\frac{1}{2}h_2$	0.044(0.020)	0.032(0.012)	0.023(0.008)	0.159(0.058)	0.111(0.037)	0.080(0.024)	0.041(0.019)	0.031(0.013)	0.023(0.009)
h_2	0.041(0.019)	0.029(0.012)	0.021(0.008)	0.142(0.052)	0.101(0.030)	0.077(0.018)	0.040(0.018)	0.030(0.013)	0.022(0.009)
$2h_2$	0.046(0.015)	0.037(0.010)	0.031(0.007)	0.210(0.039)	0.158(0.017)	0.118(0.009)	0.052(0.012)	0.042(0.009)	0.035(0.006)

Table 6. Bias and standard errors for one-step estimates in Example 2 ($h = h_1$).

n	$\beta_1(u)$ with $u = 18$			$\beta_2(u)$ with $u = 18$			$\beta_3(u)$ with $u = 18$		
	Bias	SD	SE(SD _{se})	Bias	SD	SE(SD _{se})	Bias	SD	SE(SD _{se})
500	0.132	0.145	0.115(0.024)	0.126	0.116	0.106(0.010)	9.265	10.633	9.586(1.204)
1000	0.088	0.095	0.087(0.014)	0.094	0.081	0.079(0.005)	7.091	7.522	7.212(0.668)
2000	0.073	0.075	0.066(0.008)	0.078	0.065	0.060(0.003)	5.369	5.872	5.441(0.394)

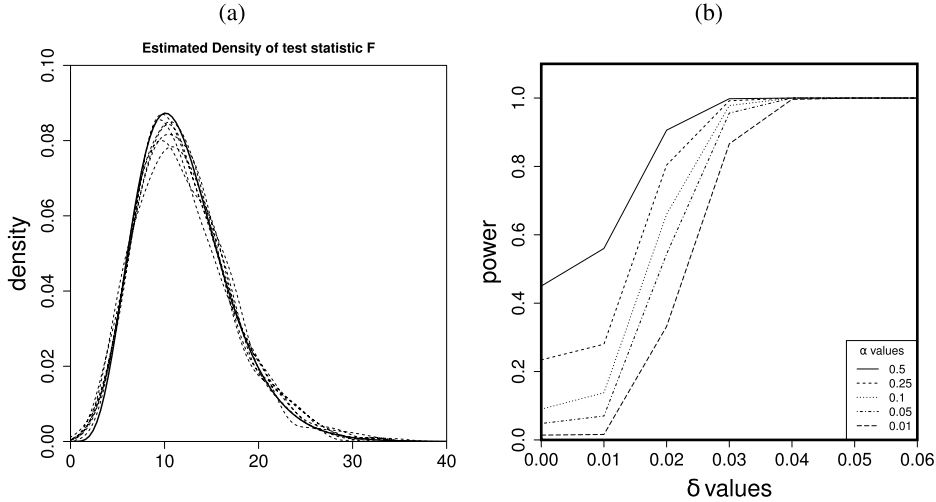


Figure 6. (a) Plots for estimated densities of the test statistic F from the proposed generalized F -test using seven different null values. In (a), the solid curve is the density of a χ^2 distribution with 12.1 degrees of freedom. (b) Estimated power functions.

Table 6 depicts the empirical bias, the SD, the SE and the SD_{se} of $\hat{\beta}_i(u)$ at $u = 18$. Note that the range of $\beta_3(\cdot)$ is much larger than those for $\beta_1(\cdot)$ and $\beta_2(\cdot)$. Thus, the bias and SD for $\beta_3(U)$ at $u = 18$ is much larger than those for the other two coefficients. The overall pattern in Tables 4 and 6 is similar to that in Tables 1 and 3. The RASE and bias decrease as the sample size increases. The SE is slightly less than the SD, but their difference is still less than two times the SD_{se} . This implies that the proposed procedures work reasonably well.

Similar to Example 1, we illustrate the Wilks phenomenon by considering

$$H_0 : \beta_j(U) = \gamma_{j0} \quad \text{vs} \quad H_1 : \beta_j(U) \neq \gamma_{j0}, \quad j = 1, 2, 3,$$

with seven different sets for $(\gamma_{10}, \gamma_{20}, \gamma_{30})$: $(\mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3})$, $(\mu_{\beta_1} \pm 2\text{std}(\beta_1), \mu_{\beta_2}, \mu_{\beta_3})$, $(\mu_{\beta_1}, \mu_{\beta_2} \pm 2\text{std}(\beta_2), \mu_{\beta_3})$, and $(\mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3} \pm 2\text{std}(\beta_3))$, where $\mu_{\beta_j} = E\{\beta_j(U)\}$ and $\text{std}(\beta_j) = \sqrt{\text{var}\{\beta_j(U)\}}$ for $j = 1, 2, 3$. Figure 6(a) depicts the kernel density estimates of the null distributions with these seven different sets for $(\gamma_{10}, \gamma_{20}, \gamma_{30})$ based on 500 bootstrap samples. The seven densities of the null distributions are very close to each other. This implies the null distribution is not sensitive to the value of $(\gamma_{10}, \gamma_{20}, \gamma_{30})$. Figure 6(a)

shows that the null density is very close to a χ^2 distribution with 12.1 degrees of freedom. Similar to the first simulation study, the value of the degrees of freedom is obtained by computing the mean of the test statistic values based on 500 bootstrap samples. This implies that the Wilks phenomenon holds for the nonlinear varying-coefficient models.

Figure 6(b) depicts the power functions at levels 0.5, 0.25, 0.10, 0.05 and 0.01 under the following alternative hypothesis:

$$H_1 : \beta_j(U, \delta) = (1 - \delta)\gamma_{j0} + \delta\beta_j^0(U),$$

where $\gamma_{j0} = E\{\beta_j(U)\}$ for $j = 1, 2, 3$ and $\delta \in [0, 0.015]$. The estimated power at $\delta = 0$ for the five significance levels are 0.450, 0.234, 0.090, 0.048 and 0.014 based on 500 Monte Carlo simulations when $n = 500$. Again, the test procedure keeps type I error rate very well for this model, and the power functions increase rapidly for increasing δ values.

4. APPLICATION TO THE ECOLOGICAL DATA

In this section, we apply the methodologies proposed in Section 2 for the ecological data set introduced in Section 1. The data set was collected by the AmeriFlux network during summer growing seasons (from June 1 to August 31) of 1993–1995. This study is a part of the U.S. Climate Change Research Initiative conducted by the North American Carbon Program (NACP) which is a component of U.S. Interagency Carbon Cycle Science Program. The objective of AmeriFlux in this project is to help NACP in modeling the causes of variation in the net exchange of CO_2 leading to general principles that can be applied more broadly. Let Y denote the response variable NEE, X denote PAR and U denote temperature. We fit the following nonlinear varying-coefficient model:

$$Y = \beta_1(U) - \frac{\beta_2(U)X}{X + \beta_3(U)} + \varepsilon,$$

using the iterative algorithm proposed in Section 2.1. Since photosynthetic activity is low at extreme temperatures, we analyze a data subset consisting of observations with temperature values between 11.5 and 24.7. The total sample size of this data subset is $n = 4636$. The bandwidth was chosen via a multi-fold cross-validation. Specifically, we minimize the cross-validation score

$$CV(h) = \sum_j \|Y_j - \hat{Y}_{-j}\|^2,$$

where \hat{Y}_{-j} denotes the fitted value with j th group of 20 subjects excluded. This results in a bandwidth of $h = 3.0$.

Figures 7(a), (b), and (c) depict the estimated regression coefficients along with the proposed 95 % point-wise asymptotic confidence intervals. From Figures 7(a), (b), and (c), we can see that the estimated coefficient functions vary with temperature; as will be described below the proposed F -test rejects the null hypothesis of a regular nonlinear model in favor of the nonlinear varying-coefficient model with a p-value of 0.00036. In our data, we have mixed species plants in a deciduous broadleaf forest. Figure 7(a) shows that these

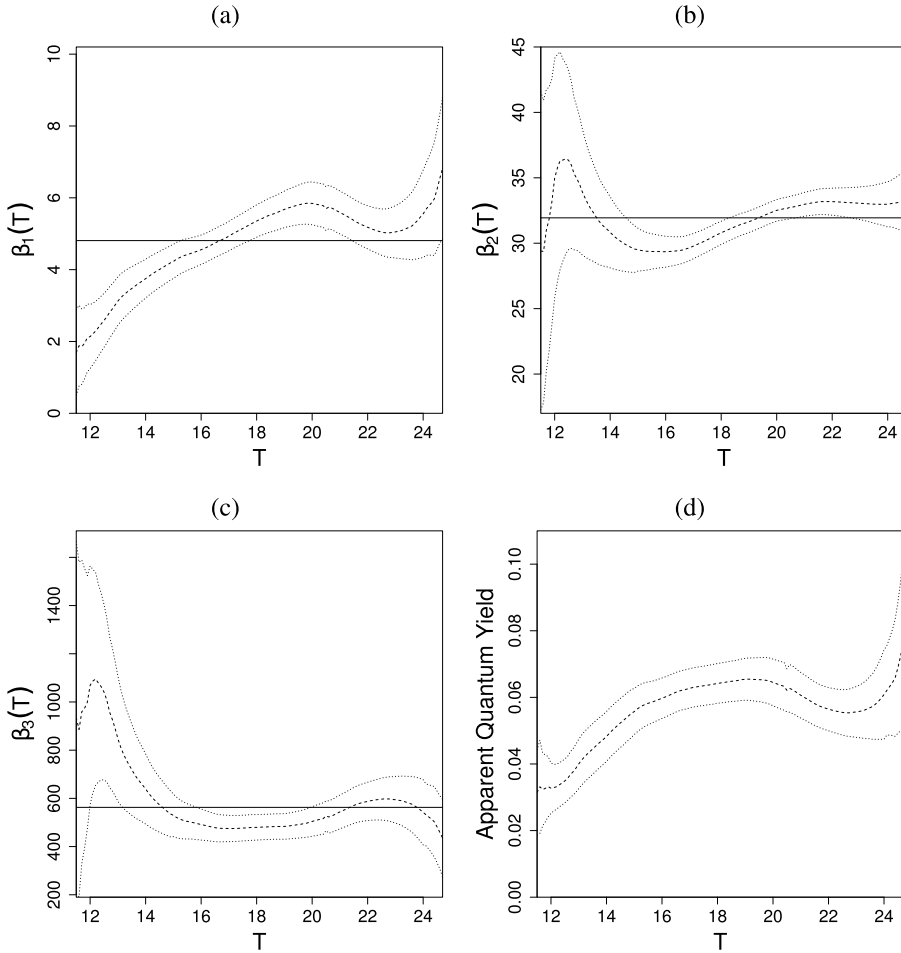


Figure 7. (a), (b), and (c) One-step estimates for the varying-coefficient functions (dashed) of the nonlinear varying-coefficient model fit to the ecological data along with 95 % point-wise asymptotic confidence intervals (dotted). The solid lines are the estimate of β_j under the null hypothesis. (d) Estimated apparent quantum yield (solid) along with 95 % point-wise asymptotic confidence intervals (dotted).

plants have higher dark respiration activity at higher temperatures ($T \geq 16$ °C) than lower temperatures ($T < 16$ °C). According to Figure 7(b), at their light-saturation point, the species in our sample achieve the maximum net photosynthetic rate $\beta_2(\cdot)$ of 36.5 around 12.5 °C, beyond which the net photosynthetic rate experiences a period of decrease and stabilizes. In other words, at their light saturation point, maximum photosynthetic capacity occurs at lower temperatures and the lowest activity is observed around 16 °C. Similarly $\beta_3(\cdot)$ achieves its maximum value shortly after 12 °C, and beyond which point it declines and stabilizes. Figure 7(d) shows the estimated apparent quantum yield $\hat{\beta}_2(\cdot)/\hat{\beta}_3(\cdot)$ along with the 95 % point-wise asymptotic confidence intervals. The apparent quantum yield increases with temperature until 19.5 °C and stabilizes for higher temperatures. Note that the proposed inference assumes i.i.d. observations and hence confidence intervals may be

narrower than those that would take into account the correlation structure over time in the data set. Extension of the proposed inference framework for correlated data is of interest.

This application illustrates the advantages of the proposed nonlinear varying-coefficient models outlined in the introduction. First, the dimension is reduced by regression NEE on only PAR in a local temperature window rather than considering a two-dimensional non-parametric regression of NEE on both PAR and temperature. Second, the proposed model reduces the modeling bias by allowing the regression coefficients to vary over temperature; the nonlinear varying-coefficient model fit leads to smaller multi-fold cross-validation error compared to the regular nonlinear regression model fit. Finally, as illustrated by the above interpretations, the fits from the nonlinear varying-coefficient model can be interpreted as describing photosynthesis via separate nonlinear regression models for each fixed temperature value. This parallels the interpretation that a varying-coefficient model can be viewed as a linear regression model at a fixed $U = u$.

For comparison, we also obtained the two-step estimates proposed in Section 2.2. The initial bandwidth is chosen to be $h_0 = 1.5$, which is half of the bandwidth obtained by using cross-validation in the one-step procedure. After obtaining the initial estimators, we apply a cross-validation bandwidth selector for each coefficient separately, and obtain the optimal bandwidths $h_2 = 2.0, 2.5$ and 3.0 for $\beta_1(\cdot)$, $\beta_2(\cdot)$, and $\beta_3(\cdot)$, respectively. Figure 8 shows the two-step estimates of regression coefficients along with 95 % percentile bootstrap confidence intervals based on 500 bootstrap samples. To have a sensible comparison between the two-step procedure and the one-step procedure, we rerun the one-step procedure with bandwidths 2.0, 2.5 and 3.0. These estimates are also shown in Figure 8. The two-step estimates are very close to the one-step estimates; however, they are smoother. To compare the two-methods, we also computed the ratio of residual sum of squares, that is, $RSS(one_step)/RSS(two_step)$, and found this ratio to be very close to 1. The two-step procedure requires 8 times more computation time than the one-step procedure implemented with CPU 2.3 GHz Intel Core i5.

We apply the testing procedure proposed in Section 2.4 to test whether the regression coefficients are actually varying over temperature. In order to do so, we fit the nonlinear model (1.1) to the data and use the resulting estimates $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (4.809, 31.940, 562.677)$ as the null value for testing whether the functions are constant over temperature. Figure 9 depicts the estimated density function of the null distribution of the test statistic F , overlaying a chi-square density. The degrees of freedom of the chi-square density is chosen close to the sample mean of test statistic values obtained from 500 bootstrap samples, which is approximately 18.5. The figure is consistent with the Wilks phenomenon stating that under the null hypothesis the chi-square distribution is a good approximation for the asymptotic behavior of F . We reject the null hypothesis at the significance level $\alpha = 0.05$ in favor of the newly proposed nonlinear varying-coefficient model with a test statistic of 46.209 and a p-value of 0.00036. In other words, the proposed nonlinear varying-coefficient model fits the data better than the nonlinear regression model (1.1).

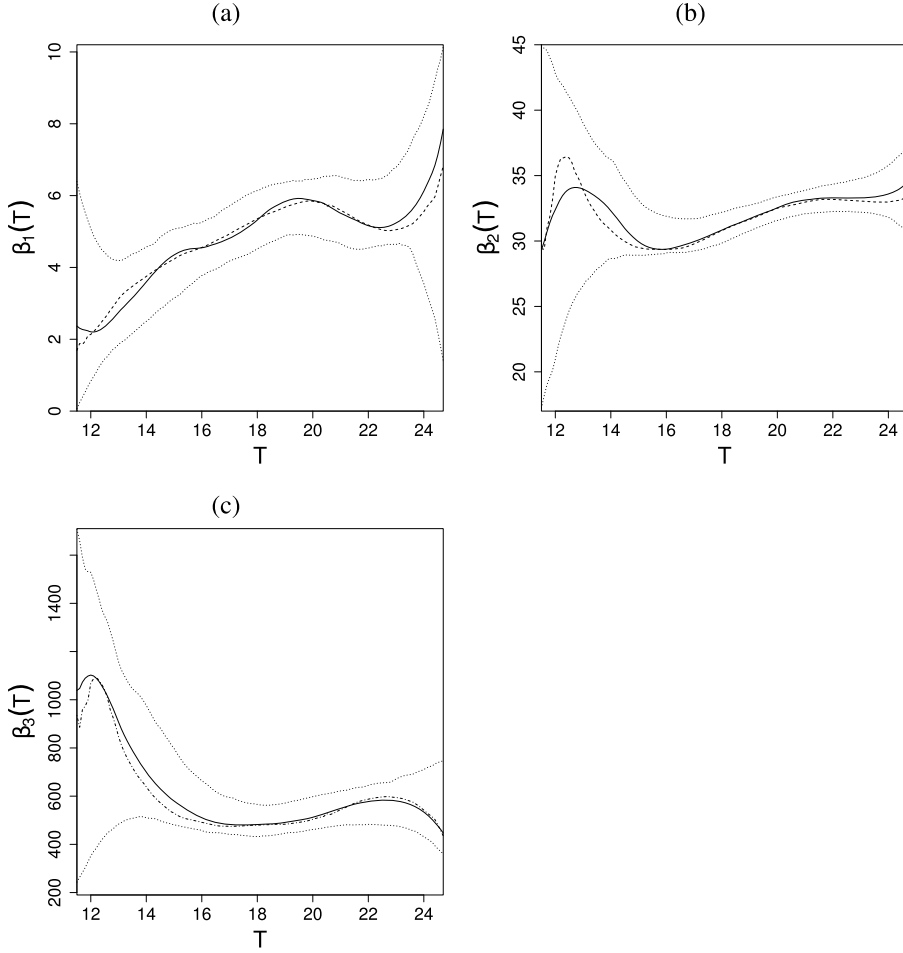


Figure 8. Two-step estimates for the varying-coefficient functions (solid) of the nonlinear varying-coefficient model fit to the ecological data along with one-step estimates (dashed) and 95 % percentile bootstrap confidence intervals based on 500 bootstrap samples (dotted).

5. DISCUSSION

Motivated by an empirical analysis of a real data set, we proposed nonlinear varying-coefficients models. We further developed two estimation procedures based on local linear regression techniques and studied their finite sample properties. Similar to estimation in regular varying-coefficient models, the two-step estimators for nonlinear varying-coefficient models also lead to efficiency gains when the varying-coefficient functions admit different degrees of smoothness, and are less computationally efficient than the one-step estimators. Asymptotic properties of the one-step estimator is established and those for the two-step estimator requires further research. Alternative estimation approaches for nonlinear varying-coefficient models may be developed via basis expansions such as smoothing splines (Hastie and Tibshirani 1993). Although splines have many advantages, some

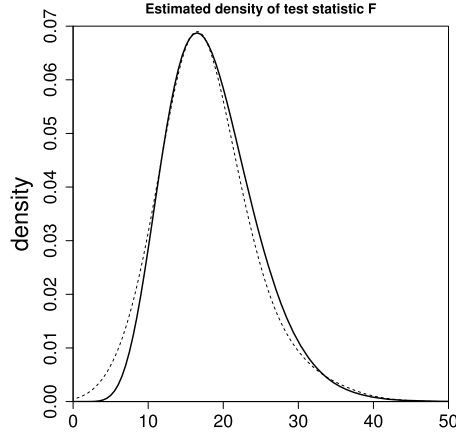


Figure 9. Estimated density of the test statistic F (dashed) based on 500 bootstrap samples, overlaying the density of a χ^2 distribution with 18.5 degrees of freedom (solid).

challenges such as the difficulty in choosing all the smoothing parameters simultaneously and their computational intensity have been reported (Fan and Zhang 1999).

The generalized F -test is proposed for testing whether the nonlinear varying-coefficient model reduces to a nonlinear regression model, i.e. whether all the varying-coefficient functions are constant over U . For assessing features of single varying-coefficient functions, we proposed asymptotic and bootstrap confidence intervals. Developing valid hypothesis testing procedures for specific features of single varying-coefficient functions would also be of interest and requires further research.

The proposal is based on a univariate index covariate U . Estimation in varying-coefficient models in higher index dimensions quickly runs into issues of ‘curse of dimensionality’. Proposals for incorporating multi-dimensional indices include Xia and Li (1999), Fan, Yao, and Cai (2003) where authors consider a linear combination of a multi-dimensional U vector to create a one-dimensional index. Similar extensions can be considered in the case of nonlinear varying-coefficient models.

As extensions of generalized varying-coefficient models and model (1.4), it may be of interest to consider a new class of following models:

$$E(Y|\mathbf{X}, U) = g[f\{\mathbf{X}, \boldsymbol{\beta}(U)\}], \quad (5.1)$$

where $f(\cdot, \cdot)$ is pre-specified, while $g(\cdot)$ is unspecified, nonparametric smooth function. This class of models include (1.4) as a special case by taking the link function $g(\cdot)$ to be the identical link. With a pre-specified link function, model (5.1) also includes generalized varying-coefficient models as special cases by taking $f(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}^T \boldsymbol{\beta}(U)$. Fan, Yao and Cai (2003) proposed adaptive varying-coefficient models, in which the link function $f(\cdot)$ is an unknown nonparametric smooth function, and $f(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}^T \boldsymbol{\beta}(U)$. The authors proposed an approach to estimating the unknown link function and $\boldsymbol{\beta}(\cdot)$ iteratively.

In practice, data may be collected over time. Thus, the data would be correlated. The proposed estimation procedures are directly applicable for correlated data. It will involve theoretical techniques in time series to establish the sampling properties of the proposed

procedure for correlated data. Thus, it may be an interesting future research topic to establish theoretical property of our proposed procedure under the setting of dependent data.

ACKNOWLEDGEMENTS

Li's research was supported by National Institute on Drug Abuse grant P50-DA10075, National Cancer Institute (NCI) grant R01 CA168676, and National Natural Science Foundation of China grant 11028103. Şentürk research was supported by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant R01 DK092232.

[Received November 2012. Accepted August 2013. Published Online September 2013.]

REFERENCES

- Breiman, L., and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580–619.
- Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.
- Cleveland, W., Grosse, E., and Shyu, W. (1992), "Local Regression Models," in *Statistical Models in S*, eds. J. M. Chambers and T. J. Hastie, Pacific Grove: Wadsworth & Brooks/Cole, pp. 309–376.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.
- Fan, J., and Li, R. (2004), "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, 99, 710–723.
- Fan, J., and Zhang, W. (1999), "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics*, 27, 1491–1518.
- Fan, J., and Zhang, W. (2008), "Statistical Methods with Varying Coefficient Models," *Statistics and Its Interface*, 1, 179–195.
- Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *Annals of Statistics*, 29, 153–193.
- Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive Varying-Coefficient Linear Models," *Journal of the Royal Statistical Society, Series B*, 65, 57–80.
- Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines," *Annals of Statistics*, 19, 1–67.
- Garrett, R. H., and Grisham, C. M. (2005), *Biochemistry* (3rd ed.), Belmont: Thomson Learning.
- Geisser, S. (1975), "The Predictive Sample Reuse Method with Applications," *Journal of the American Statistical Association*, 70, 320–328.
- Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*, London: Chapman and Hall.
- Gu, C., and Wahba, G. (1993), "Smoothing Spline Anova with Component-Wise Bayesian "Confidence Intervals"," *Journal of Computational and Graphical Statistics*, 2, 97–117.
- Härdle, W., and Stoker, T. M. (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.
- Hastie, T., and Tibshirani, R. (1990), *Generalized Additive Models*, New York: Chapman and Hall.
- Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society, Series B*, 55, 757–796.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficient Models with Longitudinal Data," *Biometrika*, 85, 809–822.

- Kauermann, G., and Tutz, G. (1999), "On Model Diagnostics Using Varying Coefficient Models," *Biometrika*, 86, 119–128.
- Li, K.-C. (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association*, 86, 316–327.
- Monteith, J. L. (1972), "Solar Radiation and Productivity in Tropical Ecosystems," *Journal of Applied Ecology*, 9, 747–766.
- Osborne, M. R. (1992), "Fisher's Method of Scoring," *International Statistical Review/Revue Internationale de Statistique*, 60, 99–117.
- Seber, G. A. F., and Wild, C. J. (2003), *Nonlinear Regression*, New York: Wiley-Interscience.
- Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997), "Polynomial Splines and Their Tensor Products in Extended Linear Modeling: 1994 Wald Memorial Lecture," *Annals of Statistics*, 25, 1371–1470.
- Wahba, G. (1984), "Partial Spline Models for the Semiparametric Estimation of Functions of Several Variables," in *Statistical Analysis of Time Series, Proceedings of the Japan U.S. Joint Seminar*, Tokyo: Institute of Statistical Mathematics, pp. 319–329.
- Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998), "Asymptotic Confidence Regions for Kernel Smoothing of a Varying-Coefficient Model with Longitudinal Data," *Journal of the American Statistical Association*, 93, 1388–1402.
- Xia, Y., and Li, W. K. (1999), "On Single-Index Coefficient Regression Models," *Journal of the American Statistical Association*, 94, 1275–1285.