A new estimation procedure for partially nonlinear model via mixed effects approach

Runze Li

Department of Statistics

Pennsylvania State University

University Park, PA 16802

Lei Nie

Department of Biostatistics and Biomathematics

Georgetown University

Washington, DC 20057

August 5, 2005

Abstract

Semiparametric regression models have been utilized extensively in the literature. Typical approaches to semiparametric regression models involve some smoothing techniques. In this paper, we shall consider a particular semiparametric nonlinear regression model, where the nonparametric component of the model will be viewed as a nuisance parameter, while the focus is on the parameters of the parametric component. The parameters will be estimated through a nonlinear mixed-effects model approach. Under certain regularity conditions, we show that the proposed estimate is consistent and it follows an asymptotic normal distribution. The proposed method was illustrated through an example from a study of ecology. A limited simulation study was conducted to investigate the finite population properties of the proposed estimators.

Keywords: Nonlinear mixed effects model, partially linear model, partially nonlinear model

Abbreviated Title: Partially Nonlinear Model.

1 Introduction

Nonlinear regression models have been widely studied in statistical and econometric literature. Many interesting examples and applications of nonlinear regression models are given in Gallant (1987) and Seber and Wild (1989). Motivated by an empirical study in the fields of ecology, we consider a partially nonlinear model

$$Y = \alpha(U) + g(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon, \tag{1.1}$$

where Y is the response variable, $\{U, \mathbf{x}\}$ is the associated covariates, $\alpha(\cdot)$ is an unknown smooth function, $g(\mathbf{x}; \boldsymbol{\beta})$ is a pre-specified function, $\boldsymbol{\beta}$ is a d-dimensional parameter vector, and ε is random error with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2$. It is also assumed that ε is independent of $(U, \mathbf{x})$.

When the baseline function $\alpha(U)$ is a constant, model (1.1) reduces to a nonlinear regression model. When $g(\mathbf{x}; \boldsymbol{\beta}) = \mathbf{x}^T\boldsymbol{\beta}$, then model (1.1) reduces to a partially linear model,

$$Y = \alpha(U) + \mathbf{x}^T\boldsymbol{\beta} + \varepsilon, \tag{1.2}$$

The partially linear model has been popular in the statistical literature (Engle, *et al.*, 1986, Heckman, 1986, Speckman, 1988 and among others). Härdle, Liang and Gao (2000) gave a systematic study of the partially linear models. Some of estimation procedures for (1.2), such as backfitting algorithm and profile likelihood approach, can be extended for model (1.1). Li and Nie (2004) proposed two estimation procedures for model (1.1) using profile nonlinear least squares approach and linear approximation approach. In this paper, we aim to develop a simple and easily implemented estimation procedure for $\boldsymbol{\beta}$.

In many situations as in the example presented in Section 3.3, our primary interest is to estimate $\boldsymbol{\beta}$, and the nonparametric function $\alpha(U)$ may be viewed as a nuissance parameter. This allows us to develop a simple estimation procedure for $\boldsymbol{\beta}$. We first establish a connection between the partially nonlinear model (1.1) and a nonlinear mixed-effects model. Our approach can be viewed as an extension or improvement of the difference-based method by Yatchew (1997) and Fan and Huang (2003). Readers are referred to Section 2 for details. Based on the nonlinear mixed-effects model, we propose a new estimation procedure for $\boldsymbol{\beta}$ by using simple mixed-effects model techniques. We will further demonstrate the effectiveness of the proposed estimation procedure. Root n consistency and asymptotic normality of the resulting estimates are established. A robust standard error formula for the resulting estimate is proposed by using a sandwich formula and is empirically tested by Monte Carlo simulation studies. The simulation results showed that our approach improved significantly over the previous difference-based methods by Yatchew (1997) and Fan and Huang (2003), for partially linear models. Simulation results and a real data analysis also demonstrate the usefulness of our approach in partially nonlinear models.

The rest of this paper is organized as follows. In Section 2, we propose an estimation procedure for $\boldsymbol{\beta}$ by employing existing estimation procedures for nonlinear mixed-effects models. In Section 3, we investigate the finite sample performance of the proposed estimator by a limited simulation study. We also illustrate the proposed methodology by analysis of an example from the study of ecology. All theoretical proofs are presented in the Appendix.

2 New Estimation Procedure for $\boldsymbol{\beta}$

Suppose that we have a random sample $(\mathbf{x}_i, u_i, y_i)$, $i = 1, \cdots, n$ from model (1.1). Reorder the data from the least to largest according to the value of the variable $\{u_i\}$. Denote $(\mathbf{x}_{(i)}, u_{(i)}, y_{(i)})$, $i = 1, \cdots, n$ to be the ordered sample. Thus,

$$y_{(i)} = \alpha(u_{(i)}) + g(\mathbf{x}_{(i)}; \boldsymbol{\beta}) + \varepsilon_{(i)}.$$

Since U and ε are independent, $\varepsilon_{(1)}, \cdots, \varepsilon_{(n)}$ are still independently and identically distributed to the distribution of ε. In practice, statistical inference for $\boldsymbol{\beta}$ will condition on

$\{(u_i, \mathbf{x}_i), i = 1, \cdots, n\}$. Thus, for ease of presentation and with a slight abuse of notation, denote $(\mathbf{x}_i, u_i, y_i)$, $i = 1, \cdots, n$, to be the ordered sample. Thus,

$$y_i = \alpha(u_i) + g(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i.$$

First of all, observe that for $i = 1, \cdots, n-1$

$$y_{i+1} - y_i = \alpha(u_{i+1}) - \alpha(u_i) + g(\mathbf{x}_{i+1}; \boldsymbol{\beta}) - g(\mathbf{x}_i; \boldsymbol{\beta}) + e_i, \tag{2.1}$$

where stochastic error $e_i = \varepsilon_{(i+1)} - \varepsilon_{(i)}$. If $\alpha(u)$ is a smooth function of u, Yatchew (1997) suggested the use of the following approximation, for partially linear models,

$$\alpha(u_{i+1}) - \alpha(u_i) \approx 0, \tag{2.2}$$

when u_i is close to u_{i+1}, and further advocated using the ordinary least squares to estimate $\boldsymbol{\beta}$. In order to obtain a better approximation, Fan and Huang (2001) proposed to approximate $\alpha(u_{i+1}) - \alpha(u_i)$ linearly:

$$\alpha(u_{i+1}) - \alpha(u_i) \approx \alpha_0 + \alpha_1(u_{i+1} - u_i), \tag{2.3}$$

and suggest using the ordinary least squares to estimate $\boldsymbol{\beta}$. Although there are many existing estimation approach for partially linear models, with the approximation (2.2) or (2.3), the unknown coefficient $\boldsymbol{\beta}$ in the partially linear models can be easily estimated by using least squares approach without involving any smoothing techniques. Simulation results presented in Fan and Huang (2003) are encouraging. However, the sampling property of Fan and Huang's proposal has not been studied yet. In this section, we explore this difference-based approach further for the partially nonlinear model (1.1).

The approximation (2.3) can be viewed as a **global** linear approximation to $\alpha(u_{i+1}) - \alpha(u_i)$, where the first derivative of $\alpha(U)$ is viewed as a constant. In general, the coefficient α_0 and α_1 in (2.3) may vary over the locations of u_i and u_{i+1}. Thus, we consider a **local** linear approximation:

$$\alpha(u_{i+1}) - \alpha(u_i) \approx \alpha_{i0} + \alpha_{i1}(u_{i+1} - u_i). \tag{2.4}$$

In the presence of ties among observation times u_i, this linear approximation to $\alpha(u_{i+1}) - \alpha(u_i)$ still holds. Thus, we have the following approximation:

$$y_{i+1} - y_i \approx \alpha_{i0} + \alpha_{i1}(u_{i+1} - u_i) + g(\mathbf{x}_{i+1}; \boldsymbol{\beta}) - g(\mathbf{x}_i; \boldsymbol{\beta}) + e_i. \tag{2.5}$$

The local linear approximation (2.4) introduces $2(n-1)$ nuisance parameters: α_{i0} and α_{i1}, $i = 1, \cdots, n-1$. However, when the space between u_i and u_{i+1} is small, and under some mild conditions, $\alpha(u_{i+1}) - \alpha(u_i)$ is negligible. Thus, to reduce the number of nuisance parameters, we view α_{i0} and α_{i1} as uncorrelated random effects: $(\alpha_{i0}, \alpha_{i1})$, $i = 1, \cdots, n-1$ are independently and identically distributed with mean (α_0^*, α_1^*) and variance (σ_0^2, σ_1^2). Now model (2.5) becomes an approximate nonlinear mixed-effects model with linear random effects, see e.g. Vonesh and Carter (1992). Hence nonlinear weighted least squares approaches can be directly applied to partially nonlinear models.

From our simulation experience, the resulting estimate with $(\alpha_0^*, \alpha_1^*) = (0, 0)$ is almost the same as that treating (α_0^*, α_1^*) as unknown parameters. Thus, we may view α_{i0} and α_{i1} as random effects with zero mean. Thus, conditioning on $\{(u_i, \mathbf{x}_i), i = 1, \cdots, n\}$

$$E(y_{i+1} - y_i) \approx g(\mathbf{x}_{i+1}; \boldsymbol{\beta}) - g(\mathbf{x}_i; \boldsymbol{\beta}). \tag{2.6}$$

Since $e_1, \cdots, e_{n-1}$ are correlated, a weighted least squares should be used to incorporate the correlation structure. Denote $\mathbf{y}_D = (y_2 - y_1, \cdots, y_n - y_{n-1})^T$ and $g_D(\boldsymbol{\beta}) = \{g(\mathbf{x}_2; \boldsymbol{\beta}) - g(\mathbf{x}_1; \boldsymbol{\beta}), \cdots, g(\mathbf{x}_n; \boldsymbol{\beta}) - g(\mathbf{x}_{n-1}; \boldsymbol{\beta})\}^T$. Consider a general weighted nonlinear least squares problem

$$S(\boldsymbol{\beta}) = \{\mathbf{y}_D - g_D(\boldsymbol{\beta})\}^T W \{\mathbf{y}_D - g_D(\boldsymbol{\beta})\}, \tag{2.7}$$

where W is a weight matrix and usually is called a working covariance matrix. Minimizing (2.7) yields a weighted nonlinear least squares estimate for $\boldsymbol{\beta}$. For example, taking W to be an identity matrix, the resulting estimate corresponds to the one with the approximation (2.2) and ignoring correlation of e_i's. Although all of these approximation models, including model (2.5), are not the actual true model, they provide some good approximations. The model (2.5) motivates us to construct a good weighted matrix $W = (w_{ij})_{n-1,n-1}$ to characterize the variance and covariance matrix of $e_i's$, which is given by

$$w_{ii} = 2\sigma_\varepsilon^2 + \sigma_0^2 + \sigma_1^2(u_{i+1} - u_i)^2, \tag{2.8}$$

and for $i \neq j$,

$$w_{ij} = -\sigma_\varepsilon^2. \tag{2.9}$$

In practical implementation, we may further assume that the random effects in (2.7) are normally distributed. Thus, the unknown parameters in (2.8) and (2.9) can be substituted by their maximum likelihood estimate (MLE) or restricted maximum likelihood estimate (REML). The root n consistency and asymptotic normality of the resulting weighted nonlinear least squares estimator are established in Theorem 1 below. In practice, statistical inference for $\boldsymbol{\beta}$ usually is conditioning on $\{(u_i, \mathbf{x}_i), i = 1, \cdots, n\}$. Therefore, to avoid a lot of tedious notations, we consider $\{(u_i, \mathbf{x}_i), i = 1, \cdots, n\}$ to be fixed in the following theorem. Denote $\boldsymbol{\beta}_0$ to be the true value of $\boldsymbol{\beta}$.

Theorem 1 *Suppose that Conditions (A)—(F) given in the Appendix hold. With probability tending to one as $n \to \infty$, there exists a minimizer $\hat{\boldsymbol{\beta}}$ of $S(\boldsymbol{\beta})$, such that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$ as $n \to \infty$, and further,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{D}} N(0, \Sigma^{-1}\Sigma^*\Sigma^{-1}),$$

where "$\xrightarrow{\mathrm{D}}$" stands for convergence in distribution, Σ and Σ^* are given in the Appendix.

The proof of Theorem 1 is given in the Appendix. The matrices Σ and Σ^* depend on the unknown parameter $\boldsymbol{\beta}_0$ and cannot be directly used to calculate the standard error of the resulting estimate. Following conventional techniques, we estimate the standard error using the sandwich formula. Denote by $S'(\boldsymbol{\beta})$ and $S''(\boldsymbol{\beta})$ the gradient vector and the Hessian matrix of $S(\boldsymbol{\beta})$, respectively. The covariance matrix of $\hat{\boldsymbol{\beta}}$ can be estimated by the following sandwich formula:

$$\{S''(\hat{\boldsymbol{\beta}}))\}^{-1}\widehat{\text{cov}}\{S'(\hat{\boldsymbol{\beta}})\}\{S''(\hat{\boldsymbol{\beta}}))\}^{-1}. \tag{2.10}$$

From our simulation studies in Section 3, the sandwich formula performs quite well with moderate sample size.

We will empirically compare efficiency of the proposed estimate with the profile likelihood estimate under the framework of partially linear models in Section 3. From our simulation studies, the proposed estimate is almost as efficient as that of profile likelihood estimates.

3 Numerical Study and Application

In Section 3.1 and 3.2, we assess the finite sample performance of the proposed estimator by Monte Carlo simulation. In Section 3.3, we further illustrate the proposed methodology by an application to a real data example. All simulations were conducted by using SAS. Programs are available upon request.

3.1 Comparison of Efficiency

We first conduct some comparison of proposed estimate in terms of efficiency. For the partially linear model,

$$Y = \alpha(U) + \mathbf{x}^T\boldsymbol{\beta} + \varepsilon,$$

where ε is a random error with mean zero, the profile likelihood approach proposed in Severini and Staniswalis (1994) yields a semiparametric efficient estimator for $\boldsymbol{\beta}$ in the sense of Bickel *et al.* (1993). The proposed estimation procedure for $\boldsymbol{\beta}$ in Section 2 can be directly applied to the partially linear model. This allows us to compare the efficiency of the resulting estimate with the profile likelihood approach by Monte Carlo simulation study. In our simulation, we consider the following two baseline functions,

$$\alpha_1(U) = \sin(2\pi U), \quad \text{and} \quad \alpha_2(U) = U^2.$$

We generate ε from $N(0,1)$ and U from $U(0,1)$, the uniform distribution over [0,1]. the covariate vector $\mathbf{x} = (x_1, x_2)^T$ was simulated from a normal distribution with mean zero and $\text{cov}(x_i, x_j) = 0.5^{|i-j|}$, and $\beta_1 = \beta_2 = 1$. We take sample size $n = 200$ and 400. For each case, we conduct 1000 Monte Carlo simulations. For the profile likelihood approach, we use local linear fitting to estimate the baseline function. Simulation results are summarized in Table 1, in which "Mean" stands for the average of the 1000 estimates of β's, and "SD" is the standard deviation of the 1000 estimates of β's and can be regarded as the true standard error of β's. From Table 1, we can see that the proposed estimation procedure for $\boldsymbol{\beta}$ is almost as efficient as a semiparametric efficient estimator.

The proposed estimation procedure is to minimize the weighted nonlinear least squares (2.7). A natural question arising here is how much efficiency would be lost if one simply uses either the approximation in (2.2) proposed in Yatchew (1997) or the global linear approximation in (2.3) proposed in Fan and Huang (2001). To address this question, we conducted 1000 simulations with data being generated from the partially linear model. The simulation results are also summarized in Table 1, in which columns labeled "Yatchew" and "FH" stand for the proposal of Yatchew (1997) and Fan and Huang (2001), respectively. From Table 1, the performances of these two proposals are similar, and the proposed mixed effects approach is more efficient than the methods of Yatchew and Fan & Huang. For example, their relative efficiency for $\hat{\beta}_1$ with $n = 200$ is $1.3506(= (0.086/0.074)^2)$. Thus, the proposed mixed effect approach gains about 35% efficiency. Furthermore, the proposed approach is less biased than the methods of Yatchew and Fan & Huang as we use local linear approximation to $\alpha(U)$.

Table 1: Comparison of efficiency

		Profile	New	Yatchew	FH
β	n	Mean (SD)	Mean(SD)	Mean(SD)	Mean(SD)
			$\alpha(U) = \sin(2\pi U)$		
β_1	200	0.996(0.072)	1.000(0.074)	1.000(0.086)	1.000(0.087)
β_2	200	0.995(0.073)	0.999(0.073)	1.002(0.086)	1.002(0.087)
β_1	400	1.000(0.049)	1.000(0.051)	1.065(0.060)	1.060(0.062)
β_2	400	1.000(0.051)	1.000(0.051)	1.083(0.061)	1.083(0.061)
			$\alpha(U) = U^2$		
β_1	200	0.997(0.0717)	0.999(0.0721)	1.000 (0.086)	1.000(0.087)
β_2	200	0.997(0.0722)	0.998(0.0727)	1.002(0.087)	1.002(0.087)
β_1	400	0.999(0.049)	1.001(0.051)	0.999(0.060)	0.999(0.060)
β_2	400	1.000(0.050)	1.002(0.051)	1.004(0.061)	1.004(0.061)

Results for other cases are similar.

3.2 Performance of $\hat{\boldsymbol{\beta}}$

Now we access the finite sample performance of the proposed estimator. In our simulation, we generate random samples from the following model

$$Y = \alpha(U) + g(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon, \tag{3.1}$$

where ε, U and $\alpha(U)$ are the same as those in Section 3.1. Here we consider two nonlinear functions $g(\cdot;\cdot)$:

$$g_1(x; \beta_1, \beta_2) = -\beta_1 x/(x + \beta_2)$$

with $\beta_1 = 18$ and $\beta_2 = 0.8$, which were chosen to be close the estimate for the real data example in Section 3.3, and $x \sim N(0, 1)$; and

$$g_2(x_1, x_2; \beta_1, \beta_2) = 10 \exp(\beta_1 x_1 + \beta_2 x_2)/\{1 + \exp(\beta_1 x_1 + \beta_2 x_2)\}$$

with $\beta_1 = \beta_2 = 1$, The covariate vector $(x_1, x_2)^T$ was simulated from a normal distribution with mean zero and $\text{cov}(x_i, x_j) = 0.5^{|i-j|}$. In our simulation, sample size $n = 200$ and 400. For each case, we conduct 1000 Monte Carlo simulation.

Simulation results for $\hat{\boldsymbol{\beta}}$ are summarized in Table 2, in which mean stands for the average of the 1000 estimates of β's, SD is the standard deviation of the 1000 estimates of β's and can be regarded as the true standard error of β's. SE and Std(SE) are the average and the standard deviation of the 1000 standard errors using the sandwich formula (2.10). We further compute the coverage probability (CP) of a 95% confidence interval. From Table 2, we can see that the bias of $\hat{\boldsymbol{\beta}}$ is very small. This means that the proposed estimation procedure works well. The SD and the SE are very close, and in comparison with the standard deviation of SE, their difference is negligible. Furthermore, the coverage probability of a 95% confidence interval is very close to 0.95. This indicates the sandwich formula performs well.

Table 2: Finite Sample Performance of $\hat{\boldsymbol{\beta}}$'s

(n, α)	Mean	SD	SE(Std(SE))	CP	Mean	SD	SE(Std(SE))	CP
				$g = g_1$				
			$\beta_1 = 18$				$\beta_2 = 0.8$	
(200, α_1)	18.29	1.51	1.44(0.46)	0.943	0.83	0.16	0.16(0.05)	0.948
(400, α_1)	18.14	1.01	0.97(0.20)	0.951	0.81	0.11	0.11(0.02)	0.952
(200, α_2)	18.27	1.50	1.42(0.46)	0.945	0.83	0.16	0.15(0.05)	0.949
(400, α_2)	18.14	1.00	0.97(0.20)	0.949	0.81	0.11	0.10(0.02)	0.951
				$g = g_2$				
			$\beta_1 = 1$				$\beta_2 = 1$	
(200, α_1)	1.002	0.060	0.058(0.005)	0.938	1.001	0.058	0.058(0.005)	0.947
(400, α_1)	1.001	0.042	0.041(0.002)	0.939	1.002	0.040	0.041(0.003)	0.960
(200, α_2)	1.002	0.061	0.059(0.005)	0.945	1.001	0.059	0.059(0.005)	0.945
(400, α_2)	1.001	0.042	0.042(0.002)	0.940	1.002	0.041	0.042(0.002)	0.965

3.3 Application

We illustrate the proposed methodology by an analysis of a real data set from the field of ecology. Of interest in this example is to study how temperature affects the relationship between the response of net ecosystem-atmosphere exchange of CO_2 (NEE) and the photosynthetically active radiation (PAR). This data set consists of 1997 observations of NEE, PAR and temperature (T), and was collected over a subalpine forest at approximately 3050 meter elevation above sea level by using three-dimensional sonic anemometers on hundreds of meter towers during parts of the growth season of 1999.

As an illustration, we consider a partially nonlinear model

$$\mathrm{NEE} = R(T) - \frac{\beta_1 \mathrm{PAR}}{\mathrm{PAR} + \beta_2} + \varepsilon. \tag{3.2}$$

The proposed estimation procedure for $\boldsymbol{\beta}$ is used to fit the data set using Model (3.2). It yields that $\hat{\beta}_1 = 17.77$ with standard error 0.415, and $\hat{\beta}_2 = 0.838$ with standard error 0.073.

To estimate the baseline function, we define partial residuals as follows

$$e_i = \mathrm{NEE}_i - \frac{\hat{\beta}_1 \mathrm{PAR}_i}{\mathrm{PAR}_i + \hat{\beta}_2}.$$

This yields a synthetic nonparametric model with single covariate

$$e_i = R(T_i) + \varepsilon_i.$$

Estimation $R(T)$ can be easily carried out by smoothing the partial residuals over temperature by various existing 1-d nonparametric smoothing procedures, such as splines smoothing and local polynomial regression. Here we direct call SAS procedure LOESS to estimate $R(T)$ with smooth parameter $s = 0.4$. The resulting estimate is depicted in Figure 1, from which

(a) Plot of $\hat{\alpha}(T)$

(b) Q-Q plot of generalized F-test

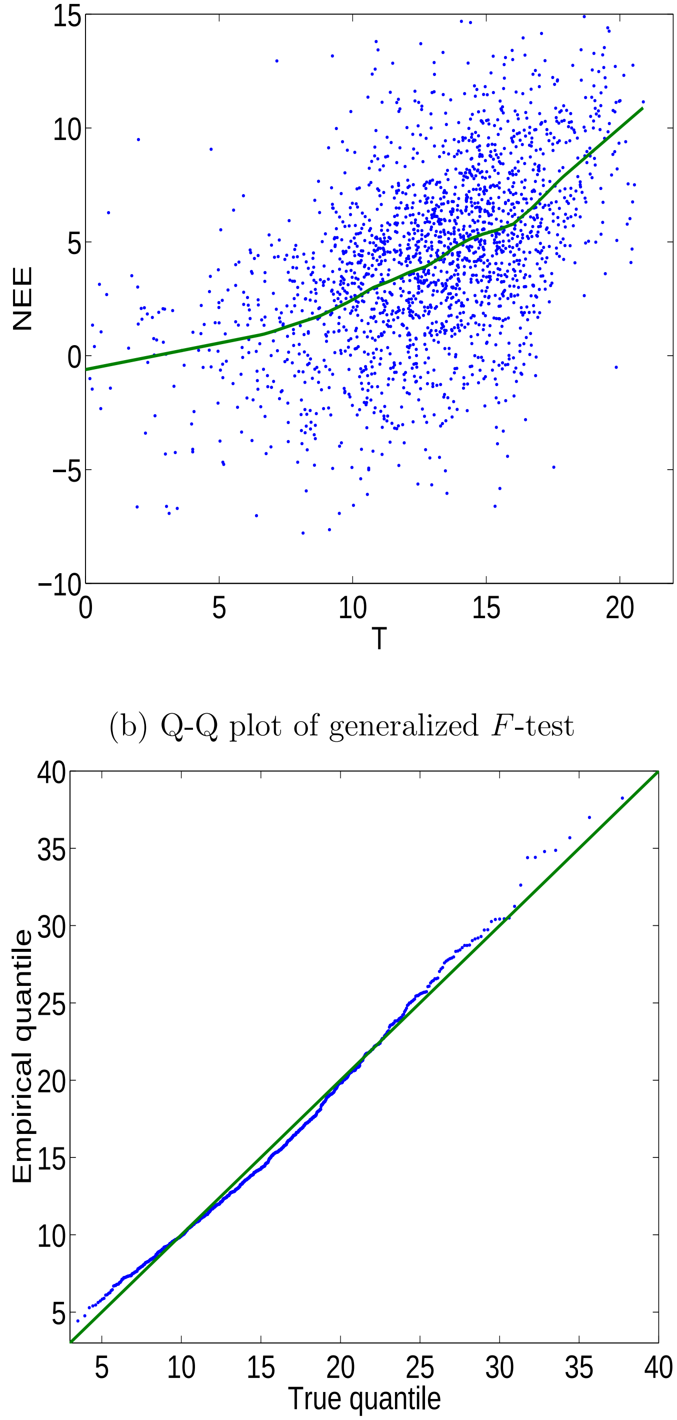

Figure 1: *(a) Plot of $\hat{\alpha}(T)$. The solid line is an estimate of the baseline function, and the dots are partial residuals: $y_i - \hat{\beta}_1 x_i/(x_i + \hat{\beta}_2)$. (b) Q-Q plot of generalized F-test.*

we can see that the partial residuals have an increasing trend over temperature. The residual sum of squares equals 24944.3486.

If $R(T)$ does not depend on T, model (3.2) reduces to a nonlinear regression model

$$\text{NEE} = R - \frac{\beta_1 \text{PAR}}{\text{PAR} + \beta_2} + \varepsilon, \tag{3.3}$$

The model (3.3) was used by Montheith (1972) and more recent work by Ruimy et al., 1999 and references therein. The model in (3.3) was originally adopted because the data were collected from a laboratory in which climate variables, such as temperature and moisture can be well controlled. However, the temperature for an ecosystem cannot be controlled, and the parameter $R(T)$ may depends on the temperature. Thus, it is of interest to compare model fitting using model (3.3) and model (3.2). Using nonlinear least squares approach, we obtain the resulting estimate for parameters in model (3.3) given by $\hat{R} = 3.0277$, $\hat{\beta}_1 = 14.0113$ and $\hat{\beta}_2 = 0.6272$. The residual sum of squares for model (3.3) is 32270.1502. Compared with model (3.3), model (3.2) dramatically reduces the residual sum of squares.

It is of scientific interest to test whether $R(T)$ really varies over the temperature. This can be formulated as a nonparametric hypothesis testing problem:

$$H_0 : R(T) = R_0 \quad \text{versus} \quad H_1 : R(T) \neq R_0 \tag{3.4}$$

for some unknown constant R_0. Motivated by F-test in theory of linear regression models, we should compare the residual sum of squares under H_0 and under H_1. We next extend an F-test for (3.4). Denote $\text{RSS}(H_0)$ and $\text{RSS}(H_1)$ to be the residual sum of squares under H_0 and under H_1, respectively. Note that the traditional F-test is not well defined since the dimension of parameter space under H_0 is finite, while it is infinite under H_1. Define a generalized F-test statistic

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

Bootstrap method can be used to calculate the critical value for the generalized F-test. In this example, the observed F-value is

$$F = \frac{32270.1502 - 24944.3486}{24944.3486/1997} = 586.4906$$

with P-value < 0.001 obtained by using 1000 bootstraps. Thus, the dark respiration rate depends on temperature.

Let us intuitively discuss the null distribution of the generalized F-test. Note that $\hat{\boldsymbol{\beta}}$ is root n consistent. Parametric convergence rate is faster than the nonparametric convergence rate of $R(T)$. Thus, the asymptotic distribution of F will be the same as that substituting with the true value of $\boldsymbol{\beta}$ in F. It can be shown that as $n \to \infty$, $n^{-1}\text{RSS}(H_1)$ tends to σ_ε^2 almost sure. Intuitively, $\text{RSS}(H_0) - \text{RSS}(H_1)$ tends to a scaled χ^2-square distribution under H_0, although its degrees of freedom may depend on the smoothing parameter used to estimate $R(T)$ and may diverge. Thus the asymptotic null distribution of the generalized F will be a scaled χ^2-square distribution. To empirically demonstrate this, we plot quantiles of a χ^2-square distribution versus the sample quantiles of a scaled generalized F-test, defined by $F_c = CF$, where C is determined by the relation between mean and variance of χ^2-square distribution, i.e, $C = 2E(F)/\text{Var}(F)$, which can be estimated by plugging-in their bootstrap sample counterpart. Figure 1(b) depicts the Q-Q plot, in which the degrees of freedom of the χ^2-square distribution is estimated by the mean of F_c and equals 15. The Q-Q plot is consistent with our intuitive analysis.

4 Concluding Remarks

A simple and effective estimation procedure was proposed for the partially nonlinear model by using weighted nonlinear least squares approach, through a nonlinear mixed-effects model. The covariate U is assumed to be univariate in previous section, however, the proposed method can be easily extended to the multivariate case where $\mathbf{U}$ is a vector. Let $\|\mathbf{U}\|^2 = \mathbf{U}^T\mathbf{U}$. Reorder the data from the least to largest according to $\|u_i\|^2$. In case $\|\mathbf{u}\|^2 = \|\mathbf{v}\|^2$, then we treat $\mathbf{u}$ and $\mathbf{v}$ be two "words" (composed using 10 numbers rather than 26 letters) and order them by the standard alphabetical order. We may extend (2.4) to

$$\alpha(u_{i+1}) - \alpha(u_i) \approx \alpha_{i0} + \boldsymbol{\alpha}_{i1}^T(\mathbf{u}_{i+1} - \mathbf{u}_i), \tag{4.1}$$

and (2.5) to

$$y_{i+1} - y_i \approx \alpha_{i0} + \boldsymbol{\alpha}_{i1}^T(\mathbf{u}_{i+1} - \mathbf{u}_i) + g(\mathbf{x}_{i+1}; \boldsymbol{\beta}) - g(\mathbf{x}_i; \boldsymbol{\beta}) + e_i. \tag{4.2}$$

Thus, the proposed estimation procedure in Section 2 is applicable for multivariate $\mathbf{U}$.

Acknowledgement

The authors thank Dr. C. Yi for providing data analyzed in Section 3.2. Li's research was supported by NSF grants DMS 0322673 and DMS-0348869.

References

Bickel, P.J., Klaassen, A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Inference in Semiparametric Models.* John Hopkins University Press, Baltimore.

Cai, Z, Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.*, **95**, 888-902.

Engle, R.F., Granger,C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Asso.*, **81**, 310-320.

Fan, J. (1992). Design-adaptive nonparametric regression. *Jour. Amer. Statist. Assoc.*, **87**, 998–1004.

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

Fan, J. and Huang, L. (2001). Goodness-of-fit test for parametric regression models, *J. Amer. Statist. Asso.*, **96**, 640-652.

Fan, J. and Huang, T. (2003). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Manuscript.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks Phenomenon, *Ann. Statist.*. **29**, 153-193.

Gallant, A. N. (1987). *Nonlinear Statistical Models*, Wiley, New York.

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

Heckman, N. (1986). Spline smoothing in partly linear models, *J. Royal Statist. Soc. B.* **48**, 244-248.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators, *Ann. Math. Statist.*, **40**, 633-643.

Li, R. and Nie, L. (2004). Statistical inferences on partially nonlinear models and their applications. Submitted for publication.

Monteith, J.L. (1972). Solar radiation and productivity in tropical ecosystems, *Journal of Applied Ecology*, **9**, 747-766.

Ruimy, A., L. Kergoat, A. Bondeau, and The Participants of The Potsdam NPP. (1999). Model Intercomparison, Comparing global models of terrestrial net primary productivity (NPP): analysis of differences in light absorption and light-use efficiency, *Global Change Biology*, **5**, 56-64.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *Jour. Amer. Statist. Assoc.*, **90**, 1257-1270.

Seber, G.A.F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.

Severini, T.A. and Staniswalis, J.G. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, **89**, 501–511.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Royal Statist. Soc. B*, **50**, 413-436.

and Carter, R. L. (1992). Mixed effects nonlinear regression for unbalanced repeated measures, *Biometrics*, **68**, 1-17.

Yatchew, A. (1997). An elementary estimator for the partial linear model, *Econometric Letters*, **57**, 135-143.

Appendix

Before proving theorems, let us introduce some notations. Let $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_d)^T$, $\epsilon_D = (e_1, \cdots, e_{n-1})$, and $\alpha_D = \{\alpha(u_2) - \alpha(u_1), \cdots, \alpha(u_n) - \alpha(u_{n-1})\}$. Denote $g_D^i(\boldsymbol{\beta})$ to be the $(n-1)$-column vector $\partial g_D(\boldsymbol{\beta})/\partial \beta_i$ and $g_D^{ij}(\boldsymbol{\beta})$ be the $(n-1)$-column vector $\partial^2 g_D(\boldsymbol{\beta})/\partial \beta_i \partial \beta_j$. Define $g'_D(\boldsymbol{\beta})$ to be the $(n-1) \times d$ matrix with the i-th column being $g_D^i(\boldsymbol{\beta})$.

Regularity Conditions

(A) There exists a positive definite matrix $\Sigma(\boldsymbol{\beta})$ in a neighborhood of $\boldsymbol{\beta}_0$ such that

$$\lim_{n\to\infty} n^{-1} g_D'(\boldsymbol{\beta})^T W g_D'(\boldsymbol{\beta}) = \Sigma(\boldsymbol{\beta}).$$

Further, denote Σ to be $\Sigma(\boldsymbol{\beta}_0)$.

(B) There exists a positive definite matrix such that

$$\lim_{n\to\infty} n^{-1} g_D'(\boldsymbol{\beta}_0)^T W W_0 W g_D'(\boldsymbol{\beta}_0) = \Sigma^*,$$

where W_0 is true variance-covariance of ϵ_D.

(C) $\lim_{n\to\infty} n^{-1} g_D^i(\boldsymbol{\beta})^T g_D^i(\boldsymbol{\beta})$, $i = 1, \cdots, d$, and $\lim_{n\to\infty} n^{-1} g_D^{ij}(\boldsymbol{\beta})^T g_D^{ij}(\boldsymbol{\beta})$, $i, j = 1, \cdots, d$, are bounded for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$.

(D) The largest eigenvalue of W is bounded.

(E) The baseline function $\alpha(\cdot)$ has a continuous 2nd derivative and is Lipschitz continuous on Ω, $u_{(i+1)} - u_{(i)} = o_P(n^{-1/2})$, where $u_{(i)}$, $i = 1, \cdots, n$ be the order statistics of $(u_1, \cdots, u_n)$.

(F) $E|\varepsilon|^4 < \infty$. The random variable U has a bound support Ω. Its density function $f(\cdot)$ is Lipschitz continuous on Ω, i.e, there exists constant $M > 0$, such that for $s, t \in \Omega$, $|f(s) - f(t)| \leq M|s - t|$.

(G) The kernel function $K(\cdot)$ is symmetric and bounded. Furthermore, the functions $u^4 K(u)$ and $u^3 K'(u)$ are bounded and $\int u^4 K(u)\, du < \infty$.

Conditions (A)—(C) are adapted from Jennrich (1969) who established asymptotic normality of nonlinear least square estimators.

Proof of Theorem 1: We first establish the root n consistency of $\hat{\boldsymbol{\beta}}$. We want to show that for any given $\eta > 0$, there exists a large constant C such that

$$P\{\inf_{\|\mathbf{t}\|=C} S(\boldsymbol{\beta}_0 + \mathbf{t}/\sqrt{n}) > S(\boldsymbol{\beta}_0)\} \geq 1 - \eta. \tag{A.3}$$

This implies with probability at least $1-\eta$ that there exists a local minimizer in the ball $\{\boldsymbol{\beta}_0 + \mathbf{t}/\sqrt{n} : \|\mathbf{t}\| \le C\}$. Hence, there exists a local minimizer such that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(1/\sqrt{n})$.

Using the Taylor expansion,

$$S(\boldsymbol{\beta}_0 + \mathbf{t}/\sqrt{n}) - S(\boldsymbol{\beta}_0) = n^{-1/2}\mathbf{t}^T S'(\boldsymbol{\beta}_0) + \frac{1}{2}n^{-1}\mathbf{t}^T S''(\boldsymbol{\beta}^*)\mathbf{t} \hat{=} I_1 + I_2,$$

where $\boldsymbol{\beta}^*$ lies between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$, $S'(\boldsymbol{\beta})$ and $S''(\boldsymbol{\beta})$ are the gradient vector and the Hessian matrix of $S(\boldsymbol{\beta})$, respectively. It is enough to show that I_2 dominates I_1 for a sufficiently large C. Let us calculate the order of I_1 first. Since

$$S'(\boldsymbol{\beta}_0) = -2g_D'(\boldsymbol{\beta}_0)^T W \epsilon_D - 2g_D'(\boldsymbol{\beta}_0)^T W \alpha_D, \tag{A.4}$$

Under Conditions (C)—(E), the second term in (A.4) is of order $o_P(\sqrt{n})$. Note that $E(\epsilon_D) = 0$. Under Condition (B) the first term is (A.4) is of order $O_P(\sqrt{n})$. Thus, $n^{-1/2}S'(\boldsymbol{\beta}_0) = O_P(1)$.

Next we deal with I_2. Let $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \mathbf{t}/\sqrt{n}$. By some calculation, the (i,j)-element of the $d \times d$ matrix $S''(\boldsymbol{\beta})$ is

$$2g_D^{ij}(\boldsymbol{\beta})^T W \epsilon_D + 2g_D^{ij}(\boldsymbol{\beta})^T W \alpha_D + 2g_D^{i}(\boldsymbol{\beta})^T W g_D^{j}(\boldsymbol{\beta}) + 2g_D^{ij}(\boldsymbol{\beta})^T W \{g_D(\boldsymbol{\beta}_0) - g_D(\boldsymbol{\beta})\},$$

Under Condition (C), it follows from Theorem 4 of Jennrich (1969) that

$$n^{-1}g_D^{ij}(\boldsymbol{\beta})^T W \epsilon_D \to 0$$

uniformly for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$. Under Conditions (C)-(E) and using Cauchy-Schwartz inequality, we have,

$$n^{-1}g_D^{ij}(\boldsymbol{\beta})^T W \alpha_D \ \to 0, \text{ and } n^{-1}g_D^{ij}(\boldsymbol{\beta})^T W \{g_D(\boldsymbol{\beta}_0) - g_D(\boldsymbol{\beta})\} \ \to 0$$

uniformly for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$. Therefore by Condition (A),

$$n^{-1}S''(\boldsymbol{\beta}^*) \to 2\Sigma \tag{A.5}$$

in probability. Note that Σ is positive definite. By choosing a sufficently large C, I_2 dominates I_1. Thus, (A.3) holds.

Now we show the asymptotic normality of $\hat{\boldsymbol{\beta}}$. Let $S_j'(\boldsymbol{\beta})$ denote the j-th component of $S'(\boldsymbol{\beta})$, and $S_j''(\boldsymbol{\beta})$ be the j-row of $S''(\boldsymbol{\beta})$. Using Taylor's expansion, for $j = 1, \cdots, d$,

$$0 = S_j'(\hat{\boldsymbol{\beta}}) = S_j'(\boldsymbol{\beta}_0) + S_j''(\boldsymbol{\beta}_j^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{A.6}$$

where $\boldsymbol{\beta}_j^*$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Using (A.5),

$$n^{-1} S_j''(\boldsymbol{\beta}_j^*) \to 2\Sigma_j,$$

in probability, where Σ_j is the j-row of Σ. Using (A.3) together with (A.6),

$$2n^{1/2}\Sigma\{1 + o_P(1)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = 2n^{-1/2} g_D'(\boldsymbol{\beta}_0)^T W \epsilon_D + o_P(1).$$

Under Condition (B)— (E),

$$n^{-1/2} g_D'(\boldsymbol{\beta}_0)^T W \epsilon_D \xrightarrow{\mathrm{D}} N(0, \Sigma^*).$$

Using the Slutsky theorem and (A.5), it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{D}} N(0, \Sigma^{-1}\Sigma^*\Sigma^{-1}).$$