Efficient Statistical Inference Procedures for Partially Nonlinear Models and their Applications

Runze Li

Department of Statistics and The Methodology Center
The Pennsylvania State University, University Park, PA 16802
email: rli@stat.psu.edu

and

Lei Nie

Department of Biostatistics, Bioinformatiocs, and Biomathematics
Georgetown University, Washington, DC20057
email: ln54@georgetown.edu

September 4, 2007

SUMMARY. Motivated by an analysis of a real data set in ecology, we consider a class of partially nonlinear models where both of a nonparametric component and a parametric component present. We develop two new estimation procedures to estimate the parameters in the parametric component. Consistency and asymptotic normality of the resulting estimators are established. We further propose an estimation procedure and a generalized F test procedure for the nonparametric component in the partially nonlinear models. Asymptotic properties of the newly proposed estimation procedure and the test statistic are derived. Finite sample performance of the proposed inference procedures are assessed by Monte Carlo simulation studies. An application in ecology is used to illustrate the proposed methods.

KEY WORDS: Local linear regression, partial linear models, profile least squares, semiparametric models

1. Introduction

Semiparametric regression models are useful for data analysis since they retain the flexibility of nonparametric models and the ease of interpretation of parametric models. Ruppert, Wand and Carroll (2003) presents various estimation procedures and many applications of semiparametric regression models. Motivated by a real data example in ecology, we consider a class of partially nonlinear models, natural extensions of partially linear models which are systematically studied in Härdle, Liang and Gao (1999).

Let us first present some background of the motivating real data example. Photosynthesis is the most important biochemical process to the dynamics of life on the Earth. Dependence of the ecosystem light response on temperature is complex. For example, because of elevated temperature and enhanced respiration, increase in radiation may lead to decrease in photosynthesis (Gu, et al., 2002). Many researches have been conducted to understand the complexity of the effects of environmental stresses on the ecosystem light response. In the literature, researchers use *Net Ecosystem CO_2 Exchange* (NEE) to measure photosynthetic rate in a natural ecosystem. It is known that sunlight intensity affects the rate of photosynthesis. In other words, the photosynthetic rate depends on the amount of *Photosynthetically Active Radiation* (PAR) available to an ecosystem. Based on empirical studies using data collected in laboratory experiments in which temperature can be well controlled, the relationship between NEE and PAR, is assumed to be nonlinear (Montheith, 1972, Ruimy et al., 1999 and references therein). See also (3.1). On the other hand, the data analyzed in Section 3.2, consisting of 1997 observations of NEE, PAR and temperature (T), were collected over a subalpine forest during parts of the growth season of 1999. Because the temperature cannot be controlled in an real ecosystem, and ecosystem respiration, which gives off CO_2, depends on temperature, we should allow the parameter associated with ecosystem respiration, i.e, R in equation (3.1), to depend on temperature. Since the parametric relationship of R and T is

not available, we assume R depends on the temperature through an unknown function $R(T)$. Thus, we consider a partially nonlinear model (3.2). See Section 3.2 for data description and more demonstrations.

The partially nonlinear model, defined in (2.1), retains the flexibility of nonparametric models for the baseline function and the model interpretability of nonlinear parametric models. As special cases of partially nonlinear models, partially linear models are popular in the literature (Engle, *et al.*, 1986, Heckman, 1986, Chen, 1988, Speckman, 1988, Ma, Chiou and Wang, 2006 for iid data, and Fan and Li, 2004, Hu, Wang and Carroll, 2004 and Wang, Carroll and Lin, 2005 for longitudinal data). As another type of special cases of partially nonlinear models, nonlinear parametric regression models have been widely used in statistical literature. Many interesting examples and applications of such models are given in Bates and Watts (1988), Seber and Wild (1989), and Huet (2004).

In this paper, we first propose an estimation procedure for parameters in the parametric component by using a profile nonlinear least squares technique. The asymptotic normality of the resulting estimate are established. We further propose an alternative estimation procedure, in which the nonlinear parametric function is approximated by a linear function. Compared to the profile approach, the linear approximation approach is computationally less intensive, however shares the same asymptotic efficiency. We further propose an estimation procedure for the nonparametric baseline function. The asymptotic properties of the estimation procedure are established. To address some scientific questions, one may have to test whether the nonparametric component has some simple forms (e.g, constant, or linear). To this end, we propose a generalized F test to the partially nonlinear model, and further derive its asymptotic null distribution. The paper is organized as follow. In Section 2, we develop two estimation procedures for parameters in the parametric component and statistical inference procedures for the baseline function. In Section 3, we conduct numerical studies

to assess finite sample performance of the proposed procedures. A real data example is used to illustrate the proposed procedures. Some discussions are given in Section 4.

2. Inference Procedures

Let Y be a response variable and $\{\mathbf{x}, U\}$ be its covariates. A partially nonlinear model is defined to be

$$Y = \alpha(U) + g(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon, \tag{2.1}$$

where $\alpha(\cdot)$ is an unknown smooth function, $g(\cdot; \cdot)$ is a pre-specified function, $\boldsymbol{\beta}$ is an unknown parameter vector, and ε is a random error with mean zero and variance σ^2. Due to the curse of dimensionality, it is assumed throughout this paper that U is a univariate continuous random variable. Suppose that $\{\mathbf{x}_i, u_i, y_i\}, i = 1, \cdots, n$ is a random sample. We next propose two estimation procedures for the parameter $\boldsymbol{\beta}$.

2.1 *Profile nonlinear least squares approach*

For model (2.1), define a nonlinear least squares function

$$Q(\alpha, \boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i - \alpha(u_i) - g(\mathbf{x}_i; \boldsymbol{\beta})\}^2. \tag{2.2}$$

We first apply the profile least squares technique to estimate parameters through (2.2) as follows. For a given $\boldsymbol{\beta}$, let $y_i^* = y_i - g(\mathbf{x}_i; \boldsymbol{\beta})$. Then, $\{y_1^*, \cdots, y_n^*\}$ can be viewed as a sample from the model

$$y^* = \alpha(u) + \varepsilon. \tag{2.3}$$

This is a 1-dimensional ordinary nonparametric regression model. Thus, $\alpha(\cdot)$ can be easily estimated using any linear smoother. Here, we will employ local linear regression (Fan and Gijbels, 1996). For any u in a neighborhood of u_0, it follows by Taylor's expansion that

$$\alpha(u) \approx \alpha(u_0) + \alpha'(u_0)(u - u_0) \widehat{=} a + b(u - u_0).$$

Let $K(\cdot)$ be a kernel function and h be a bandwidth. The local linear regression approach finds local parameters (a, b) to minimize

$$\sum_{i=1}^{n}\{y_i^* - a - b(u_i - u_0)\}^2 K_h(u_i - u_0),$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$. Note that local linear regression results in a linear estimator for $\alpha(\cdot)$ in terms of y_i^*. Let $\mathbf{y} = (y_1, \cdots, y_n)^T$, $\boldsymbol{\alpha} = \{\alpha(u_1), \cdots, \alpha(u_n)\}^T$, and $\mathbf{g}(\boldsymbol{\beta}) = \{g(\mathbf{x}_1; \boldsymbol{\beta}), \cdots, g(\mathbf{x}_n; \boldsymbol{\beta})\}^T$. Thus, the local linear estimate for $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} \hat{=} \hat{\boldsymbol{\alpha}}(\cdot; \boldsymbol{\beta}) = S_h\{\mathbf{y} - \mathbf{g}(\boldsymbol{\beta})\}, \tag{2.4}$$

where S_h is an $n \times n$ smoothing matrix depending only on $u_1, \cdots, u_n$ and the bandwidth h.

Substituting $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ in (2.2) results in the *profile nonlinear least square*:

$$Q(\boldsymbol{\beta}) \hat{=} Q\{\hat{\boldsymbol{\alpha}}(\cdot; \boldsymbol{\beta}), \boldsymbol{\beta}\} = \|(I_n - S_h)\{\mathbf{y} - \mathbf{g}(\boldsymbol{\beta})\}\|^2, \tag{2.5}$$

where I_n is the $n \times n$ identity matrix, and $\|\cdot\|$ stands for the Euclidean norm. Minimizing $Q(\boldsymbol{\beta})$ yields a nonlinear profile least squares estimator $\hat{\boldsymbol{\beta}}_P$. The idea of the nonlinear profile least squares approach stems from the generalized profile likelihood principle (Severini and Wong, 1992). Let $g'(\mathbf{x}, \boldsymbol{\beta})$ be $\partial g(\mathbf{x}; \boldsymbol{\beta})/\partial \boldsymbol{\beta}$, and $\boldsymbol{\beta}_0$ the true value of $\boldsymbol{\beta}$. Denote $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ and define

$$\mathbf{A} = E[g'(\mathbf{x}; \boldsymbol{\beta}_0) - E\{g'(\mathbf{x}; \boldsymbol{\beta}_0)|U\}]^{\otimes 2}.$$

Theorem 1. Suppose that the matrix $\mathbf{A}$ is finite and positive definite, and Conditions (A)—(H) in the Web Appendix hold. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2\mathbf{A}^{-1}),$$

where "$\xrightarrow{\mathcal{D}}$" stands for convergence in distribution.

The proof of Theorem 1 is given in the Web Appendix. From Theorem 1, the asymptotic variance of $\hat{\boldsymbol{\beta}}$, i.e. $\sigma^2 \mathbf{A}^{-1}/n$, can be estimated by

$$\frac{\hat{\sigma}^2_{\varepsilon,P}}{n} \left[\frac{1}{n} \sum_{i=1}^{n} \{g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_P) - \bar{g}(u_i)\}^{\otimes 2} \right]^{-1}, \tag{2.6}$$

where $\bar{g}(u_i)$ is the local linear estimate of $E\{g'(\mathbf{x}; \boldsymbol{\beta}_0)|U = u_i\}$ based on the sample $\{u_j, g'(\mathbf{x}_j; \hat{\boldsymbol{\beta}}_P), j = 1, \cdots, n\}$, and $\hat{\sigma}^2_{\varepsilon,P}$ is the residual mean squared error of the profile nonlinear estimator.

2.2 Linear approximation approach

Since $g(\boldsymbol{\beta})$ is nonlinear with respect to $\boldsymbol{\beta}$, there is no closed form for the minimizer of $Q(\boldsymbol{\beta})$ in (2.5). Typically, we employ the Newton-Raphson algorithm for the minimization problem. The algorithm requires to compute $S_h \mathbf{g}(\boldsymbol{\beta})$, $S_h \mathbf{g}'(\boldsymbol{\beta})$ and $S_h \mathbf{g}''(\boldsymbol{\beta})$ at each step of iterations. That is, we have to conduct nonparametric smoothing (i.e, local linear regression in this paper) many times at each step. Therefore, minimizing $Q(\boldsymbol{\beta})$ involves a heavy computation. We next propose an alternative estimation procedure which requires much less computation. This approach will need $\hat{\boldsymbol{\beta}}_I$, a consistent estimate of $\boldsymbol{\beta}$, as a starting point, while we will introduce an easy way to construct $\hat{\boldsymbol{\beta}}_I$ in (2.11). Applying Taylor expansion for $g(\cdot; \boldsymbol{\beta})$ at $\hat{\boldsymbol{\beta}}_I$, we have

$$g(\mathbf{x}; \boldsymbol{\beta}) = g(\mathbf{x}; \hat{\boldsymbol{\beta}}_I) + g'(\mathbf{x}; \hat{\boldsymbol{\beta}}_I)^T (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_I) + o_P(\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_I\|).$$

Thus, it follows that

$$y_i \approx \alpha(u_i) + g(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I) + g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I)^T (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_I) + \varepsilon_i. \tag{2.7}$$

Let $z_i = y_i - g(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I) + g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I)^T \hat{\boldsymbol{\beta}}_I$. We consider the following linear approximation model

$$z_i = \alpha(u_i) + g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I)^T \boldsymbol{\beta} + \varepsilon. \tag{2.8}$$

Existing estimation procedures can be directly employed to estimate $\boldsymbol{\beta}$ through the partial linear model (2.8). Here we use the profile least squares technique. For a given $\boldsymbol{\beta}$, let

$z_i^* = z_i - g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}_I)^T \boldsymbol{\beta}$, then (2.8) leads to

$$z_i^* = \alpha(u_i) + \varepsilon.$$

The local linear regression can be used to estimate the baseline function $\alpha(U)$. Denote $\mathbf{z}^* = (z_1^*, \cdots, z_n^*)^T$ and $\boldsymbol{\alpha} = (\alpha(u_1), \cdots, \alpha(u_n))^T$. Since the local linear regression is a linear smoother, we have

$$\hat{\boldsymbol{\alpha}} = S_h \mathbf{z}^*.$$

Substituting $\hat{\boldsymbol{\alpha}}$ for α in (2.8), then it follows that

$$(I - S_h)\mathbf{z} = (I - S_h)\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\boldsymbol{\beta} + \varepsilon,$$

where $\mathbf{z} = (z_1, \cdots, z_n)^T$ and $\mathbf{g}'(\hat{\boldsymbol{\beta}}_I) = (g'(\mathbf{x}_1; \hat{\boldsymbol{\beta}}_I), \cdots, g'(\mathbf{x}_n; \hat{\boldsymbol{\beta}}_I))^T$. We therefore estimate $\boldsymbol{\beta}$ through the ordinary least squares estimator,

$$\hat{\boldsymbol{\beta}}_L = \{\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T (I - S_h)^T (I - S_h)\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\}^{-1}\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T (I - S_h)^T (I - S_h)\mathbf{z}. \qquad (2.9)$$

Theorem 2. Suppose that the matrix $\mathbf{A}$ is finite and positive definite. Under Conditions (A)—(H) given in the Web Appendix and $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2 \mathbf{A}^{-1}).$$

The proof of Theorem 2 is given in the Web Appendix. From Theorems 1 and 2, $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_L$ share the same asymptotic distribution. Similar to (2.6), $\text{Var}\{\hat{\boldsymbol{\beta}}_L\}$ can be estimated by using its sample counterpart.

In practical implementation, one needs to obtain $\hat{\boldsymbol{\beta}}_I$ first. 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. Similar to Yatchew (1997), we obtain the following approximated model,

$$y_{(i+1)} - y_{(i)} \approx g(\mathbf{x}_{(i+1)}; \boldsymbol{\beta}) - g(\mathbf{x}_{(i)}; \boldsymbol{\beta}) + e_i, \tag{2.10}$$

the nonlinear least squares estimate will be used as an initial starting point of $\boldsymbol{\beta}$.

$$\hat{\boldsymbol{\beta}}_I = \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{n-1} \{y_{(i+1)} - y_{(i)} - g(\mathbf{x}_{(i+1)}; \boldsymbol{\beta}) + g(\mathbf{x}_{(i)}; \boldsymbol{\beta})\}^2. \tag{2.11}$$

Under some conditions, $\hat{\boldsymbol{\beta}}_I$ is root n consistent (Li and Nie, 2007).

Remark: To ensure root n consistency of $\hat{\boldsymbol{\beta}}_L$, the assumption of $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}\| = O_P(n^{-1/2})$ is critical in the linear approximation approach. Inconsistent estimate $\hat{\boldsymbol{\beta}}_I$ may result in an inconsistent estimate $\hat{\boldsymbol{\beta}}_L$. The finite sample performance of $\hat{\boldsymbol{\beta}}_L$ may be sensitive to the initial value $\boldsymbol{\beta}_I$, which is confirmed by our simulation study. The efficiency of $\hat{\boldsymbol{\beta}}_L$ depends on the intrinsic curvature of the partially nonlinear model (Zhu, Tang and Wei, 2000). If the intrinsic curvature of the partially nonlinear model at $\hat{\boldsymbol{\beta}}_I$ is big, or $\hat{\boldsymbol{\beta}}_I$ is not the root n consistent, then we suggest using the following iterative algorithm. For a given initial value $\hat{\boldsymbol{\beta}}^{(0)}$ (we may set $\hat{\boldsymbol{\beta}}^{(0)} = \hat{\boldsymbol{\beta}}_I$), we iteratively compute

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \left\{ \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)}) \right\}^{-1} \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{z}^{(m)} \quad (2.12)$$

until it converges, where the i-th element of $\mathbf{z}^{(m)}$ equals $y_i - g(\mathbf{x}_i; \hat{\boldsymbol{\beta}}^{(m)}) + g'(\mathbf{x}_i; \hat{\boldsymbol{\beta}}^{(m)})^T \hat{\boldsymbol{\beta}}^{(m)}$. As shown in the Web Appendix, algorithm (2.12) is equivalent to using the Fisher scoring algorithm to minimize $Q(\boldsymbol{\beta})$ in (2.5). As shown in the proofs of Theorems 1 and 2, both $n^{-1}\{Q''(\boldsymbol{\beta}_0)\}/2$ and $n^{-1}\{\mathbf{g}'(\boldsymbol{\beta}_0)^T (I - S_h)^T (I - S_h) \mathbf{g}'(\boldsymbol{\beta}_0)\}$ tend to $\mathbf{A}$ in probability. Therefore the resulting estimate of algorithm (2.12) shares the same efficiency of the profile nonlinear least squares estimate.

7

2.3 *Inference procedures for the baseline function*

For a given root n consistent estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, we estimate $\alpha(\cdot)$ by smoothing the partial residuals $\{r_i = y_i - g(\mathbf{x}_i; \hat{\boldsymbol{\beta}}), i = 1, \cdots, n\}$ over u_i. Here $\hat{\boldsymbol{\beta}}$ can be either $\hat{\boldsymbol{\beta}}_P$ or $\hat{\boldsymbol{\beta}}_L$. We will utilize local linear regression (Fan and Gijbels, 1996) to estimate $\alpha(u)$. Denote by $(\hat{a}, \hat{b})$ the minimizer of

$$\sum_{i=1}^{n} \{r_i - a - b(u_i - u_0)\}^2 K_h(u_i - u_0). \tag{2.13}$$

Then

$$\hat{\alpha}(u_0; \hat{\boldsymbol{\beta}}) = \hat{a} \quad \text{and} \quad \hat{\alpha}'(u_0; \hat{\boldsymbol{\beta}}) = \hat{b}.$$

The asymptotic bias and variance of $\hat{\alpha}(u_0; \hat{\boldsymbol{\beta}})$ are given in the following theorem. They are the same as that with $\boldsymbol{\beta}$ known. This is because when $\hat{\boldsymbol{\beta}}$ is root n consistent, the convergence rate of $\hat{\boldsymbol{\beta}}$ is faster than that of the nonparametric estimator. Consequently the errors in estimation of $\boldsymbol{\beta}$ are negligible in the nonparametric estimation of α.

Theorem 3. Suppose that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$. Under Conditions (A)—(C) given in the Web Appendix, if $h \to 0$ and $nh \to \infty$, then for any $u_0 \in \Omega$, the support of U, we have

$$\sqrt{nh} \left\{ \hat{\alpha}(u_0; \hat{\boldsymbol{\beta}}) - \alpha(u_0) - \frac{1}{2} \alpha''(u_0) \int t^2 K(t) dt h^2 \right\} \xrightarrow{\mathcal{D}} N(0, \sigma^2(u_0)),$$

where

$$\sigma^2(u_0) = \sigma_\varepsilon^2 \{f(u_0)\}^{-1} \int K(u)^2 du,$$

and $f(\cdot)$ is the marginal density of U.

The bias and variance expressions are the same as those in Fan (1993). Thus, the theoretic optimal bandwidth will remain the same as that for 1-dimensional nonparametric model (2.8) regarding $\boldsymbol{\beta}$ is known. The proof of Theorem 3 is similar to what is described in Fan (1993), we therefore omit the details. In practice, the bandwidth can be selected by existing data-driven methods. Specifically, let $y_i^* = y_i - g(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_I)$, and then apply an existing bandwidth

selection procedure such as the plug-in bandwidth selector (Ruppert, Sheather and Wand, 1995), for $(u_i, y_i^*), i = 1, \cdots, n$ to select a bandwidth h. This approach will be implemented in Section 3.2.

In practice, a natural question that may arise is whether $\alpha(\cdot)$ has a pre-specified parametric form. This leads us to consider the following hypothesis testing problem:

$$H_0 : \alpha(\cdot) = \alpha_0(\cdot, \boldsymbol{\theta}) \quad \text{versus} \quad H_1 : \alpha(\cdot) \neq \alpha_0(\cdot, \boldsymbol{\theta}),$$

where $\alpha_0(\cdot, \boldsymbol{\theta})$ is a pre-specified parametric function and $\boldsymbol{\theta}$ is a vector of unknown parameters. For example, if we are interested in testing whether $\alpha(\cdot)$ really depends on U variable, then we may consider the null hypothesis that $\alpha(\cdot) = \alpha_0$, where $\theta = \alpha_0$ is an unknown constant. This kind of hypothesis testing is referred to as nonparametric goodness of fit test. In this section, we propose a generalized F test for the partially nonlinear model. For simplicity of presentation, we consider only a test of linearity for $\alpha(\cdot)$:

$$H_0 : \alpha(u) = \alpha_0 + \alpha_1 u, \quad \text{versus} \quad H_1 : \alpha(u) \neq \alpha_0 + \alpha_1 u, \tag{2.14}$$

where α_0 and α_1 are unknown constants. Note that the parameter space under the null hypothesis is finite dimensional, while it is infinite dimensional under the alternative hypothesis. Thus, many traditional tests, such as the F test, cannot be directly applied for the above hypothesis. Here we propose a generalized F test to deal with this issue. Let $\hat{\boldsymbol{\beta}}$ be either $\hat{\boldsymbol{\beta}}_P$ or $\hat{\boldsymbol{\beta}}_L$, a root n consistent estimate of $\boldsymbol{\beta}$ under H_1, $(\hat{\alpha}_0, \hat{\alpha}_1)^T$ be an estimate of $(\alpha_0, \alpha_1)^T$ under H_0 and $\hat{\alpha}(\cdot)$ be an estimate of $\alpha(\cdot)$ under H_1. Denote $\text{RSS}(H_0) = \sum_{i=1}^{n} \{y_i - \hat{\alpha}_0 - \hat{\alpha}_1 u_i - g(\mathbf{x}_i; \hat{\boldsymbol{\beta}})\}^2$, the residual sum of squares under H_0, and $\text{RSS}(H_1) = \sum_{i=1}^{n} \{y_i - \hat{\alpha}(u_i) - g(\mathbf{x}_i; \hat{\boldsymbol{\beta}})\}^2$, the residual sum of squares under H_1. As in linear models, it is natural to employ an F-type test to compare the model fits under H_0 and H_1. Since the parameter space is infinite dimensional

under alternative hypothesis, we define a generalized F test statistic

$$F = \frac{\mathrm{RSS}(H_0) - \mathrm{RSS}(H_1)}{\mathrm{RSS}(H_1)/n}. \tag{2.15}$$

Intuitively, under H_0, there will be little difference between $\mathrm{RSS}(H_0)$ and $\mathrm{RSS}(H_1)$. However, under the alternative hypothesis, $\mathrm{RSS}(H_0)$ should become systematically larger than $\mathrm{RSS}(H_1)$, and hence the test statistic F will tend to take a large positive value. Hence, a large value of the test statistic F indicates that the null hypothesis should be rejected. The following theorem gives the asymptotic null distribution of F.

Theorem 4. Suppose that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$, $E|\varepsilon|^4 < \infty$ and Conditions (A)—(C) in the Web Appendix hold. If $h \to 0$ in such a way that $nh^{3/2} \to \infty$, then, under H_0 in (2.14), $(r_K/2)F$, has an asymptotic χ^2 distribution with δ_n degrees of freedom, where $\delta_n = r_K|\Omega|\{K(0) - 0.5 \int K^2(u)\,du\}/h$, and $|\Omega|$ stands for the length of the support of U, and $r_K = \{K(0) - 0.5 \int K^2(t)\,dt\}/\int\{K(t) - 0.5K * K(t)\}^2\,dt$. Here $K * K$ stands for the convolution of K and K.

The proof of Theorem 4 is given in the Web Appendix. 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. Since the F test and the likelihood ratio test are equivalent for linear model with normally distributed error, the above theorem indeed extends the generalized likelihood ratio theory (Fan, Zhang and Zhang, 2001) to model (2.1), and unveils a new Wilks phenomenon for partially nonlinear models: the asymptotic null distribution is a chi-square distribution and does not depend on the unknown parameters (α_0, β_0). We will also provide empirical justification to the null distribution. Similar to Cai, Fan and Li (2000), the null distribution of the generalized F test can be estimated by Monte Carlo simulation or a bootstrap procedure. This usually provides a better estimate than the asymptotic

10

null distribution, since the degrees of freedom tend to infinity and the results in the above theorem give only the main order of the degrees of freedom.

Remark: Note that for a χ^2 random variable with degrees of freedom r, $\mathrm{Var}\{\chi^2(r)\} = 2E\{\chi^2(r)\}$. Thus, $r_K = 4E(F)/\mathrm{Var}(F)$. Using this relationship, we can derive a better approximation to the normalizing constant r_K using the bootstrap samples of F. Specifically, let $\mathrm{mean}(F^*)$ and $\mathrm{var}(F^*)$ be the sample mean and the sample variance of bootstrap samples $\{F_{0,i}^*,\ i = 1, \cdots, N\}$ of F_0, respectively. Then r_K can be replaced by $c_K = 4\,\mathrm{mean}(F_0^*)/\mathrm{var}(F_0^*)$. This will be implemented in Section 3.

3.   Numerical study and Application

In this section, we assess the finite sample performance by Monte Carlo simulations. We further illustrate the proposed methodologies through an application to the real data example. All simulations were conducted in SAS. In our simulation, the kernel function is taken to be the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$, and $\hat{\boldsymbol{\beta}}_I$ in (2.11) is set to be the initial value for $\hat{\boldsymbol{\beta}}_P$. From our limited experience, the profile nonlinear least squares method is not sensitive to the initial value specification while a good initial value leads to fast convergence of the algorithm.

3.1   *Monte Carlo simulations*

We generate random samples from the following model

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

where $\varepsilon \sim N(0, 1)$ and $U \sim U(0, 1)$, the uniform distribution over [0,1]. In our simulation, we consider the following two baseline functions:

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

11

and 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 the estimates for the real data example in Section 3.2, and $x \sim N(0,1)$; and

$$g_2(x_1,x_2;\beta_1,\beta_2) = 20\exp(\beta_1 x_1 + \beta_2 x_2)/\{1+\exp(\beta_1 x_1 + \beta_2 x_2)\}$$

with $\beta_1 = \beta_2 = 1$, The covariate vector $(x_1,x_2)^T$ was simulated from a normal distribution with mean zero and $\mathrm{cov}(x_i,x_j) = 0.5^{|i-j|}$. The sample size n was set to 200 or 400. For each case, we conduct 1,000 Monte Carlo simulations.

Insert Table 1 here

Performance of $\hat{\boldsymbol{\beta}}$. Simulation results for $\hat{\boldsymbol{\beta}}$ 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. "SE" and "Std(SE)" are the average and the standard deviation of the 1000 estimated standard errors using the proposed formulae in Section 2. From Table 1, we can see the performance of the two estimation procedures for $\boldsymbol{\beta}$ are almost the same, and both perform well. Comparing with the standard deviation of SE's, the SE slightly under-estimate the SD for sample size $n = 200$. This is typical for sandwich formula estimator (Kauermann and Carroll, 2001). In general, the difference between SD and its corresponding SE is less than one standard deviation of SEs. This indicates that the proposed formulae in Section 2 work very well. Results in Table 1 correspond to the bandwidth $h = 0.1$, which equals approximately the optimal bandwidth selected by the plug-in method of Ruppert, Sheather and Wand (1995). To study the effects of choice of bandwidth, we consider a smaller bandwidth $h = 0.05$ and

obtained similar results (data not shown). Results for other cases are similar and omitted here. We have also conducted simulations with the sample size $n = 50$ and 100. The results are similar to those in Table 1. Thus, the results are not reported here to save space.

Performance of $\hat{\alpha}(\cdot)$. The performance of $\hat{\alpha}(\cdot)$ is assessed by the square root of average squared errors (RASE),

$$\text{RASE}^2 = n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \{\hat{\alpha}(u_k) - \alpha(u_k)\}^2,$$

where $\{u_k, k = 1, \cdots, n_{\text{grid}}\}$ are the grid points at which $\hat{\alpha}(\cdot)$ is evaluated. In our simulation, the bandwidth was set to 0.1 or 0.2. The sample averages of the RASEs based on 1000 replicates are listed in Table 2, which reports only the case with $g(x; \beta) = -\beta_1 x/(x + \beta_2)$. Results for the other case is similar, and are omitted here to save space.

<center>**Insert Table 2 here**</center>

Performance of the generalized F test. We first verify the Wilks type result in Theorem 4 empirically. To this end, we take $H_0 : \alpha(u) = a_0 + a_1 u$. where $(a_0, a_1) = \text{argmin}_{c_0, c_1} E\|\alpha(u) - c_0 - c_1 u\|^2$. Such choice of (a_0, a_1) will pose the challenge to the generalized F test in the power calculation below. Then we find the null distribution based on 1,000 bootstrap simulations. We present only results for $\alpha(u) = \alpha_1(u)$. Results for $\alpha_2(u)$ are similar. Figure 1 (a) and (c) depict the estimated density of the null distribution of $(c_K/2)F$ for $g = g_1$ and g_2, respectively. We also plot the density of the chi-square distribution in order to examine whether the null distribution is close to a chi-square distribution. The degree of freedom is chosen to be the closest integer to the sample average of $(c_K/2)F$ values from the 1,000 bootstrap samples. It is clear from Figure 1 (a) and (c) that the Wilks type results hold, i.e., the chi-square distribution is a good approximation for the asymptotic behavior of $(c_K/2)F$ under H_0.

<center>13</center>

To examine the power of the proposed nonparametric goodness of fit test, we evaluate the power of the generalized F test for the alternative model

$$\alpha(u) = (1 - c)(a_0 + a_1 u) + c\alpha(u),$$

for each given c, where $\alpha(u)$ equals either $\alpha_1(\cdot)$ or $\alpha_2(\cdot)$. We took $c = 0.2,\ 0.4,\ 0.6$. Results for $\alpha_2(\cdot)$ are similar to those of $\alpha_1(\cdot)$. We present only results for $\alpha_1(\cdot)$.

Figure 1 (b) and (d) depicts the three power functions based on 400 Monte Carlo simulations for the sample size, $n = 200$, at three different significance levels: 0.10, 0.05 and 0.01. The powers at $c = 0$ for the foregoing three significance levels are: 0.095, 0.043 and 0.018 for Figure 1 (b) and 0.090, 0.048 and 0.018 for Figure 1 (d), respectively. This shows that the bootstrap method gives us an approximately right sized test.

3.2 *An Application*

In this section, we illustrate the proposed methodologies by an analysis of a real data set in ecology. Of interest in this example is to study how temperature affects the relationship between the net ecosystem CO_2 exchange (NEE) and the photosynthetically active radiation (PAR). The data set in this example consists of 1997 observations of NEE, PAR and temperature (T), and was collected over a subalpine forest at approximately 3050 meter elevations above sea level by using three-dimensional sonic anemometers on hundreds of meter towers during parts of the growth season of 1999. When data are collected from laboratory experiments, in which the temperature can be well controlled, the following model is widely used to describe the NEE and PAR relationship.

$$\text{NEE} = R - \frac{\beta_1 \text{PAR}}{\text{PAR} + \beta_2} + \varepsilon, \tag{3.1}$$

14

where ε is a random error with zero mean, and R, β_1 and β_2 are unknown parameters with physical interpretations. Specifically, R is the dark respiration rate, β_1 is the light-saturated net photosynthetic rate, and β_1/β_2 is the apparent quantum yield.

In order to demonstrate that the model (2.1) is appropriate, we take the NEE as the response variable, and PAR and T as covariates. We first consider a fully nonparametric regression model:

$$\text{NEE} = m(\text{PAR}, \text{T}) + \varepsilon,$$

where $m(\cdot, \cdot)$ is an unspecified smoothing function, and ε is random error with mean 0 and variance σ^2. Two dimensional kernel regression was used to estimate the regression $m(\cdot, \cdot)$. To examine how temperature affects the parameters in Model (3.1), we plot $\hat{m}(\text{PAR}, \text{T})$ versus PAR for given values of T. The three lines in Figure 2 (a) depict the plot of $\hat{m}(\text{PAR}, \text{T})$ over PAR for $T = 10.76, 13.29$ and 15.41, which correspond to the three sample quartiles of temperatures. From Figure 2 (a), we can see the nonlinear relationship between PAR and NEE when temperature is fixed. The nearly parallel pattern of the three lines suggests that the parameters β_1 and β_2 do not vary over temperature, while different intercepts for the three lines show the dark respiration rate R changes over temperature. The monotone decreasing pattern of the lines implies that Model (3.1) may be appropriate when temperature is fixed.

From the above analysis, we consider a partially nonlinear model

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

The proposed estimation procedure for $\boldsymbol{\beta}$ is used to fit the data set using Model (3.2). We first compute the difference based estimate $\hat{\boldsymbol{\beta}}_I$, and then use the plug-in method of Ruppert, Sheather and Wand (1995) to select a bandwidth. The resulting optimal bandwidth is $h = 1.286$. The resulting estimates using this bandwidth are $\hat{\beta}_1 = 17.656(0.398)$ and $\hat{\beta}_2 = 0.806(0.071)$ for the profile nonlinear least square approach; and $\hat{\beta}_1 = 17.644(0.384)$

and $\hat{\beta}_2 = 0.808(0.063)$ for the linear approximation approach.

<div align="center">**Insert Figure 2 here**</div>

As proposed in Section 2.2, we estimate the baseline function based on the partial residuals, which are depicted in Figure 2(b), from which we can see that the partial residuals have an increasing trend over temperature. Figure 2(b) also depicts the estimated baseline function using local linear regression.

From Figure 2(b), the overall increasing trend might suggest a simple linear model for $R(T)$: $R(T) = a + b \times T$. Thus, it is of interest to test $H_0 : R(T) = a + b \times T$. Using nonlinear least squares approach, we obtain the resulting estimate for parameters in model (3.2) under the H_0 given by $R(T) = -2.793 + 0.554 \times T$, $\tilde{\beta}_1 = 17.556$ and $\tilde{\beta}_2 = 0.798$. It can be seen that $(\tilde{\beta}_1, \tilde{\beta}_2)$ is very close to $(\hat{\beta}_1, \hat{\beta}_2) = (17.644, 0.808)$, the $\hat{\boldsymbol{\beta}}_L$ obtained under H_1. This implies that the model under H_0 fits the data almost as well as the one under H_1, and indicates that test of linearity for this data set is challenging. Setting $(\hat{\beta}_1, \hat{\beta}_2) = (17.644, 0.808)$, we compute the generalized F test which equals 34.460 with P-value < 0.001 obtained by using 1000 bootstraps. The null distribution is depicted in Figure 2(c). Thus, the dark respiration rate is not linear in temperature. This example also demonstrate that the generalized F has good power.

4. Discussions

Motivated by the real data example analyzed in Section 3.2, we considered a class of partially nonlinear models. We proposed two estimation procedures for parameters involved in the parametric component and inference procedures for the baseline function. For the setting in which both u_i and $\mathbf{x}_i$ are fixed design points and the random error is normally distributed, Zhu and Wei (1999) derived an asymptotic efficient estimation procedure for the partially nonlinear models by using the concept of least favorable curve introduced by Severini and

Wong (1992). They focused on theoretic development rather than practical implementation. Their method is quite abstract, and they did not discuss how to implement their approach. Compared with their method, both the newly proposed methods are easy to implement, although the profile nonlinear least squares method needs more computation than the linear approximation method.

5. Supplementary Materials

Wed Appendix referenced in Section 2 is available under the Paper Information link at the *Biometrics* website http://www.tibs.org/biometrics.

References

Bates, D., Watts, D.(1988). *Nonlinear regression Analysis and Its Applications*, Wiley.

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

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, **16**, 136-146.

Engle, R., Granger, C., 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. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, **21**, 196-216.

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

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.*, **99**, 710-723.

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

Gu, L. Baldocchi, D. Verma, S. Black, T.A., Vesala, T., Falge, E. M. and Dowty, P. R. (2002). Advantages of diffuse radiation for terrestrial ecosystem productivity. *J. Geophysical Research*, **107**, 10.1029/2001JD001242.

Härdle, W. Liang, H. and Gao, J. (1999). *Partially Linear Models*, Springer, New York.

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

Hu, Z., Wang, N. and Carroll, R.J. (2004). Profile-kernel versus backfitting in partially linear models for longitudinal/clustered data. *Biometrika*, **91**, 252-261.

Huet, S. (2004). *Statistical Tools for Nonlinear Regression: a Practical Guide with S-PLUS and R Examples*, Springer, New York.

Kauermann, G. and Carroll, R. (2001). A note on efficiency of sandwich covariance matrix estimation. *J. Amer. Statist. Asso.*, **96**, 1387-1396.

Li, R. and Nie, L. (2007). A new estimation procedure for partially nonlinear model via mixed effects approach. Submitted for publication. Available at

http://www.stat.psu.edu/rli/research/partialnlnv1.pdf

Ma, Y., Chiou, J.M. and Wang, N. (2006) Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika*, **93**, 75-84.

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 Potsdam participants (1999). Model Intercomparison, Comparing global models of terrestrial net primary productivity: 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.

Ruppert, D. Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*, Cambridge University Press, New York.

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

Severini, T.A. and Wong, W.H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.*, **20**, 1768-1802.

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

Wang, N., Carroll, R.J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data, *J. Amer. Statist. Assoc.*. **100**, 147-157.

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

Zhu, Z., Tang, N. and Wei, B. (2000). On confidence regions of semiparametric nonlinear regression models (A geometric approach). *Acta Mathematica Scientia*. **20**, 68-75.

Zhu, Z. and Wei, B. (1999). Asymptotic efficient estimation in semiparametric nonlinear regression models. *J. Appl. Math. Chinese Univ.* **14B**, 43-51.

Table 1

Finite Sample Performance of $\hat{\boldsymbol{\beta}}_P$ and $\hat{\boldsymbol{\beta}}_L$ with $h = 0.1$

β	(n, α, g)	Profile Nonlinear LS			Linear Approximation		
		Mean	SD	SE(Std(SE))	Mean	SD	SE(Std(SE))
$\beta_1 = 18$	$(200, \alpha_1, g_1)$	18.294	1.524	1.417(0.452)	18.29	1.521	1.391(0.440)
	$(400, \alpha_1, g_1)$	18.145	1.004	0.963(0.193)	18.144	1.004	0.954(0.191)
	$(200, \alpha_2, g_1)$	18.294	1.524	1.416(0.452)	18.290	1.521	1.390(0.440)
	$(400, \alpha_2, g_1)$	18.145	1.004	0.963(0.193)	18.144	1.004	0.953(0.191)
$\beta_2 = 0.8$	$(200, \alpha_1, g_1)$	0.829	0.165	0.153(0.048)	0.829	0.164	0.150(0.047)
	$(400, \alpha_1, g_1)$	0.814	0.106	0.104(0.020)	0.814	0.106	0.103(0.020)
	$(200, \alpha_2, g_1)$	0.829	0.165	0.153(0.048)	0.829	0.164	0.150(0.047)
	$(400, \alpha_2, g_1)$	0.814	0.106	0.104(0.020)	0.814	0.106	0.103(0.020)
$\beta_1 = 1$	$(200, \alpha_1, g_2)$	1.001	0.030	0.029(0.002)	1.001	0.030	0.029(0.002)
	$(400, \alpha_1, g_2)$	1.000	0.021	0.021(0.001)	1.000	0.021	0.021(0.001)
	$(200, \alpha_2, g_2)$	1.001	0.030	0.029(0.002)	1.001	0.029	0.029(0.002)
	$(400, \alpha_2, g_2)$	1.000	0.020	0.021(0.001)	1.000	0.020	0.021(0.001)
$\beta_2 = 1$	$(200, \alpha_1, g_2)$	1.000	0.029	0.029(0.002)	1.000	0.029	0.029(0.002)
	$(400, \alpha_1, g_2)$	1.001	0.021	0.021(0.001)	1.001	0.021	0.021(0.001)
	$(200, \alpha_2, g_2)$	1.000	0.029	0.029(0.002)	1.000	0.029	0.029(0.002)
	$(400, \alpha_2, g_2)$	1.001	0.021	0.021(0.001)	1.001	0.021	0.021(0.001)

Table 2

Mean of RASEs for Baseline Function

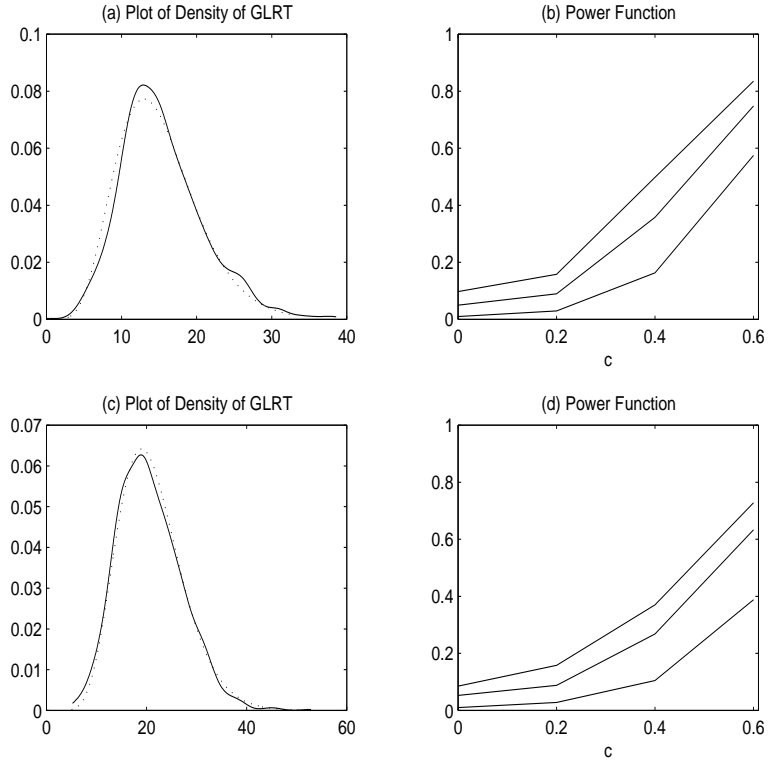| (n, g)	Profile Nonlinear LS		Linear Approximation	
	$h = 0.1$	$h = 0.2$	$h = 0.1$	$h = 0.2$
$(200, g_1)$	0.123	0.117	0.123	0.116
$(400, g_1)$	0.056	0.058	0.056	0.058
$(200, g_2)$	0.039	0.032	0.039	0.032
$(400, g_2)$	0.012	0.021	0.012	0.021

Figure 1. *Plot of Null Distribution of the Generalized F Test and its Power Function*
($n = 200$). (a) and (c) are the density of $(c_k/2)F$ under the null hypothesis. The solid
line is the estimated density curve using kernel density estimation and dash-dotted line is
the density curve of a χ^2-distribution with degrees of freedom 15 and 21 for (a) and (c)
respectively. (b) and (d) are power functions at level α=0.01, 0.05 and 0.10 from bottom to
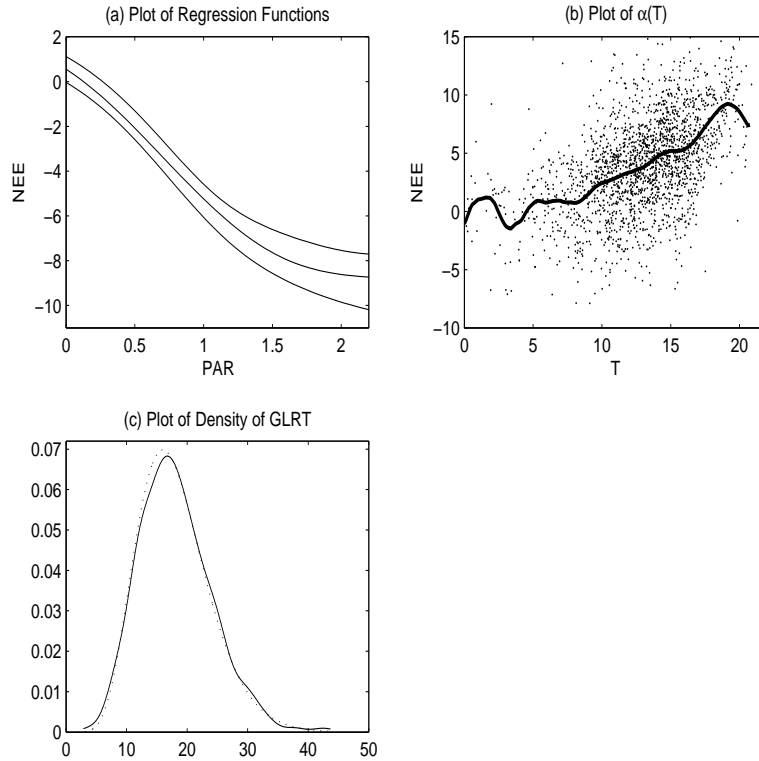*top, respectively.*

Figure 2. *Plots for Example in Section 3.2. The dashed lines in (a) are the fitted value of NEE versus PAR given temperature using 2-d kernel regression. From the bottom to top, the temperatures are 10.760, 13.290 and 15.410, which correspond to the three sample quartiles of temperatures. In (b), the solid line is an estimate of the baseline function, and the dots are partial residuals: $y_i - \hat{\beta}_1 x_i / (x_i + \hat{\beta}_2)$ with $(\hat{\beta}_1, \hat{\beta}_2) = (17.644, 0.808)$. In (c), the solid line is an estimate to the density of the $(c_k/2)F$ under H_0, and dotted line is the density of a chi-square distribution with d.f. 18.*