Statistics and Probability Letters 80 (2010) 1271–1283

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Sieve least squares estimation for partially nonlinear models

Lixin Song, Yue Zhao, Xiaoguang Wang *

School of Mathematical Sciences, Dalian University of Technology, Dalian, 116023, China

ARTICLE INFO

Article history:
Received 6 November 2009
Received in revised form 16 March 2010
Accepted 12 April 2010
Available online 22 April 2010

Keywords:
Semiparametric model
Consistency
Convergence rate
Asymptotic normality
Empirical process method

ABSTRACT

This paper considers a partially nonlinear model $E(Y|X,T) = H(\beta^{\mathrm{T}}X) + g(T)$, which is a sub-model of the general partially nonlinear model but has some particular advantages in statistical inference. We develop a sieve least squares method to estimate the parameters of the parametric part and the nonparametric part. The consistency and asymptotic normality of the estimator for the parametric part are established. Simulation results show that the sieve estimators perform quite well.

1. Introduction and motivation

The well-known general partially nonlinear model(PNLM) is

$$Y = H(\beta, X) + g(T) + \epsilon,$$

where Y is the response variable, $\{X, T\}$ are the associated covariates, $H(\cdot)$ is a pre-specified function, β is a d-dimensional parameter vector, the non-parametric part $g(\cdot)$ is an unknown smooth function and ϵ is a random error with mean 0 and constant variance. Wahba (1990), Liang (1995) and Li and Nie (2008) have carried out some research on this model. Note that $H(\cdot)$ may be an arbitrary function, so it is not easy to discuss estimation properties for statistical inference in general. For example, the identifiability condition of the model is usually given as a postulate for general PNLMs. In this paper, we consider a sub-model of general PNLMs to be

$$Y = H(\beta^{\mathrm{T}}X) + g(T) + \epsilon, \tag{1}$$

where the parametric part is $H(\beta^{\mathrm{T}}X)$ instead of $H(\beta, X)$. Model (1) can be viewed as a subfamily of the general ones that has some particular advantages in theoretical studies. As $H(\cdot)$ is a function of one variable, it is convenient to make a statistical inference in general. For example, it is not difficult to see that the model is identifiable if $H(\cdot)$ is a monotone function. Furthermore, many nonlinear functions widely used in applications can be transformed into the form of $H(\beta^{\mathrm{T}}X)$. Here are some examples. Li and Nie (2008) applied PNLM to ecology. They studied how temperature affects the relationship between the net ecosystem CO_2 exchange (NEE) and photosynthetically active radiation (PAR). The model is

$$\mathrm{NEE} = R(T) - \frac{\beta_1 \mathrm{PAR}}{\mathrm{PAR} + \beta_2} + \epsilon,$$

* Corresponding author. Tel.: +86 411 84708351 8040; fax: +86 411 84708354.
E-mail address: wangxg@dlut.edu.cn (X. Wang).

0167-7152/$ – see front matter
doi:10.1016/j.spl.2010.04.006

where β_1, β_2 are unknown parameters with physical interpretations and R is the dark respiration rate that changes over temperature T. After a simple derivation on its parametric part, we have

$$-\frac{\beta_1 \mathrm{PAR}}{\mathrm{PAR}+\beta_2} = -\frac{1}{\frac{1}{\beta_1}+\frac{\beta_2}{\beta_1}\mathrm{PAR}} \triangleq -\frac{1}{\beta_1^*+\beta_2^*\mathrm{PAR}} = H(\beta^{\mathrm{T}}X),$$

where $\beta = (\beta_1^*, \beta_2^*)^{\mathrm{T}}, X = (1, PAR)^{\mathrm{T}}, H(x) = -1/x$.

For semi-parametric models, the parametric parts are the most important of the models. Here we list three examples among a great many others to show how model (1) nests them as special cases. More interesting examples and applications of such models are given in Seber and Wild (1989) and Huet (2004).

- *Example* 1. Theoretical chemistry predicts that, for a given sample of gas kept at a constant temperature, the volume v and pressure p of the gas satisfy the relationship $pv^{\gamma} = c$. Writing $y = p$ and $x = v^{-1}$, we have

  $$y = cx^{\gamma} = \exp(\log c + \gamma \log x),$$

  where $H(\cdot)$ is derived to be an exponential function and the parameter is $(\log c, \gamma)^{\mathrm{T}}$.
- *Example* 2. In bioassays, the probability p_x of a response to a quantity x of a drug can sometimes be modeled by the logistic curve

  $$p_x = \frac{\exp(\alpha+\beta x)}{\exp(\alpha+\beta x)+1} = \frac{1}{1+\exp(-\alpha-\beta x)}.$$

  Then we get $H(x) = 1/(1+\mathrm{e}^{-x})$ and the parameter is $(\alpha, \beta)^{\mathrm{T}}$.
- *Example* 3. A famous family of models used in economics, for example in the study of demand and production functions, is the Cobb–Douglas family

  $$y = \theta_0 x_1^{\theta_1} x_2^{\theta_2} \cdots x_k^{\theta_k} = \exp(\log\theta_0 + \theta_1 \log x_1 + \theta_2 \log x_2 + \cdots + \theta_k \log x_k).$$

  Obviously, $H(x)$ is the same as *Example* 1 and the parameter is $(\log\theta_0, \theta_1, \ldots, \theta_k)^{\mathrm{T}}$.

Therefore, this paper aims to deal with the regression analysis when the concerned variables are associated in the form of model (1), i.e. the general PNLMs that can be transformed into it. We would like to mention some previous works that are related to but different from ours. Gao and Liang (1997) studied the general partially nonlinear model in a fixed design case. They used a finite series to approximate the nonparametric part. When $H(\cdot)$ is a linear function, it is the partially linear model(PLM) and there is a great deal of literature on the study of PLMs (e.g. Engle et al., 1986; Robinson, 1988; Speckman, 1988). A survey of PLMs was given by Härdle et al. (1999).

In the studies of semiparametric statistics, it is well known that dealing with the nonparametric component is often difficult. This is mainly because optimization over a large parameter space leads to undesirable properties of the estimates, such as inconsistency and roughness. One method is kernel or nearest neighbor by using a function to assign weights to nearby observations, which is in the sense of *local*. In the sense of *global*, polynomial wavelets and splines are very effective, by expressing the unknown function as a weighted sum of several curves. The estimation can be carried out by either two-stage process, profile estimators, or a direct one step process, sieve estimators (Grenander, 1981). The sieve method uses a sequence of increasing spaces (sieves) to approximate a large parameter space. Then the sieve estimator is obtained by maximizing a criterion function restricted over the approximating space. For its precise definition and theoretical results please refer to Shen (1997).

In model (1), since the nuisance parameter(nonparametric part) space is the collection of all bounded real-valued continuous functions, each sieve is chosen as a collection of continuous piecewise linear functions. Estimators of regression coefficients and the nuisance parameter for any finite sample size are obtained by maximizing the least square criterion over the product of the finite-dimensional parameter space and the sieves. Using the methods of an empirical process, we establish the consistency and asymptotic distribution of our sieve least squares estimator, and the convergence rates of both parametric and nonparametric part are acceptable.

The rest of the article is organized as follows. Baseline assumptions and sieve least squares estimation (LSE) are stated in Section 2. Based on the assumptions, the asymptotic properties of the estimation are then established in Section 3. From the properties, the estimation is asymptotic efficient. Section 4 investigates the Monte Carlo simulations of the estimators. All proofs are relegated to the Appendix.

2. Model assumptions and sieve least squares estimator

In this section we will construct the sieve LSE for model (1). We first introduce some notations and list the baseline conditions that we use explicitly throughout this article.

A1. In model (1), the link function H is a strictly monotone function, and H has a second-order bounded derivative.

A2. (i) X is bounded, that is, there exists $M_1 > 0$, such that $P\{|X_j| \le M_1\} = 1$, where $j = 1, \ldots, d$.

(ii) $\beta = (\beta_1, \ldots, \beta_d)^{\mathrm{T}} \in A$, where A is a bounded closed set in $\mathbb{R}^d$ with boundary M_2.

(iii) $T \sim U(0, 1)$, the uniform distribution, and the joint distribution of (X, T) does not depend on (β, g).