

BOOTSTRAP APPROXIMATIONS IN A PARTIALLY LINEAR REGRESSION MODEL *

Wolfgang Härdle¹, Hua Liang^{1,2} and Volker Sommerfeld¹

Institut für Statistik und Ökonometrie

and Sonderforschungsbereich 373

Humboldt-Universität zu Berlin

D-10178 Berlin, Germany¹

and

Institute of Systems Sciences

Beijing 100080, China²

Abstract

Consider the semiparametric regression model $Y_i = X_i^T \beta + g(T_i) + \epsilon_i$ ($i = 1, \dots, n$), where (X_i, T_i) are known and fixed design points, β is a p -dimensional unknown parameter, $g(\cdot)$ is an unknown function on $[0, 1]$, ϵ_i are i.i.d. random errors with mean 0 and variance σ^2 . In this paper we first construct bootstrap statistics β_n^* and σ_n^{2*} by resampling. Then we prove that, for the estimators β_n and σ_n^2 of the parameters β and σ^2 , $\sqrt{n}(\beta_n^* - \beta_n)$ and $\sqrt{n}(\beta_n - \beta)$, $\sqrt{n}(\sigma_n^{2*} - \sigma_n^2)$ and $\sqrt{n}(\sigma_n^2 - \sigma^2)$ have the same limit distributions, respectively. The advantage of the bootstrap approximation is explain. The feasibility of this approach we also show in a simulation study.

Key Words and Phrases: Semiparametric regression model, bootstrap approximation, asymptotic normality.

AMS 1991 subject classification: Primary: 62G05 Secondary: 60F15.

*This research was supported by Sonderforschungsbereich 373 “Quantifikation und Simulation Ökonomischer Prozesse”. The work of Hua Liang was partially supported by Alexander von Humboldt Foundation. The authors are extremely grateful to Dr. Michael Neumann for his many valuable suggestions and comments which greatly improved the presentation of the paper. Corresponding Author: Hua Liang, Institut für Statistik und Ökonometrie, Spandauerstr.1, D-10178 Berlin. Email:hliang@wiwi.hu-berlin.de.

1 INTRODUCTION

Consider the model given by

$$Y_i = X_i^T \beta + g(T_i) + \epsilon_i, i = 1, \dots, \quad (1)$$

where $X_i = (x_{i1}, \dots, x_{ip})^T$ ($p \geq 1$) and T_i ($T_i \in [0, 1]$) are known design points, $\beta = (\beta_1, \dots, \beta_p)^T$ is an unknown parameter vector and g is an unknown function, and $\epsilon_1, \dots, \epsilon_n$ are i.i.d. random variables with mean 0 and unknown variance σ^2 .

This model is important because it can be used in applications where one can assume that the responses Y_i and predictors X_i is linear dependence, but nonlinearly related to the independent variables T_i . Engle, et al. (1986) studied the effect of weather on electricity demand. Liang, Härdle and Werwatz (1997) used the model to investigate the relationship between income and age from German data. From the point of theory, this model generalizes the standard linear model, by restricts multivariate nonparametric regression which is subject to "Curse of Dimensionality" and is hard to interpret. Therefore, there is a lot of literature studied this model recently. Heckman (1986), Speckman (1988), Chen (1988) considered the asymptotic normality of estimators of β and σ^2 . Later Cuzick (1992a, b) and Schick (1993) discussed asymptotic properties and asymptotic efficiency for these estimators. Liang and Cheng (1993) discussed the second order asymptotic efficiency of LS estimator and MLE of β .

The technique of bootstrap is a useful tool for the approximation of an unknown probability distribution and therefore for its characteristics like moments or confidence regions. This approximation can be performed by different estimators of the true underlying distribution that should be well adapted to the special situation. In this paper we use the empirical distribution function which puts mass of $1/n$ at each residual in order to approximate the underlying error distribution (for more details see section 2). This classical bootstrap technique was introduced by B. Efron (for a review see e.g. Efron & Tibshirani, 1993). Note that for a heteroscedastic error structure a wild bootstrap procedure (see e.g. Wu, 1986 or Härdle & Mammen, 1993) would be more appropriate.

Hong and Cheng (1993) considered bootstrap approximation of the estimators for the parameters in the model (1) in the case where $\{X_i, T_i, i = 1, \dots, n\}$ are i.i.d. random variables and $g(\cdot)$ is estimated by a kernel smoother. The authors proved that their bootstrap approximation is the same as the classic methods, but failed to explain the advantage of

the bootstrap method, which will be discussed in this paper. We will construct bootstrap statistics of β and σ^2 , and studies their asymptotic normality when (X_i, T_i) are known design points and $g(\cdot)$ is estimated by general nonparametric fitting. Analytically as well as numerically we will show that the bootstrap techniques provide a reliable method to approximate the distributions of the estimates.

The effect of smoothing parameter is studied in a simulation study. Thereby it turns out that the estimators of the parametric part are quite robust against the choice of the smoothing parameter. More details can be found in section 3.

The paper is organized as follows. In the following we explain the basic idea for estimating the parameters and give the assumptions on the X_i and T_i . Section 2 constructs bootstrap statistics of β and σ^2 . In section 3 we present a simulation study in order to complete the asymptotic results. In section 4 some lemmas required later are proven. Section 5 presents the proof of the main result. For the convenience and simplicity, we shall employ $C(0 < C < \infty)$ to denote some constant not depending on n but may assume different values at each appearance.

Generally there are two methods, backfitting and local likelihood ones, to estimate the linear parameter. The asymptotic variance of the two estimates is the same. Here we adopt local likelihood method. Specifically, fix β one estimates $g(\cdot)$ as a function of β to obtain $\hat{g}(\beta)$, which is a nonparametric estimation problem. Then letting $g = \hat{g}(\beta)$, one estimates the parametric component, and this is a parametric problem. The detailed discussions can be also found in Severini and Staniswalis (1994).

To estimate g for fixed β , let $\omega_{ni}(t) = \omega_{ni}(t; T_1, \dots, T_n)$ be positive weight functions depending only on the design points T_1, \dots, T_n . Assume $\{X_i = (x_{i1}, \dots, x_{ip})^T, T_i, Y_i, i = 1, \dots, n\}$ satisfy the model (1). $\hat{g}_\beta(t) = \sum_{j=1}^n \omega_{nj}(t)(Y_j - X_j^T \beta)$ is just the nonparametric estimate of $g(t)$ for fixed β . Given the estimator $\hat{g}_\beta(t)$, an estimate of β , β_n , is obtained basing on $Y_i = X_i^T \beta + \hat{g}_\beta(T_i) + \epsilon_i$ for $i = 1, \dots, n$.

Denote $\tilde{\mathbf{X}} = (\tilde{X}_1, \dots, \tilde{X}_n)^T$, $\tilde{X}_i = X_i - \sum_{j=1}^n \omega_{nj}(T_i)X_j$, $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \dots, \tilde{Y}_n)^T$, $\tilde{Y}_i = Y_i - \sum_{j=1}^n \omega_{nj}(T_i)Y_j$. Then the estimate β_n can be expressed as

$$\beta_n = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{Y}}$$

In addition, the estimate of σ^2 , σ_n^2 is naturally defined as

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i^T \beta_n - g_n(T_i))^2,$$

which is equal to $\frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^T \beta_n)^2$. Where $g_n(t) = \sum_{j=1}^n \omega_{nj}(t)(Y_j - X_j^T \beta_n)$ is the estimate of $g(t)$.

In the following we list the sufficient conditions for our main result.

Condition 1. *There exist functions $h_j(\cdot)$ defined on $[0, 1]$ such that*

$$x_{ij} = h_j(T_i) + u_{ij} \quad 1 \leq i \leq n, \quad 1 \leq j \leq p \quad (2)$$

where u_{ij} is a sequence of real numbers which satisfy $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n u_i = 0$ and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n u_i u_i^T = B \quad (3)$$

is a positive definite matrix, and

$$\limsup_{n \rightarrow \infty} \frac{1}{a_n} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k u_i \right| < \infty \quad (4)$$

holds, where $u_i = (u_{i1}, \dots, u_{ip})^T$ and $a_n = n^{5/6} \log^{-1} n$.

Condition 2. *$g(\cdot)$ and $h_j(\cdot)$ are Lipschitz continuous of order 1.*

Condition 3. *The weight functions $\omega_{ni}(\cdot)$ satisfy the following:*

- (i) $\max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) = O(1)$,
- (ii) $\max_{1 \leq i, j \leq n} \omega_{ni}(T_j) = O(b_n)$,
- (iii) $\max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) I_{(|T_j - T_i| > c_n)} = O(c_n)$,

where $b_n = n^{-2/3}$, $c_n = n^{-1/3} \log n$.

These conditions are not more complicated than that given in related literature. They are usually needed for establishing asymptotic normality for the estimators of the parameters. Specifically, imposing Condition 1 in that we can lead $1/n \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ converges to B . In fact, (2) of Condition 1 is parallel to the case

$$h_j(T_i) = E(x_{ij}|T_i) \quad \text{and} \quad u_{ij} = x_{ij} - E(x_{ij}|T_i)$$

when (X_i, T_i) are random variables. (3) is similar to the result of the strong law of large numbers for random errors. (4) is similar to the law of the iterated logarithm. More detailed discussions may be found in Speckman (1988) and Gao et al. (1995).

The weight functions satisfied the above condition 3 are presented in Liang, Härdle and Werwatz (1997). Interested readers please find them there.

2 BOOTSTRAP APPROXIMATIONS AND MAIN RESULT

The statistics β_n and σ_n^2 have asymptotic standard normal distributions under mild assumptions. Our simulation studies indicate that the normal approximation does not work very well for small samples. Therefore in this section we propose a bootstrap method as an alternative to the normal asymptotic method.

In the semiparametric regression model the observable column n -vector $\hat{\epsilon}$ of residuals is given by

$$\hat{\epsilon} = \mathbf{Y} - \mathbf{G}_n - \mathbf{X}^T \beta_n$$

where $\mathbf{G}_n = \{g_n(T_1), \dots, g_n(T_n)\}^T$. Denote $\mu_n = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i$. Let \hat{F}_n be the empirical distribution of $\hat{\epsilon}$, centered at the mean, so \hat{F}_n puts mass $1/n$ at $\hat{\epsilon}_i - \mu_n$ and $\int x d\hat{F}_n(x) = 0$. Given \mathbf{Y} , let $\epsilon_1^*, \dots, \epsilon_n^*$ be conditional independent with common distribution \hat{F}_n ; let ϵ^* be the n -vector whose i th component is ϵ_i^* , and let

$$\mathbf{Y}^* = \mathbf{X}^T \beta_n + \mathbf{G}_n + \epsilon^*.$$

Informally, $\hat{\epsilon}^*$ is obtained by resampling the centered residuals. And \mathbf{Y}^* is generated from the data, using the regression model with β_n as the vector of parameters and \hat{F}_n as the distribution of the disturbance terms ϵ^* .

So we have reason to define the estimates of β and σ^2 as follows, respectively,

$$\beta_n^* = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{Y}}^* \quad \text{and} \quad \sigma_n^{2*} = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i^* - \tilde{X}_i^T \beta_n^*)^2,$$

where $\tilde{Y}_i^* = Y_i^* - \sum_{j=1}^n \omega_{nj}(T_i) Y_j^*$, $\tilde{\mathbf{Y}}^* = (\tilde{Y}_1^*, \dots, \tilde{Y}_n^*)^T$.

The bootstrap principle is that the distributions of $\sqrt{n}(\beta_n^* - \beta_n)$ and $\sqrt{n}(\sigma_n^{2*} - \sigma_n^2)$, which can be computed directly from the data, approximate the distributions of $\sqrt{n}(\beta_n - \beta)$ and $\sqrt{n}(\sigma_n^2 - \sigma^2)$, respectively. As will be shown later, this approximation is likely to be very good, provided n is large enough. This fact is stated as the following theorem.

Theorem 1. *Suppose conditions 1-3 hold. If $E\epsilon_1^4 < \infty$ and $\max_{1 \leq i \leq n} \|u_i\| \leq C_0 < \infty$. Then*

$$\sup_x |P^* \{ \sqrt{n}(\beta_n^* - \beta_n) < x \} - P \{ \sqrt{n}(\beta_n - \beta) < x \} | \rightarrow 0 \quad (5)$$

and

$$\sup_x \left| P^* \{ \sqrt{n}(\sigma_n^{2*} - \sigma_n^2) < x \} - P \{ \sqrt{n}(\sigma_n^2 - \sigma^2) < x \} \right| \rightarrow 0 \quad (6)$$

where and below P^* and E^* denote the conditional probability and conditional expectation given \mathbf{Y} .

Now, we outline our proof of the theorem. First we decompose $\sqrt{n}(\beta_n - \beta)$ and $\sqrt{n}(\beta_n^* - \beta_n)$ into three terms, and σ_n^2 and σ_n^{2*} into five terms, respectively. Then we will calculate the tail probability value of each term. Some additional notations are introduced. $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, $\tilde{\epsilon} = (\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n)^T$, $\tilde{\epsilon}_i = \epsilon_i - \sum_{j=1}^n \omega_{nj}(T_i)\epsilon_j$, $\tilde{g}_i = g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)g(T_k)$, $\tilde{\mathbf{G}} = (\tilde{g}_1, \dots, \tilde{g}_n)^T$. We have from the definitions of β_n and β_n^* , and σ_n^2 and σ_n^{2*}

$$\begin{aligned} \sqrt{n}(\beta_n - \beta) &= \sqrt{n}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}^T \tilde{\mathbf{G}} + \tilde{\mathbf{X}}^T \tilde{\epsilon}) \\ &= \sqrt{n}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \left[\sum_{i=1}^n \tilde{X}_i \tilde{g}_i - \sum_{i=1}^n \tilde{X}_i \left\{ \sum_{j=1}^n \omega_{nj}(T_i)\epsilon_j \right\} + \sum_{i=1}^n \tilde{X}_i \epsilon_i \right] \\ &\stackrel{\text{def}}{=} n(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}(H_1 - H_2 + H_3). \end{aligned}$$

$$\begin{aligned} \sqrt{n}(\beta_n^* - \beta_n) &= \sqrt{n}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}^T \tilde{\mathbf{G}}_{\mathbf{n}}^* + \tilde{\mathbf{X}}^T \tilde{\epsilon}^*) \\ &= \sqrt{n}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \left[\sum_{i=1}^n \tilde{X}_i \tilde{g}_{ni}^* - \sum_{i=1}^n \tilde{X}_i \left\{ \sum_{j=1}^n \omega_{nj}(T_i)\epsilon_j^* \right\} + \sum_{i=1}^n \tilde{X}_i \epsilon_i^* \right] \\ &\stackrel{\text{def}}{=} n(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}(H_1^* - H_2^* + H_3^*). \end{aligned}$$

Where $\tilde{\mathbf{G}}_{\mathbf{n}}^* = (\tilde{g}_{n1}^*, \dots, \tilde{g}_{nn}^*)^T$ with $\tilde{g}_{ni}^* = g_n(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)g_n(T_k)$ for $i = 1, \dots, n$.

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n} \tilde{\mathbf{Y}}^T \{ I - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \} \tilde{\mathbf{Y}} \\ &= \frac{1}{n} \epsilon^T \epsilon - \frac{1}{n} \epsilon^T \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \epsilon + \frac{1}{n} \tilde{\mathbf{G}}^T \{ I - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \} \tilde{\mathbf{G}} \\ &\quad - \frac{2}{n} \tilde{\mathbf{G}}^T \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \epsilon + \frac{2}{n} \tilde{\mathbf{G}}^T \epsilon \\ &\stackrel{\text{def}}{=} I_1 - I_2 + I_3 - 2I_4 + 2I_5. \end{aligned}$$

$$\begin{aligned} \sigma_n^{2*} &= \frac{1}{n} \tilde{\mathbf{Y}}^{*T} \{ I - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \} \tilde{\mathbf{Y}}^* \\ &= \frac{1}{n} \epsilon^{*T} \epsilon^* - \frac{1}{n} \epsilon^{*T} \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \epsilon^* + \frac{1}{n} \tilde{\mathbf{G}}_{\mathbf{n}}^{*T} \{ I - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \} \tilde{\mathbf{G}}_{\mathbf{n}}^* \\ &\quad - \frac{2}{n} \tilde{\mathbf{G}}_{\mathbf{n}}^{*T} \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \epsilon^* + \frac{2}{n} \tilde{\mathbf{G}}_{\mathbf{n}}^{*T} \epsilon^* \\ &\stackrel{\text{def}}{=} I_1^* - I_2^* + I_3^* - 2I_4^* + 2I_5^*. \end{aligned}$$

Here I is the identity matrix of order p . The following sections will prove that $H_{1j}, H_{2j} = o_P(1)$ and $H_{1j}^*, H_{2j}^* = o_{P^*}(1)$ and $I_i = o_P(n^{-1/2})$ and $I_i^* = o_{P^*}(n^{-1/2})$ for $j = 1, \dots, p$ and $i = 2, 3, 4, 5$.

We have up to now showed that the bootstrap method performs as least as well as the normal approximation with the error rate of $o_p(1)$ and $o(1)$, respectively. It is natural to expect that the bootstrap method should perform better than this however. Indeed, our numerical experience means that it is case. In fact, it is also true analytically as is shown in the following theorem.

Theorem 2. *Let $M_{jn}(\beta) [(\sigma^2)]$ and $M_{jn}^*(\beta) [(\sigma^2)]$ be the j -th moments of $\sqrt{n}(\beta_n - \beta) [(\sqrt{n}(\sigma_n^2 - \sigma^2))]$ and $\sqrt{n}(\beta_n^* - \beta_n) [(\sqrt{n}(\sigma_n^{2*} - \sigma_n^2))]$, respectively. Then under the conditions 1-3 and $E\epsilon_1^6 < \infty$ and $\max_{1 \leq i \leq n} \|u_i\| \leq C_0 < \infty$,*

$$M_{jn}^*(\beta) - M_{jn}(\beta) = O_P(n^{-1/3} \log n) \quad \text{and} \quad M_{jn}^*(\sigma^2) - M_{jn}(\sigma^2) = O_P(n^{-1/3} \log n)$$

for $j = 1, 2, 3, 4$.

The proof of theorem 2 can be completed by the arguments of Liang (1994) and similar procedures behind. We omit the details.

Theorem 2 indicates that the bootstrap distributions have much better approximation for the first four moments for β_n^* and σ_n^{2*} , which are most important quantities in characterizing distributions. Indeed, by Theorem 1 and Lemma 1 given later, one can only obtain that

$$M_{jn}^*(\beta) - M_{jn}(\beta) = o_P(1) \quad \text{and} \quad M_{jn}^*(\sigma^2) - M_{jn}(\sigma^2) = O_P(1)$$

for $j = 1, 2, 3, 4$, in contrast to Theorem 2.

3 NUMERICAL RESULTS

In this section we present a small simulation study in order to illustrate the finite sample behavior of the estimator. We investigate the model

$$Y_i = X_i\beta + g(T_i) + \epsilon_i \tag{7}$$

where $g(T_i) = \sin(T_i)$, $\beta = (1, 5)'$ and $\epsilon_i \sim \text{Uniform}(-0.3, 0.3)$. The independent variables $X_i = (X_i^{(1)}, X_i^{(2)})$ and T_i are realizations of a $\text{Uniform}(0, 1)$ distributed random variable. We analyze sample sizes of 30, 50, 100 and 300. For nonparametric fitting, we use a Nadaraya-Watson kernel weight function with Epanechnikov kernel. We performed the smoothing with

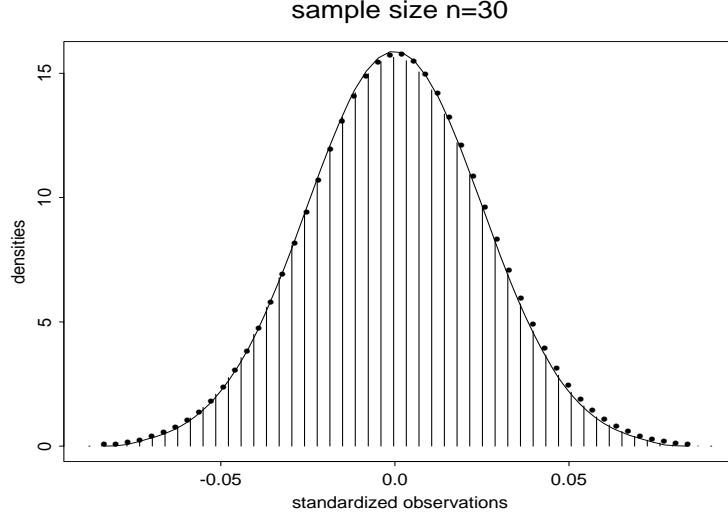


Figure 1 : Plot of of the smoothed bootstrap density (lines), the normal approximation (stars) and the smoothed true density (vertical lines).

different bandwidths using some grid search. Thereby it turned out that the results for the parametric part are quite robust against the bandwidth chosen in the nonparametric part. In the following we present only the simulation results for the parameter β_2 , those for β_1 are similar.

In figures 1 to 4 we plotted the smoothed densities of the estimated true distribution of $\sqrt{n}(\hat{\beta}_2 - \beta_2)/\hat{\sigma}$ with $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i^T \beta_n)^2$ for sample sizes 30, 50, 100, 300. Additionally we added to these plots the corresponding bootstrap distributions and the asymptotic normal distributions, where we estimated the asymptotic variance $\sigma^2 B^{-1}$ by $\hat{\sigma}^2 \hat{B}^{-1}$ with $\hat{B} = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i^T \tilde{X}_i$ and $\hat{\sigma}^2$ defined above. It turns out that the bootstrap distribution and the asymptotic normal distribution well approximates the true one even for moderate sample sizes of $n = 30$.

4 SOME LEMMAS

Under the conditions of Theorem 1, Gao et al.(1995) obtained asymptotic normalities of β_n and σ_n^2 and convergence rate of g_n , which are given in Lemma 1. Lemma 2 presents the limit of $1/n \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$. Its proof is referred to Chen (1988) and Speckman (1988). Lemma 3 provides the boundedness for $g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)g(T_k)$ and $\tilde{g}_{ni}^*(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)\tilde{g}_{nk}^*(T_k)$, whose proof is immediate. Lemma 4 shows that $\sqrt{n}H_{1j}$ and $\sqrt{n}H_{1j}^*$ are $O(n^{1/3} \log n)$ in different probability senses. Lemma 5 gives a general result for nonparametric regression,

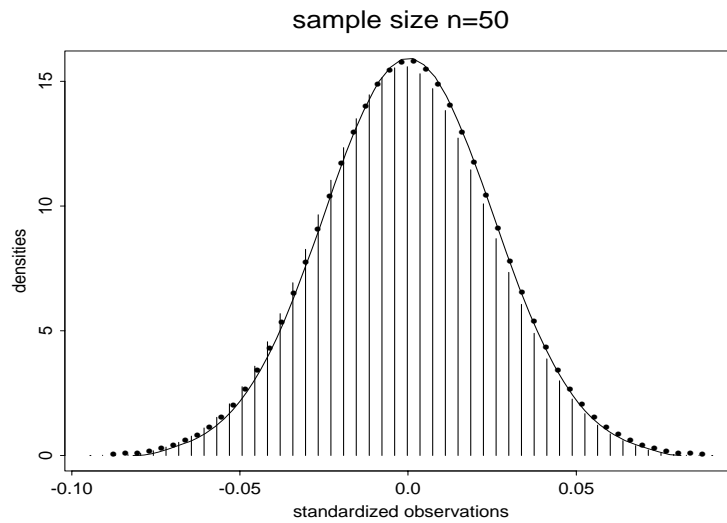


Figure 2 : Plot of of the smoothed bootstrap density (lines), the normal approximation (stars) and the smoothed true density (vertical lines).

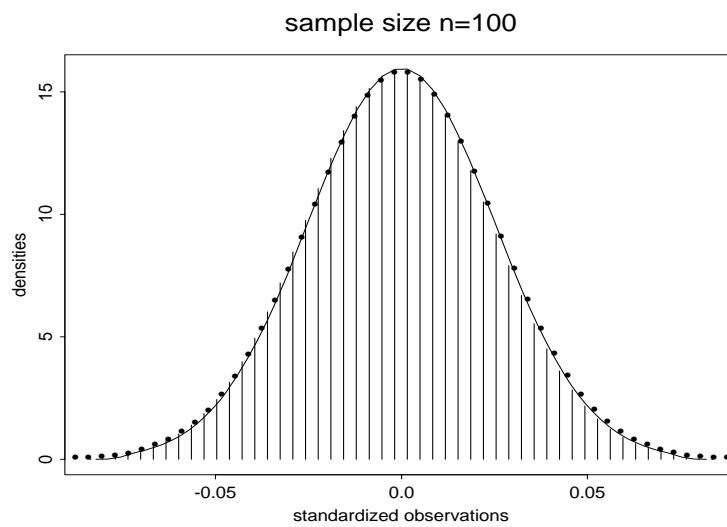


Figure 3 : Plot of of the smoothed bootstrap density (lines), the normal approximation (stars) and the smoothed true density (vertical lines).

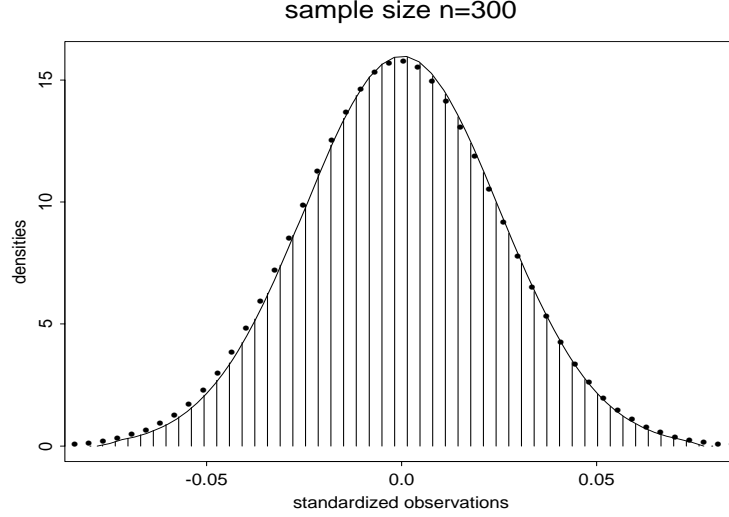


Figure 4 : Plot of the smoothed bootstrap density (lines), the normal approximation (stars) and the smoothed true density (vertical lines).

whose proof is strongly based on an exponential inequality for bounded independent random variables, that is, Bernstein's inequality. It will be used in the remainder of this section.

Lemma 1. *Suppose the conditions of Theorem 1 hold. Then*

$$\sqrt{n}(\beta_n - \beta) \rightarrow N(0, \sigma^2 B^{-1}), \quad \sup_{t \in [0,1]} |g_n(t) - g(t)| = O_p(n^{-1/3} \log n), \quad (8)$$

and

$$\sqrt{n}(\sigma_n^2 - \sigma^2) \rightarrow N(0, \text{Var}(\epsilon_1^2)) \quad (9)$$

Lemma 2. *If conditions 1-3 hold. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}} = B$$

Lemma 3. *Suppose that conditions 2 and 3 (iii) hold. Then*

$$\begin{aligned} \max_{1 \leq i \leq n} \left| g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) g(T_k) \right| &= O(n^{-1/3} \log n) \\ \max_{1 \leq i \leq n} \left| \tilde{g}_{ni}^*(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) \tilde{g}_{nk}^*(T_k) \right| &= O_P(n^{-1/3} \log n) \end{aligned}$$

The same conclusion as the first part holds for $h_j(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) h_j(T_k)$ for $j = 1, \dots, p$.

Lemma 4. *Suppose conditions 1-3 hold and $E|\epsilon_1|^3 < \infty$. Then*

$$\sqrt{n}H_{1j} = O(n^{1/2} \log^{-1/2} n) \quad \text{and} \quad \sqrt{n}H_{1j}^* = O(n^{1/2} \log^{-1/2} n) \quad \text{for } j = 1, \dots, p \quad (10)$$

Proof. Their proofs can be completed by the same methods for Lemmas 2.4 and 2.5 of Liang (1996). We omit the details.

(Bernstein's Inequality) Let V_1, \dots, V_n be independent random variables with zero means and bounded ranges: $|V_i| \leq M$. Then for each $\eta > 0$,

$$P\left\{\left|\sum_{i=1}^n V_i\right| > \eta\right\} \leq 2 \exp\left\{-\eta^2 / \left[2\left(\sum_{i=1}^n \text{var} V_i + M\eta\right)\right]\right\}.$$

Lemma 5. Assume that condition 3 holds. Let V_i be independent with mean zero and $EV_1^4 < \infty$. Then

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) V_k \right| = O_P(n^{-1/4} \log^{-1/2} n).$$

Proof. Denote $V_j' = V_j I_{(|V_j| \leq n^{1/4})}$ and $V_j'' = V_j - V_j'$ for $j = 1, \dots, n$. Let $M = Cb_n n^{1/4}$. From Bernstein's inequality

$$\begin{aligned} & P\left\{\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i)(V_j' - EV_j') \right| > C_1 n^{-1/4} \log^{-1/2} n\right\} \\ & \leq \sum_{i=1}^n P\left\{\left| \sum_{j=1}^n \omega_{nj}(T_i)(V_j' - EV_j') \right| > C_1 n^{-1/4} \log^{-1/2} n\right\} \\ & \leq 2n \exp\left\{-\frac{C_1 n^{-1/2} \log^{-1} n}{\sum_{j=1}^n \omega_{nj}^2(T_i) EV_j'^2 + 2c_n b_n \log^{-1/2} n}\right\} \\ & \leq 2n \exp\{-C_1^2 C \log n\} \leq Cn^{-1/2} \quad \text{for some large } C_1 > 0. \end{aligned}$$

This and Borel-Cantelli Lemma imply that

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i)(V_j' - EV_j') \right| = O_P(n^{-1/4} \log^{-1/2} n). \quad (11)$$

On the other hand, from condition 3(ii), we know

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i)(V_j'' - EV_j'') \right| \leq \max_{1 \leq k \leq n} \max_{1 \leq i \leq n} |\omega_{nk}(T_i)| \sum_{j=1}^n |V_j| = O_P(n^{-2/3})$$

and

$$\begin{aligned} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n \omega_{nj}(T_i) EV_j'' \right| & \leq \max_{1 \leq k \leq n} \max_{1 \leq i \leq n} |\omega_{nk}(T_i)| \sum_{j=1}^n n^{-1} E|V_j|^4 \\ & \leq Cn^{-2/3} \log n \max_{1 \leq i \leq n} E|V_i|^4 \\ & = o(n^{-1/4} \log^{-1/2} n). \end{aligned} \quad (12)$$

Combining the results of (11) to (12), we obtain

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) V_k \right| = O_P(n^{-1/4} \log^{-1/2} n). \quad (13)$$

This completes the proof of Lemma 5.

Lemma 6. *Suppose conditions 1-3 hold and $E|\epsilon_1|^3 < \infty$. Then*

$$\sqrt{n}H_{2j} = o(n^{1/2}) \quad \text{and} \quad \sqrt{n}H_{2j}^* = o(n^{1/2}) \quad \text{for } j = 1, \dots, p$$

Proof. Denote $h_{nij} = h_j(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) h_j(T_k)$. Observe the fact,

$$\begin{aligned} \sqrt{n}H_{2j} &= \sum_{i=1}^n \left\{ \sum_{k=1}^n \tilde{x}_{kj} \omega_{ni}(T_k) \right\} \epsilon_i \\ &= \sum_{i=1}^n \left\{ \sum_{k=1}^n u_{kj} \omega_{ni}(T_k) \right\} \epsilon_i + \sum_{i=1}^n \left\{ \sum_{k=1}^n h_{nkj} \omega_{ni}(T_k) \right\} \epsilon_i \\ &\quad - \sum_{i=1}^n \left[\sum_{k=1}^n \left\{ \sum_{q=1}^n u_{qj} \omega_{nq}(T_k) \right\} \omega_{ni}(T_k) \right] \epsilon_i \end{aligned}$$

Using conditions 3 (i) and (ii) and the remark in Lemma 3, we can deal with each term as (13) by letting $V_i = \epsilon_i$ in Lemma 5. The above each item can be proved to be $o_P(n^{1/2})$ by using Lemma 5 and the argument for proving Lemma 5. The same technique is also suggested to $\sqrt{n}H_{2j}^*$. We omit the details.

Lemma 7. *Under the conditions of Lemma 5. $I_n = o_P(n^{1/2})$, where*

$$I_n = \sum_{i=1}^n \sum_{j \neq i} \omega_{nj}(T_i) (V_j' - EV_j') (V_i' - EV_i').$$

Proof. Let $j_n = \lceil n^{2/3} \log^2 n \rceil$, ($[a]$ denotes the integer portion of a .) $A_j = \left\{ \left[\frac{(j-1)n}{j_n} \right] + 1, \dots, \left[\frac{jn}{j_n} \right] \right\}$, $A_j^c = \{1, 2, \dots, n\} - A_j$ and $A_{ji} = A_j - \{i\}$. Observe that I_n can be decomposed as follows,

$$\begin{aligned} I_n &= \sum_{j=1}^{j_n} \sum_{i \in A_j} \sum_{k \in A_{ji}} \omega_{nk}(T_i) (V_k' - EV_k') (V_i' - EV_i') \\ &\quad + \sum_{j=1}^{j_n} \sum_{i \in A_j} \sum_{k \in A_j^c} \omega_{nk}(T_i) (V_k' - EV_k') (V_i' - EV_i') \\ &\stackrel{\text{def}}{=} \sum_{j=1}^{j_n} U_{nj} + \sum_{j=1}^{j_n} V_{nj} \\ &\stackrel{\text{def}}{=} I_{1n} + I_{2n}. \end{aligned} \quad (14)$$

Where

$$\begin{aligned} U_{nj} &= \sum_{i \in A_j} p_{nij} (V_i' - EV_i') \stackrel{\text{def}}{=} \sum_{i \in A_j} u_{nij}, \\ V_{nj} &= \sum_{i \in A_j} q_{nij} (V_i' - EV_i') \stackrel{\text{def}}{=} \sum_{i \in A_j} v_{nij}, \end{aligned}$$

and

$$\begin{aligned} p_{nij} &= \sum_{k \in A_{ji}} \omega_{nk}(T_i)(V_k' - EV_k'), \\ q_{nij} &= \sum_{k \in A_j^c} \omega_{nk}(T_i)(V_i' - EV_i'). \end{aligned}$$

Notice that $\{v_{nij}, i \in A_j\}$ are conditionally independent random variables given $E_{nj} = \{V_k, k \in A_j^c\}$ with $E(v_{nij}|E_{nj}) = 0$ and $E(v_{nij}^2|E_{nj}) \leq \sigma^2(\max_{1 \leq i \leq n} |q_{nij}|)^2 \stackrel{\text{def}}{=} \sigma^2 q_{nj}^2$ for $i \in A_j$, and satisfy $\max_{1 \leq i \leq n} |v_{nij}| \leq 2n^{1/4} q_{nj}$ for $q_{nj} = \max_{1 \leq i \leq n} |q_{nij}|$.

On the other hand, by the same reason as that for Lemma 5,

$$\begin{aligned} q_n &= \max_{1 \leq j \leq j_n} |q_{nj}| = \max_{1 \leq j \leq j_n} \max_{1 \leq i \leq n} \left| \sum_{k \in A_j^c} \omega_{nk}(T_i)(V_k' - EV_k') \right| \\ &= O_P(n^{-1/4} \log^{-1/2} n), \end{aligned}$$

Denote the numbers of the elements in A_j by $\#A_j$. By applying Bernstein's inequality, we have, for $j = 1, \dots, j_n$,

$$P\left\{|V_{nj}| > \frac{C\sqrt{n}}{\sqrt{\log nj_n}} \middle| E_{nj}\right\} \leq C \exp\left\{-\frac{Cn(\log^{-1} n)j_n^{-2}}{\sigma^2 q_n^2 \#A_j + j_n^{-1} n^{1/4} q_n}\right\} \leq Cn^{-1/2}.$$

It follows from the bounded dominant convergence theorem, the above fact and $\#A_j \leq \frac{n}{j_n}$ that

$$\begin{aligned} P\left\{|V_{nj}| > \frac{C\sqrt{n}}{\sqrt{\log nj_n}}\right\} &\leq C \exp\left\{-\frac{Cj_n^{-2}}{\sigma^2 q_n^2 n j_n^{-1} + j_n^{-1} n^{1/4} q_n}\right\} \\ &\leq Cn^{-1/2}, \quad \text{for } j = 1, \dots, j_n. \end{aligned}$$

Then

$$I_{2n} = o_P(\sqrt{n}). \tag{15}$$

Now we consider I_{1n} . Note that $\{V_k, 1 \leq k \leq n\}$ are i.i.d. random variables, and the definition of U_{nj} , we know that

$$P\{|I_{1n}| > C\sqrt{n}(\log^{-1/2} n)\} \leq Cn(\log^{-1} n)E\left\{\sum_{j=1}^{j_n} U_{nj}\right\}^2$$

$$\begin{aligned}
&= Cn(\log^{-1} n) \left(\sum_{j=1}^{j_n} EU_{nj}^4 + \sum_{j_1 \neq j_2}^{j_n} EU_{nj_1}^2 EU_{nj_2}^2 \right) \\
&\leq Cn(\log^{-1} n) j_n^2 (\#A_j)^2 \cdot b_n^4 \left[E(V_1' - EV_1')^4 + \{E(V_1' - EV_1')^2\}^2 \right] \\
&\leq Cn^{-1/2}.
\end{aligned} \tag{16}$$

Hence $I_{1n} = o_P(\sqrt{n})$. Combining (14), (15) with (16), we complete the proof of Lemma 7.

5 PROOF OF THEOREM 1

In this section we present the proof of Theorem 1. First, we prove (5). From (10) and Lemma 6, we only need to prove $\sqrt{n}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \epsilon^*$ converges in distribution to a k -variate normal random variate with mean 0 and covariance matrix $\sigma^2 B^{-1}$.

Let q_{ii} be the i th diagonal element of the matrix $\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T$. According to proposition 2.2 of Huber (1973), if we know $\max_i q_{ii} \rightarrow 0$ as $n \rightarrow \infty$, then $\sqrt{n}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \epsilon^*$ is asymptotically normal. Since the covariance matrix of $\sqrt{n}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \epsilon^*$ is given by $n(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \int u^2 d\widehat{F}_n(u)$. Recall the definition of $\widehat{F}_n(u)$ and the result given in Lemma 2, the asymptotic variance of $\sqrt{n}(\beta_n^* - \beta_n)$ is $\sigma^2 B^{-1}$.

We now prove $\max_i q_{ii} \rightarrow 0$. Since $n^{-1}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}) \rightarrow B$ by Lemma 2, it follows from Lemma 3 of Wu (1981) that $\max_i q_{ii} \rightarrow 0$. This completes the proof of (5).

Next, we will prove (6). First we continue to give the following preliminary results. In Lemma 5, letting V_i be ϵ_i^* , E and P be E^* and P^* , then we have

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) \epsilon_k^* \right| = O_{P^*}(n^{-1/4} \log^{-1/2} n).$$

This and Lemma 3 and the fact

$$|\sqrt{n}I_3^*| \leq C\sqrt{n} \max_{1 \leq i \leq n} \left\{ \left| g_n(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) g_n(T_k) \right|^2 + \left| \sum_{k=1}^n \omega_{nk}(T_i) \epsilon_k^* \right|^2 \right\}$$

lead that $|\sqrt{n}I_3^*| = o_{P^*}(1)$.

Using the similar arguments as for proving $\sqrt{n}(\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \epsilon^* \rightarrow N(0, \sigma^2 B^{-1})$, one can conclude that

$$\sqrt{n}I_2^* = o_{P^*}(1), \quad \sqrt{n}I_4^* = o_{P^*}(1).$$

Now, we consider I_5^* . We decompose I_5^* into three terms, and prove each term tends to zero.

More precisely,

$$I_5^* = \frac{1}{n} \left\{ \sum_{i=1}^n \widetilde{g}_{ni}^* \epsilon_i^* - \sum_{k=1}^n \omega_{ni}(T_k) \epsilon_k^{*2} - \sum_{i=1}^n \sum_{k \neq i}^n \omega_{ni}(T_k) \epsilon_i^* \epsilon_k^* \right\}$$

$$\stackrel{\text{def}}{=} I_{51}^* + I_{52}^* + I_{53}^*. \quad (17)$$

From Lemma 3, using the strong law of large number to ϵ_i^* , we have

$$\sqrt{n}I_{51}^* = o_{P^*}(1), \quad (18)$$

and

$$\sqrt{n}I_{52}^* \leq b_n \sqrt{n} \sum_{i=1}^n \epsilon_i^{*2} = O_{P^*}(\log^{-2} n) = o_{P^*}(1). \quad (19)$$

Let $I_n^* = \sum_{i=1}^n \sum_{j \neq i} \omega_{nj}(T_i)(\epsilon_j^{*'} - E\epsilon_j^{*'})(\epsilon_i^* - E\epsilon_i^*)$. Observe

$$\sqrt{n}|I_{53}^*| \leq \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \sum_{k \neq i} \omega_{nk}(T_i) \epsilon_i^* \epsilon_k^* - I_n^* \right| + I_n^* \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}}(J_{1n}^* + I_n^*).$$

Simple calculations show that

$$\begin{aligned} J_{1n}^* &\leq \max_{1 \leq j \leq n} \left| \sum_{i=1}^n \omega_{nj}(T_i) \epsilon_i^* \right| \left(\sum_{i=1}^n |\epsilon_i^{*''}| + E|\epsilon_i^{*''}| \right) \\ &\quad + \max_{1 \leq j \leq n} \left| \sum_{i=1}^n \omega_{nj}(T_i) (\epsilon_i^{*'} - E\epsilon_i^{*'}) \right| \left(\sum_{i=1}^n |\epsilon_i^{*''}| + E|\epsilon_i^{*''}| \right) \\ &= o_{P^*}(1) \end{aligned}$$

By letting ϵ_i^* be V_i , E^* and P^* be E and P in Lemma 7 respectively, we have $I_n^* = O_{P^*}(\sqrt{n})$.

It follows that

$$\frac{1}{\sqrt{n}} I_{53}^* = o_{P^*}(1). \quad (20)$$

A combination (18)-(20) leads that $\sqrt{n}I_5^* = o_{P^*}(1)$.

From the above arguments and the third result of Lemma 1, the proof of (6) is equivalent to show

$$\frac{1}{\sqrt{n}} (\epsilon^{*T} \epsilon^* - \int u^2 d\hat{F}_n(u)) \rightarrow N(0, \text{Var}(\epsilon_1^2)),$$

which can be verified by the central limit theorem. Thus we complete the proof of Theorem 1.

REFERENCES

- Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for bootstrap. *Annals of Statistics*, **9**, 1196-1217.
- Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Annals of Statistics*, **16**, 136-146.
- Cuzick, J. (1992a). Semiparametric additive regression. *Journal of the Royal Statistical Society, Series B*, **54**, 831-843.
- Cuzick, J. (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Annals of Statistics*, **20**, 1129-1136.
- Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Freedman, D.A. (1981). Bootstrapping regression models. *Annals of Statistics*, **9**, 1218-1228.
- Gao, J. T., Hong, S.Y. and Liang, H. (1995). Convergence rates of a class of estimates in partly linear models. *Acta Mathematicae Sinica Series New*, **38**, 658-669.
- Härdle, W. & Mammen, E. (1993). Comparing Nonparametric Versus Parametric Regression Fits. *Annals of Statistics*, **21**, 1926-1947.
- Heckman, N.E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.
- Hong, S.Y. and Cheng, P. (1993). Bootstrap approximation of estimation for parameter in a semiparametric regression model. *Sciences in China Ser. A* **23**, 239-251.
- Huber, P.J. (1973). Robust regression: asymptotic, conjectures and Monte-Carlo. *Annals of Statistics*, **1**, 799-821.
- Liang, H. (1994). The Berry-Esseen bounds of error variance estimation in a semiparametric regression model. *Communications in Statistics, Theory & Methods*, **23**, 3439-3452.
- Liang, H. (1996). Law of the iterated logarithm for parameters in a partly linear regression model. Submitted for publication.
- Liang, H. and Cheng, P. (1993). Second order asymptotic efficiency in a partial linear model. *Statistics & Probability Letters*, **18**, 73-84.
- Liang, H., Härdle, W. and Werwatz, A. (1997). *Asymptotic properties of nonparametric regression estimation in partly linear models*. Discussion paper no. 55, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- Severini, T.A. and Staniswalis, J.G. (1994). Quasilikelihood estimation in semiparametric models, *Journal of the American Statistical Association*, **89**, 501-511.
- Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.
- Wu, C.F.J. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, **9**, 501-513.

Wu, C.F.J. (1986). Jackknife, Bootstrap and other Resampling Methods in Regression Analysis (with Discussion). *Annals of Statistics*, **14**, 1261-1295.