ASYMPTOTIC NORMALITY OF PARAMETRIC PART IN PARTIAL LINEAR HETEROSCEDASTIC REGRESSION MODELS

Hua Liang and Wolfgang Härdle *

Abstract

Consider the partial linear heteroscedastic model $Y_i = X_i^T\beta + g(T_i) + \sigma_i e_i, 1 \le i \le n$ with random variables (X_i, T_i) and response variables Y_i and unknown regression function $g(\bullet)$. We assume that the errors are heteroscedastic, i.e., $\sigma_i^2 \neq const.$ e_i are i.i.d. random error with mean zero and variance 1. In this partial linear heteroscedastic model, we consider the situations that the variances are an unknown smooth function of exogenous variables, or of nonlinear variables T_i, or of the mean response $X_i^T\beta + g(T_i)$. Under general assumptions, we construct an estimator of the regression parameter vector β which is asymptotically equivalent to the weighted least squares estimatores with known variance. In procedure of constructing the estimators, the technique of splitting the samples is adopted.

Key Words and Phrases: Nonparametric estimation, partial linear model, heteroscedastic, semiparametric model, asymptotic normality.
Short title: Heteroscedasticity

1 INTRODUCTION

Consider the semiparametric partial linear regression model, which is defined by

$$Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, i = 1, \dots, n \tag{1}$$

*Hua Liang is Associate Professor of Statistics, at Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080. Wolfgang Härdle is Professor of Econometrics, at the Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany. This research was supported by Sonderforschungsbereich 373 “Quantifikation und Simulation Ökonomischer Prozesse” . The first author was supported by Alexander von Humboldt Foundation. The authors would like to thank Dr. Ulrike Grasshoff for her valuable comments.

with $X_i = (x_{i1}, \ldots, x_{ip})^T$ and $T_i \in [0,1]$ random design points, $\beta = (\beta_1, \ldots, \beta_p)^T$ the unknown parameter vector and g an unknown Lipschitz continuous function from $[0,1]$ to $\mathbb{R}^1$. The random errors $\varepsilon_1, \ldots, \varepsilon_n$ are mean zero variables with variance 1.

This model was studied by Engle, et al. (1986) under the assumption of constant error variance. More recent work in this semiparametric context dealt with the estimation of β at a root-n rate. Chen (1988), Heckman (1986), Robinson (1988) and Speckman (1988) constructed $\sqrt{n}$–consistent estimates of β under various assumptions on the function g and on the distributions of ε and (X,T). Cuzick (1992a) constructed efficient estimates of β when the error density is known. The problem was extended later to the case of unknown error distribution by Cuzick (1992b) and Schick (1993).

Schick (1996a, b) considered the problem of heteroscedasticity, i.e., of nonconstant error variance, for model (1). He constructed root-n consistent weighted least squares estimates with random weight of the finite-dimensional parameter, and gave an optimal weight function when the variance is known up to a multiplicative constant. His model for nonconstant variance function of Y given (X,T) assumed that it is some unknown smooth function of an exogenous random vector W, which is unrelated with β and g.

The present paper focus to uniformly existed approaches in the literature and to extend some of existing results. It is concerned with the cases that σ_i^2 is some function of some independent exogenous variables; for some function of T_i; for some function of $X_i^T\beta + g(T_i)$. The aim of this paper is to present a uniformly applicable method for estimating the parameter β on the regression model (1) with heteroscedastic error, and then to prove that in large samples there is no cost due to estimating the variance function under appropriate conditions.

In our analysis it is related to the literature on attention in semiparametric models. Earlier papers are Bickel (1978), Carroll (1982), Carroll and Ruppert (1982) and Müller, et al. (1987). There are mainly two kind of theoretical analysis, that is, the parametric approach, which generally assumed $\sigma_i^2 = H(X_i, \theta)$ or $H(X_i^T\beta, \theta)$ for H being known. [See Box and Hill (1974), Carroll (1982), Carroll and Ruppert (1982), Jobson and Fuller (1980) and Mak (1992)] ; and the nonparametric approach, which assumed $\sigma_i^2 = H(X_i)$ or $H(X_i^T\beta)$ for H being unknown. [See Carroll and Härdle (1989), Fuller and Rao (1978) and Hall and Carroll (1989)]

Let $\{(Y_i, X_i, T_i), i = 1, \ldots, n\}$ denote a random sample from

$$Y_i = X_i^T \beta + g(T_i) + \sigma_i e_i, i = 1, \ldots, n, \tag{2}$$

where X_i, T_i, T_i are the same as these in model (1). e_i are i.i.d. with mean 0 and variance 1. σ_i^2 are some functions of other variables, whose specific form is discussed in later sections.

The classic approach works as follows. Assume $\{(X_i, T_i, Y_i); i = 1, \ldots, n.\}$ satisfy the model (2). Let $\{W_{ni}(t) = W_{ni}(t; T_1, \ldots, T_n), i = 1, \ldots, n\}$ be probability weight functions depending only on the design points $T_1, \ldots, T_n$.

Since $g(T_i) = E(Y_i - X_i^T \beta)$. Let β be the "true" value, and then suppose

$$g_n(t) = \sum_{j=1}^{n} W_{nj}(t)(Y_j - X_j^T \beta)$$

Replace now $g(T_i)$ by $g_n(T_i)$ in model (2), we then obtain least squares estimator of β

$$\beta_{LS} = (\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^T \widetilde{\mathbf{Y}} \tag{3}$$

where $\widetilde{\mathbf{X}}^T = (\widetilde{X}_1, \ldots, \widetilde{X}_n)$, $\widetilde{X}_i = X_i - \sum_{j=1}^n W_{nj}(T_i) X_j$; $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_n)^T$, $\widetilde{Y}_i = Y_i - \sum_{j=1}^n W_{nj}(T_i) Y_j$.

When the errors are heteroscedastic, β_{LS} is modified to a weighted least squares estimator

$$\beta_W = \Big(\sum_{i=1}^{n} \gamma_i \widetilde{X}_i \widetilde{X}_i^T\Big)^{-1} \Big(\sum_{i=1}^{n} \gamma_i \widetilde{X}_i \widetilde{Y}_i\Big) \tag{4}$$

for some weight γ_i $i = 1, \ldots, n$. In our model (2) we take $\gamma_i = 1/\sigma_i^2$.

In principle the weights γ_i (or σ_i^2)are unknown and must be estimated. Suppose $\{\widehat{\gamma}_i, i = 1, \ldots, n\}$ be a sequence of estimators of β. Naturally one can take β_W given in (4) by substituting β by $\widehat{\gamma}_i$ as our estimator of β.

In order to develop the asymptotic theory, we use the idea of splitting of sample. Let k_n be the integer part of $n/2$. $\widehat{\gamma}_i^{(1)}$ and $\widehat{\gamma}_i^{(2)}$ are the estimators of γ_i based on the first k_n sample (X_1, T_1, Y_1), ..., $(X_{k_n}, T_{k_n}, Y_{k_n})$; and the later $n - k_n$ samples $(X_{k_n+1}, T_{k_n+1}, Y_{k_n+1})$, ..., (X_n, T_n, Y_n), respectively. Define

$$\beta_{nW} = \Big(\sum_{i=1}^{n} \widehat{\gamma}_i \widetilde{X}_i \widetilde{X}_i^T\Big)^{-1} \Big(\sum_{i=1}^{k_n} \widehat{\gamma}_i^{(2)} \widetilde{X}_i \widetilde{Y}_i + \sum_{i=k_n+1}^{n} \widehat{\gamma}_i^{(1)} \widetilde{X}_i \widetilde{Y}_i\Big) \tag{5}$$

as the estimator of β.

The next step is to establish our conclusion, that is, to prove that β_{nW} is asymptotic normal. We intend to prove β_W is asymptotic normal, and then prove $\sqrt{n}(\beta_{nW} - \beta_W)$

converges to zero in probability. Some notations are introduced. $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, $\widetilde{\varepsilon} = (\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n)^T$, $\widetilde{\varepsilon}_i = \varepsilon_i - \sum_{j=1}^n W_{nj}(T_i)\varepsilon_j$; $g_{ni} = g(T_i) - \sum_{k=1}^n W_{nk}(T_i)g(T_k)$, $\widehat{G} = (g(T_1) - \widehat{g}_n(T_1), \ldots, g(T_n) - \widehat{g}_n(T_n))^T$; $h_j(t) = E(x_{ij}|T_i = t)$, $u_{ij} = x_{ij} - h_j(T_i)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. We will use the following assumptions.

Assumption 1. $\sup_{0\le t\le 1} E(\|X_1\|^3|T = t) < \infty$ *and* $\lim_{n\to\infty} 1/n \sum_{i=1}^n \gamma_i u_i u_i^T = B$ *and* B *is a positive definite matrix. Where* $u_i = (u_{i1}, \ldots, u_{ip})^T$.

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

Assumption 3. *Weight functions* $W_{ni}(\cdot)$ *satisfy the following:*

$$(i) \quad \max_{1\le i\le n} \sum_{j=1}^n W_{ni}(T_j) = O(1), \quad a.s.$$

$$(ii) \quad \max_{1\le i,j\le n} W_{ni}(T_j) = O(b_n), \quad a.s. \quad b_n = n^{-2/3},$$

$$(iii) \quad \max_{1\le i\le n} \sum_{j=1}^n W_{nj}(T_i) I(|T_j - T_i| > c_n) = O(c_n), \quad a.s. \;\; c_n = n^{-1/2}\log^{-1} n.$$

Assumption 4. *There exist constants* C_1 *and* C_2 *such that*

$$0 < C_1 \le \min_{i\le n} \gamma_i \le \max_{i\le n} \gamma_i < C_2.$$

We suppose that the estimators $\widehat{\gamma}_i$ of γ_i satisfy

$$\sup_{1\le i\le n} |\widehat{\gamma}_i - \gamma_i| = o_P(n^{-q}) \quad q \ge 1/4 \tag{6}$$

We shall construct such as estimators for several kinds of γ_i in Sections 3-5. The following theorems present general results for parameter estimate of partial linear heteroscedastic models.

Theorem 1. *Under Assumptions 1-4 .* β_W *is an asymptotically normal estimator of* β, *i.e.,*

$$\sqrt{n}(\beta_W - \beta) \longrightarrow^{\mathcal{L}} N(0, B^{-1}B_1B^{-1})$$

with $B_1 = Cov\{X_1 - E(X_1|T_1)\}$.

Theorem 2. *Under Assumptions 1-4 and (6).* β_{nW} *is an asymptotically equivalent, i.e.,* $\sqrt{n}(\beta_{nW} - \beta)$ *and* $\sqrt{n}(\beta_W - \beta)$ *have the same asymptotically normal distributions.*

Remark. 1.1. *In the case of constant error variance, i.e.* $\sigma_i^2 \equiv \sigma^2$, *Theorem 1 was obtained by many authors. See for example, Speckman (1988) and Gao et al. (1995). The point is that it has no cost neither from the adaption nor the splitting method.*

Remark. 1.2. *Assumptions 1-4 are rather general in nature, we will give concrete examples in section 3-5.*

Remark. 1.3. *Theorem 2 not only assures that our estimator given in (5) is asymptotically equivalent to the weighted LS estimator with known weights but also generalize the earlier results of related literature.*

The outline of the paper is as follows. Section 2 states some prelimary results for proving the main results. Sections 3-5 present various different variance functions and state the corresponding estimates. Section 6 gives results of simulations. The proofs of Theorems 1 and 2 are postponed in Section 7.

2 SOME LEMMAS

In this section we make some preparation for proving our main results. Lemma 2.1 provides the boundedness for $h_j(T_i) - \sum_{k=1}^n W_{nk}(T_i)h_j(T_k)$ and $g(T_i) - \sum_{k=1}^n W_{nk}(T_i)g(T_k)$. Its proof is immediate. Denote $\bar{h}_{ns}(T_i) = h_s(T_i) - \sum_{k=1}^n W_{nk}(T_i)x_{ks}$, recall that $u_{ij} = x_{ij} - h_j(T_i)$ and $u_i = (u_{i1}, \ldots, u_{ip})^T$. The variables $\{u_i\}$ are also independent identically distributed random vectors.

Lemma 2.1. *Suppose that Assumptions 2 and 3 (iii) hold. Then*

$$\max_{1\le i\le n}\Big|G_j(T_i) - \sum_{k=1}^n W_{nk}(T_i)G_j(T_k)\Big| = O(c_n) \qquad for\ j = 0, \ldots, p$$

where $G_0(\cdot) = g(\cdot)$ *and* $G_l(\cdot) = h_l(\cdot)$ *for* $l = 1, \ldots, p$.

Lemma 2.2. *Suppose Assumptions 1-3 hold. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \gamma_i \widetilde{X}_i^T \widetilde{X}_i = B$$

Proof. It follows from $x_{is} = h_s(T_i) + u_{is}$ that the $(s, m)-$th element of $\frac{1}{n}\sum_{i=1}^n \gamma_i \widetilde{X}_i^T \widetilde{X}_i$ $(s, m = 1, \ldots, p)$ is

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \gamma_i \widetilde{x}_{is}\widetilde{x}_{im} &= \frac{1}{n}\sum_{i=1}^n \gamma_i u_{is}u_{im} + \frac{1}{n}\sum_{i=1}^n \gamma_i \bar{h}_{ns}(T_i)u_{im} \\ &\quad + \frac{1}{n}\sum_{i=1}^n \gamma_i \bar{h}_{nm}(T_i)u_{is} + \frac{1}{n}\sum_{i=1}^n \gamma_i \bar{h}_{ns}(T_i)\bar{h}_{nm}(T_i) \\ &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \gamma_i u_{is}u_{im} + \frac{1}{n}\sum_{q=1}^3 R_{nsm}^{(q)} \end{aligned}$$

Assumption 1 implies $\lim_{n\to\infty}\gamma_i\sum_{i=1}^n \frac{1}{n}u_iu_i^T = B$. Lemma 2.1 and Assumption 4 imply $R_{nsm}^{(3)} = o(n)$. Then Cauchy-Schwarz inequality yields $R_{nsm}^{(1)} = o_P(n)$ and $R_{nsm}^{(2)} = o_P(n)$. These arguments prove the lemma.

Next we shall prove a rather general result on strong uniform convergence of weighted averages in Lemma 2.3, which is applied in the later proofs repeatly. First we give an exponential inequlity for bounded independent random variables, that is

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

$$P(|\sum_{i=1}^n V_i|>\eta)\le 2\exp\Big[-\eta^2\Big/\Big\{2\Big(\sum_{i=1}^n varV_i+M\eta\Big)\Big\}\Big].$$

Lemma 2.3. *Let $V_1,\dots,V_n$ be independent random variables with means zero and finite variances, i.e., $\sup_{1\le j\le n}E|V_j|^r\le C<\infty$ $(r\ge 2)$. Assume $(a_{ki},k,i=1\dots,n)$ be a sequence of positive numbers such that $\sup_{1\le i,k\le n}|a_{ki}|\le n^{-p_1}$ for some $0<p_1<1$ and $\sum_{j=1}^n a_{ji}=O(n^{p_2})$ for $p_2\ge\max(0,2/r-p_1)$. Then*

$$\max_{1\le i\le n}\Big|\sum_{k=1}^n a_{ki}V_k\Big| = O(n^{-s}\log n)\quad\text{for}\quad s=(p_1-p_2)/2.\quad a.s.$$

Proof. Denote $V_j'=V_jI(|V_j|\le n^{1/r})$ and $V_j''=V_j-V_j'$ for $j=1,\dots,n$. Let $M=Cn^{-p_1}n^{1/r}$ and $\eta=n^{-s}\log n$. By Beinstein's inequality

$$\begin{aligned}P\Big\{\max_{1\le i\le n}\Big|\sum_{j=1}^n a_{ji}(V_j'-EV_j')\Big| &> C_1\eta\Big\}\le\sum_{i=1}^n P\Big\{\Big|\sum_{j=1}^n a_{ji}(V_j'-EV_j')\Big|>C_1\eta\Big\}\\ &\le 2n\exp\Big(-\frac{C_1n^{-2s}\log^2 n}{2\sum_{j=1}^n a_{ji}^2EV_j^2+2n^{-p_1+1/r-s}\log n}\Big)\\ &\le 2n\exp(-C_1^2C\log n)\le Cn^{-3/2}\qquad\text{for some large } C_1>0.\end{aligned}$$

The last second inequality from

$$\sum_{j=1}^n a_{ji}^2EV_j^2\le\sup|a_{ji}|\sum_{j=1}^n a_{ji}EV_j^2=n^{-p_1+p_2}\qquad\text{and } n^{-p_1+1/r-s}\log n\le n^{-p_1+p_2}.$$

By Borel-Cantelli Lemma

$$\max_{1\le i\le n}\Big|\sum_{j=1}^n W_{nj}(T_i)(V_j'-EV_j')\Big| = O(n^{-s}\log n)\quad a.s. \tag{7}$$

Let $1\le p<2$, $1/p+1/q=1$ such that $1/q<(p_1+p_2)/2-1/r$. By Hölder's inequality

$$\begin{aligned}\max_{1\le i\le n}\Big|\sum_{j=1}^n a_{ji}(V_j''-EV_j'')\Big| &\le \max_{1\le i\le n}\Big(\sum_{j=1}^n|a_{ji}|^q\Big)^{1/q}\Big(\sum_{j=1}^n|V_j''-EV_j''|^p\Big)^{1/p}\\ &\le Cn^{-(p_1q-1)/q}\Big(\sum_{j=1}^n|V_j''-EV_j''|^p\Big)^{1/p}\end{aligned}\tag{8}$$

Observe that

$$\frac{1}{n}\sum_{j=1}^{n}\Big(|V_j'' - EV_j''|^p - E|V_j'' - EV_j''|^p\Big) \to 0 \qquad a.s. \tag{9}$$

and $E|V_j''|^p \le E|V_j|^r n^{-1+p/r}$, and then

$$\sum_{j=1}^{n} E|V_j'' - EV_j''|^p \le Cn^{p/r} \quad a.s. \tag{10}$$

Combining (8), (9) with (10), we obtain

$$\max_{1\le i\le n}\Big|\sum_{k=1}^{n} a_{ki}(V_k'' - EV_k'')\Big| \le Cn^{-p_1+1/q+1/r} = o(n^{-s}) \quad a.s. \tag{11}$$

Lemma 2.3 follows from (7) and (11) directly.

Let $r = 3$, $V_k = e_k$ or u_{kl}, $a_{ji} = W_{nj}(T_i)$, $p_1 = \frac{2}{3}$ and $p_2 = 0$. We obtain the following formulas, which will play critical roles in the process of proving the theorems.

$$\max_{i\le n}\Big|\sum_{k=1}^{n} W_{nk}(T_i)e_k\Big| = O(n^{-1/3}\log n) \quad a.s. \tag{12}$$

and

$$\max_{i\le n}\Big|\sum_{k=1}^{n} W_{nk}(T_i)u_{kl}\Big| = O(n^{-1/3}\log n) \ \text{ for } l = 1,\ldots,p \quad a.s.$$

3 VARIANCE IS A FUNCTION OF OTHER RANDOM VARIABLES

This section is devoted to the nonparametric heteroscedasticity structure

$$\sigma_i^2 = H(W_i),\ H \text{ unknown Lipschitz continuous}$$

where $\{W_i; i = 1,\ldots,n\}$ are also design points, which are assumed to be independent of e_i and (X_i, T_i) and defined on $[0,1]$.

Define

$$\widehat{H}_n(w) = \sum_{j=1}^{n} \widetilde{W}_{nj}(w)(Y_j - X_j^T\beta_{LS} - \widehat{g}_n(T_i))^2$$

as the estimator of $H(w)$. Where $\{\widetilde{W}_{nj}(t); i = 1,\ldots,n\}$ is a sequence of weight functions satisfying also the same assumptions on $\{W_{nj}(t); j = 1,\ldots,n\}$.

Theorem 3.1. *Under our assumptions,*

$$\sup_{1\le i\le n} |\widehat{H}_n(W_i) - H(W_i)| = o_P(n^{-1/3}\log n)$$

Proof. Set $\varepsilon_i = \sigma_i e_i$ for $i = 1, \ldots, n$. Note that

$$\begin{aligned}\widehat{H}_n(W_i) &= \sum_{j=1}^n \widetilde{W}_{nj}(W_i)(\widetilde{Y}_j - \widetilde{X}_j^T \beta_{LS})^2 \\ &= \sum_{j=1}^n \widetilde{W}_{nj}(W_i)\{\widetilde{X}_j^T(\beta - \beta_{LS}) + \widetilde{g}(T_i) + \widetilde{\varepsilon}_i\}^2 \\ &= (\beta-\beta_{LS})^T \sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{X}_j\widetilde{X}_j^T(\beta - \beta_{LS}) + \sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{g}^2(T_i) + \sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{\varepsilon}_i^2 \\ &\quad +2(\beta-\beta_{LS})\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{X}_j^T\widetilde{g}(T_i) + 2(\beta-\beta_{LS})\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{X}_j^T\widetilde{\varepsilon}_i \\ &\quad +2\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{g}(T_i)\widetilde{\varepsilon}_i \end{aligned} \tag{13}$$

Since $\sum_{j=1}^n \widetilde{X}_j\widetilde{X}_j^T$ is a symmetric matrix, and $0 < \widetilde{W}_{nj}(W_i) \le Cn^{-2/3}$,

$$\sum_{j=1}^n \{\widetilde{W}_{nj}(W_i) - Cn^{-2/3}\}\widetilde{X}_j\widetilde{X}_j^T$$

is a $p\times p$ nonpositive matrix. Recall that $\beta_{LS} - \beta = O(n^{-1/2})$. These arguments mean the first term of (13) is $O_P(n^{-2/3})$. The second term of (13) is easily shown to be order $O_P(n^{1/3}c_n^2)$.

Now we want to show that

$$\sup_i \Big|\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\widetilde{\varepsilon}_i^2 - H(W_i)\Big| = O_P(n^{-1/3}\log n) \tag{14}$$

This is equivalent to prove the following three items

$$\sup_i \Big|\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\Big\{\sum_{k=1}^n W_{nk}(T_j)\varepsilon_k\Big\}^2\Big| = O_P(n^{-1/3}\log n) \tag{15}$$

$$\sup_i \Big|\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\varepsilon_i^2 - H(W_i)\Big| = O_P(n^{-1/3}\log n) \tag{16}$$

$$\sup_i \Big|\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\varepsilon_j\Big\{\sum_{k=1}^n W_{nk}(T_j)\varepsilon_k\Big\}\Big| = O_P(n^{-1/3}\log n) \tag{17}$$

(12) assures that (15) holds. Lipschitz continuity of $H(\bullet)$ and assumptions on $\widetilde{W}_{nj}(\bullet)$ entails (16), i.e.,

$$\sup_i \Big|\sum_{j=1}^n \widetilde{W}_{nj}(W_i)\varepsilon_i^2 - H(W_i)\Big| = O_P(n^{-1/3}\log n) \tag{18}$$

whose proof is similar as that of Lemma 2.1.

By taking $a_{ki} = \widetilde{W}_{nk}(W_i)H(W_k)$ and $V_k = e_k^2 - 1$ and $r = 2$ and $p_1 = 2/3$ and $p_2 = 0$ in Lemma 2.3, we have

$$\sup_i \Big| \sum_{j=1}^{n} \widetilde{W}_{nj}(W_i)H(W_j)(e_j^2 - 1) \Big| = O_P(n^{-1/3} \log n) \tag{19}$$

A combination of (19) and (18) means (15). (16) and (15) and Cauchy-Schwarz inequality imply (17). Thus we proved (14).

The later three terms of (13) are all of order $o_P(n^{-1/3} \log n)$ by Cauchy-Schwarz inequality and the conclusions for the first three terms of (13). Thus we complete the proof of Theorem 3.1.

4 VARIANCE IS A FUNCTION DESIGN T_i

In this section we consider the case in which we suppose the variance σ_i^2 is a function of the design points T_i, i.e.

$$\sigma_i^2 = H(T_i) \ H \text{ unknown Lipschitz continuous}$$

Similar as in section 3, we define our estimator of $H(\bullet)$ as

$$\widehat{H}_n(t) = \sum_{j=1}^{n} \widetilde{W}_{nj}(t)\{Y_j - X_j^T \beta_{LS} - \widehat{g}_n(T_i)\}^2$$

Theorem 4.1. *Under our assumptions,*

$$\sup_{1 \le i \le n} |\widehat{H}_n(T_i) - H(T_i)| = o_P(n^{-1/3} \log n)$$

Proof. The proof of Theorem 4.1 is similar to that of Thorem 3.1 and is omitted.

5 VARIANCE IS A FUNCTION OF THE MEAN

Here we consider the model (2) with

$$\sigma_i^2 = H\{X_i^T \beta + g(T_i)\}, \ H \text{ unknown Lipschitz continuous}$$

which means that the variance is a unknown function of mean response. A related situations in linear and nonlinear models are discussed by Carroll (1982), Box and Hill (1974), Bickel

(1978), Jobson and Fuller (1980) and Carroll and Ruppert (1982). Engle et al. (1986), Green et al. (1985) and Wahba (1984) and others studied the estimator for the regession function $X^T\beta + g(T)$.

Since $H(\cdot)$ is assumed completely unknown, the standard method is to get information about $H(\cdot)$ by replication, i.e., we consider the following "improved" partial linear heteroscedastic model

$$Y_{ij} = X_i^T\beta + g(T_i) + \sigma_i e_{ij}, \quad j = 1, \ldots, m_i; i = 1, \ldots, n$$

Here Y_{ij} is the response of the jth replicate at the design point (X_i, T_i), e_{ij} are i.i.d. with mean 0 and variance 1, β, $g(\cdot)$ and (X_i, T_i) are the same as that in model (2).

We will borrow the idea of Fuller and Rao (1978) for linear heteroscedastic model to construct an estimate of σ_i^2. That is, to compute predicted value $X_i^T\beta_{LS} + \widehat{g}_n(T_i)$ by fit least squares estimate β_{LS} and nonparametric estimate $\widehat{g}_n(T_i)$ to the data, and residuals $Y_{ij} - \{X_i^T\beta_{LS} + \widehat{g}_n(T_i)\}$ and estimate

$$\widehat{\sigma}_i^2 = \frac{1}{m_i}\sum_{j=1}^{m_i}[Y_{ij} - \{X_i^T\beta_{LS} + \widehat{g}_n(T_i)\}]^2. \tag{20}$$

When each m_i stays bounded, Fuller and Rao (1978) concluded that the weighted estimate based on (20) and the weighted least squares estimates based on the true weights have different limiting distributions results from the fact that $\widehat{\sigma}_i^2$ do not converge in probability to the true σ_i^2.

Theorem 5.1. *Let $m_i = a_n n^{2q} \stackrel{\text{def}}{=} m(n)$ for some sequence a_n converging to infinite. Under our assumptions,*

$$\sup_{1\le i\le n} |\widehat{\sigma}_i^2 - H\{X_i^T\beta + g(T_i)\}| = o_P(n^{-q}) \quad q \ge 1/4$$

Proof. We only outline the proof of the theorem. In fact

$$|\widehat{\sigma}_i^2 - H\{X_i^T\beta + g(T_i)\}| \le 3\{X_i^T(\beta - \beta_{LS})\}^2 + 3\{g(T_i) - g_n(T_i)\}^2 + \frac{3}{m_i}\sum_{j=1}^{m_i}\sigma_i^2(e_{ij}^2 - 1)$$

The first two items are obviously $o_P(n^{-q})$. Since e_{ij} are i.i.d. with mean zero and variance 1, after taking $m_i = a_n n^{2q}$, $\sum_{j=1}^{m_i}(e_{ij}^2 - 1)$ is equivalent to $\sum_{j=1}^{m(n)}(e_{1j}^2 - 1)$. Using the law of the iterated logarithm and the boundedness of $H(\cdot)$ one know that

$$\frac{1}{m_i}\sum_{j=1}^{m_i}\sigma_i^2(e_{ij}^2 - 1) = O\{m(n)^{-1/2}\log m(n)\} = o_P(n^{-q})$$

Thus we derive the proof of Theorem 5.1.

6 SIMULATION

We present a small simulation study to explain the behaviour of the previous results. We took the following model with different variance functions.

$$Y_i = X_i^T \beta + g(T_i) + \sigma_i \varepsilon_i, \quad i = 1, \ldots, n = 300$$

Here $\{\varepsilon_i\}$ are standard normal random variables, $\{X_i\}$ and $\{T_i\}$ are both of uniform random variables on $[0,1]$. $\beta = (1, 0.75)^T$ and $g(t) = \sin(t)$. The simulation number for each situation is 500.

Three models for the variance functions are considered. LSE and WLSE represent the least squares estimator and the weighted least squares estimator given in (3) and (5), respectively.

- Model 1: $\sigma_i^2 = T_i^2$;
- Model 2: $\sigma_i^2 = W_i^3$; where W_i are i.i.d. uniformly distributed random variables.
- Model 3: $\sigma_i^2 = a_1 \exp[a_2 \{X_i^T \beta + g(T_i)\}^2]$, where $(a_1, a_2) = (1/4, 1/3200)$. This model is mentioned by Carroll (1982) without the item $g(T_i)$.

TABEL 1: Simulation results ($\times 10^{-3}$)

Estimator	Variance	$\beta_0 = 1$		$\beta_1 = 0.75$	
	Model	Bias	MSE	Bias	MSE
LSE	1	8.696	8.7291	23.401	9.1567
WLSE	1	4.230	2.2592	1.93	2.0011
LSE	2	12.882	7.2312	5.595	8.4213
WLSE	2	5.676	1.9235	0.357	1.3241
LSE	3	5.9	4.351	18.83	8.521
WLSE	3	1.87	1.762	3.94	2.642

From tabel 1, one can find that our estimator (WLSE) is better than LSE in the sense of both bias and MSE for above each model.

By the way, we also mention the behaviour of the estimate for nonparametric part, that is

$$\sum_{i=1}^{n} \omega_{ni}^*(t)(\widetilde{Y}_i - \widetilde{X}_i^T \beta_{nW})$$

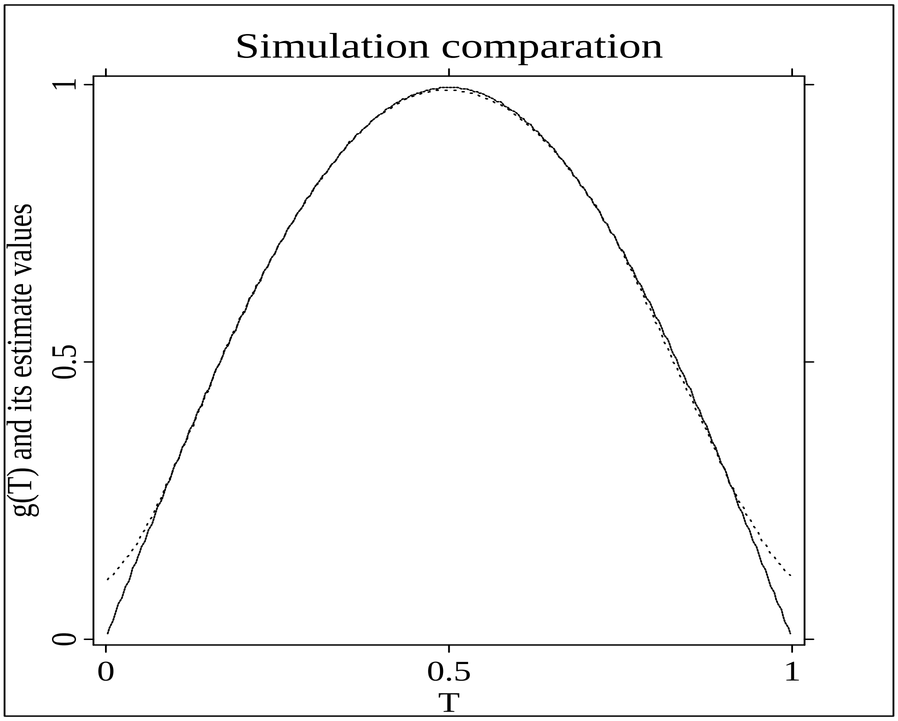

Figure 1: *Estimates of the function $g(T)$ for the first model*

$\omega^*_{ni}(\cdot)$ are other weight functions which also satisfy the Assumption 3. In procedure of simulations, we take Nadaraya-Watson weight function with quartic kernel $(15/16)(1-u^2)^2I(|u| \leq 1)$ and use Cross-Validation criterion to select bandwidth. Figures 1,2,3 are devoted to the simulation results of the nonparametric parts for the models 1, 2, 3, respectively. In the following figures, solid-lines for real values and dished-lines for our estimate values. The figures indicate that our estimators for nonparametric part perform also well except the neighbourhoods of the points 0 and 1.

7 PROOFS OF THEOREMS

First two notations are introduced.

$$\widehat{A}_n = \sum_{i=1}^n \widehat{\gamma}_i \widetilde{X}_i \widetilde{X}_i^T, \quad A_n = \sum_{i=1}^n \gamma_i \widetilde{X}_i \widetilde{X}_i^T$$

For any matrix S, $s(j,l)$ denotes the (j,l)-th element of S.

Proof of Theorem 1. It follows from the definition of β_W that

$$\beta_W - \beta = A_n^{-1}\Big\{\sum_{i=1}^n \gamma_i \widetilde{X}_i \widetilde{g}(T_i) + \sum_{i=1}^n \gamma_i \widetilde{X}_i \widetilde{\varepsilon}_i\Big\}$$

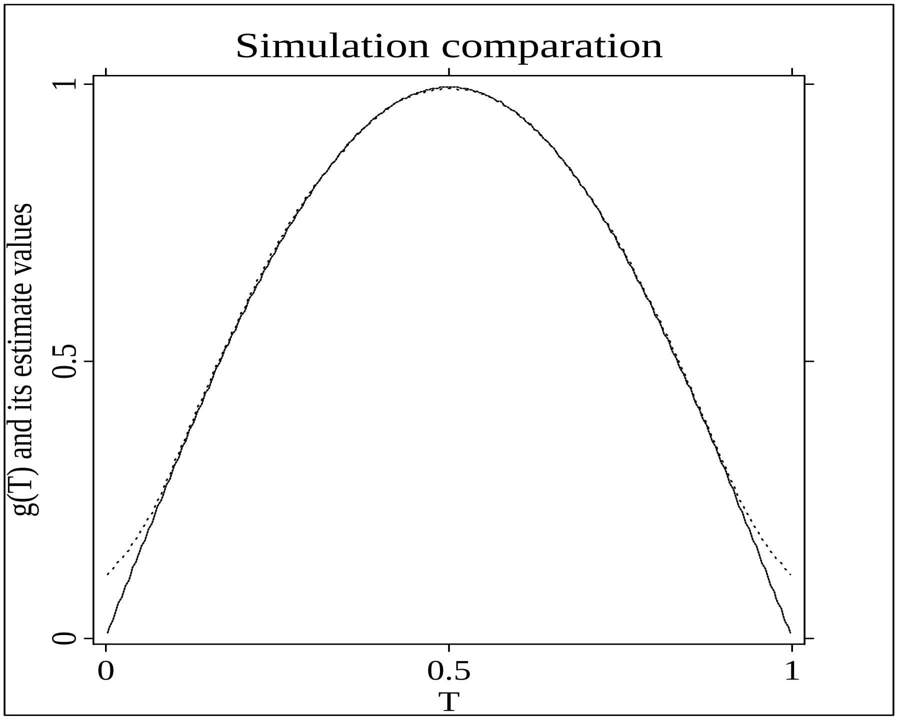

Figure 2: *Estimates of the function $g(T)$ for the second model*

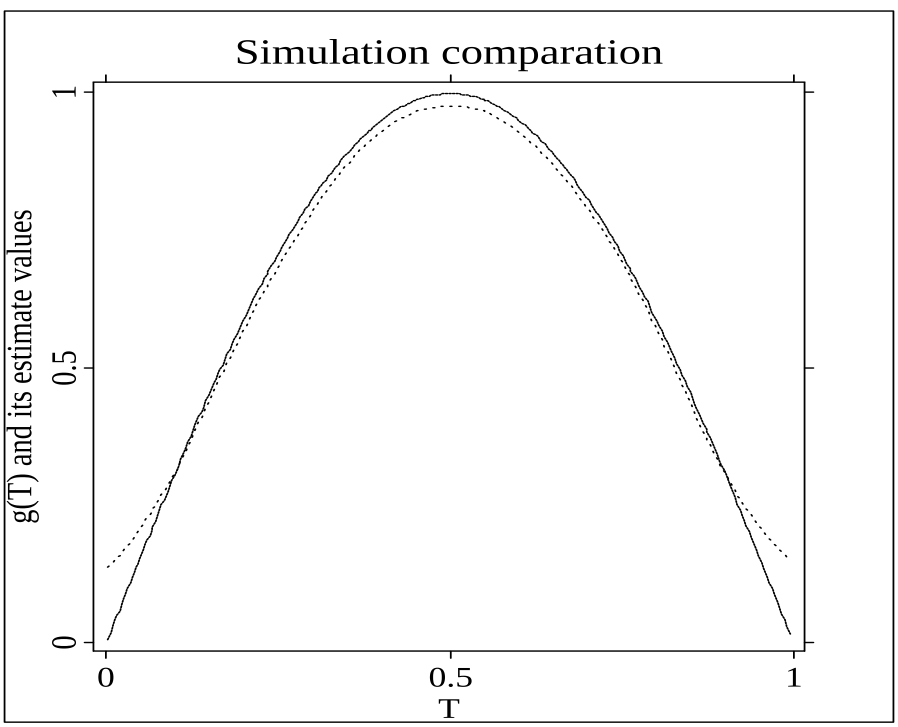

Figure 3: *Estimates of the function $g(T)$ for the third model*

We will complete the proof by the following three steps, for $j = 1, \ldots, p$,

(i) $H_{1j} = 1/\sqrt{n}\sum_{i=1}^n \gamma_i \widetilde{x}_{ij}\widetilde{g}(T_i) = o_P(1)$;

(ii) $H_{2j} = 1/\sqrt{n}\sum_{i=1}^n \gamma_i \widetilde{x}_{ij}\Big\{\sum_{k=1}^n W_{nk}(T_i)e_k\Big\} = o_P(1)$;

(iii) $H_3 = 1/\sqrt{n}\sum_{i=1}^n \gamma_i \widetilde{X}_i e_i \longrightarrow^{\mathcal{L}} N(0, B^{-1}B_1B^{-1})$.

The proof of (i) is mainly based on lemmas 2.1 and 2.3. Denote $h_{nij} = h_j(T_i) - \sum_{k=1}^n W_{nk}(T_i)h_j(T_k)$. Note

$$\sqrt{n}H_{1j} = \sum_{i=1}^n \gamma_i u_{ij} g_{ni} + \sum_{i=1}^n \gamma_i h_{nij} g_{ni} - \sum_{i=1}^n \gamma_i \sum_{q=1}^n W_{nq}(T_i) u_{qj} g_{ni} \tag{21}$$

In Lemma 2.3 we take $r = 2$, $V_k = u_{kl}$, $a_{ji} = g_{nj}$, $\frac{1}{4} < p_1 < \frac{1}{3}$ and $p_2 = 1 - p_1$. Then the first term of (21) is

$$O_P(n^{-\frac{2p_1-1}{2}}) = o_P(n^{1/2})$$

The second term of (21) can be easily shown to be order $O_P(nc_n^2)$ by using Lemma 2.1.

The third term of (21) can be handled by using Abel's inequality and lemmas 2.1 and 2.3. Hence

$$\Big|\sum_{i=1}^n\sum_{q=1}^n \gamma_i W_{nq}(T_i)u_{qj}g_{ni}\Big| \le C_2 n \max_{i\le n}|g_{ni}| \max_{i\le n}\Big|\sum_{q=1}^n W_{nq}(T_i)u_{qj}\Big| = O(n^{2/3}c_n \log n).$$

Thus we complete the proof of (i).

We now show (ii), i.e., $\sqrt{n}H_{2j} \to 0$. Notice that

$$\begin{aligned}
\sqrt{n}H_{2j} &= \sum_{i=1}^n \gamma_i\Big\{\sum_{k=1}^n \widetilde{x}_{kj}W_{ni}(T_k)\Big\}e_i \\
&= \sum_{i=1}^n \gamma_i\Big\{\sum_{k=1}^n u_{kj}W_{ni}(T_k)\Big\}e_i + \sum_{i=1}^n \gamma_i\Big\{\sum_{k=1}^n h_{nkj}W_{ni}(T_k)\Big\}e_i \\
&\quad - \sum_{i=1}^n \gamma_i\Big[\sum_{k=1}^n\Big\{\sum_{q=1}^n u_{qj}W_{nq}(T_k)\Big\}W_{ni}(T_k)\Big]e_i
\end{aligned} \tag{22}$$

The order of the first term of (22) is $O(n^{-\frac{2p_1-1}{2}}\log n)$ by letting $r = 2$, $V_k = e_k$, $a_{li} = \sum_{k=1}^n u_{kj}W_{ni}(T_k)$, and $\frac{1}{4} < p_1 < \frac{1}{3}$ and $p_2 = 1 - p_1$ in Lemma 2.3.

It follows from Lemma 2.1 and (12) that the second term of (22) is bounded by

$$\Big|\sum_{i=1}^n \gamma_i\Big\{\sum_{k=1}^n h_{nkj}W_{ni}(T_k)\Big\}e_i\Big| \le n\max_{k\le n}\Big|\sum_{i=1}^n W_{ni}(T_k)e_i\Big|\max_{j,k\le n}|h_{nkj}| = O(n^{2/3}c_n\log n) \quad a.s. \tag{23}$$

The same argument as that for (23) yields that the thrid term of (22) cab be dealt with as

$$\begin{aligned}\Big|\sum_{i=1}^n \gamma_i\Big[\sum_{k=1}^n\Big\{\sum_{q=1}^n u_{qj}W_{nq}(T_k)\Big\}W_{ni}(T_k)\Big]e_i\Big| &\le n\max_{k\le n}\Big|\sum_{i=1}^n W_{ni}(T_k)e_i\Big|\max_{k\le n}\Big|\sum_{q=1}^k u_{qj}W_{nq}(T_j)\Big| \\ &= O(n^{\frac{1}{3}}\log^2 n) = o(n^{1/2}) \quad a.s. \qquad (24)\end{aligned}$$

A combination (22)–(24) entails (ii).

Finally the central limit theorem and Lemma 2.2 derive that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i\widetilde{X}_i e_i \longrightarrow^{\mathcal{L}} N(0, B_1),$$

which and the fact that $A_n \to B$ imply that

$$\frac{1}{\sqrt{n}}A_n^{-1}\sum_{i=1}^n \gamma_i\widetilde{X}_i e_i \longrightarrow^{\mathcal{L}} N(0, B^{-1}B_1B^{-1}).$$

This completes the proof of Theorem 1.

Proof of Theorem 2. In order to complete the proof of Theorem 2, we prove

$$\sqrt{n}(\beta_{nW} - \beta_W) = o_P(1)$$

First we state a fact, whose proof is immediately derived by (6) and Lemma 2.2,

$$\frac{1}{n}|\widehat{a}_n(j,l) - a_n(j,l)| = o_P(n^{-q}) \qquad (25)$$

for $j, l = 1, \ldots, p$. This will be used later repeatly.

It follows that

$$\begin{aligned}\beta_{nW} - \beta_W &= \frac{1}{2}\Big\{A_n^{-1}(A_n - \widehat{A}_n)\widehat{A}_n^{-1}\sum_{i=1}^n \gamma_i\widetilde{X}_i\widetilde{g}(T_i) \\ &+\widehat{A}_n^{-1}\sum_{i=1}^{k_n}(\gamma_i - \widehat{\gamma}_i^{(2)})\widetilde{X}_i\widetilde{g}(T_i) + \widehat{A}_n^{-1}(A_n - \widehat{A}_n)\widehat{A}_n^{-1}\sum_{i=1}^n \gamma_i\widetilde{X}_i\widetilde{e}_i \\ &+\widehat{A}_n^{-1}\sum_{i=1}^{k_n}(\gamma_i - \widehat{\gamma}_i^{(2)})\widetilde{X}_i\widetilde{e}_i + \widehat{A}_n^{-1}\sum_{i=k_n+1}^{n}(\gamma_i - \widehat{\gamma}_i^{(1)})\widetilde{X}_i\widetilde{g}(T_i) \\ &+\widehat{A}_n^{-1}\sum_{i=k_n+1}^{n}(\gamma_i - \widehat{\gamma}_i^{(1)})\widetilde{X}_i\widetilde{e}_i\Big\} \qquad (26)\end{aligned}$$

By Cauchy-Schwarz inequality, for any $j = 1, \ldots, p$,

$$\Big|\sum_{i=1}^n \gamma_i\widetilde{x}_{ij}\widetilde{g}(T_i)\Big| \le C\sqrt{n}\max_{i\le n}|\widetilde{g}(T_i)|\Big(\sum_{i=1}^n \widetilde{x}_{ij}^2\Big)^{1/2}$$

This is $o_P(n^{3/4})$ by lemmas 2.1 and 2.2. Thus each element of the first term of (26) is $o_P(n^{-1/2})$ by watching the fact that each element of $A_n^{-1}(A_n - \widehat{A}_n)\widehat{A}_n^{-1}$ is $n^{-5/4}$. The similar argument shows that each element of the second and fifth terms is also $o_P(n^{-1/2})$.

Recall that the proofs for $H_{2j} = o_P(1)$ and H_3 converges to normal distribution, we conclude that the third term of (26) is also $o_P(n^{-1/2})$. Thus we see that the difficult problem is to show that the fourth and the last terms of (26) are both $o_P(n^{-1/2})$. Since their proofs are the same, we only show that, for $j = 1, \ldots, p$,

$$\Big\{\widehat{A}_n^{-1} \sum_{i=1}^{k_n} (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{X}_i \widetilde{e}_i \Big\}_j = o_P(n^{-1/2})$$

or equivalently

$$\sum_{i=1}^{k_n} (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{x}_{ij} \widetilde{e}_i = o_P(n^{1/2}) \tag{27}$$

Let $\{\delta_n\}$ be a sequence numbers converge to zero but satisfy $\delta_n > n^{-1/4}$. Then for any $\mu > 0$ and $j = 1, \ldots, p$,

$$P\Big\{\sum_{i=1}^{k_n} (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{x}_{ij} e_i I(|\gamma_i - \widehat{\gamma}_i^{(2)}| \geq \delta_n) > \mu n^{1/2}\Big\} \leq P\Big\{\max_{i \leq n} |\gamma_i - \widehat{\gamma}_i^{(2)}| \geq \delta_n\Big\} \to 0 \tag{28}$$

The last step is due to (6).

Next we shall deal with the term

$$P\Big\{\sum_{i=1}^{k_n} (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{x}_{ij} e_i I(|\gamma_i - \widehat{\gamma}_i^{(2)}| \leq \delta_n) > \mu n^{1/2}\Big\}$$

by Chebyshv's inequality. Since $\widehat{\gamma}_i^{(2)}$ is independent of e_i for $i = 1, \ldots, k_n$, we can easily calculate

$$E\Big\{\sum_{i=1}^{k_n} (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{x}_{ij} e_i\Big\}^2$$

This is why we use splitting techniqne to estimate γ_i by $\widehat{\gamma}_i^{(2)}$ and $\widehat{\gamma}_i^{(1)}$. In fact,

$$\begin{aligned} P\Big\{\sum_{i=1}^{k_n} & (\gamma_i - \widehat{\gamma}_i^{(2)}) \widetilde{x}_{ij} e_i I(|\gamma_i - \widehat{\gamma}_i^{(2)}| \leq \delta_n) > \mu n^{1/2}\Big\} \\ &\leq \frac{\sum_{i=1}^{k_n} E\{(\gamma_i - \widehat{\gamma}_i^{(2)}) I(|\gamma_i - \widehat{\gamma}_i^{(2)}| \leq \delta_n)\}^2 E\|\widetilde{X}_i\|^2 E e_i^2}{n\mu^2} \\ &\leq C \frac{k_n \delta_n^2}{n\mu^2} \to 0 \end{aligned} \tag{29}$$

Thus, by (28) and (29),

$$\sum_{i=1}^{k_n}(\gamma_i - \widehat{\gamma}_i^{(2)})\widetilde{x}_{ij}e_i = o_P(n^{1/2})$$

Finally

$$\begin{aligned}\Big|\sum_{i=1}^{k_n}(\gamma_i - \widehat{\gamma}_i^{(2)}) \quad \widetilde{x}_{ij} \quad \Big\{\sum_{k=1}^{n} W_{nk}(T_i)e_k\Big\}\Big| \\ \leq \quad \sqrt{n}\Big(\sum_{i=1}^{k_n}\widetilde{X}_{ij}^2\Big)^{1/2}\max_{i\leq n}|\gamma_i - \widehat{\gamma}_i^{(2)}|\max_{i\leq n}\Big|\sum_{k=1}^{n} W_{nk}(T_i)e_k\Big|\end{aligned}$$

This is $o_P(n^{1/2})$ by using (25) and (12) and Lemma 2.2, which and (29) entail (27). We complete the proof of Theorem 2.

REFERENCES

Bickel, P.J. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Annals of Statistics*, **6** 266-291.

Bickel, Peter J., Klaasen, Chris A.J., Ritov, Ya'acov and Wellner, Jon A. (1993). *Efficient and adaptive estimation for semiparametric models.* The Johns Hopkins University Press.

Box,G.E.P. and Hill, W.J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16**, 385-389.

Carroll, R.J. and Härdle, W. (1989). Second order effects in semiparametric weighted least squares regression. *Statistics* **2**, 179-186.

Carroll, R.J. (1982). Adapting for heteroscedasticity in linear models, *Annals of Statistics*, **10**, 1224-1233.

Carroll, R.J.and Ruppert, D. (1982). Robust estimation in heteroscedasticity linear models, *Annals of Statistics*, **10**, 429-441.

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.

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

Fuller, W.A. and Rao, J.N.K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Annals of Statistics*, **6**, 1149-1158.

Green, P., Jennison, C. and Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B*, **47**, 299-315.

Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society, Series B*, **51**,3-14.

Heckman, N.E.(1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.

Jobson, J.D. and Fuller, W.A. (1980). Least squares estimation when covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Association*, **75**, 176-181.

Mak, T. K. (1992). Estimation of parameters in heteroscedastic linear models. *Journal of the Royal Statistical Society, Series B*, **54**, 648-655.

Müller, H. G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Annals of Statistics*, **15,** 610-625.

Robinson, P.W. (1987). Asymptotically efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica*, **55**, 875-891.

Robinson, P.M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

Schick, A. (1987). A note on the construction of asymptotically linear estimators. *Journal of Statistical Planning & Inference*, **16**, 89-105. Correction **22**, (1989) 269-270.

Schick, A. (1996a). Weighted least squares estimates in partly linear regression models. *Statistics & Probability Letters*, **27**, 281-287.

Schick, A. (1996b). Efficient estimates in linear and nonlinear regression with heteroscedastic errors. To appear by *Journal of Statistical Planning & Inference.*

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.

Wahba, G. (1984). Partial spline models for the semi-parametric estimation of several variables. *In Statistical Analysis of Time Series,* 319-329.