ASYMPTOTIC PROPERTIES OF THE NONPARAMETRIC PART IN PARTIAL LINEAR HETEROSEDASTIC REGRESSION MODELS

Hua Liang, Wolfgang Hardle and Axel Werwatz

Abstract

This paper considers estimation of the unknown function \( g \) in the partial linear regression model

\[
Y_i = X_i^T \beta + g(T_i) + \epsilon_i
\]

with heteroscedastic errors. We first construct a class of estimates \( \hat{g}_n \) of \( g \) and prove that, under appropriate conditions, \( \hat{g}_n \) is weakly mean square consistent. Rates of convergence and asymptotic normality for the estimator \( \hat{g}_n \) are also established.

Key Words and Phrases: Key words and phrases: Asymptotic normality, consistency, heteroscedasticity, kernel estimation, rates of convergence, partial linear model, semiparametric models, nonparametric models, model selection, penalization.

1 INTRODUCTION

Semiparametric models combine the flexibility of nonparametric modeling with structural parametric components. One such model that has received a lot of attention in the literature is the semiparametric partial linear regression model

\[
Y_i = X_i^T \beta + g(T_i) + \epsilon_i
\]

where \( \epsilon_i \) is the error term. The estimation of the unknown function \( g \) is a challenging problem. The authors would like to thank Mr. Knut Bartels for his valuable comments which greatly improved the presentation of this paper.

Hua Liang is Associate Professor of Statistics, Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, China. Wolfgang Hardle is Professor of Econometrics, Axel Werwatz is Dr. of Economics, both of them are at the Institute for Statistical and Econometric Methods of Humboldt-Universität zu Berlin, D-10117 Berlin, Germany. This research was supported by Sonderforschungsbereich 373 "Quantifikation und Simulation ökonomischer Prozesse"/Quanti
cation und Simulation ökonomischer Prozesse/ and by the Alexander von Humboldt Foundation. The authors would like to thank Mr. Knut Bartels for his valuable comments which greatly improved the presentation of this paper.

\[
\beta \cdots i \beta + (i \beta)^T \theta + \epsilon_i^T X = Y
\]
\[ Y_i = \beta' X_i + \varepsilon_i \quad (i=1, \ldots, n) \]

Specifically, we consider estimation of \( g(T) \) and \( \beta \) in the following partial linear model:

\[ Y_i = \beta' X_i + g(T_i) + \varepsilon_i \quad (i=1, \ldots, n) \]

The remainder of this paper is organized as follows. In the following section we will briefly study and an empirical illustration in the final section of the paper. The purpose of the asymptotic distribution results for applied work by a small-scale Monte Carlo study of the estimator of \( g(t) \) in section 3 and 4. We illustrate the usefulness of the estimator and describe methods for estimation of \( g(t) \) and \( \beta \). We prove consistency and asymptotic normality of the estimator of \( g(t) \) and \( \beta \). We consider its consistency, weak convergence rate and rate of estimation of \( g(t) \) and \( \beta \). Several authors have considered estimation of \( g(t) \) and \( \beta \) under various assumptions. Here, we consider estimation of \( g(t) \) and \( \beta \) in the following partial linear model:

\[ Y_i = \beta' X_i + g(T_i) + \varepsilon_i \quad (i=1, \ldots, n) \]

Under various assumptions, several authors have considered estimation of \( g(t) \) and \( \beta \). For instance, Czichon and Schwamb (1961) considered estimation of \( g(t) \) and \( \beta \) in a partially linear model with a possibly non-linear relationship between \( X \) and \( T \). The wage equation is obtained by including the local unemployment rate as an additional regressor with a possibly non-linear influence. Rendle and Schwamb (1986) studied the effect of the local unemployment rate on the wage curve and labor market experience, which makes the wage curve play the role of the variable \( T \). In both cases, log of human capital and measures of personal characteristics (sex, marital status and education, and the wage curve (Blanchflower and Oswald (1994))). In both cases, log of human capital is estimated as a function of \( T \) and \( X \). Well-known applications in the economics literature that can be put in the form of (1) are the human capital earnings function (Wills (1986)) and the wage curve (Blanchflower and Oswald (1994)).
Given the estimator \( \hat{g} \) (as defined in (3)) we may estimate \( g \) by the least squares method.

is the kernel function and \( h \) denotes the bandwidth. See Remark 2 below for details.

Here \( n^u \cdot \cdots \cdot n^{(u)} = 1 \) for \( s = 0, \ldots, u - 1 \), \( s \).

\[
\frac{u}{s} \int_{s}^{u} \frac{1}{y} = (i)^{m_0}
\]

For instance, a Gasser-Müller-type kernel estimator takes

where \( c \) are weight functions that depend on the observations

\[
\sum_{u = i}^{i + 1} \left( \frac{1}{i} \right)^{m_0}
\]

write (still assuming that \( \hat{g} \) is known):

\[
\hat{g}(\hat{X}) = \sum_{i = 1}^{n} \left( \frac{1}{i} \right)^{m_0}
\]

the variation in \( X \) not accounted for by the linear component (0) of \( g \) is

Suppose we knew \( \beta \). Then we may estimate \( \hat{g} \) by nonparametric regression of \( \hat{g} \) as a function of \( \hat{X} \).

We focus on nonparametric estimation of \( \hat{g} \), we consider straightforward least squares estimation of \( \hat{g} \) with \( \hat{g} \) and \( \hat{g} \) has been considered in Silverman (1986) but the

The heteroscedastic version of (1) with \( \hat{g} \) with the partial integration of the variance function by

\[
\hat{g}(\hat{X}) = \sum_{i = 1}^{n} \left( \frac{1}{i} \right)^{m_0}
\]

model \( \gamma = 0 \), Miller et al. (1987) proposed an estimator of the variance function by

where \( \hat{g} \) is an unknown, smooth function \( g \) with unknown dimension vector. We allow the variance

\[
\hat{g} + (\hat{L}) \hat{g} + (\hat{L}) \hat{X} = \hat{Y}
\]
\[ d \cdot \ldots \cdot 1 = u \text{ of} \quad \infty > \left( \frac{\max_{i=1}^{u}}{u} \right) \frac{1}{\sum_{i=1}^{u}} u \sum_{i=1}^{u} \infty - u \lim_{n \to \infty} \]

is a positive definite matrix, and

\[ G = \sum_{i=1}^{\infty} \sum_{i=1}^{u} u \sum_{i=1}^{u} \infty - u \]

\[ 0 = \frac{1}{u} \sum_{i=1}^{u} \sum_{i=1}^{u} \infty - u \]

where \( u \) is a sequence of real numbers which satisfies \( \lim \frac{n}{u} \to 1 \).

For \( d \leq \ell \leq 1 \), \( u \geq 1 \), \( \ell n + (\ell)^{\ell} = \ell !x \)

\begin{align}
\text{assumption I} \quad & \text{The class continuous functions of } \ell \text{ defined on } [0, 1], \text{ each had each constant not depending on } u. \text{ We will use the following assumptions, appendix G, seminars, for continuity and simplicity we always let } C \text{ denote some positive number.} \\
& \text{Further motivation for the estimators defined in (5) and (6) is given in Spickman (1988).} \\
& \text{(6) } (\frac{s}{\ell}Y, \frac{s}{\ell}X - \lambda)(\frac{s}{\ell}u - \lambda) = (\frac{s}{\ell})^{\star} \lambda \text{ unknown in (3) (6) } \lambda \text{ variables.} \\
& \text{In the final step we obtain the feasible estimator of } \lambda \text{ by substituting the (6) for the} \\
& \text{variables are the proportional design and response } X \text{ where } Z = \frac{1}{X} \sum_{n=1}^{X} X \frac{1}{n} \text{ and } Z = \frac{1}{X} \sum_{n=1}^{X} X \frac{1}{n} \text{ is given in Spickman (1988).} \\
& \text{That is, we estimate by the generalized least squares estimator.} \\
& \lambda + \frac{\ell}{\ell} \frac{1}{\lambda} X = \lambda \\
& \frac{\ell}{\ell} \left\{ \frac{1}{\lambda} X (\frac{\ell}{\ell} - \lambda) \right\} \sum_{n=1}^{X} X \frac{1}{n} = \frac{1}{\lambda} \sum_{n=1}^{X} X \sum_{n=1}^{X} X - \lambda \\
\end{align}
\[(01) \quad (L)^{\varepsilon} \equiv X_1(\varepsilon X_1)^{\varepsilon} \sum_{u=1}^{n} - \left\{ (L) - (L)^{\varepsilon} \right\} = (L) - (L)^{\varepsilon}\]

where \((L)^{\varepsilon} = (L)^{\varepsilon} \equiv X_1(\varepsilon X_1)^{\varepsilon} \sum_{u=1}^{n} - \left\{ (L) - (L)^{\varepsilon} \right\} = (L) - (L)^{\varepsilon}\)

Remark. Assumption 1 is a common requirement for proving consistency of the partial linear model. In fact, from Assumption 1, it follows that\[(a)\quad \lim_{n \to \infty} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\]

Under Assumptions 1 and 2, Theorem 1. Discussions may be found in Speckman (1998) and Gao et al. (1999). More detailed proofs of the following results are similar to the result of the strong law of large numbers for random errors. A similar result is similar to the result of the strong law of large numbers for random variables.

Proof. Decompose the difference \((a)\quad \lim_{n \to \infty} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\)

Theorem 1. Under Assumptions 1 and 2, discussions may be found in Speckman (1998) and Gao et al. (1999).

When are random variables, \((a)\quad \lim_{n \to \infty} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\)

Remark 1. Assumption 1 is a common requirement for proving consistency of the partial linear model. In fact, from Assumption 1, it follows that\[(a)\quad \lim_{n \to \infty} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\]

Assumption 2. \[\frac{1}{n} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\]

\[\frac{1}{n} \sum_{i=1}^{n} \left\{ (L)^{\varepsilon} \cdot \sum_{u=1}^{n} \right\} = (L) - (L)^{\varepsilon}\]
We shall now prove the fourth form also converges to zero. Denote $\rho(2n)$ as our representation.

The first term tends to zero by Lemma A1. By Lemma A2 and Cauchy, the second can be shown to be order of $1/n$, hence it can be replaced by a term $O(1/n)$. The second can be shown to be of order $1/n^2$, and we complete the proof of Theorem 1.

Theorem 2 assumes the conditions of Theorem 1 hold except Assumption 2 which is also mean-square consistent.

Theorem 2 will demonstrate that $\bar{\rho}(2n)$ is also an asymptotically unbiased estimator of $\bar{\rho}(n)$ for every

Thus

\[ (\tau^1_1)_{\rho} \approx x(1)^{\tau/2} \]

The same proof as one for Lemma A1. Moreover, by Kolmogorov's inequality and Assumption (i) $\gamma$, the second term of the continuity point of $\bar{\rho}(n)^{\tau/2}$ converges to $\bar{\rho}(n)^{\tau/2}$ as $n \to \infty$. Since $\bar{\rho}(n)^{\tau/2}$ is continuous, (i) $\gamma$ implies

\[ \rho(n) + (\tau^2_1)_{\rho} \approx x(1)^{\tau/2} \]

It suffices to show that every element of $\bar{\rho}(n)^{\tau/2}$. Observe that

\[ d \cdots d = d \cdots d \]

\((u \delta_{1/2} - u) dO = (L \delta_{1/2} X - (X_{1/2} X_{1/2}^\top)_{1/2} X_{1/2}^\top)^{\frac{1}{u}}\)  \\

The similar arguments as that for (11) and (12) yield  
\[(u \delta_{1/2} - u) dO = \tau^2 \delta_{1/2}^u \frac{1}{u}\]

Since \(\tau^2 \delta_{1/2}^u \frac{1}{u}\) and Chebyshev's inequality, we have  
\[(u \delta_{1/2} - u) dO = (L \delta_{1/2}^u \frac{1}{u}) \frac{1}{u}\]

**Proof.** By Lemma A.1

\[(u \delta_{1/2} - u) dO = (\delta - \delta)\]

Furthermore, we can easily show that  
\[(u \delta_{1/2} - u) dO = \tau (\delta - \delta)\]

The following result gives the weak convergence rate of \(\delta\) under stronger assumptions.  

**Theorem 2.** The weight functions satisfy  
\[(\delta \tau_{1/2}^2 \delta_{1/2}) \sim \text{max } |\frac{1}{u} + \frac{1}{u} - \frac{1}{u}| \text{ for } \theta \]

\[(\delta \tau_{1/2}^2 \delta_{1/2}) \sim \text{max } |\frac{1}{u} + \frac{1}{u} - \frac{1}{u}| \text{ for } \theta \]

Theorem 2. The weight functions satisfy  
\[(\delta \tau_{1/2}^2 \delta_{1/2}) \sim \text{max } |\frac{1}{u} + \frac{1}{u} - \frac{1}{u}| \text{ for } \theta \]

The following result gives the weak convergence rate of \(\delta\) under stronger assumptions.  

# The proof of Theorem 2

The following result gives the weak convergence rate of \(\delta\) under stronger assumptions.  

Combining (11) and (12) ensures that the fourth term of (1) is \(o(1)\) and thus completes  

\[(\theta_{1/2}^2 \theta_{1/2}) \sim \text{max } |\frac{1}{u} + \frac{1}{u} - \frac{1}{u}| \text{ for } \theta \]

Furthermore, we can easily show that  
\[(\theta_{1/2}^2 \theta_{1/2}) \sim \text{max } |\frac{1}{u} + \frac{1}{u} - \frac{1}{u}| \text{ for } \theta \]

It follows from the arguments for (12) that this equals to  
\[\frac{1}{u} + \frac{1}{u} - \frac{1}{u} \text{ for } \theta \]
Theorem 4. Under Assumptions 1, 2, and 3, Theorem 4 gives the asymptotic variance of \( \xi \). This completes the proof of Theorem 3.

\[ \limsup_{n \to \infty} \frac{1}{n} \int \frac{\partial}{\partial u} T u = \{ (i)^\delta \} \text{ for } i > 0. \]

**Proof.** Under Assumptions 1, 2, and 3, Theorem 4 gives the asymptotic variance of \( \xi \). We can conclude from the above arguments that

\[ (u, \log \xi) dO = \frac{\partial}{\partial u} T u = \{ (i)^\delta \} \text{ for } i > 0. \]

This completes the proof of Theorem 4.

\[ (1-u)O = \int \frac{\partial}{\partial u} T u = \{ (i)^\delta \} \text{ for } i > 0. \]

Thus, by the arguments for (1), \( n \) tends to zero for (1), \( n \) satisfies the Cauchy-Schwarz inequality for (1), and then (1) tends to zero.

Finally, observe that

\[ (1-u)O = \int \frac{\partial}{\partial u} T u = \{ (i)^\delta \} \text{ for } i > 0. \]
Assumption 2. The weight function \( w \) satisfies \( (i)^{m_n} \) condition.

Remark 1. If \( \varepsilon \) is a small positive constant, then the result of Theorem 1 follows from the conditions \( (\varepsilon_i) \) and \( \varepsilon < (i) \).

Remark 2. The proof of Theorem 2 follows immediately from the conditions (61) and (61).
(1) \( \phi = np(n) \nu^{\frac{n-1}{2}} \int_{1}^{\infty} \gamma \, d\nu \)
Then Assumption (c) hold also. In fact

\[ \begin{align*}
\ln Y_i &= X_i' \beta + g(T_i) + \epsilon_i, \\
\end{align*} \]

In the model estimate and \( \hat{\beta} \) in the model individually skills, personal characteristics and labor market conditions. Specifically, we particularly relates the logarithm of earnings to a set of explanatory variables describing an known economic application that can be put into the form of a partial linear model.

In the introduction we already mentioned the human-capital earnings function as a well-

### NUMERICAL EXAMPLES

This is just the classical conclusion in nonparametric regression estimation.

\[ \begin{align*}
\alpha \leftarrow \sup \{ \beta' \alpha \} \text{ with } \text{if } \beta' \alpha \leq \sup \{ \beta' \alpha \} \text{ then } \beta' \alpha \leq \sup \{ \beta' \alpha \} \\
\end{align*} \]

Then we can take \( \alpha < \varepsilon \), and Assumptions 3 and 6 imply that

\[ \frac{u}{C} > \frac{1 - \beta}{\beta' \beta} \sup_{\alpha \in O} \frac{u}{C} \geq \frac{1 - \beta}{\beta' \beta} \left( \sup_{\alpha \in O} \frac{u}{C} \right) \geq \frac{u}{C} \]

There exist constants \( C > 0 \) such that

\[ (u \log \varepsilon)^2 \Omega = n \rho(n) \frac{\rho(n)^2}{\varepsilon^2} \]

Now let us take \( \alpha' = \varepsilon \), for some \( \alpha' < \varepsilon \), and suppose \( 0 < \alpha' < \varepsilon \).

\[ (1) \alpha = \sup \{ \beta' \alpha \} \left( 1 - \beta \right) \sup_{\alpha \in O} \frac{u}{C} \]

\[ s \rho \left( \frac{u}{C} \right) \sup \{ \beta' \alpha \} \left( 1 - \beta \right) \sup_{\alpha \in O} \frac{u}{C} \]

\[ (d < |L| - 1) I_{(d)} \sup \left( \frac{u}{C} \right) \left( 1 - \beta \right) \sup_{\alpha \in O} \frac{u}{C} \]

Then Assumption (c) hold also. In fact
We also conducted a small simulation study to get further insights into the small-sample estimation by econometric theory and often confirmed by parametric model fitting. To allow for concavity, parametric specifications of the earnings function typically include \( T \) and \( T^2 \) in the model and obtain a positive estimate for the coefficient of \( T \) and a negative estimate for the coefficient of \( T^2 \). 

For nonparametric fitting, we use a Nadaraya-Watson weight function with quartic kernel. To allow for concavity, parametric specifications of the earnings function typically include \( T \) and \( T^2 \) in the model and obtain a positive estimate for the coefficient of \( T \) and a negative estimate for the coefficient of \( T^2 \). 

To allow for concavity, parametric specifications of the earnings function typically include \( T \) and \( T^2 \) in the model and obtain a positive estimate for the coefficient of \( T \) and a negative estimate for the coefficient of \( T^2 \). 

To allow for concavity, parametric specifications of the earnings function typically include \( T \) and \( T^2 \) in the model and obtain a positive estimate for the coefficient of \( T \) and a negative estimate for the coefficient of \( T^2 \). 

To allow for concavity, parametric specifications of the earnings function typically include \( T \) and \( T^2 \) in the model and obtain a positive estimate for the coefficient of \( T \) and a negative estimate for the coefficient of \( T^2 \). 

Figure 1: Relationship of log-earnings and labor-market experience

![Figure 1: Relationship of log-earnings and labor-market experience](image-url)
Simulation comparison

\[ g(T) \text{ and its estimate values} \]

Figure 2: Estimates of the function \( g(T) \) per performance of the estimator of \( g(T) \).

We consider the model

\[ Y_i = X_T_i + \sin(T_i) + \sin(X_{T_i} + T_i) \]

where \( \varepsilon \) is standard normally distributed and \( X_i, T_i \) are sampled from a uniform distribution.

\[ \frac{1}{n} \sum_{i=1}^{n} g(T_i) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon(X_i + \varepsilon X_i) \sin(T_i) + \varepsilon(X_i + \varepsilon X_i) \sin(T_i) = X \]

Theorem A1. Suppose that Assumption A hold and \( g \) are continuous.

In this appendix we state some useful lemmas.

6 APPENDIX

Lemma A1.1. Suppose that Assumption A holds and \( g \) and \( h \) are continuous.

Then

\[ (1) = \frac{1}{n} \sum_{i=1}^{n} g(T_i) \leq \frac{1}{n} \sum_{i=1}^{n} h(T_i) \]

Performance of the estimator of \( g(T) \)
Furthermore, if \( (u) = \frac{\alpha}{(\beta)} \) and \( (u) = \frac{\alpha}{(\beta)} \), then

\[ \max \{ \gamma, \delta \} \leq \frac{\alpha}{(\beta)} \]
The following Lemma is a slight version of Theorem 1.8.1 of Chow and Teicher (1988).

Lemma A.3. Let $\mathcal{A} = \max_{i \leq n} \{ Y_i \}$ be independent random variables with $E[\mathcal{A}] = 0$ and $E[\mathcal{A}^2] < \infty$. Assume that

$$\lim_{n \to \infty} \mathbb{P}(\mathcal{A} = 0) = 0,$$

and for some constant $c > 0$, $\mathbb{P}(\max_{i \leq n} \{ Y_i \} > c) \to 0$ as $n \to \infty$. Then

$$\mathbb{P}(\max_{i \leq n} \{ Y_i \} > c) \to \mathbb{P}(\mathcal{A} > c)$$

as $n \to \infty$. Therefore, the following Lemma is therefore derived from Lemma A.3.

Lemma A.4. Let $X_1, \ldots, X_n$ be independent random variables with $E[X_i] = 0$ and $E[X_i^2] < \infty$. Assume that

$$\lim_{n \to \infty} \mathbb{P}(\max_{i \leq n} |X_i| > c) = 0,$$

for some constant $c > 0$. Then

$$\mathbb{P}(\max_{i \leq n} |X_i| > c) \to 0,$$

as $n \to \infty$. Therefore, we have

$$(\mathcal{A})^{1/2}/\mathcal{A} \to \log \mathbb{E}[\max_{i \leq n} |X_i|]$$

as $n \to \infty$. Also assume that

$$\sup_{\alpha > 0} \mathbb{E}[\max_{i \leq n} \{ X_i \}] = \sup_{\alpha > 0} \mathbb{E}[\max_{i \leq n} \{ X_i \}] < \infty.$$

Then

$$\max_{i \leq n} \{ X_i \} \to \mathcal{A}$$

as $n \to \infty$. Therefore, the following Lemma is therefore derived from Lemma A.4.
REFERENCES


Wills, R. J. (1986). Wage Determinants: A Survey and Reinterpretation of Human Capital


cal Society, Series B*, 50, 413-436.
